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

Á. Halász, V. Kumar, M. Imielinski, 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.

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 chemical and molecular kinetics of the system, but offer a simple description, preferably in terms of discrete states. Abstractions or meta-models that are rooted in the full biological network can be then modelled combining such simple abstractions. This path is taken in most of the biological

© The Institution of Engineering and Technology 2007
doi:10.1049/iet-syb:20060035
Paper first received 20th March and in revised form 8th December 2006
Á. Halász, V. Kumar, M. Imielinski, 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. Imielinski 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
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 piece-wise 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 the 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 (typical 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 behaviours 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] analyzed 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 realize 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 \( x = [x_1, x_2, \ldots, x_n]^T \) 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

\[
\dot{x} = f(x)
\]

where \( f(x) \) is a vector-valued, nonlinear function of the state \( x \), 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 factorized 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 \([u^{(i)}, u^{(i+1)}]\) 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 \( \Delta_k \) but these formulas 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

\[
C_{jk} = [x_k^{(j_k-1)}, x_k^{(j_k)}] \times \cdots \times [x_n^{(j_n-1)}, x_n^{(j_n)}]
\]

\( 1 \leq j_k = 1, \ldots, d_k, \quad (\forall)k = 1, \ldots, n \) (2)

where \( [x_k^{(j_k-1)}, x_k^{(j_k)}] \) are the lower and upper limits for the \( k \)th state variable in the \( j_k \)th 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 \( g(x) \), 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
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

$$ A + B \xrightarrow{k_c} C \quad \frac{d[C]}{dt} = (k_c[A][B] - k_{-c}[C]) \quad (3) $$

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]

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

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

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-variable 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 $k$th interval is then given by a simple linear interpolation

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

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 $I \subset \mathbb{R}$, 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 $\delta$ 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 $f_i$ 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, $X = [x^T, p^T]^T$, has the state equation

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

with the hyper-rectangles now being defined as

$$ C_{j_1 \ldots j_n} = \begin{bmatrix} x_1^{(j_1-1)}, x_1^{j_1} \times \cdots \times x_n^{(j_n-1)}, x_n^{j_n} \\ p_1^{l_1}, p_2^{l_2} \times \cdots \times p_m^{l_m} \end{bmatrix} \times \begin{bmatrix} p_1^{u_1}, p_2^{u_2} \times \cdots \times p_m^{u_m} \end{bmatrix} \times 1 \leq j_k = 1, \ldots, d_k, \forall k = 1, \ldots, n \quad (7) $$

where $p_k^{l}$ and $p_k^{u}$ denote the lower and upper limits for the $k$th 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 face. 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, \( S \in [0, 2 k_u] \) and \( S \in [2 k_u, \infty) \). 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 \( \beta \)-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 \( \beta \)-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 of \( mRN A \) transcribed from both genes, while \( B \) and \( P \) denote the \( \beta \)-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

\[
x = [M, B, A, L, P]^T
\]

while \( L_e \) is viewed as an input to the system that is controlled outside the cell. The differential equations describing the evolution of the system can be written as follows

\[
\frac{dM}{dt} = \alpha_M f_1(e^{-\mu t} A \tau_y + \Gamma_0 - \gamma_M M)
\]

\[
\frac{dB}{dt} = \alpha_B e^{-\mu t} M \tau_y - \gamma_B B
\]

\[
\frac{dA}{dt} = \alpha_A B g_2(L) - \beta_A f_2(A) - \gamma_A A
\]

\[
\frac{dL}{dt} = \alpha_P h(L_e) - \beta_P g_2(L) - \alpha_B g_2(L) - \gamma_L L
\]

\[
\frac{dP}{dt} = \alpha_P e^{-\mu t} (\tau_y + \tau_p) M \tau_y + \tau_p - \gamma_P P
\]

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

\[
f_1(A) = \frac{1 + K_A A^2}{K_A + K_A^2}
\]

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

\[
f_2(A) = \frac{A}{K_A + A}, \quad g_1(L) = \frac{L}{K_L + L}
\]

\[
g_2(L) + \frac{L}{K_{L1} + L}, \quad h(L_e) = \frac{L_e}{K_{Le} + L_e}
\]

As in [5], \( \alpha \) and \( \beta \) denote rate constants for the different processes, the \( \gamma \) are coefficients for terms representing decay of species, and \( \tau \) denotes delays associated with the finite time required to complete transcription (\( \tau_y \)) and translation (\( \tau_p \)). For example, if \( \tau_A \) is the finite time required for the transcription process, \( \Lambda_A \) denotes the concentration of \( \Lambda_A \) seconds prior to the current time. \( \Gamma_0 \) 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 \( \alpha_M \) in (8) accounts for the \( mRNA \) transcription rate controlled by allolactose. In the argument of the rate function \( f_1 \), \( A \tau_y \) is the concentration of allolactose at time \( t - \tau_y \), 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 \( \tau_0 \) term accounts for the basal rate of \( mRNA \) transcription. The term containing \( \alpha_B \) in (9) and the one containing \( \alpha_P \) in (12) for the synthesis of \( \beta \)-galactosidase and permease, respectively. Here we take into account the translation times by introducing delays of \( \tau_B \) for the transcription of lacZ, and \( \tau_B + \tau_P \) for the transcription of lacY which is downstream from lacZ. The term containing \( \alpha_A \) accounts for the decomposition of lactose into allolactose, facilitated by \( \beta \)-galactosidase. The term containing \( \beta_A \) accounts for the conversion of allolactose to glucose and galactose which also takes place via \( \beta \)-galactosidase. Similarly, the term containing \( \alpha_L \) describes the inbound transport of lactose from outside the cell, facilitated by permease, while the term containing \( \beta_L \) describes the reverse transport process facilitated by the same enzyme. Finally, the term \( \gamma_S S \in \{M, B, A, L, P\} \) denotes the net effect of the degradation of the species \( S \) and its dilution due the growth rate. Thus, \( \gamma_S = \mu + \gamma_T \), where \( \gamma_T \) 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 linearising the state equations at the steady state, and check the eigenvalues of the Jacobian matrix for stability.

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 correspond 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 $\beta$-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 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 intervals. 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 first 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. 3](image)

**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%
These piecewise linear functions reproduce the steady-state solution of allolactose and mRNA as a function of time that agrees with the trends observed for both functions that depend on each of the variables \( L \) and \( A \). 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

\[
\begin{align*}
\mathbf{x} = [x_1, x_2, x_3, x_4, x_5]^T &= [M, B, A, L, P]^T \\
\end{align*}
\]

The state equations are

\[
\begin{align*}
\frac{dx_1}{dt} &= a_M f_1^{(1)} x_3 - \gamma_M x_1 + a_M f_1^{(0)} + \gamma_0 \\
\frac{dx_2}{dt} &= a_B e^{-\rho t} x_1 - \gamma_B x_2 \\
\frac{dx_3}{dt} &= a_A g_1^{(1)} x_2 x_4 - \beta_A f_1^{(1)} x_3 x_3 - \gamma_A x_3 \\
&\quad+ (a_A g_1^{(0)} - \beta_A f_1^{(0)}) x_2 \\
\frac{dx_4}{dt} &= -\beta_L g_2^{(1)} x_3 x_4 - a_d g_1^{(1)} x_2 x_4 - (\alpha_d h(L_c) \\
&\quad- \beta_L g_2^{(0)} x_3 x_4 - a_d g_1^{(0)} x_2 - \gamma_L x_4 \\
\frac{dx_5}{dt} &= a_p e^{-\omega(t+\tau)} x_1 - \gamma_p x_5
\end{align*}
\]

Here the constants \( g_1^{(1)}, g_1^{(0)}, \ldots \), 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 where 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
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.4a, correspond to the results of [34] and [35], of time evolution of β-galactosidase at $I_c = 8.0$ mM, during a diauxic shift from glucose to lactose [34] respectively, induction with the lactose analog IPTG [35]. In Fig.4b, 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 $\bar{T}$ of simulated feeding which drives the growth rate: $\dot{\mu}(t) = \bar{\mu} - \alpha \ln(T(t) \bar{T})$ 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 $I_c$, 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 hyper-planes and the parallelepiped becomes a hyper-rectangle. We are interested in the reachability properties of these hyper-rectangles. We further assume that the function $f(\cdot)$ 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, 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 C1 by linear interpolation along x for fixed y between B11 and B14. The same is true for the value at C2, at the point on the opposite side in the y-direction. Finally, the value at D is obtained by linear interpolation along y.
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.

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 $f(i)$ and $g(i)$ 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 $\alpha_p / \gamma_p$, 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 $[0.00025, 0.0005] \times [0.0002, 0.0004] \times [0.05011, 0.07517] \times [0.1, 0.2] \times [0.004, 0.008]$ 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 $[M, B, A, L, P] | M \leq 2.5 \times 10^{-3}$ mM, $B \leq 2.0 \times 10^{-3}$ mM, $A \leq 2.0$ mM, $L \leq 0.7$ mM, $P \leq 4.0 \times 10^{-2}$ 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
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 \(L_0\) and \(T_0\) 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 condition. 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 \(L_0\) and \(T_0\) values.

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. 11](image1.png)

**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.

![Fig. 12](image2.png)

**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 above the threshold value of external lactose required to effect the jump to the upper branch. 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.

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 \( \Gamma_0 \) 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 \( \Gamma_0 \), 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 \( \beta \)-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 \( \mu \). As \( \mu \) 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 \( L_c \), 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, \( p = [\Gamma_0, L_c]^T \). We performed a forward reachability calculation for various combinations of \( \Gamma_0 \) and \( L_c \) 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 \( L_c = 0.04 \) 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 \( L_c \)) are higher than the maximum values in partition B. In order to reach the high steady-state starting from a hyper-rectangle 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 \( L_c \) and \( \Gamma_0 \) values under consideration.

In Fig. 14 we plot the line in the \( L_c - \Gamma_0 \) plane corresponding to the location of the threshold \( L_c \) (which would need to be cleared for guaranteed induction). For several combinations of \( L_c \) and \( \Gamma_0 \) 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 partition B, 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 \( L_c \) 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.

The low results for the low basal rate values for all \( L_c \) values show that for these values the system cannot evolve into the high \( A \) range no matter how \( L_c \) changes. This proves that it is impossible to simply increase \( L_c \) 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 \( \Gamma_0 \) is linear. Since \( h(L_c) \) 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 \( L_c \), then it also holds for any \( L_c \) 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, implementing different abstractions and reachability algorithms 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 \( \Delta x \) 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 inexact 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 \( \Gamma_0 \) 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 \( \Gamma_0 \). 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. Finn 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

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: \( \gamma_c = \mu + \gamma_r \), etc. Throughout this paper we used the above listed value for \( \mu \).

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 essentially. Solving first for \( M \) from its own equation of motion, then for \( B \) and \( P \), we obtain

\[
\tilde{M}(A) = \frac{\alpha_M + \tilde{K}_1 A^2 + \Gamma_0}{\gamma_M K + \tilde{K}_1 A^2 + \gamma_M}
\]

\[
\tilde{B}(A) = \frac{\tilde{\alpha}_B \alpha_M + \tilde{K}_1 A^2 + \Gamma_0}{\gamma_B \gamma_M K + \tilde{K}_1 A^2 + \gamma_M}
\]

\[
\tilde{P}(A) = \frac{\tilde{\alpha}_P \alpha_M + \tilde{K}_1 A^2 + \Gamma_0}{\gamma_M K + \tilde{K}_1 A^2 + \gamma_M}
\]

where we introduced

\[
\tilde{K}_1 = K_1 e^{-2\mu t}\quad \tilde{\alpha}_B = \alpha_B e^{-\mu t}\quad \tilde{\alpha}_P = \alpha_P e^{-\mu (t_0 + \gamma)}
\]

\[
f = \frac{\alpha_M + \Gamma_0 K}{(\alpha_M + \Gamma_0) K} ; \quad Q = \sqrt{\frac{K}{K_1}}
\]

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

\[
\alpha_L \frac{L}{K_L + L} = \beta_A \frac{A}{K_A + A} + \frac{\gamma_A \gamma_B \gamma_M}{\alpha_B (\alpha_M + \Gamma_0)} A (Q^2 + A^2)
\]

\[
\alpha_A \frac{L_A}{K_L + L_A} = \beta_L \frac{A}{K_L + A} + \frac{\gamma_A \gamma_B \gamma_M}{\alpha_P \gamma_B K_L + L_A} L
\]

The first of the above equations is easily solved for the steady state function \( T(A) \) that describes the value of the steady-state value of \( L \) on the steady-state value of \( A \). One more substitution, of \( T(A) \) 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 \( T \) 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), \( \alpha_L L_e/(K_{L_e} + L_e) \) 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 \( \alpha_L \), 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 \( L(A^*) \) 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 \( A_{high} = 0.0961567 \) (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 \( \text{mM} \) and 0.02777 \( \text{mM} \), respectively; see Fig. 5). It follows that in the interval \( \text{mM} \leq A \leq 2.0 \text{mM} \), the \( L(A^*) \) 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 corresponding \( L_e(A^*) \) 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 \( 0 \leq A \leq 3.0 \text{ mM} \) interval. These roots are identical to the two zeros \( \{A_{low}, A_{high}\} \) of the derivative of the
$L_c(A^*)$ function. The determinant is negative outside the $[A_{\text{low}}, A_{\text{high}}]$ 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 $A_{\text{low}} < A < A_{\text{high}}$. This is so because there can be one, three, or 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 $L_c$ between $L_{c\text{min}} = 0.02777 \text{ mM}$ and $L_{c\text{max}} = 0.06201 \text{ mM}$, 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 reaffirmed the observation that the steady states parameterised by $A^*$ and given by the curve $L_c(A^*)$ with $0 < A^* < 0.5 \text{ mM}$ are stable if and only if $dL_c(A^*)/dA^* > 0$, 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 $L_c$ with the expression $L_c(A^*)$. We calculated the eigenvalues of this matrix for 5000 equally spaced $A^*$ values ranging from 0 to 0.25. The points with $A^* \in (A_{\text{low}}, A_{\text{high}})$ 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.

$$f_1(A) = f_1^{(0)} + f_1^{(1)} A \quad f_2(A) = f_2^{(0)} + f_2^{(1)} A$$

$$g_1(L) = g_1^{(0)} + g_1^{(1)} L \quad g_2(L) = g_2^{(0)} + g_2^{(1)} L$$

(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

<table>
<thead>
<tr>
<th>$A_{\text{low}}$</th>
<th>$A_{\text{high}}$</th>
<th>$f_1^{(0)}$</th>
<th>$f_1^{(1)}$</th>
<th>$f_2^{(0)}$</th>
<th>$f_2^{(1)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.000</td>
<td>0.008</td>
<td>0.139,000</td>
<td>0.027,990</td>
<td>0.000,000</td>
<td>0.510,725</td>
</tr>
<tr>
<td>0.008</td>
<td>0.015</td>
<td>-0.280,000</td>
<td>0.080,407</td>
<td>0.000,031</td>
<td>0.506,827</td>
</tr>
<tr>
<td>0.015</td>
<td>0.030</td>
<td>-0.001,427</td>
<td>0.156,860</td>
<td>0.000,116</td>
<td>0.501,195</td>
</tr>
<tr>
<td>0.030</td>
<td>0.050</td>
<td>-0.005,021</td>
<td>0.276,681</td>
<td>0.000,379</td>
<td>0.492,424</td>
</tr>
<tr>
<td>0.050</td>
<td>0.075</td>
<td>-0.012,452</td>
<td>0.425,272</td>
<td>0.000,926</td>
<td>0.481,481</td>
</tr>
<tr>
<td>0.075</td>
<td>0.150</td>
<td>-0.034,243</td>
<td>0.715,818</td>
<td>0.002,646</td>
<td>0.458,554</td>
</tr>
<tr>
<td>0.150</td>
<td>0.500</td>
<td>-0.095,560</td>
<td>1.124,602</td>
<td>0.014,577</td>
<td>0.379,009</td>
</tr>
<tr>
<td>0.500</td>
<td>1.000</td>
<td>-0.155,673</td>
<td>0.622,136</td>
<td>0.069,180</td>
<td>0.269,803</td>
</tr>
<tr>
<td>1.000</td>
<td>2.000</td>
<td>-0.622,275</td>
<td>0.155,534</td>
<td>0.171,637</td>
<td>0.167,346</td>
</tr>
<tr>
<td>2.000</td>
<td>5.000</td>
<td>-0.896,436</td>
<td>0.018,453</td>
<td>0.364,266</td>
<td>0.071,032</td>
</tr>
</tbody>
</table>

### Table 3: Steady-state values in hybrid model

<table>
<thead>
<tr>
<th>Significance</th>
<th>mRNA (M)</th>
<th>$\beta$-galactosidase (B)</th>
<th>Allolactose (A)</th>
<th>Lactose (L)</th>
<th>Permease (P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uninduced (low)</td>
<td>$2.43083 \times 10^{-6}$</td>
<td>$1.64591 \times 10^{-6}$</td>
<td>$6.72331 \times 10^{-3}$</td>
<td>$1.32672 \times 10^{-1}$</td>
<td>$3.39017 \times 10^{-5}$</td>
</tr>
<tr>
<td>Induced (high)</td>
<td>$4.05508 \times 10^{-4}$</td>
<td>$2.74568 \times 10^{-4}$</td>
<td>$2.38213 \times 10^{-1}$</td>
<td>$1.72064 \times 10^{-1}$</td>
<td>$5.65543 \times 10^{-3}$</td>
</tr>
<tr>
<td>Unstable</td>
<td>$1.21031 \times 10^{-6}$</td>
<td>$8.19497 \times 10^{-6}$</td>
<td>$3.65486 \times 10^{-2}$</td>
<td>$1.62600 \times 10^{-1}$</td>
<td>$1.68796 \times 10^{-4}$</td>
</tr>
</tbody>
</table>

### Table 4: Details of Partition A

<table>
<thead>
<tr>
<th>Variable</th>
<th>Partitions</th>
<th>Delimiter values</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M$</td>
<td>11</td>
<td>0.000 2.50 $\times 10^{-5}$ 5.00 $\times 10^{-5}$ 1.00 $\times 10^{-4}$ 2.50 $\times 10^{-4}$ 5.00 $\times 10^{-4}$ 1.00 $\times 10^{-3}$ 2.50 $\times 10^{-3}$ 5.00 $\times 10^{-3}$</td>
</tr>
<tr>
<td>$B$</td>
<td>11</td>
<td>0.000 2.00 $\times 10^{-5}$ 4.00 $\times 10^{-5}$ 8.00 $\times 10^{-5}$ 2.00 $\times 10^{-4}$ 4.00 $\times 10^{-4}$ 8.00 $\times 10^{-4}$ 2.00 $\times 10^{-3}$ 4.00 $\times 10^{-3}$ 8.00 $\times 10^{-3}$</td>
</tr>
<tr>
<td>$A$</td>
<td>11</td>
<td>0.000 8.00 $\times 10^{-3}$ 2.00 $\times 10^{-2}$ 5.00 $\times 10^{-2}$ 1.00 $\times 10^{-1}$ 5.00 $\times 10^{-1}$ 10.0 50.0 7.00 $\times 10^{-1}$</td>
</tr>
<tr>
<td>$L$</td>
<td>12</td>
<td>0.000 1.00 $\times 10^{-2}$ 2.00 $\times 10^{-2}$ 5.00 $\times 10^{-2}$ 1.00 $\times 10^{-1}$ 5.00 $\times 10^{-1}$ 3.00 $\times 10^{-1}$ 7.00 $\times 10^{-1}$ 2.00</td>
</tr>
<tr>
<td>$P$</td>
<td>11</td>
<td>0.000 4.00 $\times 10^{-4}$ 8.00 $\times 10^{-4}$ 1.60 $\times 10^{-3}$ 4.00 $\times 10^{-3}$ 8.00 $\times 10^{-3}$ 1.60 $\times 10^{-2}$ 4.00 $\times 10^{-2}$ 8.00 $\times 10^{-2}$</td>
</tr>
</tbody>
</table>
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 $L_c = 0.04$ 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>
<thead>
<tr>
<th>Variable</th>
<th>Partitions</th>
<th>Delimiter values, mM</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M$</td>
<td>14</td>
<td>$6.124 \times 10^{-7}$</td>
</tr>
<tr>
<td></td>
<td>3.062 $\times 10^{-6}$</td>
<td>$3.675 \times 10^{-6}$</td>
</tr>
<tr>
<td></td>
<td>8.574 $\times 10^{-6}$</td>
<td>$1.225 \times 10^{-5}$</td>
</tr>
<tr>
<td>$B$</td>
<td>14</td>
<td>$4.147 \times 10^{-7}$</td>
</tr>
<tr>
<td></td>
<td>2.073 $\times 10^{-6}$</td>
<td>$2.488 \times 10^{-6}$</td>
</tr>
<tr>
<td></td>
<td>5.806 $\times 10^{-6}$</td>
<td>$8.294 \times 10^{-6}$</td>
</tr>
<tr>
<td>$A$</td>
<td>18</td>
<td>$1.743 \times 10^{-3}$</td>
</tr>
<tr>
<td></td>
<td>8.714 $\times 10^{-3}$</td>
<td>$1.046 \times 10^{-2}$</td>
</tr>
<tr>
<td></td>
<td>3.00 $\times 10^{-2}$</td>
<td>$3.485 \times 10^{-2}$</td>
</tr>
<tr>
<td></td>
<td>6.971 $\times 10^{-2}$</td>
<td>$7.842 \times 10^{-2}$</td>
</tr>
<tr>
<td>$L$</td>
<td>24</td>
<td>$3.50 \times 10^{-2}$</td>
</tr>
<tr>
<td></td>
<td>1.663 $\times 10^{-1}$</td>
<td>$1.750 \times 10^{-1}$</td>
</tr>
<tr>
<td></td>
<td>2.10 $\times 10^{-1}$</td>
<td>$2.188 \times 10^{-1}$</td>
</tr>
<tr>
<td></td>
<td>3.282 $\times 10^{-1}$</td>
<td>$3.50 \times 10^{-1}$</td>
</tr>
<tr>
<td></td>
<td>4.376 $\times 10^{-1}$</td>
<td>$4.682 \times 10^{-1}$</td>
</tr>
<tr>
<td>$P$</td>
<td>14</td>
<td>$8.542 \times 10^{-6}$</td>
</tr>
<tr>
<td></td>
<td>4.271 $\times 10^{-5}$</td>
<td>$5.125 \times 10^{-5}$</td>
</tr>
<tr>
<td></td>
<td>1.196 $\times 10^{-4}$</td>
<td>$1.708 \times 10^{-4}$</td>
</tr>
</tbody>
</table>