Analysis of lactose metabolism in E.coli using reachability analysis of hybrid systems

Á. Halász, V. Kumar, M. Imieliński, C. Belta, O. Sokolsky, S. Pathak and H. Rubin

Abstract: We propose an abstraction method for medium-scale biomolecular networks, based on hybrid dynamical systems with continuous multi-affine dynamics. This abstraction method follows naturally from the notion of approximating nonlinear rate laws with continuous piecewise linear functions and can be easily automated. An efficient reachability algorithm is possible for the resulting class of hybrid systems. An efficient reachability algorithm is possible for the resulting class of hybrid systems. An approximation for an ordinary differential equation model of the *lac* operon is constructed, and it is shown that the abstraction passes the same experimental tests as were used to validate the original model. The well studied biological system exhibits bistability and switching behaviour, arising from positive feedback in the expression mechanism of the *lac* operon. The switching property of the *lac* system is an example of the major qualitative features that are the building blocks of higher level, more coarse-grained descriptions. The present approach is useful in helping to correctly identify such properties and in connecting them to the underlying molecular dynamical details. Reachability analysis together with the knowledge of the steady-state structure are used to identify ranges of parameter values for which the system maintains the bistable switching property.

1 Introduction

The functionality of genetic networks is ultimately determined by the kinetics of the underlying biomolecular processes. A full dynamical description of the kinetics at the genome scale would be computationally prohibitive. Another obstacle to a full dynamic description lies in the fact that many of the kinetic parameters are currently inaccessible to direct *in vivo* measurements, and often the *in vitro* kinetics is very different. On the other hand such a detailed description is often unnecessary from a genome-scale perspective, where one is only interested in the high-level behaviour of molecular subsystems.

Consider for instance a subsystem such as the one controlling the induction of the *lac* operon [1]. One would be interested in correctly describing the conditions for switching between the induced and uninduced states while ignoring kinetic details which are irrelevant to this simplified description. In order to be meaningful, the simplified description must retain the dependence of the switching thresholds on the variations of the parameter values.

doi:10.1049/iet-syb:20060035

Paper first received 20th March and in revised form 8th December 2006

Á. Halász, V. Kumar, M. Imieliński, O. Sokolsky and S. Pathak are with the GRASP Laboratory, University of Pennsylvania, 402 Levine Hall, 3330 Walnut St., Philadelphia, PA 19104-6228, USA

C. Belta is with the Center for Information and Systems Engineering, Boston University, 15 St Mary's St., Brookline, MA 02246, USA

H. Rubin is with the School of Medicine, University of Pennsylvania, 522 Johnson Pavilion, 3610 Hamilton Walk, Philadelphia, PA 19104, USA

Á. Halász and V. Kumar are also with the Department of Mechanical Engineering and Applied Mechanics, 229 Towne Building, 220 S. 33rd St., Philadelphia, PA 19104-6315, USA

M. Imieliński is also with the School of Medicine, University of Pennsylvania, 608 Levine Hall, 3330 Walnut St., Philadelphia, PA 19104-6228, USA

O. Sokolsky is also with the Department of Computer and Information Science, 522 Johnson Pavilion, 3610 Hamilton Walk, Philadelphia, PA 19104, USA

E-mail: halasz@grasp.upenn.edu

This example illustrates the task of building consistent abstractions or meta-models that are rooted in the full biochemical and molecular kinetics of the system, but offer a simple description, preferably in terms of discrete states. In pursuing a whole cell or genome-scale model, there are many subsystems which are characterised by fairly simple high-level behaviour and are amenable to this type of description. Our goal is to investigate how simplified descriptions or abstractions can be obtained and/or verified with less analytical work. We also want to address the related issue of unaccessible or poorly known parameter values, by identifying parameter ranges for which a certain high-level behaviour is expected. We hope to be able to offer an approach that can be at least partially automated, leading to what may be called high throughput model validation.

In this paper we propose an approach to biomolecular networks using switched hybrid dynamical systems. Biological systems exhibit both discrete and continuous behaviour. Individual genes are often described as being induced ('on') or uninduced ('off'); each configuration of genes triggers specific continuous dynamics. Discontinuous or switching behaviour also results from features of continuous dynamics, such as multi-stability and bifurcations. Dynamical systems with both continuous evolution and discrete transitions are described mathematically as hybrid systems.

One way of obtaining a hybrid abstraction of a genetic network is to *a priori* identify qualitatively different behaviours or modes, typically by treating genes or operons as discrete switches whose position depends on the values of other state variables, such as concentrations of various substances. As discussed above, such a model for the *lac* operon would have a single discrete state variable, which would switch between the induced and uninduced states based on critical values of the external lactose concentration. More complicated susbystems, of the genetic network can be then modelled combining such simple abstractions. This path is taken in most of the biological

applications of hybrid dynamical systems to date, notably [2–4].

In this paper, we propose a method to address the connection between high-level behaviour such as switching and the underlying detailed dynamics. We use a specific type of hybrid system to build a faithful approximation of a continuous model of the *lac* operon, closely reproducing its full dynamics[5]. We then use reachability analysis on the hybrid abstraction to investigate the parameter dependence of the switching behaviour.

1.1 Hybrid systems and simplified models for biomolecular networks

The traditional approaches to modelling genetic networks lead to highly nonlinear systems of differential equations for which analytical solutions are not normally possible. One way to work around the difficulties of the nonlinearities is to use simplified, approximate models. Decoupled piecewise linear differential equations (PLDE) are considered in [6–8], where gene regulation is modelled as a discontinuous step function and chemical reactions are ignored. An even more radical idealisation is obtained if the state of a gene is abstracted to a Boolean variable and the interaction among elements to Boolean functions [9]. Other types of simplified approaches combining logical and continuous aspects include generalised logical formalisms [10] and qualitative differential equations [11]. The highest level of abstraction is achieved in the knowledge-based, or rule-based formalism [12]. While amenable for interesting analysis, the methods mentioned above are based on assumptions that cannot accommodate biochemical phenomena where continuous variations of concentrations or gene expression levels are important.

The modelling approach based on differential equations implicitly assumes that the concentrations of the species in the network vary continuously and deterministically. The continuity assumption is compromised in cases when the number of molecules of a certain species is small or due to fluctuations in the timing of cellular events. Methods for he simulation of discrete stochastic models have been developed [13–15]. However, these methods are typically computationally more expensive than ordinary differential equation (ODE) simulations. In constructing our method we will assume that the continuity assumption holds for all molecular species of interest.

In this paper, we model genetic networks as hybrid systems, which can be thought of as collections of traditional dynamical subsystems, each with its own continuous dynamics. The dynamics of each component subsystem define a made or a behaviour of the system. A hybrid system can transition between modes depending on its state (concentrations of different species) and inputs (environmental stimuli), thus exhibiting changes in its continuous dynamics. Hybrid system models arise naturally in systems biology, where continuous time evolution (typically described by differential equations) is often intertwined with discrete transitions, such as the induction of a gene. Hybrid systems are also obtained as a result of a modelling approach when nonlinear rate laws are approximated by several linear regimes, as is the case in this paper (see also [16, 17]). Finally, hybrid systems models are attractive because they provide a global description of a biological system, with local descriptions derived from simple dynamic models of behavious that are easier to analyse [18–20]. Thus, one can build global approximations starting from models of local behaviour in a bottom-up fashion.

The ideas of using hybrid systems to model genetic networks and formal analysis to specify and check their properties are not new. Piecewise affine hybrid systems were used to model lateral inhibition through delta-notch signalling [21], nutritional stress response [4], and initiation of sporulation [3]. Our own work has applied multi-affine hybrid systems modelling to phenomena such as quorum sensing [22] and the stringent response [23]. Stochastic hybrid systems have been used [24–25]. Reachability analysis of hybrid systems has been used to check biological qualitative properties [21, 23, 26]. Even more generally, temporal logic and model checking have been used [4, 27, 28].

As in [23], we too use rectangular multi-affine hybrid systems as models for genetic networks. Such systems are defined in hyper-rectangular regions of the state space and have specific dynamics in each hyper-rectangle. The dynamics are specified in terms of functions which are affine in each scalar variable, hence the term multi-affine. There are two main advantages of using such systems. First, as opposed to all the approaches based on simplified models reviewed above, this class of systems can accommodate chemical reactions of unitary stoichiometry. Secondly, as shown in [29], qualitative questions such as reachability analysis can be answered by searches on finite transition graphs.

1.2 Modelling bistability in the lac system

The *lac* operon (for more background see [1]) is induced in the presence of lactose and inhibited in the presence of glucose. It is a well studied example of autocatalytic gene expression. Two of the three genes in the operon code for enzymes which contribute to the synthesis of allolactose, which in turn acts as an inducer for the operon itself. The phenomenon of all-or-none gene expression was discovered by Novick and Weiner [30]. They studied enzymatic activity under various circumstances and found that individual cells in an isogenic population of *E.coli* are either fully induced or not induced at all, while the total activity of the population exhibited a continuous variation. This behaviour was immediately assigned to bistability resulting from the positive feedback in the expression mechanism.

Mathematical models of gene expression with a positive feedback have been constructed since the early 1960s (see [31] for a review of earlier work). The steady states and their stability for several three to five dimensional models have been worked out, for example in [32] a three-dimensional model is shown to have two stable equilibria separated by the two-dimensional stable manifold of a third steady-state which is a saddle point. Mahaffy and Simeonov [31] considered the effect of time delays due to transcription and translation in a four-dimensional model, and concluded the existence of a Hopf bifurcation. However, this only occurs when the degradation rates of permease and β-galactosidase are precisely equal, which is not the case in reality.

Wong *et al.* [33] have integrated detailed experimental information in a comprehensive model. In addition to the elements of the lactose induction mechanism, they also take into account processes that inhibit induction in the presence of glucose; however, the issue of bistability is not addressed. Using similar, experimentally justified parameter values, Yildirim and Mackey [5] have constructed a five-dimensional model of the induction of the *lac* operon explicitly including time delays. The model was validated by comparing simulations with experimental datasets due to Knorre, [34], Pestka *et al.* [35] and Goodwin [36].

The steady states of the model and their stability are investigated numerically, with the identification of ranges of values of the external lactose concentration where bistability is present. In a subsequent work, Yildirim *et al*. [37] analysed a reduced three-dimensional version of the original model, offering additional analytical insight. Another three-dimensional version, obtained via a quasi-steady-state approximation, was investigated in [38]. The mechanism described in [33], complete with the effect of glucose (with some simplifying assumptions), as well as a statistical physics-based description of the transcription initiation process, were integrated in a six-dimensional model [39]. Here, dynamic simulations were shown to match experimental results and to exhibit a step-like behavior indicative of bistability.

Stochasticity is a very important factor in the phenomenology of the *lac* operon. Bistability in this system was discovered by Novick and Weiner [30] in a stochastic context. It is important to realise that the model of Yildirim and Mackey [5] is calibrated using bulk biochemical data, and should be regarded primarily as a model of the biochemistry of a bacterial population consisting of a very large number of individual cells. Clearly, a definitive model of the induction of the *lac* operon must address the processes on the level of single cells, including an appropriate treatment of stochastic effects due to small numbers of molecules and other sources of natural variability. There has been significant work in this direction, both theoretically [40] and experimentally [41]. In the latter study, the authors showed that the continuous variations of β-galactosidase activity during induction are in fact variations of the relative sizes of an induced and an uninduced sub-population of cells. This confirms the early conclusions of [30] and implies that the timescale of induction on the level of single cells is much faster than the macroscopic induction time, which is in fact the time it takes for the whole cell population to (un)induce.

In the present study we do not explicitly model stochastic phenomena. Instead we focus on the deterministic description of the processes that control the expression of the *lac* operon, as described by the model defined in [5]. We construct an approximate version of this model, based on piecewise affine approximations of the nonlinear rate laws (resulting in multi-affine equations of motion) and the concept of hybrid dynamical systems. We wish to show that the approximate model is a reasonable surrogate for the exact one, by subjecting it to the same experimental validation tests and comparing its steady-state structure with the exact model. We wish to investigate the dependence of the bistability property on the particular values of selected model parameters, as an illustration of a type of analysis that is possible for the approximate model.

2 Mathematical preliminaries

The state of the biochemical system is represented by a state vector $x = [x_1, x_2, \ldots, x_n]^{\mathrm{T}}$ which is an n-dimensional vector of concentrations of all mRNAs, proteins, and other species that are significant descriptors of the behaviour of the system. We assume the system to be well mixed and described by a set of deterministic equations of motion. The nonlinear equations governing the evolution of the system, called the state equations, can be written in the form

$$\dot{x} = f(x) \tag{1}$$

where $f(x)$ is a vector-valued, nonlinear function of the state describing the rate of change of the variables in the system. Typically, the components of $f(x)$ are sums of nonlinear rate functions, as illustrated by (8)–(13).

Assume for now that the nonlinear terms can be factorised into functions of one variable, such as those in (13). We approximate these functions with continuous piecewise affine functions. The approximating functions follow a possibly different affine function (of the general form $h(u) = au + b$ on every interval $[u^{(l)}, u^{(l+1)}]$ generated by a sequence of dividing values of their variable: $\{u^{(0)}, \ldots, u^{(m)}\}$. Note that a product of several functions of different variables will give rise to a product of affine functions in different variables. Sums of such terms will be multi-affine functions of several variables.

Generally, the same state variable x_k will be involved in several nonlinear rate functions, and each of the functions will be approximated using a different set of dividing values. Let $\Delta_k = \{x_k^{(0)}, \ldots, x_k^{(d_k)}\}$ be the (ordered) union of all these sets, where d_k is the number of intervals of variable x_k. Each rate function will be approximated by one affine formula within every interval defined by Δ_k but these formulae will not necessarily differ between adjacent intervals. The collection of all n dividing sequences defines a division of the state space into $D = \prod_{k=1}^{N} d_k$ hyper-rectangles of the form

$$c_{j_i,\ldots,j_n} = [x_1^{(j_1-1,)}, x_1^{(j_1)}] \times \cdots \times [x_n^{(j_n-1,)}, x_n^{(j_n)}]$$
$$1 \le j_k = 1, \ldots, d_k, \quad (\forall) k = 1, \ldots, n \tag{2}$$

where $[x_k^{(j_k-1)}, x_k^{j_k}]$ are the lower and upper limits for the kth state variable in the j_kth interval, as illustrated in Fig. 1. Within each hyper-rectangle, the components of $f(x)$ that define the original system (1) are replaced by multi-affine functions that arise from approximating the one-variable nonlinear rate functions in (1). A (multivariate) function $g(x)$ is multi-affine in x if $g(x_i)$, the function derived by holding the $n-1$ state variables excluding x_i constant, is affine in x_i. Thus a multi-affine function can include product type nonlinearities. Note that we assumed that each term in $f(x)$ is a product of factors that depend each on a different single scalar variable. In fact, due to the interpolation property of multi-affine functions, the construction of piecewise multi-affine approximations to multivariate functions is very similar to constructing piecewise linear approximations to univariate functions.

The choice of continuous piecewise multi-affine functions differentiates our approach [23, 42, 43] from other

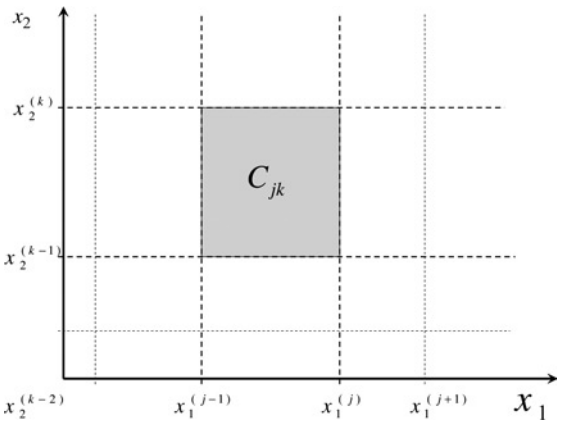

Fig. 1 *State space is partitioned into hyper-rectangular sets (rectangles in two dimensions, rectangular parallelepipeds or cuboids in three dimension), each hypr-rectangle characterised by mutli-affine dynamics*

methods that rely on hybrid abstractions of biomolecular networks [3, 4, 8, 21, 44]. There are two important arguments for the use of multi-affine rather than affine functions. First, multi-affine functions arise naturally both in the exact and approximate description of biomolecular networks. Secondly, in our approximation method they give rise to a class of hybrid systems which have continuous dynamics and allow for an efficient reachability algorithm.

Mass-action kinetics with unitary stoichiometry are readily described by multi-affine rate functions. For example, the rate through the binary reaction

$$A + B \underset{k_{-C}}{\overset{k_c}{\rightleftharpoons}} C \qquad \frac{\mathrm{d}[C]}{\mathrm{d}t} = (k_c[A][B] - k_{-c}[C]) \tag{3}$$

can be seen to be multi-affine in the concentrations of A, B and C. In principle, higher-order mass-action kinetics as well as more complicated enzymatic rate laws can be re-obtained as quasi-steady-state approximations of a larger set of elementary reactions. However, this approach increases the number of system variables and may not be practical.

Enzymatic reaction laws are typically given by rational functions, such as the Michaelis–Menten law describing the rate $R([S],[E])$ of a reaction catalysed by a single enzyme [45]

$$R([S],[E]) = k_{cat}\frac{[E][S]}{K_m + [S]}$$

$$r([S],[E]) = \begin{cases} k_{cat}[E][S]/2K_m & \text{if } [S] \in [0, 2K_m] \\ k_{cat}[E] & \text{if } [S] \in [2K_m, \infty] \end{cases} \tag{4}$$

where $[S]$ and $[E]$ are concentrations of the substrate and enzyme, respectively, and k_{cat} and k_m are constants.The Michaelis–Menten rate law is often approximated by a piecewise linear function similar to $r([S],[E])$ of (4). The approximation is a bilinear function in $[E]$ and $[S]$ on its first interval. This is an example of a simple multi-affine function arising from the linearisation of a nonlinear single-variate function multiplied with another variable. It approaches the exact function asymptotically in the limit of $s/(K_m) \to \infty$. It is a continuous function of both $[S]$ and $[E]$, but is not smooth; its derivative with respect to $[S]$ is discontinuous at $[S] = 2K_m$.

We propose the use of continuous piecewise affine functions to approximate all univariate nonlinear rate functions in a biomolecular network model. Such approximations are straightforward to construct by specifying a set of value pairs $\{(x_k, y_k)\}_{k=0,\ldots,m}$. The approximating function on the kth interval is then given by a simple linear interpolation

$$h(u) = \left(\frac{u - x_{k-1}}{x_k - x_{k-1}}\right)y_k + \left(\frac{x_k - u}{x_k - x_{k-1}}\right)y_{k-1} \tag{5}$$

The set of divisions $\Delta = \{x_0, \ldots, x_m\}$ and the corresponding y values can be chosen to satisfy the precision requirements. For example, if the exact function $H(x)$ is Lipschitz on an interval $I \subset R$, then one can always find a set of divisions $\Delta \subset I$ so that $\min_{x \in I} |H(x) - h(x)| \leq \delta$ where δ is a maximal acceptable deviation. In this paper we choose $y_k = h(x_k)$ and select the divisions so as to ensure a precision of 5% for all approximated function. The approximation method is discussed further in Section 3.2.

The one-dimensional construction can be generalised to the case of multivariate functions which are not expressed as a product of single-variable functions by using the interpolation property of multi-affine functions [42]. A multi-affine function of n real variables is uniquely defined by the values it takes on the vertices of an n-dimensional hyper-rectangle such as those in (2). Furthermore, the values it takes on the inside of the hyper-rectangle are positive linear combinations of the values taken on the vertices. It follows that given an n-dimensional rectangular grid (as defined by a collection of dividing values for each variable) and a prescription of the value of the function on each grid point, we can construct a unique continuous piecewise multi-affine function that matches the set of prescribed values at each grid point.

This approach allows us to model any network of reactions in a high-dimensional state space divided into non-overlapping hyper-rectangles (as shown in Fig.1) with multi-affine state equations in each hyper-rectangle. The resulting system is hybrid [46] because the dynamics of the system change at discrete points in time as the trajectory of the system crosses the boundaries of the hyper-rectangle. Because the state space is divided into non-overlapping cells with a unique set of state equations assigned to each hyper-rectangle, this type of hybrid system is called a switched system. Thanks to the interpolation property of multi-affine functions, our approximations of all rate functions and hence of all components of the velocity function (the vector function $f()$ of (1) can be chosen to be continuous. This property eliminates the need to handle the boundaries between hyper-rectangles as separate modes.

In this paper we are interested in determining which of the hyper-rectangles (which may be regarded as discrete states) are reachable under initial conditions that lie in a given hyper-rectangle. In other words, we consider an entire set of initial conditions as opposed to conducting simulation experiments for many initial conditions in a given set. This is referred to as forward reachability analysis in hybrid systems. We will also be interested in the backward reachability which refers to the determination of the hyper-rectangles (sets of initial conditions) from which trajectories of the state equations lead to a given hyper-rectangle. We can also consider uncertainties in parameter values in this framework by augmenting the n-dimensional state vector by a m-dimensional parameter vector, p. The augmented state vector, $X = [x^{\mathrm{T}}, P^{\mathrm{T}}]^{\mathrm{T}}$, has the state equation

$$\dot{X} = \begin{bmatrix} \dot{x} \\ \dot{p} \end{bmatrix} = \begin{bmatrix} f(x) \\ 0 \end{bmatrix} \tag{6}$$

with the hyper-rectangles now being defined as

$$\begin{aligned} C_{j_1,\ldots j_n} = &\left[x_1^{(j_1-1)}, x_1^{j_1}\right] \times \cdots \left[x_n^{(j_n-1)}, x_n^{(j_n)}\right] \\ &\times \left[p_1^{(l)}, p_1^{(u)}\right] \times \cdots \left[p_m^{(l)}, p_m^{(u)}\right] \\ &1 \leq j_k = 1, \ldots, d_k, \quad (\forall)k = 1, \ldots, n \end{aligned} \tag{7}$$

where $p_i^{(l)}$ and $p_i^{(u)}$ denote the lower and upper limits for the ith parameter. Note that in this framework each original hyper-rectangle of the type shown in (2) is augmented with the same hyper-rectangle from parameter space. The dynamical conclusions drawn from the model apply to all parameter values inside this additional hyper-rectangle.

We also refer to the collection of all sets of the type defined in (7) as one partition. The collection of all sets of the type defined in (7) does not constitute a proper set partition of the state space since any two adjacent hyper-rectangles share a common facet. However, this is a measure zero set which does not require a separate treatment; the collection of hyper-rectangles is turned into a proper partition by using semi-open intervals in (7). This

choice would have no further consequence for our analysis. The minimal number of sets is dictated by the piecewise linear approximation scheme we choose for nonlinear rate laws. For example, if we use the approximation of (4) with two linear segments, then we have two intervals, $S \in [0, 2\,k_m]$ and $S \in [2\,k_m, \infty]$. In general, the minimal partitions resulting from only these types of approximations will be coarse, with only several divisions along some variables and none for the rest. We may also refine our partition adding additional divisions without further refining our approximation of the rate laws. However, these additional divisions will not correspond to changes in dynamic behaviour. Instead the addition of these divisions serves to refine the partition leading to a finer discrete representation of the state space for reachability analysis.

3 Model

3.1 Yildirim–Mackey model

Our model of the *lac* system is adapted from Yildirim and Mackey [5]. Central to this system is the *lac* operon which consists of three genes, *lacZ, lacY,* and *lacA* [47]. The enzyme β-galactosidase, the product of the *lacZ* gene, cleaves lactose, a necessary first step to metabolise lactose. The *lacY* gene encodes the permease that brings lactose into the cell. Finally, *lacA* encodes an acetylase that is not relevant to our discussion here. In the absence of lactose, the *lac* repressor binds the *lac* operator turning down the transcription of its component genes to a low basal level. However, in the presence of lactose, the decomposition of lactose results in the formation of allolactose, which binds to the *lac* repressor, causing it to undergo an allosteric change, greatly reducing the DNA binding of the repressor. This activates the transcription of the genes to a level that may be as much as 40-fold higher than the basal rate. The presence of glucose decreases the synthesis of cyclic AMP which in turn affects the ability of catabolite activator protein (CAP) to bind DNA and recruit RNA polymerase to the promoter. Thus, in the presence of glucose, this increase is significant, but much less than 40-fold. We do not consider this feedback loop in our system.

The lac model is illustrated in Fig. 2. The main species are the two genes, *lacZ* and *lacy*, their products β-galactosidase and permease, the lactose and allolactose inside the cell, and external lactose that is outside the cell. We denote the concentrations of the different species as follows: M is the concentration *mRNA* transcribed from both genes, while B and P denote the β-galactosidase and permease. The external lactose is L_e, while the intra-cellular lactose is denoted by L. Finally, the concentration of allolactose is denoted by A. The state of the system is given by

$$x = [M, B, A, L, P]^{\mathrm{T}}$$

while L_e is viewed as an input to the system that is controlled outside the cell. The differential equations describing

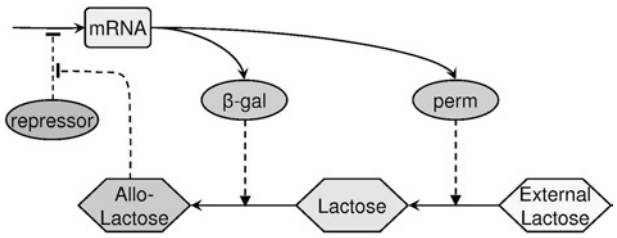

Fig. 2 *Lac model network*

the evolution of the system can be written as follows

$$\frac{\mathrm{d}M}{\mathrm{d}t} = \alpha_M f_1(\mathrm{e}^{-\mu\tau_M} A_{\tau_M}) + \Gamma_0 - \tilde{\gamma}_M M \tag{8}$$

$$\frac{\mathrm{d}B}{\mathrm{d}t} = \alpha_B \mathrm{e}^{-\mu\tau_B} M_{\tau_B} - \tilde{\gamma}_B B \tag{9}$$

$$\frac{\mathrm{d}A}{\mathrm{d}t} = \alpha_A B_{g1}(L) - \beta_A B f_2(A) - \tilde{\gamma}_A A \tag{10}$$

$$\frac{\mathrm{d}L}{\mathrm{d}t} = \alpha_L P h(L_e) - \beta_L P g_2(L) - \alpha_A B_{g1}(L) - \tilde{\gamma}_L L \tag{11}$$

$$\frac{\mathrm{d}P}{\mathrm{d}t} = \alpha_P \mathrm{e}^{-\mu(\tau_P+\tau_B)} M_{\tau_P+\tau_B} - \tilde{\gamma}_P P \tag{12}$$

where $f_1(A)$ is a function that captures the effector–repressor dynamics of the transcription enhanced by allolactose,

$$f_1(A) = \frac{1 + K_1 A^2}{K + K_1 A^2}$$

and f_2, g_1, g_2 and h are rates of irreversible Michaelis–Menten reactions

$$f_2(A) = \frac{A}{K_A + A}, \quad g_1(L) = \frac{L}{K_L + L}$$
$$g_2(L) + \frac{L}{K_{L1} + L}, \quad h(L_e) = \frac{L_e}{K_{Le} + L_e} \tag{13}$$

As in [5], α and β denote rate constants for the different processes, the $\tilde{\gamma}$ are coefficients for terms representing decay of species, and τ denotes delays associated with the finite time required to complete transcription (τ_M) and translation (τ_P and τ_B). For example, if τ_M is the finite time required for the transcription process, A_{τ_M} denotes the concentration of A_{τ_M} seconds prior to the current time. Γ_0 is a fixed basal rate of *mRNA* transcription.

The significance of the terms in the model equations,(8)–(12) is as follows. The term containing α_M in (8) accounts for the *mRNA* transcription rate controlled by allolactose. In the argument of the rate function f_1, A_{τ_M} is the concentration of allolactose at time $t - \tau_M$, to take into account the time delay due to transcription. The exponential in the argument is due to growth-driven dilution during that time. The τ_0 term accounts for the basal rate of *mRNA* transcription. The term containing α_B in (9) and the one containing α_P in (12) for the synthesis of β-galactosidase and permease, respectively. Here we take into account the translation times by introducing delays of τ_B for the transcription of *lacZ*, and $\tau_B + \tau_P$ for the transcription of *lacy* which is downstream from *lacZ*. The term containing α_A accounts for the decomposition of lactose into allolactose, facilitated by β-galactosidase. The term containing β_A accounts for the conversion of allolactose to glucose and galactose which also takes place via β-galactosidase. Similarly, the term containing α_L describes the inbound transport of lactose from outside the cell, facilitated by permease, while the term containing β_L describes the reverse transport process facilitated by the same enzyme. Finally, the term $\tilde{\gamma}_S S (S \in \{M, B, A, L, P\})$ denotes the net effect of the degradation of the species S and its dilution due the growth rate. Thus, $\tilde{\gamma}_S = \mu + \gamma_S$, where γ_S is the half-life of S.

Before proceeding further it is worth recalling the main results of an analysis of the model proposed in [5]. The lactose model exhibits three steady states (equilibrium points) for a range of values of external lactose. One of these three steady states is unstable, while the other two

correspond to the non-induced and induced steady states. This result is stated both in [5] and [37] (the reduced model has a similar steady-state structure and characteristic equation). In [5] there is a plot of an S-shaped steady-state curve, identical to the one in Fig. 3. The stability of the corresponding steady states is tested by ODE simulations. A more complete stability analysis is performed in [37], but for a three-dimensional reduced version of the present model.

Indeed it is straightforward to solve for the steady state for A, M, B, P, and L symbolically, calculate the Jacobian by linearising the state equations at the steady state, and check the eigenvalues of the Jacobian matrix for stability. For the parameters in Table 1, this check allows us to conclude that the system has one steady state for $L_e > 0.062$ and $L_e < 0.02777$ and three for L_e between these two values. The non-induced states corresponds to low intracellular concentrations of the five substances of interest, corresponding to a low or basal level of lactose metabolism. By contrast, the induced steady states are characterised by higher values of all of internal lactose, allolactose, mRNA, permease and β-galactosidase. These states correspond to a state of active lactose metabolism.

3.2 Hybrid model

Thus far, our model is identical to the one in [5]. We now introduce two approximations that will reduce this model to a simpler, hybrid system and allow us to use a computational technique to investigate reachability.

Because of the time-delayed terms the state of the system is determined by the present values of the variables along with their time history going back in time to the longest delay time, rendering it infinite-dimensional. This makes symbolic analysis difficult. On the other hand, the time delays associated with transcription and translation in prokaryotes are relatively small. The time delays cannot affect the location of the steady states. However, they may modify their regions of attraction and even change their stability properties. There is some indication [37] that this is the case for the present model, but the effect is likely to be small.

We found the dynamical effects of the time delays to be relatively small in our simulations (see Fig. 4), with the discrepancies between the traces for the original model with time delays never differing from the traces of the model

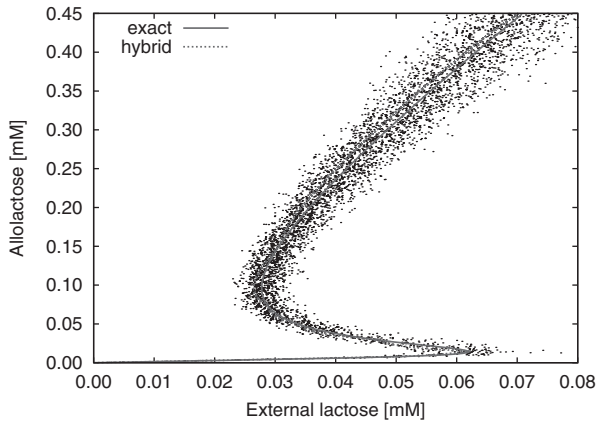

Fig. 3 *Steady-state values of allolactose as a function of external lactose, for the original continuous model from [5] and for the hybrid model in (14)*

Black dots indicate a scatter plot of steady states for parameter sets with values modified randomly by up to 5%

Table 1: Values of constants used in model

γ_M	0.411 min^{-1}	γ_B	8.33×10^{-4} min^{-1}
γ_A	0.52 min^{-1}	Γ_0	7.25×10^{-7} mM/min
K	7.2×10^3	α_M	9.97×10^{-4} mM/min
τ_B	2.0 min	α_A	1.76×10^4 min^{-1}
K_{L1}	1.81 mM	α_B	1.66×10^{-2} min^{-1}
K_A	1.95 mM	β_A	2.15×10^4 min^{-1}
τ_M	0.1 min	K_L	0.97 mM
γ_L	0.0 min^{-1}	γ_P	0.65 min^{-1}
α_L	2.88×10^3 min^{-1}	γ_P	10.0 min^{-1}
τ_P	0.83 min	β_L	2.65×10^3 min^{-1}
K_{Le}	0.26 mM	K_1	2.52×10^4 mM^{-2}
μ	0.0226 min^{-1}		

without time delays by more than 5%. We will not consider the time delays any further and we will confine ourselves to building a hybrid model based on the Yildirim–Mackey model with no (functional) time delays. We retain the exponential factors due to the time delays whose inclusion reduces the error due to ignoring the time delays and does not imply technical difficulties.

There are four instances of nonlinear functions of state variables in the continuous *lac* model, given by the functions in (13). We will approximate each of them with piecewise affine functions and ignore the time-delays. We start by partitioning the variables $[A]$ and $[L]$ into a set of non-overlapping interals. On a given interval, each rate function is approximated by linear interpolation between the exact function values at the ends of the interval, corresponding to (5) with $y_k = h(x_k)$.

The dividing values along with the interpolation procedure uniquely define the piecewise approximation of the nonlinear rate laws. While the construction here was performed 'manually', in automated applications this issue needs to be treated as a search or optimisation problem. The manual procedure we used here can be summarised as follows: given the rate function, a finite interval of interest, and a criterion for acceptable difference; (1) construct the fist linear approximation using the ends of the interval; (2) locate the points where the difference between the approximation and the exact function are locally maximal; (3) add new division points at the approximate location of the maxima which exceed the acceptable difference; (4) construct the piecewise linear approximation using the expanded set of dividing points; (5) repeat until the acceptable difference is never exceeded. This rough sketch of an algorithm has to be refined to take into account two issues. First the 'exact' rate laws themselves may not be continuous up to their second derivative. Secondly, when several rate functions depend on the same variable, the reunion of division points obtained for two different functions may not be the most economical. One may find sets of fewer dividing points that provide a satisfactory approximation for both functions simultaneously. Finally, the issue of convergence or algorithm termination may be non-trivial. Therefore we refrain from putting forward a general algorithmic solution to this problem, which can be framed as a search or an optimisation problem in the space of dividing values and will be addressed in future, more technical contributions.

The exact and approximate versions of the two rate laws that depend on A are plotted in Fig. 5. Our starting point was the Yildirim–Mackey model without time delays. We partitioned A into a total of ten, and L into six non-overlapping intervals. We constructed piecewise linear approximations

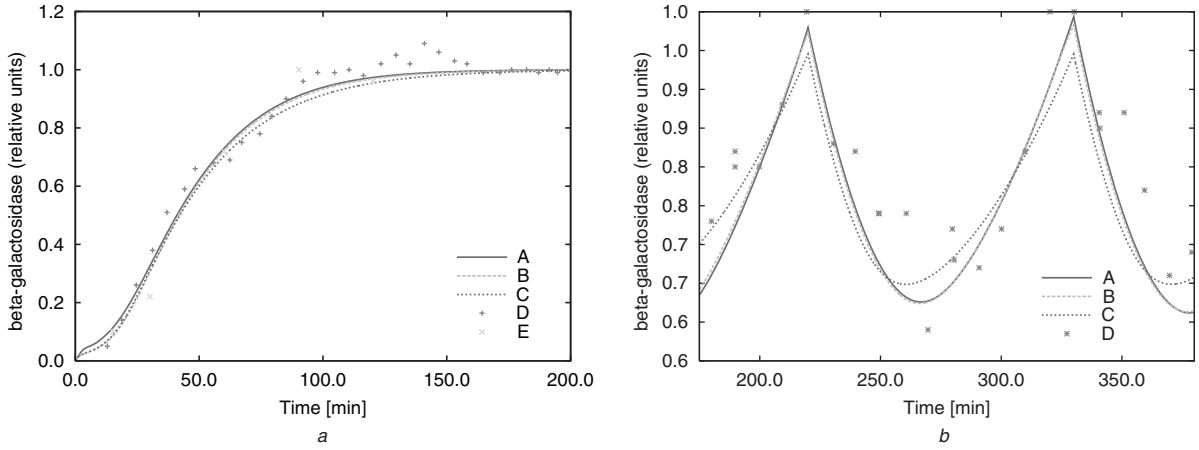

Fig. 4 *Time evolution of beta-galactosidase during diauxic shift fir $L_e = 8.0$ mM*

a Initial condition close to the unstable steady state
b During periodic phosphate feeding
A Time traces obtained with the original, continuous model defined in [5]
B Time tracks obtained in continuous model without times delays
C Results from hybrid model

to the four nonlinear functions in (13) using dividing points chosen to provide an error of less than 5% simultaneously for both functions that depend on each of the variables A and L. The locations of the dividing values $[A] \in \{0.008, 0.015, 0.03, 0.05, 0.075, 0.15, 0.5, 1.0, 2.0, 5.0\}$ mM are indicated by vertical dotted lines. The dividing values used for L are $[L] \in \{0.15, 0.3, 0.7, 1.0, 2.0, 5.0\}$ mM. These piecewise linear functions reproduce the steady-state structure with reasonable accuracy over a range of concentrations of interest (corresponding to $[A] \in [0, 0.4]$ mM and $[L] \in [0, 0.25]$ mM).

We now summarise our piecewise multi-affine, hybrid model. We define the state vector

$$x = [x_1, x_2, x_3, x_4, x_5]^{\mathrm{T}} = [M, B, A, L, P]^{\mathrm{T}}$$

The state equations are

$$\begin{aligned}
\frac{\mathrm{d}x_1}{\mathrm{d}t} &= \alpha_M f_1^{(1)} x_3 - \bar{\gamma}_M x_1 + \alpha_M f_1^{(0)} + \gamma_0 \\
\frac{\mathrm{d}x_2}{\mathrm{d}t} &= \alpha_B \mathrm{e}^{-\mu\tau_B} x_1 - \bar{\gamma}_B x_2 \\
\frac{\mathrm{d}x_3}{\mathrm{d}t} &= \alpha_A g_1^{(1)} x_2 x_4 - \beta_A f_2^{(1)} x_x x_3 - \bar{\gamma}_A x_3 \\
&\quad + (\alpha_A g_1^{(0)} - \beta_A f_2^{(0)}) x_2 \\
\frac{\mathrm{d}x_4}{\mathrm{d}t} &= -\beta_L g_2^{(1)} x_5 x_4 - \alpha_A g_1^{(1)} x_2 x_4 - (\alpha_L h(L_e) \\
&\quad - \beta_L g_2^{(0)}) x_5 - \alpha_A g_1^{(0)} x_2 - \bar{\gamma}_L x_4 \\
\frac{\mathrm{d}x_5}{\mathrm{d}t} &= \alpha_p \mathrm{e}^{-\mu(\tau_B + \tau_P)} x_1 - \bar{\gamma}_P x_5
\end{aligned} \tag{14}$$

Here the constants $g_1^{(1)}, g_1^{(0)}, \cdots$, are the slopes and intercepts from the linear approximations of the respective functions $g_1, f_1, \ldots$. Their values depend on the specific combination of intervals or hyper-rectangle in which the state vector is found and can be readily derived from the two sets of dividing values of the variable x_3 (A) and x_4 (L), respectively. Note that the partitions listed in the Appendix section include further divisions that are added for reachability analysis as discussed later.

3.3 Comparison between the exact model, its hybrid approximation and experimental results

To assess whether our approximation is acceptable, we need to answer the following four questions. (1) Are the qualitative predictions of the exact model reproduced by the hybrid one? (2) Do the quantitative predictions of the hybrid model compare to experimental results the same way those of the exact model do? (3) Is the hybrid description of the underlying molecular processes within the precision with which these processes are characterised? (4) Does the hybrid model introduce any features that are not present in the exact model?

Before addressing these questions specifically, it is worth recalling that this approximation is easily controlled and can be refined to any desired precision by including more intervals. It is possible to optimise the approximation to keep the difference between the exact and approximating function to specifications. In the present case, since we know the exact steady states we can also check that the hybrid model reproduces the steady-state structure within specifications. In cases when this is not possible, it is definitely feasible to employ a significant set of numerical simulations of the two dynamic models to ensure that the simulation traces are sufficiently similar. Indeed, it is possible to derive Lipschitz constants for the hybrid approximation that can be used to establish analytical guarantees on the closeness of the approximation.

3.3.1 Qualitative predictions: As seen in Fig. 5, the general shape of the steady-state curve is reproduced by the hybrid model. We performed the same stability calculations on the hybrid model as we did on the exact one without time delays. We calculated the eigenvalues of the Jacobian for a number of equally spaced values of $0 < A < 0.5$ mM and found the same pattern of stability: the states on the positive sloping branches are stable and those on the middle, negative sloping branch are unstable.

3.3.2 Quantitative comparison with experimental datasets: In Fig.4 we show simulation plots for the evolution of allolactose and *mRNA* as a function of time that shows the effects of first, ignoring the time delays; and secondly the hybrid approximation. The simulations reproduce

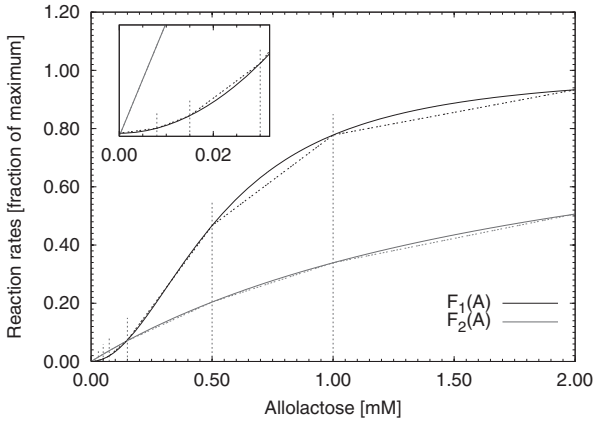

Fig. 5 *Piecewise linear approximation of the two rate laws depend on [A]*

those described in [5]. They are plotted alongside the same three experimental datasets, used in [5] to validate the exact model. The points in Fig.4*a*, correspond to the results of [34] and [35], of time evolution of β-galactosidase at $L_e = 8.0$ mM, during a diauxic shift from glucose to lactose [34]respectively, induction with the lactose analog IPTG [35]. In Fig.4*b*, we attempted to reproduce the effect of periodic phosphate feeding according to [36] Based on the information in [5] we were unable to perfectly reproduce the calculated curve (corresponding to the exact Yildirim–Mackey model. complete with time delays) presented in Fig. 4. of that paper. From the location of the maxima (at approximately 220, respectively 330 min) we assumed that the period T of simulated feeding which drives the growth rate: $\mu(t) = \bar{\mu} - \alpha \bmod (t, T)$ was in fact 110 min rather than 100 as stated. Correspondingly, we used $\alpha = 1/110$ min^{-1} for the constant to keep the growth rate from becoming negative. The fit to the experimental data is only marginally worse in the non-time-delay and the hybrid case than in that of the full Yildirim–Mackey model.

3.3.3 Precision in describing individual molecular processes: A test of the validity of our approximation of the molecular reaction rate functions is to compare the discrepancy between the hybrid model and the exact one with the variations that would result from small changes in the model parameters. The model parameters are mostly kinetic constants. Their values were inferred based on experiments, whose precision is typically worse than 5%. Fig. 5. we compare is the steady states from the original continuous model with the approximate hybrid model. We generated parameter sets where each parameter value was modified with a random value of up to 5%. For each parameter set we calculated one steady-state point and in the right-hand panel of the figure we superimposed it on the two curves (the exact one and the hybrid one). By comparing the extent of the cloud to the difference between the two curves it is clear that the steady states of the hybrid model are closer to those of the exact model than its expected deviation under a 5% uncorrelated change of all parameters. Hence, a prediction made with the approximate model is closer to the prediction of the exact model than the expected deviation due to a very reasonable parametric uncertainty.

3.3.4 Absence of spurious features: Our piecewise approximation closely reproduces the steady-state structure of the exact model, including stability properties. This is to be expected for a close enough approximation of the rate laws with piecewise functions. However, there is no guarantee that stability properties or even steady states are not lost or that spurious steady states, without a correspondent in the exact model, are not introduced by the approximation. The Presence of the correct steady-state structure must be verified on a case-by-case basis when applying this method.

However, other potentially damaging artifacts are guaranteed to be excluded. Consider the calculation of steady-states. They are solutions of the equations of motion (8)–(12) with the time derivatives set to zero, as discussed in the Appendix Section 9.1. Ultimately one obtains an equation that uniquely determines the steady-state values for four of the state variables and L_e, given the value of one state variable, A. In the piecewise approximation we replace the functional dependence with a different one in each mode (partition or hyper-rectangle). The steady-state equation using the approximation may now have a different solution for each mode. How do we end up with the same number of solutions as in the original model?. The simple answer to this is that the steady state of the functional dependence of a mode is likely to be outside its support, unless it corresponds to a steady state of the exact model. To illustrate with a one-dimensional example, if we approximate a nonlinear function with n linear pieces, we introduce n potential zeros, since each linear expression will likely have one zero. Of course, this is not the case, since these zero crossings occur outside the interval where the respective linear approximation is used.

A more subtle problem is that of situations which occur when the flow lines from both sides of a mode switch are directed towards the delimiting surface. The system would be 'stuck' on the dividing surface, or on the intersection of several such surfaces, without actually being kept there by the dynamics on either side. Such situations cannot occur if the functions describing the dynamics are continuous, as is the case in our approach.

4 Reachability analysis

4.1 Method

Consider a hybrid system with non-overlapping rectangular partitions as shown in Fig. 1. When the dimension $n = 3$, these rectangles become rectangular parallelepipeds, each face obtained by holding one variable constant and varying the other two variables within specified lower and upper bounds. When $n > 3$, the faces are replaced by hyperplanes and the parallelepiped becomes a hyper-rectangle. We are interested in the reachability properties of these hyper-rectangles. We further assume that the function $f()$ defining the dynamics $\dot{x} = f(x)$ is multi-affine within each hyper-rectangle and is continuous everywhere some of these restrictions may be relaxed. The definitions and results are presented here for completeness and are discussed in more detail in [29, 23].

Definition 1 (Reachability): For two hyper-rectangles A and B, we say that hyper-rectangle B is reachable from hyper-rectangle A if there exists an initial condition in hyper-rectangle A such that a trajectory of the system with this initial condition enters hyper-rectangle B.

According to this definition, if B is not reachable from A then there exist no points in A from which trajectories can evolve into B. However, if there exists even one point in A from which the trajectory of the state equations leads to B, we say B is reachable from A. An important observation is that reachability is not transitive. If B is reachable from A

and C is reachable from B, it does not imply that C is also reachable from A.

The reachability relation between two hyper-rectangles in a partition of the states of a biomolecular system is equivalent to knowing the outcomes of an infinite number of dynamical simulations. In the general case that is precisely what one would need to perform in order to find the exact answer. On the other hand, reachability between two adjacent hyper-rectangles can be checked more easily, by focusing on the direction of the velocity vector on the common facet of the two hyper-rectangles.

Definition 2 (Direct reachability): For two hyper-rectangles A and B, we say that B is directly reachable from A if: (1) A and B are adjacent (they have a common facet); and (2) there exists a system trajectory originating in A which enters B, and which is contained in the reunion A ∪ B.

We will use a weaker notion of reachability which can be traced back to the reachability of pairs of adjacent hyper-rectangles and transitivity.

Definition 3 (Weak reachability): For two hyper-rectangles A and B, we say that B is weakly reachable from A if at least one of the following is true: (1) B is reachable from A(2) There exists a hyper-rectangle C so that C is weakly reachable from *A* and *B* is weakly reachable from C.

It is clear that weak reachability as defined above is an extension of reachability to chains of hyper-rectangles where the intermediates are reachable from one another. A more practical characterisation of weak reachability is obtained by restricting the first clause of the definition to direct reachability.

Proposition 1 (Weak reachability based on pairs of direct reachable of hyper-rectangles): For two hyper-rectangles A and B, B is weakly reachable from A if and only if at least one of the following is true: (1) B is directly reachable from A and B; and (2) There exists a hyper-rectangle C such that B is weakly reachable from C and C is weakly reachable from A.

This proposition means that weak reachability between two hyper-rectangles always implies the existence of a continuous chain of directly reachable intermediate hyper-rectangles. In order to prove it we need to show that if a hyper-rectangle B is reachable from A then there is a chain of direct reachable pairs of hyper-rectangles connecting them. We illustrate the idea of the proof in Fig. 6. and refer to [29, 23] for a detailed discussion. If B is reachable from A there must exist a trajectory *t* which originates in a point inside A and reaches a point inside B. The set of hyper-rectangles intersected by the trajectory form a chain of adjacent modes which are directly reachable from one another: I is directly reachable from A, J from I and so on. By transitivity (the second clause of the definition of weak reachability) it follows from here that J is weakly reachable from A and so on to B.

4.1.1 Basis for the reachability algorithm: We may summarise the above results as follows. Weak reachability, as described in proposition 1 is a necessary condition for reachability, and weak reachability between two hyper-rectangles implies the existence of a chain of direct reachable hyper-rectangles connecting them. Our reachability procedure described below checks for a local property,

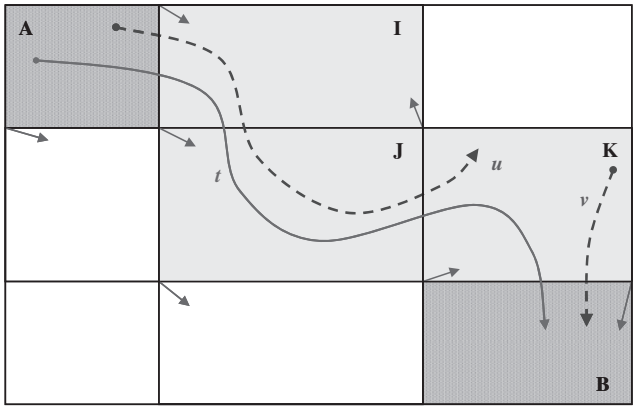

Fig. 6 *Reachability against weak reachability through pairs of adjacent direct reachable hyper-rectangles.*

Necessity: by definition, hyper-rectangle B is reachable from hyper-rectangle A of and only if there is a trajectory, such as t, which originates in A and traverses B. The existence of this trajectory implies the existence of a sequence of rectangle (AIJKB) with the following property. The second one in each ordered pair (AI,IJ,JK,KB) is direct reachable from the first one. Direct reachability between two adjacent hyper-rectangles can be verified efficiently by checking the signs of the relevant components of the velocities on thevertices
Sufficiency: the dashed trajectories show that weak reachability does not imply reachability

direct reachability between adjacent hyper-rectangles. However, an unbroken chain of adjacent reachable hyper-rectangles is a necessary condition for the global property of reachability between any two hyper-rectangles in the system. This way, our analysis procedure can provide global insight using a sequence of local checks of reachability between adjacent hyper-rectangles. Exclusion of weak reachability guarantees exclusion of reachability.

The dashed trajectories in Fig. 6 illustrate that weak reachability is not a sufficient condition for reachability. If some initial condition in A enters K, and some initial condition in K leads to B, there is no guarantee that there exists an initial condition in A that leads to B. Thus, our approach leads to computed reachable sets that are conservative approximations of the exact reachable set. Conservative approximations this theory of conservative approximation has been developed in [48, 49] are known to be safe in the sense that they preclude false negatives. That is, a point not in the computed reachable set does not belong to the exact reachable set. However, the converse is not true. A point that is included in the over-approximation may or may not belong to the exact reachable set.

4.1.2 Checking for direct reachability: In order to determine if a hyper-rectangle is directly reachable from an adjacent hyper-rectangle, we use the fact that the right-hand side of (1) and (6) are multi-affine in x and p respectively. Consider the example in Fig. 7. The arrows represent the velocity vectors at the four vertices of the top facet of a three-dimensional hyper-rectangle (a mode of our switched system). To determine whether there are trajectories that exit or enter this mode through the top facet we only need to establish whether there are points on the facet where the vertical component of the velocity vector is positive (directed upwards) or negative (directed downwards), respectively. Since this component of the velocity vector is a multi-affine function of the co-ordinates, we can obtain its value at C_1 by linear interpolation along x for fixed y between B_{11} and B_{21}. The same is true for the value at C_2, at the point on the opposite side in the y-direction. Finally, the value at D is obtained by linear interpolation along y

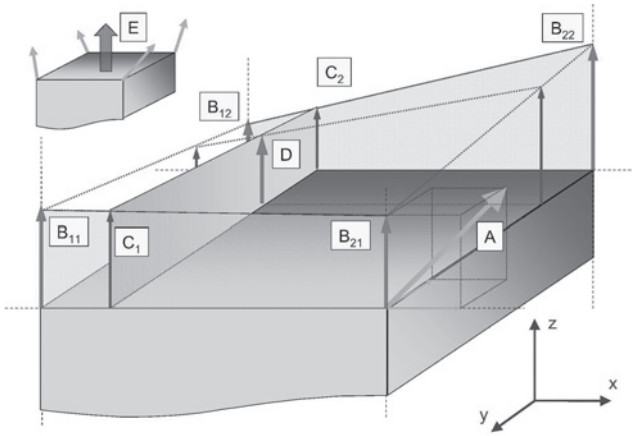

Fig. 7 *Illustration of the determination of direct reachability between adjacent hyper-rectangles using the interpolation property of multi-affine functions*

There is a trajectory starting in the lower hyper-rectangle that enters its upper neighbour if and only if the a z-component of the velocity vector is positive at any point on the common facet. This can be determined by only considering the signs of the velocity vectors on the vertices of this facet. The value of the vertical component D at some point on the facet id a convex combination of the corresponding values at the vertices: the vertical component B_{21} of the vector A together with B_{11} determine the values of the vertical component along the edge; therefore C_1 is a convex combination of C_1 and C_2. Similarly, C_2 is a convex combination of B_{12} and B_{22} and finally D is determined by C_1 and C_2. In this case we conclude that the vertical component is positive everywhere on the facet, therefore all trajectories through this facet point upwards, so direct reachability is true inly in the positive z-direction(E)

with fixed x, between C_1 and C_2. Note that the result of each such interpolation is a positive linear combination of the values at the ends of the respective intervals. It follows that if the z-component values of the velocity at all four vertices are positive (as in Fig. 7.), then the values of the z-component will be positive everywhere on the inside of the facet, implying that all trajectories exit this mode through the top facet. In other words, this mode is not reachable from the mode adjacent to it in the positive z-direction, but the adjacent mode is reachable from this one (inset in Fig. 7). Conversely, if at least one of the z-components on the four vertices were negative, we would have to conclude reachability in the opposite direction as well.

To summarise, a hyper-rectangle (or mode) A is directly reachable from an adjacent mode B with which it shares a common facet if there are trajectories crossing this facet which originate in B and travel into A. For piecewise continuous multi-affine dynamics this is possible if and only if the component of the velocity that is orthogonal to the common facet points towards A on at least one of the vertices of the facet. The notion of the velocity on the common facet is unambiguous in our case, since the velocity function is continuous. Otherwise each of the two modes involved would define a different velocity pattern on the common facet. In n dimensions, checking for reachability between two adjacent modes is hence achieved by examining the signs of 2^{n-1} velocity component values. We can thus establish reachability for all pairs of adjacent hyper-rectangles and construct a reachability graph such as the one illustrated in Fig. 8. Establishment of reachability through transitivity for any two modes is then achieved by a graph search.

We refer the reader to [42, 43, 29] for a more rigorous treatment of the algorithm and its properties. In summary, there are four basic steps to the reachability analysis.

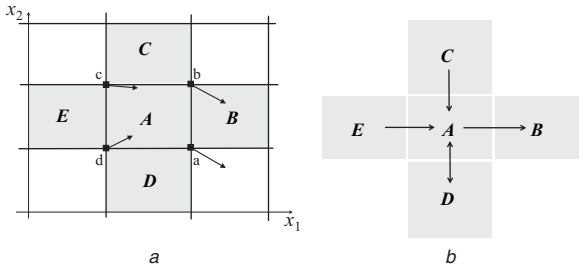

Fig. 8 *Construction of the reachability graph*

a Vector field evaluated at the four vertices (a, b, c, and d) of the rectangle A

b Edges of reachability graph obtained from the four evaluations

1. Partition the n-dimensional state space into m hyper-rectangles as necessary to approximate nonlinear (non-multi-affine) reactions. If necessary introduce additional hyper-planes to improve the granularity of the partition.
2. Evaluate the vector field at each vertex. Each hyper-rectangle has 2^n vertices.
3. Use the properties of the multi-affine vector field to infer properties of direct reachability for adjacent hyper-rectangles.
4. Use transitivity to generate a (weak) reachability graph for all the hyper-rectangles in the state space.

This algorithm calculates weak reachability relations between different hyper-rectangles, which is a necessary but not sufficient condition for reachability. The 'reach set' that is obtained is an over-approximation of the actual reach set. This is the case for any reachability algorithm[2, 3, 26, 50]. In the rest of this section and Section 4.2 we will use the terms 'reached', 'reachability', and so on loosely to indicate weak reachability as calculated by our algorithm. The difference between the over-approximation and the exact reachability properties is mentioned throughout the discussion.

There are many questions that can now be asked. For example, is there a subset of hyper-rectangles from which there are no additional reachable hyper-rectangles? The answer to this can provide information about 'sinks' in the state space. Stable steady states behave like sinks with respect to their own region of attraction. Trajectories originating in a point of the region of attraction of a steady state tend asymptotically to that steady state. Using the definition of reachability, the region of attraction is contained in the set of all points from which the steady state is reachable. Another possible question: is there a subset of hyper-rectangles that is not reachable from other hyper-rectangles? We will show that the hyper-rectangles corresponding to the induced state cannot be reached from many plausible initial conditions. Finally, we are always interested in sets of initial conditions that might lead to a specified condition (sets of hyper-rectangles). We illustrate these questions on the *lac* system in the next section.

4.2 Reachability results

4.2.1 Partitions for reachability analysis: The partition defined by the piecewise approximations of the rate functions $f()$ and $g()$ leads to only $6 \times 10 = 60$ different hyper-rectangles. This is too coarse to resolve any significant dynamical detail. We may refine this partition by simply adding further subintervals to the two variables that are already partitioned and by partitioning the

remaining variables into intervals. We use the partition schemes described in the Appendix (Section 9.2). Each one of the five variables is partitioned into a number of intervals, determined by the divider values given. We refer to a specific hyper-rectangle by its 'integer coordinates' which specify the index of each interval that defines it. For example, hyper-rectangle [1, 2, 1, 1, 1] is the Cartesian product of the first interval for x_1 (the lowest one), the second interval for x_2, and the first interval for x_3, x_4 and x_5.

The partitions listed in the Appendix (Section 9.2) have been each constructed for a specific purpose, to illuminate a certain aspect of the dynamics of the model under consideration. Partition A is designed to identify rough limits on the range of concentration values that are attainable to the system starting from plausible initial states. Here we chose division sequences that start from a relatively small value, and increase quasi-exponentially to span up to three orders of magnitude. For instance, the values for the three non-divided variables, M, B, and P, were chosen {0.05, 0.1, 0.2, 0.5, 1.0, 2.0, 5.0, 10.0, 20.0, 50.0, 100.0} times a 'scale' that results from the equations of motion such as α_P/γ_P, the maximum production rate of a substance divided by its degradation rate. The role of this 'scale' is to set a typical order of magnitude for the values the respective variables may take. Partition B is constructed using the knowledge of the location of the steady states, with the specific purpose of investigating the possibility of induction starting from an initial state comparable to the low (uninduced) steady state.

4.2.2 Forward reachability: determination of a closed set in state space: The first example illustrates the result of a forward reachability calculation for determining the range of possible concentrations that can be encountered in the system. We construct a partition with intervals spanning several orders of magnitude, Partition A, in the Appendix (Section 9.2). This type of partition does not require detailed a priori knowledge of the dynamics of the system, other than a 'scale' or an order of magnitude of plausible values of each concentration. Such a rough estimate for the highest possible concentration value can be inferred from minimal experimental knowledge, or the ratios of the saturation value of the synthesis rate and the minimal degradation rate of a substance. Because it is impossible to pictorially show a five-dimensional reachable set on paper, we show the projection of this set onto all 10 $(= \binom{5}{2})$ (x_i, x_j) planes. As illustrated in Fig. 9, for an external lactose concentration of $L_e = 0.04$ mM, starting from initial conditions in the [2, 5, 8, 1, 3] hyper-rectangle, which corresponds to the concentration ranges [0.00025, 0.0005] × [0.0002, 0.0004] × [0.05011, 0.07517] × [0.1, 0.2] × [0.004, 0.008] in the five-dimensional state space, the system does not exceed the limiting values of $M_{\max} = 0.0025$, $B_{\max} = 0.002$, $A_{\max} = 2.0$, $L_{\max} = 0.7$ and $P_{\max} = 0.04$. We performed many such reachability calculations originating from a mode inside the integer five-dimensional hyper-rectangle delimited by [1, 1, 1, 1, 1] and [7, 7, 8, 7, 7], and none of them lead to the reachability of a mode outside this region. As a final check, reachability calculations initialised in the hyper-rectangle $\{[M, B, A, L, P] | M \leq 2.5 \times 10^{-3}\text{mM}, B \leq 2.0 \times 10^{-3}$ mM, $A \leq 2.0\text{mM}$, $L \leq 0.7\text{mM}$, $P \leq 4.0 \times 10^{-2}\text{mM}\}$ showed that this set does not have any system trajectories exiting it.

Another example of reachability result is illustrated in Fig. 10. This calculation is part of a sweep of parameter values, where we investigated whether induction was possible starting from states in the (known) region of attraction of the uninduced state. Partition B used in this calculation spans the space between the origin and three times the unstable steady state (at $L_e = 0.04$), to ensure that the uninduced state is included, and that a significant part of the modes are in the region of attraction of the induced state. Thus if the reachability initiated from a mode terminates without exceeding the unstable steady-state, we may conclude that the induced steady-state is not reachable from the respective mode. This is the case in Fig. 10. We find that the steady state is weakly reachable, consistent with the fact that the original mode is inside its region of attraction. However, the reached set does not exceed the index of the unstable state in any direction, and in fact

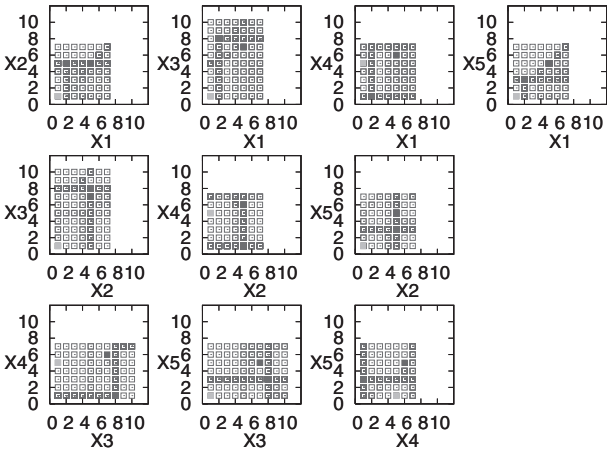

Fig. 9 *Forward reachability for the [2,5,8,1,3] hyper-rectangle using partition A defined in Section 9, with the hyper-rectangles projected onto 10 planes*

Reachable set outlines reasonable ranges for values of five variables
grey rectangles indicate the location of induced states
Mid grey rectangles indicate the initial hyper-rectangle or mode
Pale grey rectangles indicate the hyper-rectangle containing the uninducted steady state
Black rectangles indicate the hyper-rectangle containing the induced steady state

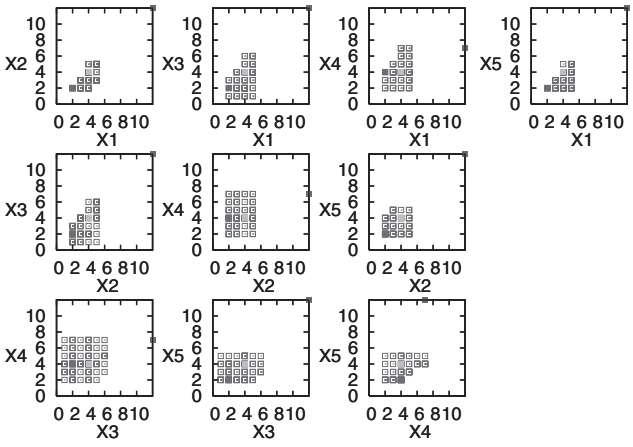

Fig. 10 *Forward reachability for the [2,2,2,4,2] hyper-rectangle (Shown solid),using partition B defined in Section 9, with the hyper-rectangles projected onto 10 planes*

Reachable set outlines reasonable ranges for values of five variables
grey rectangles indicate the location of induced states
Mid grey rectangles indicate the initial hyper-rectangle or mode
Pale grey rectangles indicate the hyper-rectangle containing the uniduced steady state
Black rectangles indicate the hyper-rectangle containing the induced steady state

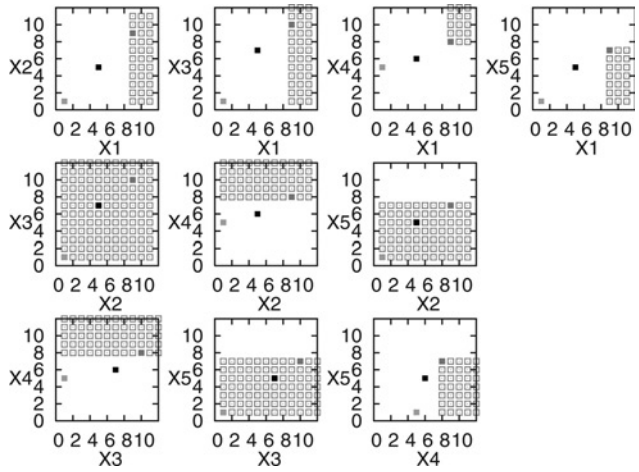

Fig. 11 *Backward reachability for hyper-rectangle [9, 9, 10, 8, 7] in the partition used in Fig. 9*

Large sections of state space are excluded, correctly indicating that this particular hyper-rectangle cannot be reached from most initial conditions
Modes from the closed set identified previously (see text) cannot reach this hyper-rectangle
Mid grey rectangles indicate the initial hyper-rectangle or node
Pale grey rectangles indicate the hyper-rectangle containing the uninduced steady state
Black rectangles indicate the hyper-rectangle containing the induced steady state

does not reach the unstable steady state. We conclude that it cannot reach the induced state, hence induction is not possible from this mode. In Section 5 we perform a parameter sweep where we repeat this reachability calculation for many L_e and Γ_0 values.

4.2.3 Backward reachability: The normal or forward reachability algorithm identifies a conservative approximation of the set of modes which are reachable from a given initial mode or hyper-rectangle. If the direction of all velocities is reversed, this algorithm will provide a similarly conservative approximation of the set of all initial conditions which may reach the mode of interest. Ideally, one would be readily able to identify the region of attraction of a steady state. This is not possible with the current algorithm (see below). However, many modes, typically corresponding to large imbalances of substances (i.e. far away from a steady state) are found to have small backward reachable sets, as illustrated in Fig. 11, on the same partition used in Fig. 9. Only modes outside the closed set can reach this particular mode. This type of result can offer additional insight by further restricting the set of plausible states.

4.2.4 Reachability analysis is conservative: As discussed earlier, reachability analysis may lead to overly conservative results. Indications of this problem can be found most clearly in Fig. 9. The plot indicates that seemingly both steady states are reachable from the initial mode we considered. This is, of course, possible since it does not imply that both steady states are reached starting from the same point. However, under some circumstances one may find even more counter-intuitive results, for instance, that the modes containing the two stable-steady states are reachable from one another. This problem can be partially alleviated by choosing finer partitions, but cannot be completely eliminated.

4.2.5 Comparison with simulations: A comparison of system trajectories and the computed reached set is given in Fig. 12. We show 40 system trajectories that start from the initial set (denoted by thick lines). Only the contours of the reached hyper-rectangles are shown. Fig. 12 illustrates both the conservative nature of the analysis and the fact that in spite of that, useful information can be obtained. This particular reachability calculation takes less than 1 min on the same computer on which obtaining 100 system trajectories took more than 20 min.

The cost of the calculation and the overestimation of the reach set vary depending strongly on the choice of the partition. The length of the reachability calculation is set by the number of hyper-rectangles that are checked. This number may become extremely high (in the millions, in our case), if the majority of the hyper-rectangles in a partition are reached. At the low end, obviously the closed box identified earlier in this section is immediately verifiable, so the whole reachability calculation consists of checking the signs of the velocities on the vertices of one hyper-rectangle. The choice of an appropriate partition is therefore very important for the successful application of this type of algorithm. Useful applications, such as the one illustrated in the next section, should start with the formulation of a question on system dynamics that can be answered using reachability, and followed by the definition of an appropriate partition.

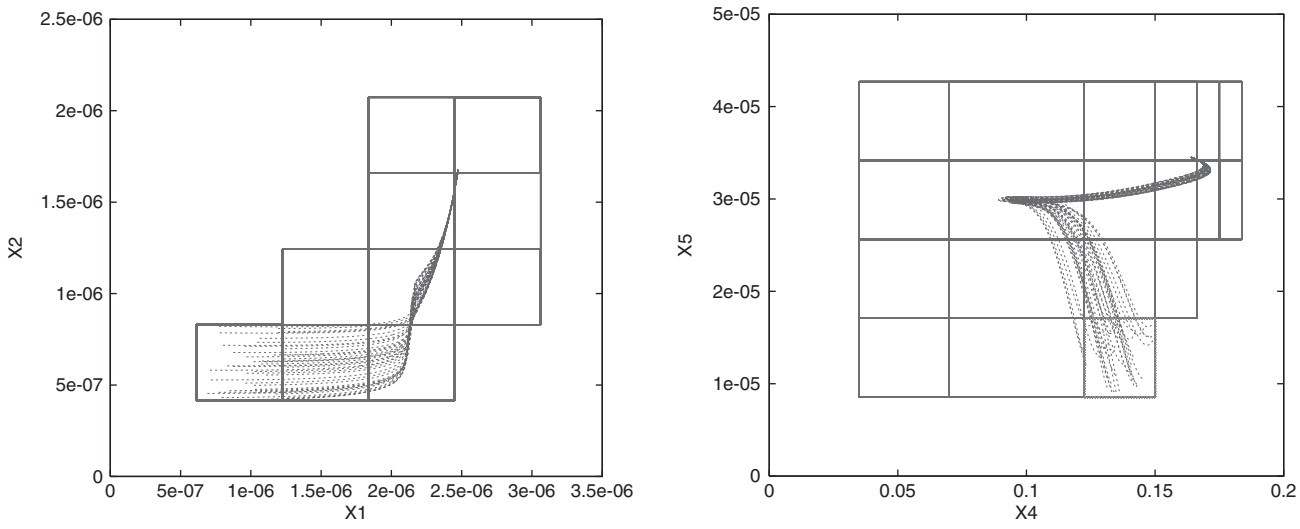

Fig. 12 *Comparison of reachability results and system trajectories for two of the projections of Fig.10 initial hyper-rectangle*

40 system trajectories shown take approximately 8 min on the same computer on which the corresponding reachability calculation takes less than 1 min

This should not be confused with the construction of the hybrid model, where the intervals for piecewise approximations are dictated by the system equations and can easily be constructed automatically. These modes are often too large for analysis purposes, and further partitions are necessary to study the system dynamics. The choice of these partitions is the choice of the user and is discussed here.

5 Effects of parameter variations and robustness

The system defined by (8)–(12) is bi-stable for certain external lactose concentration values, an important result of Yildirim and Mackey [5]. This is illustrated by the S-shaped curve in Fig. 5, which shows the steady-state values of allolactose for given external lactose concentrations. We choose A as a measure of the lactose processing activity. Given a putative steady-state value of A, the steady-state equations corresponding to (8)–(12) uniquely determine all four other variables M,B,L,P, as well as the corresponding level of external lactose.

From Fig. 5 it is easy to see that for a given fixed external lactose concentration we can have either one or three steady states. The negatively sloped portion of the S-shaped curve in the figure contains unstable states. Thus, in practice, there are at most two steady states for a fixed external lactose concentration, one corresponding to a high metabolic activity level (high A) and the other a low-activity level (low A). For very low or very high levels of external lactose, there is only one steady state for A, which is stable.

This specific structure allows the cell to switch between the two steady states to adapt to the changing availability of external lactose. To illustrate this, consider Fig. 13. We recalculated the line of steady states for different values of the basal transcription rate $\Gamma_0' = b\Gamma_0$. Referring to the basal rate factor value of b=1 (the original value in the [5] model), suppose the cell is initially in an environment characterised by $L_e = 0.04$ mM, hence in the steady state denoted by point A. An increase from $L_e = 0.04$ to 0.065 mM will move the system past the 'knee' of the curve to the right, after which the steady-state will move to the upper branch toward the steady state B. Decreasing L_e back to $L_e = 0.04$ mM will move the system to the steady-state C. Thus, it is possible to move the steady-state metabolic level in the cell from a low level to a high level by increasing the external lactose to beyond the level corresponding to the knee of the curve, before reducing it back to the original value.

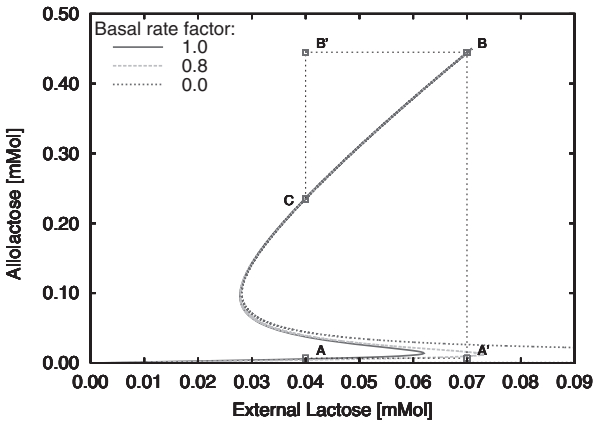

Fig. 13 *Switching functionality provided by the model and of the effect of reducing the mRNA transcription basal rate Γ_0. T_0 spontaneously switch from the lower branch to the upper branch of steady states, the external lactose concentration has to temporarily exceed a threshold value. Below a minimal basal rate value of 0.03 Γ_0 the steady state curve in the L_e-A Plane breaks up into two branches which never meet and the critical L_e becomes singular*

This thought experiment demonstrates the crucial switching property of the *lac* operon and of the model under consideration. In asking how sensitive this property is to parameter fluctuations, it is useful to recall the plot from Fig. 5 in which the steady-state characteristics are reproduced for variations in parameter values of up to 5 % of the nominal values. In fact, variations up to 20% produce results that preserve the overall shape of the curve and the bi-stability property. While this result by no means constitutes a mathematical proof, it does provide evidence that this switching property is robust to changes in parameters. Similar results for other systems have been reported in the literature [51].

However, there is one parameter that is critical to the bi-stability property. Small changes in the basal *mRNA* transcription rate Γ_0 lead to significant changes in the position of the knee of the curve as seen in Fig. 13. Scaling down the basal transcription rate by a factor from 1.0 to 0.0 results in the knee shifting to the right. This upper threshold or the knee is plotted in Fig. 13. As this rate decreases, the level of external lactose required to effect the jump to the upper branch increases dramatically. In fact, the model shows if this rate were reduced below approximately 10% of the original value of Γ_0, the knee effectively moves to infinity. In this case, it appears impossible to effect a switch from a low steady state to the high steady-state. The basal rate of transcription plays the role of a pilot light, keeping a minimal amount of permease and β-galactosidase available even in the total absence of external lactose.

A closer examination of the possibility of induction shows that the above conclusion is not completely straightforward. Indeed, if the basal rate is reduced to the point where the threshold L_e becomes infinite, there will be two stable-steady states, an induced one and an uninduced one, for all external lactose concentrations. This fact in itself does not guarantee that induction is impossible for all external lactose concentrations. To see this, consider the thought experiment illustrated in Fig. 13. We start with the system in an uninduced state (A), and suddenly increase L_e (point A' on the plot). If the basal rate is at its full value (solid curve), the only available steady state is the induced one (B), so the system will evolve towards it. If the basal rate is lowered, a stable uninduced state is available; however, there is no mathematical guarantee that the location of the initial state (A) is inside the region of attraction of the uninduced state for the modified L_e.

It is important to note that the basal rate has not been directly measured. It was introduced in [5] in order to ensure a realistic ratio between the allolactose concentrations in the high and low steady-states. Yet it has a strong influence on the switching property of the model. If we can establish that induction from an uninduced stable steady state is possible only when the external lactose is increased to a value where there is no uninduced steady-state, then we have proved that a finite basal transcription rate is necessary for induction, and we can use the value of the threshold L_e to calculate this minimal rate.

Another parameter that effects this threshold is the growth rate μ. As μ increases, the threshold increases and the knee of the curve shifts to the right. It is possible to calculate $\mu - \Gamma_0$ combinations that result in specified threshold value of L_e. This may be interesting from an experimental point of view as a means of constraining the values of these parameters, along with the observed growth rate and the steady-state allolactose ratios.

5.1 Reachability and the possibility of induction

To clarify the connection between induction and the threshold L_e, we now turn our attention back to the reachability analysis technique to analyse the effect of these parameters. We augment the state space with the parameters of interest as in (6) and (7). Specifically we consider the parameter vector, $p = [\Gamma_0, L_e]^T$. We performed a forward reachability calculation for various combinations of Γ_0 and L_e values, using partition B described in the Appendix. (Section 9.3) For each of the five variables, the divisions extend from zero to three times the value corresponding to the unstable steady-state, starting with smaller intervals close to zero which gradually increase. Typical values are at {0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.4, 0.5, 0.6, 0.7, 1.0, 1.2, 1.5, 3.0} times the value of the respective variable at the unstable steady-state for $L_e = 0.04$ mM. Therefore the upper end of the area covered by this partition contains points that converge to the upper steady-state.

The concentrations value in the high steady-state (for any L_e) are higher than the maximum values in partition B. In order to reach the high steady-state starting from a hyperrectangle inside the partition, the system must pass through the highest value of one of the variables inside the partition (in other words, it must exit the partition in some direction). If the reachability algorithm starting from a mode terminates inside partition B, then we have proved that the high steady-state is not reachable from that mode for the combination of L_e and Γ_0 values under consideration.

In Fig. 14 we plot the line in the $L_e - \Gamma_0$ plane corresponding to the location of the threshold L_e (which would need to be cleared for guaranteed induction). For several combinations of L_e and Γ_0 values, we performed reachability calculations starting from mode [2, 2, 2, 4, 2] in partition B, corresponding to approximately 50% of the concentrations at the uninduced steady-state. We plot those parameter combinations for which the reachability terminated inside the partition with stars, and the others with crosses. As we discussed above, if the reachability algorithm terminates, then induction is not possible. We can see that indeed, the majority of parameter combinations where L_e is below the direct induction threshold are also uninducible in the sense that reachability analysis proves that the induced state is not accessible from the set of initial states under consideration. This result is stronger for initial states closer to the origin and weakens as one approaches the unstable steady-state.

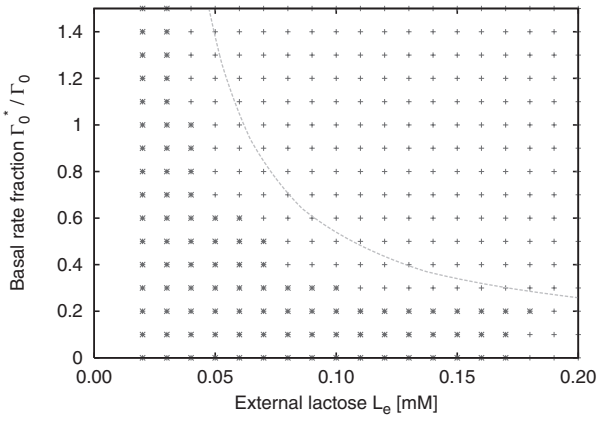

Fig. 14 *Plot in $L_e - \Gamma_0$ plane*

— upper threshold value of external lactose as a function of the basal mRNA transcription rate
*$L_e - \Gamma_0$ combinations for which induction is not possible from an initially uninduced state
+ other parameter combinations

The low results for the low basal rate values for all L_e values show that for these values the system cannot evolve into the high A range no matter how L_e changes. This proves that it is impossible to simply increase L_e to achieve the switch in the allolactose level. Thus a lowered basal rate adversely affects the ability of the cell to induce.

Note that this calculation is equivalent to a reachability calculation performed in the state space augmented with the two parameters of interest. The dependence of Γ_0 is linear. Since $h(L_e)$ is convex. For any x, x_1, $x_2 \geq 0$, if $x \in (x_1, x_2)$ then $h(x) \in (h(x_1), h(x_2))$, even though it is not linear, if a reachability property holds for two values of L_e, then it also holds for any L_e value between them. Finally, it should be remembered that the reachability analysis has been performed on the piecewise linear approximation of the original system, which also exhibits a behaviour very similar to that illustrated in Fig. 13.

6 Conclusions

One important function of quantitative modelling of biochemical networks is the validation of qualitative models proposed to explain observed phenomena such as bistability and hysteresis in the case of the *lac* operon. A question that typically arises is 'is there a parameter set for which a certain observed (qualitative) behaviour results from a given model structure?' In this paper we illustrate the use of a reachability analysis technique on the lactose utilisation network of *E. coli*. The notion of reachability is particularly well suited to address questions of this type, due to the potential to (in)validate whole ranges of possible parameter values. The potential gain over performing many direct ODE simulations is very important, especially since biochemical networks tend to have many elements and many relevant timescales which are not easily disentangled. Hybrid systems are the natural mathematical framework to model the co-existence of smooth dynamics with sharp qualitative changes that are typical in biological systems.

There have been several efforts in this direction [2, 26], implementing different abstractions and reachability algorithms. These methods construct abstractions with discontinuous dynamics, relying on a priori identification of qualitatively different regimes of the system dynamics. By contrast, our approach allows for abstractions that approximate a traditional dynamical description with any desired precision. The class of hybrid systems that are used in our approach are obtained quite naturally from biochemical network models by simply building piecewise linear approximations to their nonlinear rate laws.

6.1 Abstraction and analysis method

Our proposed computational approach consists of first building a piecewise approximation to the nonlinear rate laws of a given model, thus transforming it into a hybrid system with multi-affine modes; and secondly applying algorithms specific to the latter in order to identify the salient properties of the system, such as the location and properties of its steady states as well as reachability issues. While the idea of a global approximation by a collection of well behaved local approximations is widespread successful examples go from spline functions to the finite element method used in predicting the weather, we believe that our choice of using hybrid systems, piecewise linear approximations, and reachability analysis on the resulting multi-affine systems is novel and is well positioned to meet both challenges associated with biochemical networks, namely, parameter uncertainty and high dimensionality.

The computational cost of our method increases exponentially with the number of variables. Performing simulations starting from a set of initial points uniformly covering a region of state space, also leads to exponential scaling. However, simulations are significantly more intensive, since each point corresponds to a complete simulation (at a minimum, $N\,D/\Delta x$ function evaluations, where Δx is the size of a simulation step); in our case, each grid point requires $N \times 2^N$ function evaluations. For the comparison shown in Fig. 11, the reachability calculation takes less than 1 min while 40 system trajectories (a reasonable number that can provide reliable coverage of the whole initial set) take more than 8 min.

Our algorithm, like most other reachability algorithms, has significant limitations, in that under some circumstances it tends to be overly conservative, hence overestimating the reached set. We were able to use our algorithm to identify closed subsets of state space, and identify regions of state space corresponding to grossly imbalanced states which cannot occur. We also showed that reachability analysis can well approximate the combinations of external lactose and basal transcription rate for which induction is possible.

Both the cost and the conservativeness of the method depend strongly on the choice of hyper-rectangles for analysis. At this level of development of the method, useful applications should start by identifying a dynamical question about the system under consideration (such as: is induction always possible? Can a certain response mechanism activate before irreversible damage is done by an external condition?), which can be formulated in terms of reachability. The nature of the question and biological and mathematical insights can then lead to the definition of an analysis partition that can provide significant results at a reasonable cost. The payoff for this effort is that results showing no reachability connection between modes of interest can then be taken as proof that certain outcomes are not possible.

Further refinements are needed for this reachability algorithm. The direction of interest is somewhat different from the traditional engineering approach: in the case of biological models identifying the typical or most likely behaviours is more important than rigorous safety verification. This is true especially for the type of mechanism validation problems described above. Automated model analysis is a desirable goal and our approach could provide a useful element in the set of tools that will make that possible. For now, the most rational approach is to use this and other advanced methods in conjunction with traditional model analysis, including direct ODE simulations and symbolic manipulation.

There has been recent progress on reachability analysis of linear systems, including piecewise linear switched systems. Reachability analysis in a linear system is greatly simplified by the fact that a time evolution under linear dynamics is a convex mapping, that is, the interior of a line segment is always mapped to a line segment. For a linear system with no perturbations and an initial set defined as the convex hull of a set of points, the reached set is simply the reunion of the convex hulls of the locations of these points as they evolve under the linear dynamics [50]. The linear systems approach also requires the construction of a piecewise linear replacement to the original system [52]. This line of investigation is complementary to the one presented here. The dynamics of multi-affine (sometimes inexactly called, multi-linear) systems is more complicated, but the procedure of approximating the original biological model with a piecewise continuous hybrid system is more straightforward.

In general, the equations describing biomolecular networks are neither linear nor multi-linear or multi-affine. While parameter uncertainty and lack of dynamical knowledge make them prime candidates for reachability analysis, these models in their original form are not readily amenable to current methods of reachability analysis. Until the advent of general-purpose reachability algorithms, such nonlinear models have to be replaced by simpler approximations.

Our method of approximation is generally applicable to nonlinear rate laws and can achieve any desired precision in approximating them. We did not provide an analysis of the error on the system trajectories, but our comparisons showed that the approximate model matched the experimental data with a precision comparable to that achieved by the original model. This approach could potentially be seen as a viable alternative as mathematical model building and analysis become better integrated with experimental work.

6.2 Importance of the basal rate for induction in the Yildirim–Mackey model

Our analysis of the lactose system yields new insight into the role of various parameters in ensuring that the cell can switch between the high metabolism and the low metabolism modes depending on the availability of external lactose. By checking the sensitivity of the steady states of the model to small variations in the model parameters, we found that the basal transcription rate Γ_0 is important for both the existence of bistability, and more importantly, for the possibility of induction. We found that the threshold L_e, the lowest external lactose level for which there is no uninduced stable state, is sensitive to the value of Γ_0. If this basal rate falls below a finite minimal value, the threshold L_e tends to infinity.

The basal transcription rate has a function similar to a pilot light. It ensures the existence of a minimal amount of permease and beta-galactosidase necessary for induction when lactose is provided. Without these minimal supplies induction is not possible. This situation, when a stable induced state exists but induction is not possible for any level of lactose, is not discussed in the biological literature. If this situation were realised experimentally, stochastic transitions to the 'forbidden' state would still occur, but the phenomenology of induction would be significantly different.

We have found that the basal transcription rate, a parameter whose value in [5] is not the result of a measurement, has an important role in determining this threshold value. This suggests an indirect way of determining the basal transcription rate. Using reachability analysis we could prove that for low values of the basal transcription rate the system can never be caused to switch from the low metabolism mode to the high metabolism mode, even if the external lactose level is increased. Since the same effect could be achieved by modifying other parameters, it would be interesting to see if this situation can be realised experimentally.

7 Acknowledgments

This work was supported in part by the DARPA BioComp grant F30602-01-2-0562. AMH was supported by NIH Individual Biomedical Informatics Fellowship award F37-LM-008343-1.

The authors wish to thank A. Hsieh for useful discussions and several critical readings of the manuscript, O. Maler for useful discussions and pointing us to other relevant work in the field of reachability analysis, and N. Samatova for useful

suggestions. The contribution of P. Finin in the early stages of this project is acknowledged. The authors are also grateful for the constructive comments of the anonymous reviewers which helped to improve our presentation.

8 References

1 Beckwith, J.: 'Escherichia coil and Salmonella typhmurium: cellular and molecular microbiology' (American Society for Microbiology, Washington, D.C., USA, 1987) Ch 4, pp. 1439–1443

2 Ghosh, R., and Tomlin, C.: 'Symbolic reachable set computation of piecewise affine hybrid automata and its application to biological to biological modeling: Delta-notch protein signaling'. *Syst. Biol.*, 2004, **1**, (1), pp. 170–182

3 de Jong, H., Geiselman, J., Batt, G., Hernandez, C., and Page, M.: 'Qualitative simulation of the initiation of sporualation in Bacillus subtilis'. *Bull. Math. Biol.*, 2003, **66**, pp. 261–300

4 Batt, G., Ropers, D., de Jong, H., Geiselman, J., Mateescu, R., Page, M., and Schneider, D.: 'Validation of qualitative models of genetic regulatory networks by model checking: Analysis of the nutritional stress response in Escherichia coil'. *Bioinformatics*, 2005, **21**, (Supp1.1), pp. 119–128

5 Yildirim, N., and Mackey, M.C.: 'Feedback regulation in the lactose operon: A mathematical modeling study and comparison with experimental data'. *Biophys. J.*, 2003, **84**, pp. 2841–2851

6 Glass, L.: 'Classification of biological networks by their qualitative dynamics'. *J. Theor. Biol.*, 1975, **54**, pp. 85–107

7 Mestl, T., Plathe, E., and Omholt, S.W.: 'Periodic solution in system of piecewise-linear differential equations'. *Dyn. Stability Sys.*, 1995, **10**, (2), pp. 179–193

8 de Jong, H., Gouze, J.L., Hernez, C., Page, M., Sari, T., and Geiselmann, J.: 'Hybrid modeling and simulation of genetic regulatory networks: a qualitative approach'. In 'Hybrid Systems: Computation and Control (HSCC 2003), Lecture Notes in Computer Science, 2623' (Springer-Verlag, New York, 2003), pp. 267–282

9 Kauffmann, S.A.: 'Metabolic stability and epigenesis in randomly constructed genetic nets'. *J. Theor. Biol.*, 1969, **22**, pp. 437–469

10 Thomas, R.: 'Regulatory networks seen as asynchronous automata: a logical description'. *J. Theor. Biol.*, 1991, **153**, pp. 1–23

11 Kuipers, B.: 'Qualitative simulation'. *Art. Intell.*, 1981, **29**, pp. 289–388

12 Brutlag, D.L., Galper, A.R., and Millis, D.H.: 'Knowledge-based simulation of dna metabolism: prediction of enzyme action'. *Comput. Appl. Biosci.*, 1991, **7**, (1), pp. 9–19

13 Gillespis, D.T.: 'Exact stochastic simulation of coupled chemical reaction'. *J. Phys. Chem.*, 1977, **81**, pp. 2340–2361

14 McAdams, H.M., and Arkin, A.: 'stochastic mechanisms in gene expression'. *Proc. Natl. Acad. Sci.*, USA, 1997, **94**, pp. 814–819

15 Gillespie, D.T.: 'Fluctuation and dissipation in Brownian motion'. *Am. J. Phys.*, 1993, **61**, pp. 1077–1083

16 Glass, L., and Kaufman, S.A.: 'The logical analysis of continuous non-linear biochemical control networks'. *J. Theor. Biol.*, 1973, **39**, pp. 103–129

17 de Jong, H.: 'Modeling and simulation of genetic regulatory systems'. *J. Comput. Biol.*, 2002, **9**, (1), pp. 69–105

18 Savageau, M.A.: 'Biochemical systems analysis: a study of function and design in molecular biology' (Addison-Wesley, Reading, MA, 1976)

19 Savageau, M.A.: 'Power-law formalism: A canonical approach to modeling and analysis' in Lakh, V. (Ed.): 'World Congress of Nonlinear Analysts 92' (Walter de Gruyter, Berlin, 1996), vol. 4, pp. 3323–3334

20 Savageau, M.A.: 'Alternative designs for a power switch: Analysis of switching times using the power-law representation'. *Math. Biosci.*, 2002, **180**, pp. 237–253

21 Ghosh, R., Tiwari, A. and Tomlin, S.: 'Automated symbolic reachability analysis with application to delta-notch signaling automata' in 'Hybrid systems: Computation and control (HSCC 2003), Lecture Notes in Computer Science, 2623' (Springer-Verlag, New York, 2003), pp. 233–248

22 Belta, C., Schug, J., Dang, T., Kumar, V., Pappas, G.J., Rubin, H., and Dunlap, P.V.: 'Stability and reachablity analysis of a hybrid model of luminescence in the marine bacterium Vibrio fischeri'. 40th IEEE CDE, Orlando, FL, 2001

23 Belta, C., Finin, P., Habets, L.C.G.J.M., Halasz, A., Imielinski, M., Kumar, V., and Rubin, H.: 'Understanding the bacterial stringent response using reachability analysis of hybrid systems'. 7th Int. Conf. on Hybrid systems: Computation and Control, Philadelphia, PA, USA, March 2005.

24 Hu, J., Wu, W.C., and Sastry, S.: 'Modeling subtilin production in bacilus subtilis using stochastic hybrid systems'. Hybrid systems Computation and control (HSCC 2004), lecture Notes in Computer Science, 2993, (Springer-Verlag, 2004), p. 417

25 singh, A., and Hespanha, J.: 'Models for gene regulatory networks using polynomial stochastic hybrid system'. Proc. 44th Conf. on Decision and Control, Seville, Spain, December 2005

26 de Jong, H., Gouz, J.-L., Hernandez, C., Page, M., Sari, T., and Geiselmann, J.: 'Qualitative simulation of genetic regulatory networks using piecewise-linear models'. *Bull. Math. Biol.*, 2004, **66**, (2), pp. 301–340

27 Chabrier-River, N., Chiaverini, M., Danos, V., Fages, F., and Schachter, V.: 'Modeling querying biomolecular interaction networks'. *Theor. Comput. Sci.*, 2004, **325**, (1), pp. 25–44

28 Antoniotti, M., Piazza, C., Policriti, A., Simeoni, M., and Mishra, B.: 'Taming the complexity of bio-chemical models through bisimulation and collapsing: Theory and practice'. *Theor. Comput. Sci.*, 2004, **325**, (1), pp. 45–67

29 Kloetzer, M., and Belta, C.: 'Reachability analysis of multi-affine systems'. 9th Int.Workshop on Hybrid systems: Computation and control, Santa Barbara, CA, 2006. In 'Hybrid Systems: Computation and Control (HSCC 2006), Lecture Notes in Computer Science, 3927' (Springer-Verlag, New York, 2006), pp. 348–362

30 Novick, A., and Weiner, M.: 'Enzyme induction as an all-or-none phenomenon' *Proc. Natl. Acad. Sci.*, 1957, USA, **43**, pp. 553–566

31 Mahaffy, J.M., and Simeonov, E.: 'Stability analysis for a mathematical model of the lac operon'. *Q. Appl. Math.*, 1999, **57**, pp. 37–53

32 Selgrade, J.F.: 'Mathermatical analysis of a cellular control process with positive feedback'. *SIAM J. Appl. Math.*, **36**, pp. 219–229

33 Wong, P., Gladney, S., and Keasling, J.D.: 'Mathematical model of the lac operaon: inducer exclusion, catabolite repression, diauxic growth on lactose/glucose'. *Biotechnon. Prog.*, 1997, **13**, pp. 132–143

34 Knorre, W.A.: 'Oscillation of the rate of synthesis of beta-galactosidase in escherichia coli ml 30 and ml 308'. *Biochem. Biophys. Res. Commun.*, 1968, **30**, pp. 1248–1290

35 Pestka, S., Daugherty, B.L., Jung, V., Hotta, K., and Pestka, R.K.: 'Anti-mrna: Specific inhibition of translation of single mrna molecules'. *Proc. Natl. Acad. Sci.,* USA, 1984, **81**, pp. 7527–7528

36 Goodwin, B.C.: 'Control dynamics of β-galacsidase in relation to the baterial cell cycle'. *Eur. J. Biochem*, 1969, **10**, pp. 515–522

37 Yildirim, N., Santillan, M., Horike, D., and Mackey, M.C.: 'Dynamics and bistability in a reduced Model of the la operon'. *Chasos*, 2004, **14**, (2), pp. 279–292

38 Ahmadzadeh, A., Halasz, A., Prajna, S., Jadbabaie, A., and Kumar, V.: 'Analysis of the lactose metabolism in e. coli using sum-of-squares decomposition.'. Proc 44th Con. on Decision and Control, Seville, Spain, December 2005

39 Santillan, M., and Mackey, M.C.: 'Influence of catabolite repression and inducer exclusion on the bistable behavior of the lac operon'. *Biophys. J.*, 2004, **86**, pp. 1282–1292

40 Carrier, T.A., and Keasling, J.D.: 'Investigating autocatalytic gene expression systems through mechanistic modeling'. *J. Theor. Biol.*, 1999, **201**, pp. 25–36

41 Ozbudak, E.M., Thattai, M., Lim, H.N., Shraiman, B.I., and Oudenarden, van: 'Multistability in the lactose utilization network of escherichia coli'. *Nature*, 2004, **427**, pp. 737–740

42 Belta, C., Habets, L., and Kumar, V.: 'Control of multi-affine systems on rectangles with applications to hybrid biomolecular networks'. 41st IEEE Conf. On Decision and Control, Las Vegas, NV, USA, 2002

43 Belta, C., Esposito, J., Kim, J.M., and Kumar, V.: 'Computational techniques for analysis of genetic network dynamics'. *In., J. Robot. Res.*, 2005, **24**, pp. 219–235

44 Twari, A., and Khanna, G.: 'Series of abstractions for hybrid automata'. Fifth Int. Workshop on Hybrid Systems: Computation and Control, Stanford, CA, USA, 2002

45 Heinrich, R., and Schuster, S.: 'The regulation of cellular systems' (Chapman & Hall, 1996). In 'Hybrid Systems: Computation and Control (HSCC 2002), Lecture Notes in Computer Science, 2289' (Springer-Verlag, New York, 2002), pp. 465–478

46 Alur, R., Henzinger, T.A., and Ho, P.-H.: 'Automatic symbolic verification of embedded systems'. *IEEE Tran. Soft. Eng.*, 1996, **22**, pp. 181–201

47 Ptashne, M., and Gann, A.: 'Genes and signals' (Cold Spring Harbor Laboratory Press, New York, USA, 2002)

48 Cousot, P., and Cousot, R.: 'Abstract interpretation: a unified lattice model for static analysis of Programs by construction or approximation of fixpoints'. Conference Record of the Fourth

Annual ACM SIGPLAN-SIGACT Sym. On Principles of Programming Languages, (ACM Press, New York, NY, USA, 1977, pp. 238–252
49 Cousot, P.: 'Verification by abstract interpretation' in Dershowitz, N. (Ed.): 'Proc. Int. Symp. on Verification theory and practice - honoring zohar Manna's 64th birthday, Taormia, Italy, June–July 2003, Springer-Verlag, pp. 243–268
50 Girard, A., Le Guernic, C., and Maler, O.: 'Efficient computation of reachable sets of linear time- invariant systems with inputs'. Ninth Int. Workshop on Hybrid Systems: Computation and Control, Santa Barbara, LA, USA, 2006. In 'Hybrid Systems: Computation and Control (HSCC 2006), Lecture Notes in Computer Science, 3927' (Springer-Verlag, New York, 2006), pp. 257–271
51 von Dassow, G., Meir, E., Munro, E.M., and Odell, G.M.: 'The segment polarity network is a robust developmental module'. *Nature*, 2000, **406**, (6792), pp. 131–132
52 Asarin, E., Dang, T., and Girard, A.: 'Reachability analysis of nonlinear systems using conservative approximation'. Proc. 6th Int. Workshop on Hybrid Systems: Computation and Control, New York, USA, (Springer-Verlag, 2003), pp. 20–35
53 Sontag, E.: 'Molecular systems biology and control'. *Eur. J. Control*, 2005, **11**, pp. 396–435

9 Appendix

9.1 Analysis of the exact model

Table 1 lists the constant values we used, which are given in [5]. The decay constants in the model equations, (8)–(12) include the growth rate: $\tilde{\gamma}_X = \mu + \gamma_X$, etc. Throughout this paper we used the above listed value for μ.

9.1.1 Steady-state structure: The easiest way to identify the steady states of the system described by (8)–(12) is to solve the steady-state versions of the equations of motion equentially. Solving first for M from its own equation of motion, then for B and P, we obtain

$$\bar{M}(A) = \frac{\alpha_M}{\gamma_M}\frac{1+\tilde{K}_1A^2}{K+\tilde{K}_1A^2} + \frac{\Gamma_0}{\gamma_M} = \frac{(\alpha_M+\Gamma_0)}{\gamma_M}\left(\frac{fQ^2+A^2}{Q^2+A^2}\right) \quad (15)$$

$$\bar{B}(A) = \frac{\tilde{\alpha}_B\alpha_M}{\gamma_B\gamma_M}\frac{1+\tilde{K}_1A^2}{K+\tilde{K}_1A^2} + \frac{\Gamma_0}{\gamma_M} = \frac{\tilde{\alpha}_B(\alpha_M+\Gamma_0)}{\gamma_B\gamma_M}\left(\frac{fQ^2+A^2}{Q^2+A^2}\right) \quad (16)$$

$$\bar{P}(A) = \frac{\tilde{\alpha}_B\alpha_M}{\gamma_p\gamma_M}\frac{1+\bar{K}_1A^2}{K+K_1A^2} + \frac{\Gamma_0}{\gamma_M} = \frac{\tilde{\alpha}_p(\alpha_M+\Gamma_0)}{\gamma_p\gamma_M}\left(\frac{fQ^2+A^2}{Q^2+A^2}\right) \quad (17)$$

where we introduced

$$\bar{K}_1 = K_1e^{-2\mu\tau M} \quad \tilde{\alpha}_B = \alpha_Be^{-\mu\tau B} \quad \tilde{\alpha}_P = \alpha_pe^{-\mu(\tau_B+\tau_p)}$$
$$f = \frac{\alpha_M+\Gamma_0K}{(\alpha_M+\Gamma_0)K}; \quad Q = \sqrt{\frac{K}{\tilde{K}_1}} \quad (18)$$

It is useful to rewrite the steady-state conditions for the L and A equations of motion as follows

$$\alpha_A\frac{L}{K_L+L} = \beta_A\frac{A}{K_A+A} + \frac{\gamma_A\gamma_B\gamma_M}{\tilde{\alpha}_B(\alpha_M+\Gamma_0)}\frac{A(Q^2+A^2)}{fQ^2+A^2} \quad (19)$$

$$\alpha_L\frac{L_e}{K_{Le}+L_e} = \beta_{L_1}\frac{A}{K_{L_1}+L} + \frac{\alpha_A\tilde{\alpha}_B\gamma_P}{\tilde{\alpha}_P\gamma_B}\frac{L}{K_L+L} + \frac{\gamma_L\gamma_P\gamma_M}{(\tilde{\alpha}_P(\alpha_M+\Gamma_0))}L\left(\frac{Q^2+A^2}{fQ^2+A^2}\right) \quad (20)$$

The first of the above equations is easily solved for the steady state function $\bar{L}(A)$ that describes the dependence of the steady-state value of L on the steady-state value of A. One more substitution, of $\bar{L}(A)$ into the second equation, yields an equation involving only model constants and the steady-state value of A. This equation is a ratio of polynomials of degrees exceeding four; therefore a closed-form solution for $\bar{A}$ is not possible. However, it is easy to formally solve for the value of the external lactose concentration, L_e. The left-hand side of (20), $\alpha_{L_e}L_e/(K_{L_e}+L_e)$ can be written as a proper single-valued function of A. This implies that for any non-negative value A^* of the internal allolactose concentration, we can find at most one value of the external lactose for which A^* corresponds to a steady state of the system. However, only those A^* values correspond to steady states for which the right-hand side of (20) is less than α_{L_e} so that the corresponding L_e value is non-negative.

We would like to identify all physically and biologically meaningful steady states of the model. The $L_e(A^*)$ function is obtained using Maple as the ratio of two polynomials of degree 14. It vanishes at $A = 0$ and its denominator has three real positive roots, at $A = \{3.7794, 5.3328, 354.22\}$, where the function becomes infinite and changes signs. The derivative of the function has only two real positive roots, at $A_{low} = 0.014167$ and at $A_{high} = 0.0961567$ (all units are in milli-molars). These two values give the S-shape to the steady-state characteristic (the corresponding L_e values are 0.06201 mM and 0.02777 mM, respectively; see Fig. 5). It follows that in the interval mM $\leq A \leq$ 2.0 mM, the $L_e(A^*)$ function is well defined and there may be only one, two or three steady states depending on the value of L_e.

9.1.2 Stability: In the previous section we found closed form expressions for the steady-state values of *M, B, L, P* as a function of A as well as for the possible value of L_e for which the given A corresponds to a steady state. To investigate the stability of the resulting steady states, we constructed the Jacobian of the system defined by (8)–(12) without time delays. We then substituted the analytical expressions from (15)–(20), obtaining a matrix which only depends on A and model constants (other than L_e). This is the Jacobian for the (unique) possible steady states corresponding to values of A^* and the correponding $L_e(A^*)$ from the steady state curve discussed above.

The determinant of this matrix can be calculated directly, and is also a function of A and model constants. The formula cannot be simplified by Maple in its algebraic form; however, this becomes possible after substitution of the numerical values of the model constants. We obtained the determinant of the Jacobian as a ratio of a polynomials in A, of degree 38, respectively 40 in the numerator and the denominator. We calculated all the roots of the numerator and found exactly two in the $0 \leq A \leq 3.0$ mM interval. These roots are identical to the two zeros $\{A_{low}, A_{high}\}$ of the derivative of the

$L_e(A^*)$ function. The determinant is negative outside the $[A_{low}, A_{high}]$ interval and positive inside it.

From here we conclude that the Jacobian at the steady states must have at least one real positive eigenvalue for $A_{low} < A < A_{high}$. This is so because there can only be one, three, of five real eigenvalues (since they are the roots of a fifth-degree polynomial with real coefficients). If they were all negative, their product should be negative. Therefore the steady states corresponding to this interval, and corresponding to the middle steady-states for L_e between $L_e^{\min} = 0.02777$ mM and $L_e^{\max} = 0.06201$ mM, are unstable.

This result can be obtained in a similar way for various other parameter values. We have shown here the instability of the respective states for the system without time delays. This has been shown for a reduced version of the Yildirim–Mackey model in [37], where time delays are also taken into account. The numerical results in [5] are consistent with the instability of the middle states.

The upper and lower branches are believed to be stable, consistent with the numerical simulations in [5]. An algebraic analysis using results on monotone systems by Enciso, Angeli and Sontag (see [53]) and references therein) seems to confirm this assertion at least partially; however that is beyond the scope of this work. Here we reconfirmed the observation that the steady states parameterised by A^* and given by the curve $L_e(A^*)$ with $0 < A^* < 0.5$ mM are stable if and only if $dL_e(A^*)/dA^* > 0$, based on the following. We constructed the Jacobian matrix for the system of differential equations in (8)–(12) and substituted the numerical values of the model parameters (Table 1). Then, we substituted the variables *M, B, L, P* with their steady state values as a function of A^* (as described above). Finally, we replaced L_e with the expression $L_e(A^*)$. We calculated the eigenvalues of this matrix for 5000 equally spaced A^* values ranging from *0* to *0.25*. The points with $A^* \in (A_{low}, A_{high})$ had one positive real root and no other roots with a positive real part. The points outside this range had no eigenvalues with a positive real part, implying that the respective states are stable.

9.2 Numerical details of the hybrid model

Details on the piecewise affine approximation to the rate laws are given below.

$$f_1(A) = f_1^{(0)} + f_1^{(1)}A \quad f_2(A) = f_2^{(0)} + f_2^{(1)}A$$
$$g_1(L) = g_1^{(0)} + g_1^{(1)}L \quad g_2(L) = g_2^{(0)} + g_2^{(1)}L \tag{21}$$

The values of the slopes and intercepts we used are listed in Table 2.

Table 2: Numerical values of slopes and intercepts used in piecewise linear approximation

A_{low}	A_{high}	$f_1^{(0)}$	$f_1^{(1)}$	$f_2^{(0)}$	$f_2^{(1)}$
0.000	0.008	0.139,000	0.027,990	0.000,000	0.510,725
0.008	0.015	−0.280,000	0.080,407	0.000,031	0.506,827
0.015	0.030	−0.001,427	0.156,860	0.000,116	0.501,195
0.030	0.050	−0.005,021	0.276,661	0.000,379	0.492,424
0.050	0.075	−0.012,452	0.425,272	0.000,926	0.481,481
0.075	0.150	−0.034,243	0.715,818	0.002,646	0.458,554
0.150	0.500	−0.095,560	1.124,602	0.014,577	0.379,009
0.500	1.000	−0.155,673	0.622,136	0.069,180	0.269,803
1.000	2.000	−0.622,275	0.155,534	0.171,637	0.167,346
2.000	5.000	−0.896,436	0.018,453	0.364,266	0.071,032
L_{low}	L_{high}	$g_1^{(0)}$	$g_1^{(1)}$	$g_2^{(0)}$	$g_2^{(1)}$
0.00	0.15	−0.00	0.892,857	0.00	0.510,204
0.15	0.30	0.031,637	0.681,946	0.010,881	0.437,663
0.30	0.70	0.099,015	0.457,353	0.039,652	0.341,761
0.70	1.00	0.212,772	0.294,842	0.099,247	0.256,625
1.00	2.00	0.341,828	0.165,786	0.186,809	0.169,062
2.00	5.00	0.563,987	0.547,707	0.385,414	0.069,760

Table 3: Steady-state values in hybrid model

Significance	mRNA (M)	β-galactosidase (B)	Allolactose (A)	Lactose (L)	Permease (P)
Uninduced (low)	2.43083×10^{-6}	1.64591×10^{-6}	6.72331×10^{-3}	1.32672×10^{-1}	3.39017×10^{-5}
Induced (high)	4.05508×10^{-4}	2.74568×10^{-4}	2.38213×10^{-1}	1.72064×10^{-1}	5.65543×10^{-3}
Unstable	1.21031×10^{-5}	8.19497×10^{-6}	3.65486×10^{-2}	1.62600×10^{-1}	1.68796×10^{-4}

Table 4: Details of Partition A

Variable	Partitions	Delimeter values							
M	11	0.000	2.50×10^{-5}	5.00×10^{-5}	1.00×10^{-4}	2.50×10^{-4}	5.00×10^{-4}	1.00×10^{-3}	2.50×10^{-3}
		5.00×10^{-3}	1.00×10^{-2}	2.50×10^{-2}	5.00×10^{-2}				
B	11	0.000	2.00×10^{-5}	4.00×10^{-5}	8.00×10^{-5}	2.00×10^{-4}	4.00×10^{-4}	8.00×10^{-4}	2.00×10^{-3}
		4.00×10^{-3}	8.00×10^{-3}	2.00×10^{-2}	4.00×10^{-2}				
A	11	0.000	8.00×10^{-3}	1.50×10^{-2}	3.00×10^{-2}	5.00×10^{-2}	7.50×10^{-4}	1.50×10^{-1}	3.00×10^{-1}
		0.50	0.70	1.000	2.000	5.00	10. 0	50.0	
L	12	0.000	1.00×10^{-2}	2.00×10^{-2}	5.00×10^{-2}	1.00×10^{-1}	1.50×10^{-1}	3.00×10^{-1}	7.00×10^{-1}
		1.000	2.000	5.000	10.0	20.0			
P	11	0.000	4.00×10^{-4}	8.00×10^{-4}	1.60×10^{-3}	4.00×10^{-3}	8.00×10^{-3}	1.60×10^{-2}	4.00×10^{-2}
		8.00×10^{-2}	1.60×10^{-1}	4.00×10^{-1}	8.00×10^{-1}				

9.3 Steady-state values and partitions used in reachability calculations

The choice of additional partitions (those not corresponding to hybridisation intervals) is necessary to increase the granularity of our results. These partitions should be chosen to the extent possible according to the remarkable features of the system under consideration. In the present study the location of the steady states, both stable and not, can be calculated directly. Table 3 lists the corresponding values in the hybrid model, for an external lactose concentration of $L_e = 0.04$ mM. All values are in mM (milli-molar).

We used four partition schemes, depending on the level of detail desired. The respective divider values for each of the five variables of the model, are listed in Tables 4 and 5.

Table 5: Details of Partition B

Variable	Partitions	Delimiter values, mM				
M	14	0.00	6.124×10^{-7}	1.225×10^{-6}	1.837×10^{-6}	2.450×10^{-6}
		3.062×10^{-6}	3.675×10^{-6}	4.90×10^{-6}	6.124×10^{-6}	7.349×10^{-6}
		8.574×10^{-6}	1.225×10^{-5}	1.470×10^{-6}	1.837×10^{-6}	3.675×10^{-5}
B	14	0.0000	4.147×10^{-7}	8.294×10^{-7}	1.244×10^{-6}	1.659×10^{-6}
		2.073×10^{-6}	2.488×10^{-6}	3.317×10^{-6}	4.147×10^{-6}	4.976×10^{-6}
		5.806×10^{-6}	8.294×10^{-6}	9.952×10^{-6}	1.244×10^{-6}	3.675×10^{-5}
A	18	0.0000	1.743×10^{-3}	3.485×10^{-3}	5.228×10^{-3}	8.00×10^{-3}
		8.714×10^{-3}	1.046×10^{-2}	1.50×10^{-2}	1.743×10^{-2}	2.091×10^{-2}
		3.00×10^{-2}	3.485×10^{-2}	4.183×10^{-2}	5.00×10^{-2}	5.925×10^{-2}
		6.971×10^{-2}	7.842×10^{-2}	8.714×10^{-2}	9.585×10^{-2}	
L	24	0.000	3.50×10^{-2}	7.001×10^{-2}	1.225×10^{-1}	1.50×10^{-1}
		1.663×10^{-1}	1.750×10^{-1}	1.838×10^{-2}	1.925×10^{-1}	2.013×10^{-1}
		2.10×10^{-1}	2.188×10^{-1}	2.275×10^{-1}	2.625×10^{-1}	3.00×10^{-1}
		3.282×10^{-1}	3.50×10^{-1}	3.719×10^{-2}	3.938×10^{-1}	4.157×10^{-1}
		4.376×10^{-1}	4.682×10^{-1}	4.813×10^{-1}	5.032×10^{-1}	5.251×10^{-1}
P	14	0.000	8.542×10^{-6}	1.708×10^{-5}	2.562×10^{-5}	3.417×10^{-5}
		4.271×10^{-5}	5.125×10^{-5}	6.833×10^{-5}	8.542×10^{-5}	1.025×10^{-4}
		1.196×10^{-4}	1.708×10^{-4}	2.050×10^{-4}	2.562×10^{-4}	5.125×10^{-4}