

THE LACTOSE UTILIZATION NETWORK OF E. COLI AS A HYBRID SYSTEM: EXPLORING BISTABILITY WITH REACHABILITY ANALYSIS

Ádám M. Halász^{1*}, Marcin Imieliński^{1,2}, Peter Finin¹, Oleg Sokolsky¹, Harvey Rubin² and Vijay Kumar¹

¹GRASP Lab and ²School of Medicine, University of Pennsylvania, Philadelphia, PA 19104-6228

Abstract

We present an approach to the analysis of biochemical networks based on the paradigm of hybrid systems and reachability analysis, and apply it to model and analyze the lactose induction mechanism in E.coli. By solving the steady state equations directly, we identify parameter ranges for which the system exhibits bistability. This property is necessary for the crucial feature we call ‘inducibility’, the phenomenon of reaching the induced state from a given set of initial states. We use reachability analysis to identify conditions on initial states and parameter variations which are necessary for inducibility.

Keywords

Lactose metabolism, Reachability, Induction, lac operon

Introduction

With the increase in experimental data and an increasing interest in modeling it makes sense to formulate biologically relevant queries as reachability questions, for at least two reasons. On one hand, the corresponding dynamical systems are typically high dimensional and have nonlinear interactions, making analytical and intuitive insights difficult. On the other hand, the quantitative data is subject to significant uncertainties, due to both natural variability and experimental limitations.

Reachability allows us to delineate the reachable states from a given set of plausible initial states, or to determine the set of plausible initial states that lead to a set of final states. In this paper, we apply reachability to a model of the Lactose metabolism of E.coli.

Our method (Belta et al., 2002), relies on dividing the state-space into hyper-rectangles and replacing the nonlinear functions describing the kinetics with different multi-affine approximations in each hyper-rectangle. The linear interpolation property of multi-affine functions allows a computationally efficient procedure to identify

necessary conditions for reachability between hyper-rectangular sets in state space.

The lactose-glucose metabolism of E.coli is particularly interesting because it exhibits bistability, as seen in the phenomenon of induction of the lac operon. During exponential growth, the bacterium can process lactose at either a high or at a low rate, for a range of external lactose concentrations. Induction or switching to the high rate occurs when a threshold value of lactose is exceeded. A mathematical model (Yildirim and Mackey, 2003) successfully reproduced this feature. In our previous work (Halász et al, 2004) we showed that while the model was robust under variations of parameter values, the threshold for induction was sensitive to the basal rate of transcription.

A more complete model (Santillán and Mackey, 2004) takes into account the effect of glucose. Lactose is actually an alternative source of energy to the bacterium, secondary to glucose. When the latter is present, the lactose metabolism is kept in its low state via catabolite repression

* To whom all correspondence should be addressed (email: halasz@grasp.upenn.edu).

(enhancement of lac transcription via a signaling molecule, cAMP) and inducer exclusion (blocking of lactose influx by glucose). Catabolite repression is described mathematically by a sophisticated stochastic model of the lac promoter and its interactions with the lac repressor, allolactose, and the cAMP-CRP complex.

Reachability and Piecewise Affine Systems

Multi-affine functions are affine (of the form $f(x)=ax+b$) in each of their variables, and can be represented as a sum of monomials containing at most the first power of each variable. They have a linear interpolation property, over hyper-rectangles (Cartesian products of intervals of each of their variables). The value of the function at any point inside a hyper-rectangle can be written as a convex combination of the values taken on the vertices of the hyper-rectangle.

Consider a dynamical system described by a state vector X , and a (matrix) velocity function, $f(X)$:

$$\dot{X} = f(X) \quad (1)$$

Each element $g(x_1, \dots, x_k)$ of $f(X)$ is a multivariate function. We choose a convenient partition of the space of X into hyper-rectangles and replace the original function with a multi-affine function on each hyper-rectangle, so that the resulting approximating function is continuous. After doing this for each entry in $f(X)$, we merge the partitions resulting from the individual functions. Each hyper-rectangle in the joint partition has a complete set $f_H(X)$ of multi-affine functions $g_H(x_1, \dots, x_k)$.

Suppose a trajectory crosses the common facet between two adjacent hyper-rectangles, A and B , in the direction of B . The velocity vector at the crossing point points towards the interior of B , ie, its component orthogonal to the facet must have the appropriate sign. The orthogonal component is a multi-affine function on the common facet, and its value inside the facet is bounded by its values on the vertices (because of the interpolation property). It can have a given sign somewhere on the facet if and only if it also does so on one of the vertices.

This reduces the problem of checking reachability between two adjacent hyper-rectangles to checking the sign of a multi-affine function at 2^d points. Since the vertex is shared by just as many hyper-rectangles, the reachability verification between adjacent hyper-rectangles requires on average one function evaluation. Using this criterion we identify the existence of crossing trajectories in either direction between all pairs of adjacent hyper-rectangles, resulting in a graph whose vertices represent hyper-rectangles, with directed edges between pairs of hyper-rectangles for which crossing trajectories exist. The existence of a path between two nodes is a necessary condition for the existence of a trajectory linking the two corresponding hyper-rectangles.

Since we exploit a necessary condition reachability analysis using this procedure is conservative. Because of this it is possible that the configuration of hyper-rectangles resulting from the piecewise approximation may be too coarse for a given reachability question. In practice, it is useful to subdivide these rectangles to provide additional granularity.

If the dependence on parameters is approximated with piecewise affine functions, the parameters can be treated as variables whose time derivative is zero. One may perform reachability analyses for intervals of parameter values, particularly useful when parameter uncertainty is an issue.

The Glucose-Lactose System

A diagram of the model network is given in Figure 1. The lac operon codes for permease and beta-galactosidase. Permease brings external lactose into the cell and beta-galactosidase converts it into allolactose. Allolactose blocks the lac repressor which otherwise inhibits transcription of the lac operon. In the absence of glucose, the signaling substance cAMP is produced. cAMP binds to CRP to form a complex, cAMP·CRP, which enhances lac transcription. Glucose also inhibits the inbound transport of lactose.

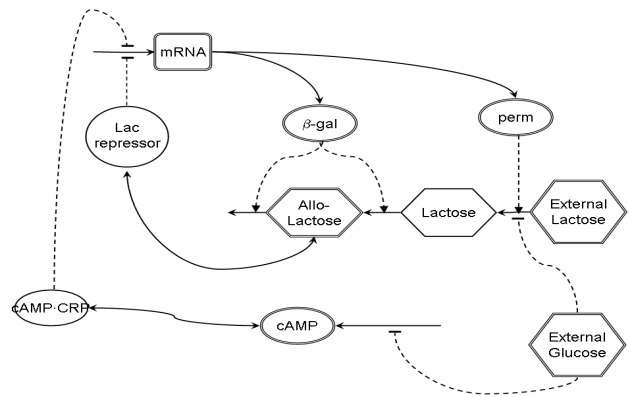


Figure 1: Diagram of the lactose-glucose network. The external glucose and external lactose are the external inputs while the dynamics determine the concentrations of the other quantities.

We follow closely the model by (Santillán and Mackey, 2004), referring the reader there for further details and references. The reactions of cAMP with CRP are assumed to be fast and the concentrations of cAMP-CRP and of free cAMP are algebraic functions of the total concentration of cAMP. The activity of the lac promoter is described by a thermodynamical model which gives the transcription rate as a function of the concentrations of complex and (free) lac repressor. The dependence of the free lac repressor on the concentration of allolactose is also given algebraically. The amount of allolactose is set to one half of that of lactose inside the

cell. The above dependencies are summarized by two functions, $\eta([A_i], [cAMP])$ for the activity of the lac promoter and $\omega([cAMP])$ for the concentration of free cAMP. The dynamical part of the model describes the evolution of six variables representing the concentrations of β -galactosidase, mRNA, permease mRNA, β -galactosidase, permease, total allolactose, and total cAMP, $\{[M_b], [M_p], [B], [P], [A_i], [cAMP]\}$:

$$\begin{aligned}
\dot{[M_b]} &= [P]k_m\eta([cAMP], [A_i])_{\tau_b} - (\mu + \xi_M)[M_b] \\
\dot{[M_p]} &= [P]k_m\eta([cAMP], [A_i])_{\tau_p} - (\mu + \xi_M)[M_p] \\
\dot{[B]} &= \frac{1}{4}\kappa_B[M_b]_{\tau_B} - (\mu + \xi_B)[B] \\
\dot{[P]} &= \kappa_P[M_p]_{\tau_P} - (\mu + \xi_P)[P] \\
\dot{[A_i]} &= \frac{1}{2}\varphi_{L_1}[P]\left(\frac{[L_E]}{\Phi_{L_1} + [L_E]}\Phi_{G_1} + [G_E] + \frac{[A_i]}{\Phi_{L_1} + [L_E]}\right) \\
&\quad - \frac{1}{2}\varphi_{L_2}[B]\left(\frac{[A_i]}{[A_i] + \Phi_{L_2}/2}\right) \\
\dot{[cAMP]} &= \varphi_C\frac{\Phi_C}{\Phi_C + [G_E]} - \xi_C\omega([cAMP]) - \mu[cAMP]
\end{aligned} \tag{2}$$

The concentrations of external glucose $[G_E]$ and lactose $[L_E]$ are inputs. The remaining symbols are constants taken from (Santillán and Mackey, 2004). The first four equations contain time delays which we ignore as they do not modify the steady state structure and we do not expect them to significantly impact the reachability properties.

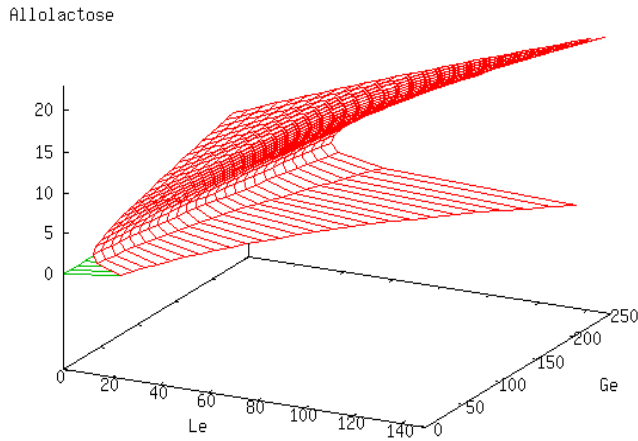


Figure 2: Steady state values of allolactose as a function of external glucose and lactose. All units are μM .

The steady states of the system (2) can be calculated by setting the equations of motion to zero. A diagram of the steady state values of allolactose is given in Figure 2. The surface is 'folded' for a section of the lactose-glucose plane, found between two threshold values of Le (Figure

4). For these parameter values there are three solutions to the steady-state equations, of which only the outer two are stable (as shown numerically).

We are interested in the possibility of induction, when the system is caused to evolve into a high internal lactose state. It can be achieved by temporarily increasing lactose, decreasing glucose or a combination of the two.

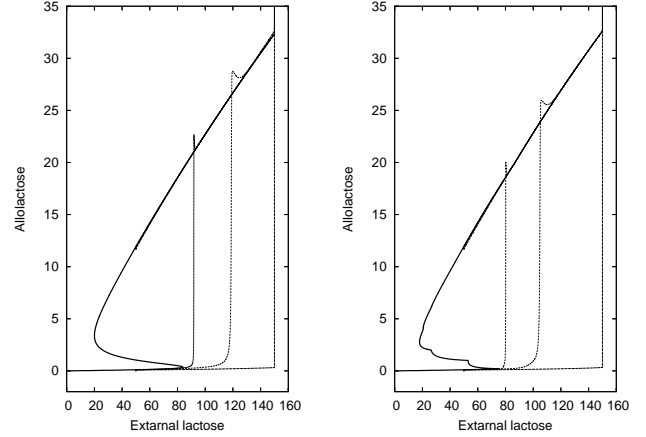


Figure 3: Steady state values and simulation traces for allolactose, in the exact (left) and in the piecewise linear model (right). All units are μM .

Figure 3 shows several upward switching trajectories obtained by gradually increasing Le over a time T at fixed Ge, as well as the corresponding steady state curves. The trajectories turn upward for some value of Le *exceeding* the location of the kink in the steady state curve. If the increase of Le is fast, the upturn occurs for Le values significantly higher than the threshold, so this method is not reliable in predicting bistability and induction.

We expect that the likely 'inducible' region in the Le-Ge plane is the one with only the high steady state, to the right of the curve corresponding to the edge of the fold in the steady state surface (Figure 4). This intuition needs to be verified using calculations of Jacobians and eigenvalues in a six dimensional state-space. Figure 2 only shows one of the relevant dimensions of the system. The regions of attraction of the two steady states can have a complicated geometry, so there is no *mathematical guarantee* that the external lactose and glucose concentrations cannot be manipulated to achieve induction, even without leaving the bi-stable region. Reachability analysis can in fact delineate the region from which an increase in external lactose may lead to induction.

Reachability Analysis

The first step in our procedure is the piecewise multi-affine approximation of the equations of motion (2). The functions that need to be approximated are $\eta([A_i], [cAMP])$ and $\omega([cAMP])$ and the nonlinear terms involving $[A_i]$.

We defined a grid of values for the two variables and computed tables of values. The approximating functions are (bi-)linear interpolations between the tabulated values. The steady states and trajectories for this system closely reproduce the exact ones, as illustrated in Figure 3.

This linearization procedure only partitioned the ranges of two model variables. In order to gain insight from reachability calculations, we need to have partitions that delineate regions of interest for the analysis. The partition we used in the calculations below had $12^4 \times 22 \times 9$ hyper-rectangles. It occupies a hyper-rectangular region spanning from zero to 2-10 times the typical unstable steady state value, for all variables, except for cAMP, where we chose intervals to cover the range of its steady state values for the glucose concentrations we considered.

We performed reachability calculations using this grid, for various values of external glucose and lactose. In each calculation we evaluated the set of all hyper-rectangles reachable from the hyper-rectangle including the steady state value of cAMP for the given external glucose and the lowest ranges for all other variables.

Our partition does not contain the high steady state. If the reached set does not include the boundary of the partition then we can conclude that the high steady state is not reachable from the given initial set of configurations hence induction is forbidden.

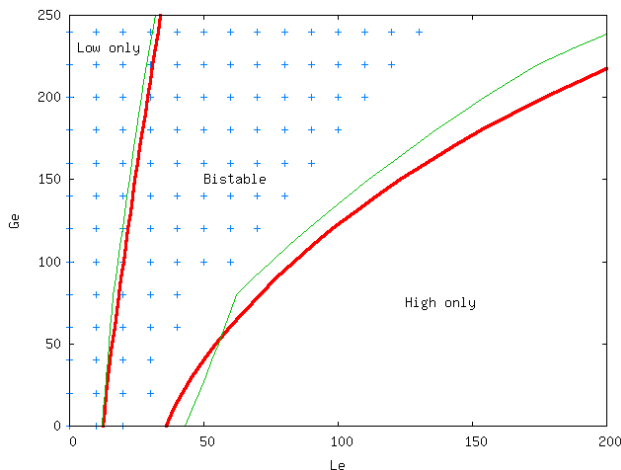


Figure 4: Bistability region in the Le - Ge plane, in the exact model and the piecewise approximation. Superimposed are points where reachability forbids upward switching. All units are μM .

Figure 4 summarizes the reachability results. Points signify values for which induction is not possible according to reachability. The plot also shows the boundary of the bi-stable region. The non-inducible points are inside the bistability region, as expected intuitively.

A similar calculation can be performed with the external lactose and glucose dependence also piecewise approximated. These calculations delineate a set of hyper-rectangles from which it is impossible for the cell to

evolve to a high internal lactose concentration, regardless of the external lactose and glucose concentration.

Conclusions

We applied reachability analysis to a model of the glucose-lactose metabolism of *E.coli* in order to investigate the possibility of induction under different combinations of glucose and lactose availability. The reachability analysis delineates (a) sets of hyper rectangles that lead to the induced state; and (b) sets of hyper rectangles from which it is not possible to evolve to the induced steady state. These results (exact for a piecewise linear approximation of the model) are obtained without performing extensive simulations or a detailed stability analysis.

This example illustrates how reachability analysis can be used in conjunction with other methods. We used traditional techniques for steady states and simulations. However, numerical simulations only provide snapshots of system behavior. Based on these simulations we can formulate global hypotheses which can be verified by reachability calculations.

Our analysis of this system shows that the introduction of glucose into the model of the lactose operon does not significantly change the mathematical picture of lactose-driven induction. Catabolite repression due to external glucose has an effect similar to that of varying the basal lac transcription rate. The bistability property and its relation to inducibility are preserved. In both cases, induction is possible for those parameter sets for which there is only one (high) stable steady state. This finding suggests a direction for future wet lab experiments.

Acknowledgments

This work was supported in part by DARPA IPTO through grant no. AF F30602-01-2-0563. ÁMH is supported by an NIH-NLM Individual Bioinformatics Fellowship.

References

- Belta, C., L. Habetz, and V. Kumar. 2002. Control of multi-affine systems on rectangles with applications to hybrid biomolecular networks. *In 41st IEEE Conference on Decision and Control*. Las Vegas, NV.
- Halász, A. M., V. Kumar, M. Imielinski, O. Sokolsky, P. Finin, C. Belta and H. Rubin. 2004. Analysis of Lactose Metabolism in *E.coli* using Reachability Analysis of Hybrid Systems. *Submitted for publication*.
- Santillán, M., M. C. Mackey (2004). Influence of Catabolite Repression and Inducer Exclusion on the Bistable Behavior of the *lac* Operon. *Biophys. J.*, 86, 1282.
- Yildirim, N., M. C. Mackey (2003). Feedback Regulation in the Lactose Operon: A Mathematical Modeling Study and Comparison with Experimental Data. *Biophys. J.*, 84, 2841.