Simplification of Stochastic Chemical Reaction Models with Fast and Slow Dynamics

Guang Qiang Dong · Luke Jakobowski ·
Marco A. J. Iafolla · David R. McMillen

Received: 15 November 2006 / Accepted: 29 May 2007 /
Published online: 5 September 2007
© Springer Science + Business Media B.V. 2007

Abstract Biological systems often involve chemical reactions occurring in low-molecule-number regimes, where fluctuations are not negligible and thus stochastic models are required to capture the system behaviour. The resulting models are generally quite large and complex, involving many reactions and species. For clarity and computational tractability, it is important to be able to simplify these systems to equivalent ones involving fewer elements. While many model simplification approaches have been developed for deterministic systems, there has been limited work on applying these approaches to stochastic modelling. Here, we describe a method that reduces the complexity of stochastic biochemical network models, and apply this method to the reduction of a mammalian signalling cascade and a detailed model of the process of bacterial gene expression. Our results indicate that the simplified model gives an accurate representation for not only the average numbers of all species, but also for the associated fluctuations and statistical parameters.

Keywords Stochastic biochemical modelling (modeling) · Model reduction · Signalling (signaling, signal) cascade · Gene expression · Slow manifold · Simplification

1 Introduction

Recent advances in both experimental and computational methods have combined to offer the prospect of developing a quantitative, systematic understanding of

G. Q. Dong · L. Jakobowski · M. A. J. Iafolla · D. R. McMillen (✉)
Institute for Optical Sciences and Department of Chemical and Physical Sciences,
University of Toronto Mississauga, 3359 Mississauga Rd N,
Mississauga, ON L5L 1C6, Canada
e-mail: mcmillen@utm.utoronto.ca

biological systems, and this has driven a surge in recent interest in the formulation of mathematical models in biology, especially of the molecular-level details of biochemical kinetics [1, 2]. The massive accumulation of experimental data in molecular and cellular biology has made modelling an increasingly vital tool in the effort to understand and analyze biological processes [3]. Many methods of modelling the internal states of cells have been proposed [3, 4]. Deterministic modelling treats concentrations of biochemical species as continuous state variables, uses standard chemical kinetics to obtain a set of differential equations describing the system dynamics, and solves these equations to analyze the behaviour of the system. Stochastic models treat each species as an integer number of discrete molecules, and treat reactions as discrete, random events. Both of these approaches neglect spatial information, approximating cells as well-mixed systems, and we will maintain this approximation here, while noting that other modelling approaches are being developed to address spatial inhomogeneities in cells [5, 6].

There is extensive evidence for stochasticity in cellular processes [1, 7–11]. This randomness manifests itself, for example, as variability in the numbers of proteins (or other gene products such as messenger RNA) expressed in cells that are genetically identical: cells with identical genomes, under identical conditions, still display a wide range of expression levels of a given gene. Generally, stochasticity in cells is partitioned into two sources: intrinsic and extrinsic noise. Intrinsic noise is inherent to a system, arising from the non-negligible fluctuations in reaction rates in regimes where small numbers of reactant molecules are interacting (as is often taken to be the case inside cells). Extrinsic noise represents random variations in environmental factors outside the system of interest, but varying between cells. The boundary between intrinsic and extrinsic effects depends on the extent of the system being modelled: to a single-gene model, variation in a second regulatory gene may be an extrinsic source of noise, while a two-gene model explicitly representing the production of the regulatory gene would view that same variability as intrinsic to the system. Commonly cited sources of extrinsic noise include the levels of transcriptional and translational enzymes (RNA polymerases and ribosomes, for example) available within a given cell: these proteins may be expressed to different degrees even within genetically identical cells, as a result of complex chains of stochastic effects. Variation of these levels will directly affect how much a gene of interest is able to express its products, and this type of net cell-to-cell variability is gathered under the label of extrinsic noise.

Formulating biochemical models by writing out reaction systems and translating them into deterministic sets of ordinary differential equations (ODEs) does not capture this biological variability, so stochastic modelling approaches are often applied to intracellular biological models [1, 8]. By simulating the fluctuations in small-number regimes, such models can capture the intrinsic noise in biological systems, while extrinsic noise sources are not explicitly handled and must be added by other means. Stochastic models are computationally expensive, and cellular processes tend to be complicated, with detailed models quickly growing to include large numbers of interacting species and their associated chemical reactions. Beyond their computational expense, such models are difficult to understand and to analyze, and thus any reduction in model complexity is desirable, provided it can be achieved without substantially altering the system's behaviour [12].

Three main methods have been developed for model reduction: lumping, sensitivity analysis, and time-scale analysis [13]. Lumping is useful when only limited measurements and information are available about specific reaction kinetics and detailed chemical composition [14]. Sensitivity analysis seeks to determine and eliminate insignificant reactions and species on the basis of their impact on designated important species [15]. As the number of important species increases, sensitivity analysis is less likely to provide substantial model order reduction, and thus it is often used in conjunction with the quasi-steady-state approximation. Time-scale analysis identifies the different time scales over which species react, and the fast-timescale reactions and species are assumed to be at steady state [16, 17]. The slow variables remain in the reduced model, but their motion is constrained to a lower-dimensional space on the basis of their relationship to the fast variables. Thus, order reduction is possible since the differential equations for the fast species are replaced by algebraic relations. This is the basis for the quasi-steady-state approximation, mathematically formalized by singular perturbation theory. Time-scale analysis requires manual manipulation of the model and significant a priori knowledge of the behaviour of various species; in practice, this means that it is feasible to simplify only compact models. However, all three methods have been developed for deterministic modelling, whereas reduction approaches for stochastic models remain largely unexplored.

Here, we propose a systematic method for reducing the complexity of stochastic biochemical models while keeping their statistical properties unchanged based on time-scale analysis [18, 19]. Signalling cascades and gene expressions are examples of this complexity, involving many participating species interacting in highly branched networks [12]. We apply the method to these two models, reducing them to substantially fewer reactions and species while maintaining the same overall behavior.

2 Methods

In this section, we address the algorithm used to find the simplified reaction system based on a reduction method designed for non-isothermal systems [18]; we then extend this technique by adding a final step in which the system is returned to a set of biochemical reactions, suitable for stochastic simulation by any desired method. The key problem of time-scale analysis lies in determining the lower-dimensional space on which the nonlinear differential equations for the slow species are constrained to evolve. This whole procedure is automated using Mathematica (Wolfram Research; Mathematica script available online). The methodology is as follows:

Consider a spatially homogeneous biochemical reaction system with relatively fast and slow reactions. The reaction dynamics can be written as differential equations:

$$\frac{d\mathbf{x}}{dt} = \mathbf{v}(\mathbf{x}) \cdot \mathbf{r}(\mathbf{x}) \quad (1)$$

where $\mathbf{x}$ denotes the species vector, $\mathbf{v}(\mathbf{x})$ is the stoichiometric matrix and $\mathbf{r}(\mathbf{x})$ is the reaction rate vector.

Example To illustrate the method, we will use a simplified model of gene transcription in the bacterium *E. coli* [20] as a running example, showing, for each step in the

process, how the method is applied to this system. As shown in reactions (2) to (5) below, transcription starts with the binding of RNA polymerase (A) to a region of the DNA (B). This RNA/wound-DNA structure is referred to as the closed complex (C). Then the DNA is unwound and becomes single-stranded (“open”) in the vicinity of the binding site. This RNA/unwound-DNA structure is called the open complex (D). The RNA polymerase will then move along DNA and transcribe the information from DNA to form mRNA (E), which then degrades (vanishing from the model) at a specified rate.

where $(k_1, k_{-1}, k_2, k_3, k_4) = (4 \times 10^{-6}, 1, 1, 5.79, 5 \times 10^{-4})$.

The differential equations for this system can be written as:

$$\begin{pmatrix} \dot{A} \\ \dot{B} \\ \dot{C} \\ \dot{D} \\ \dot{E} \end{pmatrix} = \begin{pmatrix} -1 & 1 & 0 & 1 & 0 \\ -1 & 1 & 0 & 1 & 0 \\ 1 & -1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} k_1 AB \\ k_{-1} C \\ k_2 C \\ k_3 D \\ k_4 E \end{pmatrix} \quad (6)$$

where A, B, C, D , and E represent the concentrations of each species, and the right-hand side has been written as a column vector containing each of the five reaction rates in the system, pre-multiplied by a stoichiometry matrix representing how each species is affected by each reaction. For example, the entries in the first row of the stoichiometric matrix indicates that species A loses one molecule when the forward reaction (2) occurs, gains one when the reverse reaction (2) occurs, and gains one when reaction (4) occurs. In general, for a system with r reactions and s species, the stoichiometric matrix will be $s \times r$; this matrix is post-multiplied by the $r \times 1$ column of reaction rates, yielding the desired $s \times 1$ vector of rates of change for the species. For our example, $r = s = 5$, since the forward and reverse directions in (2) are treated as distinct reactions.

Step 1 First we need to identify fast and slow reactions, which can be selected by examining the rate constants for the target system. By treating different sets of reactions as slow, the model may be reduced to varying degrees, resulting in reduced models with different final numbers of reactions.

After finding the fast and slow reactions, we can separate the vector $\mathbf{r}(\mathbf{x})$ into fast reactions $\mathbf{r}_f(\mathbf{x})$ and slow reactions $\mathbf{r}_s(\mathbf{x})$, and select the corresponding $\mathbf{v}_f(\mathbf{x})$ and $\mathbf{v}_s(\mathbf{x})$. Then we get differential equations:

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}) + \mathbf{v}_f(\mathbf{x}) \cdot \mathbf{r}_f(\mathbf{x}) \quad (7)$$

where $\mathbf{f}(\mathbf{x}) = \mathbf{v}_s(\mathbf{x}) \cdot \mathbf{r}_s(\mathbf{x})$ represents the dynamics for slow reactions. $\mathbf{v}_f(\mathbf{x})$ will be a $s \times p$ matrix, where s is the number of species in the model and p is the number of fast reactions.

Note that the user is free to choose the set of fast reactions based on any desired criterion. The most common basis for selection would be either the rate constants associated with each reaction, or the net reaction rates. Reaction rates, taking into account both the rate constant and the abundance of the participating species, can vary over time, which may in some cases make it difficult to identify a single rate as a basis for selection. Empirically, we have found that better matches between the full and reduced models are obtained by choosing based on rate constants than by using the reaction rates. However, we have no rigorous proof that one method or the other is superior, and users may find that their model reduction works better by choosing fast reactions based on reaction rates, or through some other scheme of their own devising.

Example (continued) Here, we select two reactions as slow and the remaining three as fast, since k_1 and k_4 are much smaller than all other rate constants. Picking out the columns corresponding to fast reactions from $\mathbf{v}(\mathbf{x})$ (indicated in bold on the left, below), we have:

$$\begin{pmatrix} -1 & \mathbf{1} & \mathbf{0} & \mathbf{1} & 0 \\ -1 & \mathbf{1} & \mathbf{0} & \mathbf{1} & 0 \\ 1 & -\mathbf{1} & -\mathbf{1} & \mathbf{0} & 0 \\ 0 & \mathbf{0} & \mathbf{1} & -\mathbf{1} & 0 \\ 0 & \mathbf{0} & \mathbf{0} & \mathbf{1} & -1 \end{pmatrix} \Rightarrow \mathbf{v}_f(\mathbf{x}) = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ -1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix}. \quad (8)$$

Selecting the corresponding terms in $\mathbf{r}_f(\mathbf{x})$ and putting the remaining terms into $\mathbf{f}(\mathbf{x})$, we obtain:

$$\begin{pmatrix} \dot{A} \\ \dot{B} \\ \dot{C} \\ \dot{D} \\ \dot{E} \end{pmatrix} = \begin{pmatrix} -k_1 AB \\ -k_1 AB \\ k_1 AB \\ 0 \\ -k_4 E \end{pmatrix} + \begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ -1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} k_{-1} C \\ k_2 C \\ k_3 D \end{pmatrix}. \quad (9)$$

Step 2 To find the slow manifold, we need to make the stoichiometric vectors of the fast reactions linearly independent, which is equivalent to ensuring that the generalized stoichiometric matrix for the fast dynamics, $\mathbf{v}_f(\mathbf{x})$, has full column rank. If $\mathbf{v}_f(\mathbf{x})$ has full column rank, set $\mathbf{v}'_f(\mathbf{x}) = \mathbf{v}_f(\mathbf{x})$ and $\mathbf{r}'_f(\mathbf{x}) = \mathbf{r}_f(\mathbf{x})$.

Otherwise, we need select the independent columns of $\mathbf{v}_f(\mathbf{x})$ as $\mathbf{v}'_f(\mathbf{x})$, and let

$$\mathbf{r}'_f(\mathbf{x}) = \left\{ [\mathbf{v}'_f(\mathbf{x})]^T [\mathbf{v}'_f(\mathbf{x})] \right\}^{-1} [\mathbf{v}'_f(\mathbf{x})]^T [\mathbf{v}_f(\mathbf{x})] \mathbf{r}_f(\mathbf{x}). \quad (10)$$

The differential equations will become:

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}) + \mathbf{v}'_f(\mathbf{x}) \cdot \mathbf{r}'_f(\mathbf{x}). \quad (11)$$

Example (continued) Since $\mathbf{v}_f(\mathbf{x})$ has full column rank of 3, $\mathbf{v}'_f(\mathbf{x})$ will be identical to $\mathbf{v}_f(\mathbf{x})$ and $\mathbf{r}'_f(\mathbf{x})$ will be identical to $\mathbf{r}_f(\mathbf{x})$.

Step 3 Next, we need to make the elements in $\mathbf{r}'_f(\mathbf{x})$ independent. To test whether the elements in $\mathbf{r}'_f(\mathbf{x})$ are independent, we must calculate the Jacobian matrix of the vector $\mathbf{r}'_f(\mathbf{x})$. Let the number of elements of $\mathbf{r}'_f(\mathbf{x})$ be p' , and let s be the number of species. We will have:

$$J = \begin{pmatrix} \frac{\partial r'_{f_1}(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial r'_{f_1}(\mathbf{x})}{\partial x_j} & \cdots & \frac{\partial r'_{f_1}(\mathbf{x})}{\partial x_s} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{\partial r'_{f_k}(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial r'_{f_k}(\mathbf{x})}{\partial x_j} & \cdots & \frac{\partial r'_{f_k}(\mathbf{x})}{\partial x_s} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{\partial r'_{f_{p'}}(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial r'_{f_{p'}}(\mathbf{x})}{\partial x_j} & \cdots & \frac{\partial r'_{f_{p'}}(\mathbf{x})}{\partial x_s} \end{pmatrix}. \quad (12)$$

Define the row rank of the Jacobian matrix as $p^* = \text{rank}(J)$. If $p^* = p'$, let $\mathbf{v}^*_f(\mathbf{x}) = \mathbf{v}'_f(\mathbf{x})$ and $\mathbf{r}^*_f(\mathbf{x}) = \mathbf{r}'_f(\mathbf{x})$. Alternatively, we can construct a nonsingular matrix $\mathbf{E}(\mathbf{x})$ such that $\mathbf{r}^*_f(\mathbf{x}) = \mathbf{E}(\mathbf{x}) \cdot \mathbf{r}'_f(\mathbf{x})$ has the first p^* rows with independent scalar functions, and the last $p' - p^*$ rows identically equal to 0. The construction of $\mathbf{E}(\mathbf{x})$ is also automated in our Mathematica script using the following method. First, find the independent rows from $\mathbf{r}'_f(\mathbf{x})$ and construct the first p^* rows of $\mathbf{E}(\mathbf{x})$ to select the independent scalar functions from $\mathbf{r}'_f(\mathbf{x})$. Then, to zero out the last $p' - p^*$ rows of $\mathbf{r}^*(\mathbf{x})$, the dependent scalar functions are decomposed in terms of the independent scalar functions in $\mathbf{r}'_f(\mathbf{x})$. By putting -1 in the last $p' - p^*$ rows of $\mathbf{E}(\mathbf{x})$ to select the dependent rows and putting the corresponding decomposition coefficients inside, we can make the last $p' - p^*$ rows of $\mathbf{E}(\mathbf{x})$ equal to zero. From the $\mathbf{E}(\mathbf{x})$ we construct, we obtain a transformed stoichiometric matrix $\mathbf{v}^*_f(\mathbf{x}) = \mathbf{v}'_f(\mathbf{x}) \cdot \mathbf{E}(\mathbf{x})^{-1}$. Since the last $p' - p^*$ rows of $\mathbf{r}^*_f$ are zero, we can discard the last $p' - p^*$ rows of $\mathbf{r}^*_f$ and the last $p' - p^*$ columns of $\mathbf{v}^*_f$. The differential equation will become:

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}) + \mathbf{v}^*_f(\mathbf{x}) \cdot \mathbf{r}^*_f(\mathbf{x}). \quad (13)$$

Example (continued) The Jacobian matrix of $\mathbf{r}'_f(\mathbf{x})$ will be:

$$J = \begin{pmatrix} \frac{\partial r'_{f_1}(\mathbf{x})}{\partial A} & \frac{\partial r'_{f_1}(\mathbf{x})}{\partial B} & \frac{\partial r'_{f_1}(\mathbf{x})}{\partial C} & \frac{\partial r'_{f_1}(\mathbf{x})}{\partial D} & \frac{\partial r'_{f_1}(\mathbf{x})}{\partial E} \\ \frac{\partial r'_{f_2}(\mathbf{x})}{\partial A} & \frac{\partial r'_{f_2}(\mathbf{x})}{\partial B} & \frac{\partial r'_{f_2}(\mathbf{x})}{\partial C} & \frac{\partial r'_{f_2}(\mathbf{x})}{\partial D} & \frac{\partial r'_{f_2}(\mathbf{x})}{\partial E} \\ \frac{\partial r'_{f_3}(\mathbf{x})}{\partial A} & \frac{\partial r'_{f_3}(\mathbf{x})}{\partial B} & \frac{\partial r'_{f_3}(\mathbf{x})}{\partial C} & \frac{\partial r'_{f_3}(\mathbf{x})}{\partial D} & \frac{\partial r'_{f_3}(\mathbf{x})}{\partial E} \end{pmatrix} = \begin{pmatrix} 0 & 0 & k_{-1} & 0 & 0 \\ 0 & 0 & k_2 & 0 & 0 \\ 0 & 0 & 0 & k_3 & 0 \end{pmatrix}. \quad (14)$$

The row rank of the Jacobian matrix of the vector $\mathbf{r}'_f(\mathbf{x})$ is $p^* = 2$. It is less than $p' = 3$, thus we need construct a nonsingular matrix $\mathbf{E}(\mathbf{x})$ to make $\mathbf{r}^*_f(\mathbf{x}) = \mathbf{E}(\mathbf{x}) \cdot \mathbf{r}'_f(\mathbf{x})$ have the first 2 rows with independent scalar functions and the last row of $\mathbf{r}^*_f(\mathbf{x})$ be zero. Clearly the first and third rows of $\mathbf{r}^*_f(\mathbf{x})$ are independent scalar functions, while the second row can be written as a function of the first. In this case, we can construct a matrix to select the first and third items from $\mathbf{r}'_f(\mathbf{x})$ by setting $(1 \ 0 \ 0)$ as the first row of $\mathbf{E}(\mathbf{x})$ and $(0 \ 0 \ 1)$ as the second row. To zero out the last row of $\mathbf{r}^*(\mathbf{x})$, we multiply

the second row by k_{-1}/k_2 and subtract it from the first row, creating a third row of $\mathbf{E}(\mathbf{x})$ equal to $(1 - \frac{k_{-1}}{k_2} \ 0)$, so that the desired matrix becomes

$$\mathbf{E}(\mathbf{x}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 1 - \frac{k_{-1}}{k_2} & 0 & 0 \end{pmatrix}. \quad (15)$$

In this case, $\mathbf{E}(\mathbf{x})$ contains only constants, but in general it may be a function of the state variables $\mathbf{x}$. Applying $\mathbf{E}(\mathbf{x})$ on $\mathbf{v}'_{\mathbf{f}}(\mathbf{x})$ and $\mathbf{r}'_{\mathbf{f}}(\mathbf{x})$, we will get:

$$\mathbf{r}^*_{\mathbf{f}}(\mathbf{x}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 1 - \frac{k_{-1}}{k_2} & 0 & 0 \end{pmatrix} \times \begin{pmatrix} k_{-1}C \\ k_2C \\ k_3D \end{pmatrix} = \begin{pmatrix} k_{-1}C \\ k_3D \\ 0 \end{pmatrix}, \quad (16)$$

$$E^{-1}(\mathbf{x}) = \begin{pmatrix} 1 & 0 & 0 \\ \frac{k_2}{k_{-1}} & 0 & -\frac{k_2}{k_{-1}} \\ 0 & 1 & 0 \end{pmatrix}, \quad (17)$$

and

$$\mathbf{v}^*_{\mathbf{f}}(\mathbf{x}) = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ -1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} 1 & 0 & 0 \\ \frac{k_2}{k_{-1}} & 0 & -\frac{k_2}{k_{-1}} \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ -1 - \frac{k_2}{k_{-1}} & 0 & \frac{k_2}{k_{-1}} \\ \frac{k_2}{k_{-1}} & -1 & -\frac{k_2}{k_{-1}} \\ 0 & 1 & 0 \end{pmatrix}. \quad (18)$$

Discarding the zeros in $\mathbf{r}^*(\mathbf{x})$ and the corresponding columns in $\mathbf{v}^*(\mathbf{x})$, we will obtain the differential equations written as:

$$\begin{pmatrix} \dot{A} \\ \dot{B} \\ \dot{C} \\ \dot{D} \\ \dot{E} \end{pmatrix} = \begin{pmatrix} -k_1AB \\ -k_1AB \\ k_1AB \\ 0 \\ -k_4E \end{pmatrix} + \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ -1 - \frac{k_2}{k_{-1}} & 0 \\ \frac{k_2}{k_{-1}} & -1 \\ 0 & 1 \end{pmatrix} \times \begin{pmatrix} k_{-1}C \\ k_3D \end{pmatrix}. \quad (19)$$

Note that, to this point, no approximations have been made. Equation (19) represents precisely the same dynamics as the original equation (6), rearranged into a form suitable for the next step, where we begin to reduce the model, introducing approximations in the process.

Step 4 With $\mathbf{r}^*_{\mathbf{f}}(\mathbf{x})$ now guaranteed to consist of linearly independent terms, we proceed to set $\mathbf{r}^*_{\mathbf{f}}(\mathbf{x}) = \mathbf{0}$ to obtain the equilibrium manifold. The reduced differential equations will become [21]:

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}) - \mathbf{v}^*_{\mathbf{f}}(\mathbf{x}) \cdot (\mathbf{L}_{\mathbf{v}\mathbf{f}}\mathbf{r}^*_{\mathbf{f}}(\mathbf{x}))^{-1}(\mathbf{L}_{\mathbf{f}}\mathbf{r}^*_{\mathbf{f}}(\mathbf{x})) \quad (20)$$

where

$$(\mathbf{L}_{\mathbf{f}}\mathbf{r}^*_{\mathbf{f}}(\mathbf{x}))_i = \sum_{j=1}^s \left(\frac{\partial (\mathbf{r}^*_{\mathbf{f}}(\mathbf{x}))_i}{\partial \mathbf{x}_j} \right) \cdot (\mathbf{f}(\mathbf{x}))_j \quad (21)$$

for $i = 1 \dots p^*$, and

$$(\mathbf{L}_{\mathbf{v}\mathbf{f}}^*(\mathbf{x}))_{ij} = \sum_{k=1}^s \left(\frac{\partial(\mathbf{r}_{\mathbf{f}}^*(\mathbf{x}))_i}{\partial \mathbf{x}_{\mathbf{k}}} \right) \cdot (\mathbf{v}_{\mathbf{f}}^*(\mathbf{x}))_{kj} \quad (22)$$

for $i, j = 1 \dots p^*$.

The matrix $\mathbf{L}_{\mathbf{v}\mathbf{f}}^*(\mathbf{x})$ is assumed not to be singular, so that its inverse can be taken. In certain cases (not shown here), this matrix can become singular, in which case the method breaks down and the reduction will not be able to be completed. There do not appear to be any general results specifying the conditions under which $\mathbf{L}_{\mathbf{v}\mathbf{f}}^*(\mathbf{x})$ is guaranteed to be non-singular, and we do not currently have any such results to present. This potential singularity represents a possible limitation in the utility of the algorithm. In practice, it is often possible to resolve the singularity issue by selecting a slightly different set of fast reactions and reapplying the steps.

Example (continued) Solving $\mathbf{r}_{\mathbf{f}}^*(\mathbf{x}) = \mathbf{0}$, we find that $C = 0$ and $D = 0$. Calculating $\mathbf{L}_{\mathbf{f}}^*(\mathbf{x})$ and $\mathbf{L}_{\mathbf{v}\mathbf{f}}^*(\mathbf{x})$ according to (21–22), we find:

$$\mathbf{L}_{\mathbf{f}}^*(\mathbf{x}) = \begin{pmatrix} k_1 k_{-1} AB \\ 0 \end{pmatrix} \quad (23)$$

$$\mathbf{L}_{\mathbf{v}\mathbf{f}}^*(\mathbf{x}) = \begin{pmatrix} -(k_{-1} + k_2) & 0 \\ k_3 \frac{k_2}{k_{-1}} & -k_3 \end{pmatrix} \quad (24)$$

$$(\mathbf{L}_{\mathbf{v}\mathbf{f}}^*(\mathbf{x}))^{-1} = \begin{pmatrix} -\frac{1}{k_{-1} + k_2} & 0 \\ -\frac{k_2}{k_{-1}(k_{-1} + k_2)} - \frac{1}{k_3} \end{pmatrix} \quad (25)$$

$$(\mathbf{L}_{\mathbf{v}\mathbf{f}}^*(\mathbf{x}))^{-1}(\mathbf{L}_{\mathbf{f}}^*(\mathbf{x})) = \begin{pmatrix} \frac{k_1 k_{-1}}{k_{-1} + k_2} AB \\ \frac{k_1 k_2}{k_{-1} + k_2} AB \end{pmatrix}. \quad (26)$$

The reduced differential equations then become:

$$\begin{pmatrix} \dot{A} \\ \dot{B} \\ \dot{C} \\ \dot{D} \\ \dot{E} \end{pmatrix} = \begin{pmatrix} -k_1 AB \\ -k_1 AB \\ k_1 AB \\ 0 \\ -k_4 E \end{pmatrix} + \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ -1 - \frac{k_2}{k_{-1}} & 0 \\ \frac{k_2}{k_{-1}} & -1 \\ 0 & 1 \end{pmatrix} \times \begin{pmatrix} \frac{k_1 k_{-1}}{k_{-1} + k_2} AB \\ \frac{k_1 k_2}{k_{-1} + k_2} AB \end{pmatrix}. \quad (27)$$

Resolving the matrix multiplication and simplifying, this becomes

$$\Rightarrow \begin{pmatrix} \dot{A} \\ \dot{B} \\ \dot{C} \\ \dot{D} \\ \dot{E} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \frac{k_1 k_2}{k_{-1} + k_2} AB - k_4 E \end{pmatrix}. \quad (28)$$

Substituting the values of the rate constants, the single nonzero rate becomes

$$\dot{E} = 2 \times 10^{-6} AB - 5 \times 10^{-4} E. \quad (29)$$

Step 5 We construct simplified reaction systems yielding the dynamics of the reduced set of differential equations. This process is difficult to formalize into an algorithm, but is generally straightforward by inspection of the equations, where first and second-order terms indicate the need for first- and second-order reactions.

Example (continued) From the single differential equation (29), we note that the product AB implies a second-order reaction, while the zero rates of change for species A and B indicates that these species must appear on both sides of this reaction. The negative term depending only on E implies a first-order decay of species E , so the reduced set of chemical reactions is:

where $k'_1 = 2 \times 10^{-6}$ and $k'_2 = 5 \times 10^{-4}$.

Step 6 In the reduced model, the fast dynamics are set to equilibrium instantaneously, whereas in fact equilibrium is expected to be reached quickly but not instantly. This approximation can lead to deviation between the reduced model and the original, full model. The degree of this deviation will depend on a number of factors, including the number of reactions treated as fast, and the details of relative speeds of the slow and fast time scales. The deviation in the reduced model can be significant when the reduction is substantial (that is, when many reactions are treated as fast), or when the difference between fast and slow rates is not large. To obtain a better fit between the original model and the reduced version, the rate constants in the reduced system can be adjusted, using the match between the models as a goodness-of-fit parameter in some form of optimization algorithm. For simple cases this may be possible using standard linear or nonlinear least-squares fitting, but in general there will be multiple rate constants to be adjusted, affecting multiple outputs. We compare the trajectories generated by the differential equations of the original deterministic model to the reduced deterministic model (neglecting stochastic effects in each case), and use simulated annealing [22] to adjust the rate constants. Simulated annealing is a probabilistic global optimization technique formulated in analogy with the cooling of real materials. When a physical substance is gradually cooled (“annealed”) in an effort to reach the global minimum energy state, it starts at high temperatures where the system is relatively free to visit higher-energy states, but as it is slowly cooled, lower energies are increasingly favoured, until eventually it solidifies into a minimum state that should be a good approximation to the global minimum; if the substance is cooled too rapidly, it can instead be caught in a local minimum far from the global minimum. In simulated annealing, any desired criterion may be used to represent the energy (here, we use the sum of the square of the difference between the original and reduced model at each time point), and the temperature represents the probability of making transitions between two states with energy difference ΔE , according to the Boltzmann distribution: $P(\text{transition}) = \exp(-\Delta E/kT)$. A number of simulated annealing implementations are freely available for download; we use one by Vandekerckhove [23] that provides a script in MATLAB (The MathWorks).

Example (continued) As shown in Fig. 1a, in this case the reduced model gives a good match to the original model, without any fitting process. The more complex

Fig. 1 Comparisons of the ODE and stochastic solutions for species E from the simple gene transcription system, reactions (2) to (5), used as a running example in the text. Curves are shown for the original model (6) (solid line) and the reduced model (29) (dashed line). **a** Time series for the deterministic versions of the original and reduced models. **b** Frequency histogram for species E, for the original and reduced models. The shape of the histogram for the reduced model is faithfully reproduced, with a small shift (about 0.5%) in the mean, arising from the elimination of intermediate steps which, in the original model, slow down the transcription process and reduce the steady-state level of E

gene expression system considered in Section 3, below, required fitting to obtain a similarly good match.

Step 7 The reduced reaction system obtained in Step 5 is simulated using a stochastic chemical kinetics algorithm. A number of methods exist for implementing such simulations [24]. Here, we use the Biochemical Network Stochastic simulator (BioNetS), software that was developed for the purpose of efficiently and accurately simulating stochastic models of biochemical networks [25]. BioNetS is able to simulate reactions using either an implementation of the exact stochastic simulation method formulated by Gillespie [26, 27], a continuum approximation treating the species numbers as continuous rather than discrete quantities, or a hybrid of the two. Here, we have used only the exact, discrete method to simulate both the original and reduced models.

Calculations required in the simplification process were carried out using Mathematica (Wolfram Research). Our reaction models are read directly into Mathematica from BioNetS files, simplifying the process of applying the method to a variety of biochemical systems.

Example (concluded) The simplified chemical reactions (30) and (31) were simulated in BioNetS, and the results compared to a stochastic simulation of the original reactions (2) to (5). Figure 1b compares the steady-state frequency histogram for species E for the two models, and shows that the reduced model reproduces the distribution of the original, with a slight shift in the mean introduced by the reduction process. In the reduction, the intermediate steps of forming species C and D have been eliminated completely, but in fact these reactions, while fast, do take some finite time to occur, slowing down the process of transcription slightly in the original

model. This approximation leads to the roughly 0.5% increase in the mean seen in the reduced model.

3 Applications

3.1 Signal Cascade Model

The signalling cascade of interest in this study begins with epidermal growth factor (EGF) and its corresponding epidermal growth factor receptor (EGFR); a

Fig. 2 Schematic of the ShcA signalling cascade. Binding of epidermal growth factor (EGF) to a membrane-bound receptor triggers a chain of activations, starting with phosphorylation of the ShcA protein and leading to extracellular signal-regulated kinase (ERK) and activated ERK (ERKPP). The actual cascade continues past this point, with nuclear translocations leading to transcription factor activation, but the model halts at ERKPP

schematic of the cascade is shown in Fig. 2. The EGFR belongs to a family of receptors possessing an intrinsic tyrosine kinase activity, called receptor tyrosine kinases (RTKs) [28]. Receptor tyrosine kinases are characterized by containing an extracellular ligand (hormone) binding domain, a single transmembrane helix and a cytoplasmic protein tyrosine kinase domain [28]. The EGFR exists as a monomer in the plasma membrane, and the binding of EGF to the extracellular ligand binding domain of the EGFR induces a conformational change that allows two EGF/EGFR complexes to dimerize [29]. Dimerization then activates the cytoplasmic tyrosine kinase domain and the EGFR autophosphorylates itself at tyrosine residues in a trans fashion, whereby each subunit of the dimer phosphorylates the other [30]. These phosphorylated tyrosines now have the ability to serve as docking sites for downstream modular proteins with the appropriate SH2 or PTB domains since phosphotyrosine residues exist on the EGFR in both contexts [28]. The SchA protein binds to the phosphotyrosine residues on the EGFR through its PTB domain and it consequently becomes phosphorylated at three tyrosine residues [31]. These new phosphotyrosine residues, as well as specific phosphotyrosine residues on the EGFR, now serve as docking sites for the adaptor protein Grb2 [31]. Grb2 contains a central SH2 domain that is flanked by N-terminal and C-terminal SH3 domains, which serve to bind specific proline-rich regions [32]. As a result, via its SH2 domain,

Fig. 3 Comparisons of statistical parameters as a function of time for the ERKPP species, from the original (*solid line*) and reduced (*dashed line*) models. In each case, statistics were accumulated over 10,000 runs. **a** Mean number of ERKPP. **b** Standard deviation. **c** Skewness. **d** Kurtosis

Grb2 binds to phosphotyrosine residues and recruits SOS to the complex via its SH3 domains [28]. This brings SOS in close proximity to the membrane-bound Ras protein [28], a GTP-hydrolase that is inactive in its GDP-binding state and active in its GTP-binding state [33]. Once in its active GTP bound state, Ras interacts with Raf and consequently activates it through serine phosphorylation [28]. Activated Raf then catalyzes the phosphorylation of MEK at two serine residues [28]. The activated MEK then in turn activates ERK by catalyzing the phosphorylation of a threonine and a tyrosine residue [28]. Finally, activated ERK is readily translocated into the nucleus, where it phosphorylates and activates several transcription factors [28]. The modelling of this signal cascade will end at activated ERK and does not include the subsequent translocation to the nucleus followed by transcription factor activation. The chemical reactions involved are shown in Fig. 12 below, as entered into the BioNetS user interface. The values of the rate constants are available in the BioNetS files, but for completeness we list them in the Appendix.

The original deterministic model [34] has 63 reactions and 41 species. To begin the reduction of the original model, we identify 19 reactions as fast, selecting based on the magnitude of their rate constants. Separating the fast and slow reactions, we will get $\mathbf{v}_f(\mathbf{x})$ with 41 rows and 19 columns. The column rank of this matrix is 18, less than the number of columns 19. So independent items are picked out of the

Fig. 4 Comparisons of statistical parameters as a function of time for the ERK species, from the original (solid line) and reduced (dashed line) models. In each case, statistics were accumulated over 10,000 runs. **a** Mean number of ERK. **b** Standard deviation. **c** Skewness. **d** Kurtosis

Fig. 5 Comparisons of statistical parameters as a function of time for the ERKP species, from the original (solid line) and reduced (dashed line) models. In each case, statistics were accumulated over 10,000 runs. **a** Mean number of ERKP. **b** Standard deviation. **c** Skewness. **d** Kurtosis

$\mathbf{v}_f(\mathbf{x})$ to form $\mathbf{v}'_f(\mathbf{x})$, and $\mathbf{r}'_f(\mathbf{x})$ is calculated correspondingly. After this we calculate the Jacobian matrix of $\mathbf{v}'_f(\mathbf{x})$, and the rank of this Jacobian matrix is 14, less than the number of columns, 18. Thus we need construct a nonsingular matrix $\mathbf{E}(\mathbf{x})$ to make $\mathbf{r}^*_f(\mathbf{x}) = \mathbf{E}(\mathbf{x}) \cdot \mathbf{r}'_f(\mathbf{x})$ have the first 14 rows with independent scalar functions and the last 4 rows of $\mathbf{r}^*_f(\mathbf{x})$ be zero. To construct $\mathbf{E}(\mathbf{x})$, we need to find the 14 independent rows from $\mathbf{r}'_f(\mathbf{x})$ first. then we can construct the first 14 rows of $\mathbf{E}(\mathbf{x})$ to pick out the 14 independent scalar functions from $\mathbf{r}'_f(\mathbf{x})$. To zero out the last 4 rows of $\mathbf{r}^*_f(\mathbf{x})$, the 4 dependent scalar functions are decomposed in terms of the 14 independent scalar functions in $\mathbf{r}'_f(\mathbf{x})$. Then, by putting -1 in the last four rows of $\mathbf{E}(\mathbf{x})$ to pick out the dependent rows and put the corresponding decomposition coefficients inside, we can

Fig. 6 a Transcription: 1 RNA polymerase (Rpoly) has the potential to bind and unbind to the promoter. 2 Upon binding to the promoter, Rpoly opens the DNA strands in a process called isomerization. 3 After opening the RNA, Rpoly clears the promoter site and begins transcribing the mRNA. Clearing the promoter site allows additional Rpoly to transcribe the same gene simultaneously (not shown). 4 After completing transcription, the Rpoly is ejected from the gene and releases the completed transcript. **b Translation:** 1 Ribosomes (ribo) both bind and unbind to the ribosome binding site (RBS) on a complete mRNA or nascent mRNA. 2 Upon binding to the RBS, 3 the ribo clears the RBS and begins translation. Clearing the RBS allows additional ribos to translate the same transcript simultaneously (not shown). 4 After completing translation, the ribo is ejected from the transcript and releases the completed peptide

make the last 4 rows of $\mathbf{E}(\mathbf{x})$ to be zero. From the $\mathbf{E}(\mathbf{x})$ we construct, we get the final version of $\mathbf{v}_f^*(\mathbf{x})$ and $\mathbf{r}_f^*(\mathbf{x})$. Setting $\mathbf{r}_f^*(\mathbf{x})$ to zero and following the calculations in step 4, we get the final simplified differential equations for this system, from which we deduce the reduced reaction system as shown below in Fig. 13. For this system, no fitting of the rate constants was necessary: the results shown use the original rate constants.

a Transcription

b Translation

The reduced model contains 27 species and 34 reactions with an overall reduction in computational time of 50% as compared to the original model. For an accurate comparison of the fluctuations in the original and reduced models, both models are run 10,000 times with varying random number generator seeds, with a simulated extracellular signal applied to the network at time zero. This yields 10,000 realizations of the random process, and statistics (mean, standard deviation, skewness, and kurtosis) are calculated over this ensemble at each point in the time series, as shown in Fig. 3. All random sample paths converge to a consistent final value of ERKPP after an initial transient, and the standard deviation drops to near zero after this transient. The distribution of ERKPP values over the ensemble converges to a Gaussian distribution for large time values, with a skewness of zero and a kurtosis of three. The reduced model gives an accurate representation for all statistical parameters in the species of interest. The comparison of ERK and ERKP is also shown in Fig. 4 and Fig. 5. Some species that have been removed in the model reduction existed only in numbers fewer than ten; interestingly, the statistical information present in the original model is satisfactorily preserved despite these excisions.

3.2 Detailed Gene Expression Model

We have also applied the methodology in this paper to the process of gene expression (transcription and translation) in a dividing *E. coli* bacterial cell. Please see Iafolla and McMillen [20] for a more detailed discussion of the model; here we present an overview of the processes included. The whole model includes RNA polymerases (RNAPs) transcribing both stable RNA (sRNA) and messenger RNA (mRNA) from a dividing genome. Closed and open complexes between the promoter sequences and RNA polymerases are required prior to elongation events. The set of reactions is designed to allow multiple RNAPs binding to the same gene, mimicking events in a real cell. The nascent and complete mRNA are subject to translation

Fig. 7 Comparisons of protein numbers in the gene expression system, from the original model with 71 reactions and 47 species (*solid line*), the reduced model with 28 reactions and 29 species (*dashed line*), and the reduced model with 10 reactions and 10 species (*dotted line*). **a** Before fitting, using rate constants from the original model directly in the reduced models. **b** After fitting, where the rate constants have been adjusted using simulated annealing

Fig. 8 Comparisons of statistical parameters as a function of time for protein numbers in the gene expression system. Shown are the original model with 71 reactions and 47 species (*solid line*), the reduced model with 28 reactions and 29 species (*dashed line*), and the reduced model with 10 reactions and 10 species (*dotted line*). **a** Mean number of proteins. **b** Coefficient of Variation. **c** Skewness. **d** Kurtosis

by ribosomes. The model ensures that nascent transcripts are converted into their complete species, accounting for those species that are complexed with a ribosome. Analogous to RNA polymerases, the ribosomes form a complex with the ribosome-binding-site (RBS) prior to elongation and also offer the possibility for multiple elongations from more than one ribosome bound to a single mRNA. A schematic view of the main processes is presented in Figs. 6a and b, and the complete set of reactions is shown below in Fig. 14.

The original model has 71 reactions and 47 species, with a range of rate constants that makes the division into fast and slow reactions somewhat flexible. By treating different numbers of reactions as fast (29 and 59), we reduce the original model to different degree. For each case (29 fast reactions or 59 fast reactions), separating the fast and slow reactions, we will get $\mathbf{v}_f(\mathbf{x})$ with 29 columns and 59 columns respectively. The column rank of these matrices are 28 and 41 respectively, less than the number of columns (29 and 59). So independent terms must be picked out of the $\mathbf{v}_f(\mathbf{x})$ to form $\mathbf{v}'_f(\mathbf{x})$, and $\mathbf{r}'_f(\mathbf{x})$ is calculated as described in Section 2. After this we calculate the Jacobian matrices of $\mathbf{v}'_f(\mathbf{x})$, and the ranks of these matrices are 18 and 37, less than the number of columns (28 and 41). Thus we follow step 3 and use the same methods as used in the example in the Section 2 above to construct $\mathbf{E}(\mathbf{x})$, from which we get the final version of $\mathbf{v}^*_f(\mathbf{x})$ and $\mathbf{r}^*_f(\mathbf{x})$. Setting $\mathbf{r}^*_f(\mathbf{x})$ to zero and following calculations

Fig. 9 Comparisons of statistical parameters as a function of time for mRNA numbers in the gene expression system. Shown are the original model with 71 reactions and 47 species (*solid line*), the reduced model with 28 reactions and 29 species (*dashed line*), and the reduced model with 10 reactions and 10 species (*dotted line*). **a** Mean number of mRNA. **b** Coefficient of Variation (CV). **c** Skewness. **d** Kurtosis

on step 4, we get the final simplified differential equations for this system, which are then converted to the reduced reaction systems displayed in Figs. 15 and 16.

As shown in Fig. 7a, the deterministic behavior of the reduced models shows significant deviations from the original model for both versions of the reduced model, but especially for the severely reduced ten reaction, ten species model. As in the illustrative example in Section 2, the reduction process tends to yield an over-expression of final products, by removing fast but not instantaneous intermediate steps that serve to slow down the net reaction rate in the original model. To obtain a better fit for the reduced models, we follow Step 6 from Section 2: solve the original and reduced models deterministically, then use simulated annealing to find the optimized rate constants for each reduced model. Plugging the fitted reaction rates into the reduced model, we get a good match between the original model and two reduced models as seen in Fig. 7b.

Comparing Figs. 7 and 8, we note that the average behaviour of the stochastic simulations shows a significant difference from the deterministic behaviour: rather than maintaining the “sawtooth” pattern generated by the growth and division of the simulated cells, the stochastic mean shows a damped, decaying oscillation converging toward a flat steady state. This effect arises from the stochastic nature of the cell

Fig. 10 Comparisons of the steady-state frequency histogram for mRNA and protein in the gene expression system. Shown are the original model with 71 reactions and 47 species (*solid line*), the reduced model with 28 reactions and 29 species (*dashed line*), and the reduced model with 10 reactions and 10 species (*dotted line*). **a** Frequency histogram for mRNA, for the original and reduced models. **b** Frequency histogram for protein, for the original and reduced models

Fig. 11 Comparisons of statistical parameters as a function of time for protein numbers in the gene expression system. Shown are the original model with 73 reactions and 48 species (*solid line*), and the reduced model with 11 reactions and 12 species (*dashed line*). **a** Mean number of mRNA. **b** Coefficient of Variation (CV). **c** Skewness. **d** Kurtosis

Left	Right	Forward	Backward
iR + L	iRL	k1	k_1
iRL + iRL	iRLRL	k2	k_2
iRLRL	RP	k3	
RP + PT1	RP_PT1	k4	k_4
RP_PT1	iRLRL_PT1	k5	
iRLRL_PT1	iRLRL + PT1	k6	
RP + SH	RP_SH	k7	k_7
RP_SH	RP_SHP	k8	
RP_SHP + GR	RP_SHP_GR	k9	k_9
RP_SHP_GR + SO	RP_SHP_GR_SO	k10	k_10
RP_SHP_GR_SO	RP + SHP_GR_SO	k11	k_11
RP_SHP	RP + SHP	k12	k_12
SHP + PT2	SHP_PT2	k13	k_13
SHP_PT2	SH_PT2	k14	
SH_PT2	SH + PT2	k15	
RP + GR	RP_GR	k16	k_16
RP_GR + SO	RP_GR_SO	k17	k_17
RP_GR_SO	RP + GR_SO	k18	k_18
SHP + GR	SHP_GR	k19	k_19
SHP_GR + SO	SHP_GR_SO	k20	k_20
SHP_GR_SO	SHP + GR_SO	k21	k_21
GR_SO	GR + SO	k22	k_22
RP_SHP_GR	RP + SHP_GR	k23	k_23
RP_SHP + GR_SO	RP_SHP_GR_SO	k24	k_24
RP_SHP_GR_SO + RD	RP_SHP_GR_SO_RD	k25	k_25
RP_SHP_GR_SO_RD	RP_SHP_GR_SO + RasGTP	k26	
RasGTP + iRaf	Raf_RasGTP	k27	k_27
Raf_RasGTP	aRaf + RasGTP	k28	
aRaf + MEK	aRaf_MEK	k29	k_29
aRaf_MEK	MEKP + aRaf	k30	
aRaf + MEKP	aRaf_MEKP	k31	k_31
aRaf_MEKP	MEKPP + aRaf	k32	
MEKPP + ERK	MEKPP_ERK	k33	k_33
MEKPP_ERK	ERKP + MEKPP	k34	
MEKPP + ERKP	MEKPP_ERKP	k35	k_35
MEKPP_ERKP	ERKPP + MEKPP	k36	
RP_GR_SO + RD	RP_GR_SO_RD	k37	k_37
RP_GR_SO_RD	RP_GR_SO + RasGTP	k38	

Fig. 12 Full signal cascade model, as entered into the BioNetS user interface [25]. Each line represents one reaction. Species shown on the *left column* appear on the left side of a chemical reaction, while species shown on the *right column* appear on the right. The *Forward* and *Backward* columns give the rate constants for the forward and reverse reactions

division itself in the simulations: the timing of the cell division events is itself a stochastic variable, with division triggered when a “dummy” species representing cellular volume crosses a designated threshold. Each realization of the stochastic process will trigger cell divisions at slightly different times, and over long enough times the phase information is erased, meaning that at a given instant each cell may be at any phase of the cell division cycle, and the oscillatory behaviour will not be seen in the ensemble average.

Using BioNetS, a single simulation run of the original model takes 650 s on a 2.5 GHz G5 CPU, while the reduced model with 29 species takes 80 s and the reduced

Left	Right	Forward	Backward
iR+L	iRL	k1	k_1
iRL+iRL	RP	k2	
PT1+RP	iRLRL_PT1	k3	k_3
iRLRL_PT1	PT1+iRL+iRL	k4	
RP+SH	SHP+RP	k5	
PT2+SHP	SHP_PT2	k6	k_6
SHP_PT2	SH_PT2	k7	
SH_PT2	SH+PT2	k8	
GR+SHP	SHP_GR	k9	k_9
GR+RP	RP_GR	k10	k_10
GR+SO	GR_SO	k11	k_11
SHP_GR+SO	SHP+GR_SO	k12	
RP_GR+SO	RP_GR_SO	k13	k_13
GR_SO+RP	RP_GR_SO	k14	k_14
RD+RP_GR_SO	RasGTP+RP_GR_SO	k15	
iRaf+RasGTP	aRaf+RasGTP	k16	
aRaf+MEK	aRaf+MEKP	k17	
aRaf+MEKP	aRaf+MEKPP	k18	
ERK+MEKPP	ERKP+MEKPP	k19	
ERKP+MEKPP	ERKPP+MEKPP	k20	

Fig. 13 Reduced signal cascade model, again shown in the BioNetS user interface, with columns labelled as in Fig. 12

model with 10 species takes 22 s, representing an improvement in computational speed of between 8-fold and 30-fold, depending on the degree of reduction. Despite the simplicity of the reduced models, they still give a good representation in terms of the statistical parameters, as seen in Figs. 8 and 9. Figure 10 shows the steady-state histogram for the mRNA and protein species, and again the match between the original and reduced models is evident, though the mRNA results match better than the proteins. The statistical parameters for *protein_small*, *mRNA_small*, *operon_ns*, and *operon_s* also yield good matches between the original and reduced models (results not shown). One limitation, however, is that the reduced model with ten species can only capture the steady-state behaviour; during the initial transient before steady-state is reached, this model displays large deviations from the original model. The reduced model with 29 species, on the other hand, can capture both the steady-state and the transient behaviour. Depending on the goals of a particular modelling effort, different degrees of reduction may be appropriate; some investigations may consider only steady-state values, for example, while others may wish to capture transients.

To show how this method works in reaction networks with feedback (where product species interact with earlier steps in the reaction scheme), we add a simple feedback reaction to the original gene expression model: *protein_small* + *operon_ns* $\leftrightarrow$ *protein_small_operon_ns*. This represents the small proteins produced in the model binding to operator sites, occupying them and preventing polymerase binding, and thus also preventing transcription; the result is a negative feedback, wherein the species *protein_small* inhibits its own production. The augmented model has 73 reactions and 48 species; the protein operon complex adds one species to the

Left	Right	Forward	Backward
operon_ns + Rpoly	closed_Rpoly_prom_ns	k_on_Rpoly_prom_ns/(v/1000)	k_off_Rpoly
closed_Rpoly_prom_ns	open_Rpoly_prom_ns	k_isomerization	
open_Rpoly_prom_ns	operon_ns + Rpoly_operon_ns_1 + incom_mRNA_1	k_prom_clearance	
Rpoly_operon_ns_1	Rpoly_operon_ns_2 + mRNA + incom_mRNA_2	k_transcription_ns	
Rpoly_operon_ns_2	Rpoly_operon_ns_3 + mRNA + incom_mRNA_3	k_transcription_ns	
Rpoly_operon_ns_3	Rpoly_operon_ns_4 + mRNA + incom_mRNA_4	k_transcription_ns	
Rpoly_operon_ns_4	Rpoly_operon_ns_5 + mRNA + incom_mRNA_5	k_transcription_ns	
Rpoly_operon_ns_5	Rpoly_operon_ns_6 + mRNA + incom_mRNA_6	k_transcription_ns	
Rpoly_operon_ns_6	Rpoly_operon_ns_7 + mRNA + incom_mRNA_small	k_transcription_ns	
Rpoly_operon_ns_7	Rpoly + mRNA_small	k_transcription_ns/0.871794871794871	
operon_s + Rpoly	closed_Rpoly_prom_s	k_on_Rpoly_prom_s/(v/1000)	k_off_Rpoly
closed_Rpoly_prom_s	open_Rpoly_prom_s	k_isomerization	
open_Rpoly_prom_s	operon_s + Rpoly_operon_s	k_prom_clearance	
Rpoly_operon_s	Rpoly + stable_RNA	k_transcription_s	
v	2v	k_cell_div	
	operon_ns	k_rep_operon_ns	
	operon_s	k_rep_operon_s	
	Rpoly	k_rep_Rpoly	
	ribo	k_rep_ribo	
incom_mRNA_1		k_transcription_ns	
incom_mRNA_2		k_transcription_ns	
incom_mRNA_3		k_transcription_ns	
incom_mRNA_4		k_transcription_ns	
incom_mRNA_5		k_transcription_ns	
incom_mRNA_6		k_transcription_ns	
incom_mRNA_small		k_transcription_ns/0.871794871794871	
ribo + incom_mRNA_1	open_ribo_incom_RBS_1	k_on_ribo_RBS/(v/1000)	k_off_ribo
open_ribo_incom_RBS_1	ribo_incom_mRNA_1 + incom_mRNA_1	k_RBS_clearance	
ribo_incom_mRNA_1	ribo + protein	k_translation	
ribo + incom_mRNA_2	open_ribo_incom_RBS_2	k_on_ribo_RBS/(v/1000)	k_off_ribo
open_ribo_incom_RBS_2	ribo_incom_mRNA_2 + incom_mRNA_2	k_RBS_clearance	
ribo_incom_mRNA_2	ribo + protein	k_translation	
ribo + incom_mRNA_3	open_ribo_incom_RBS_3	k_on_ribo_RBS/(v/1000)	k_off_ribo
open_ribo_incom_RBS_3	ribo_incom_mRNA_3 + incom_mRNA_3	k_RBS_clearance	
ribo_incom_mRNA_3	ribo + protein	k_translation	
ribo + incom_mRNA_4	open_ribo_incom_RBS_4	k_on_ribo_RBS/(v/1000)	k_off_ribo
open_ribo_incom_RBS_4	ribo_incom_mRNA_4 + incom_mRNA_4	k_RBS_clearance	
ribo_incom_mRNA_4	ribo + protein	k_translation	
ribo + incom_mRNA_5	open_ribo_incom_RBS_5	k_on_ribo_RBS/(v/1000)	k_off_ribo
open_ribo_incom_RBS_5	ribo_incom_mRNA_5 + incom_mRNA_5	k_RBS_clearance	
ribo_incom_mRNA_5	ribo + protein	k_translation	
ribo + incom_mRNA_6	open_ribo_incom_RBS_6	k_on_ribo_RBS/(v/1000)	k_off_ribo
open_ribo_incom_RBS_6	ribo_incom_mRNA_6 + incom_mRNA_6	k_RBS_clearance	
ribo_incom_mRNA_6	ribo + protein	k_translation	
ribo + incom_mRNA_small	open_ribo_incom_RBS_small	k_on_ribo_RBS/(v/1000)	k_off_ribo
open_ribo_incom_RBS_small	ribo_incom_mRNA_small + incom_mRNA_small	k_RBS_clearance	
ribo_incom_mRNA_small	ribo + protein_small	k_translation/0.871794871794871	
open_ribo_incom_RBS_1	ribo_mRNA	k_ribo_conversion	
open_ribo_incom_RBS_2	ribo_mRNA	k_ribo_conversion	
open_ribo_incom_RBS_3	ribo_mRNA	k_ribo_conversion	
open_ribo_incom_RBS_4	ribo_mRNA	k_ribo_conversion	
open_ribo_incom_RBS_5	ribo_mRNA	k_ribo_conversion	
open_ribo_incom_RBS_6	ribo_mRNA	k_ribo_conversion	
open_ribo_incom_RBS_small	ribo_mRNA_small	k_ribo_conversion/0.871794871794871	
ribo + mRNA	open_ribo_RBS	k_on_ribo_RBS/(v/1000)	k_off_ribo
open_ribo_RBS	ribo_mRNA + mRNA	k_RBS_clearance	
ribo_mRNA	ribo + protein	k_translation	
ribo + mRNA_small	open_ribo_RBS_small	k_on_ribo_RBS/(v/1000)	k_off_ribo
open_ribo_RBS_small	ribo_mRNA_small + mRNA_small	k_RBS_clearance	
ribo_mRNA_small	ribo + protein_small	k_translation/0.871794871794871	

Fig. 14 Full gene expression model (71 reactions, 47 species), shown in the BioNetS user interface, with columns labelled as in Fig. 12

original model, and its binding and unbinding steps add two reactions. By making the binding constant of *protein_small* to *operon_ns* (0.0004) much larger than the binding constant of *Rpoly* to *operon_ns* (3.99×10^{-6}), most of the *operon_ns* will be occupied by *protein_small*, preventing *Rpoly* from binding to most of the *operon_ns* sites and thus greatly reducing the level of gene expression. Following the reduction method, we obtain a reduced model with 11 species and 12 reactions, consisting essentially of the 10 species/10 reactions reduced model, plus a reaction representing the feedback. A comparison of the reduced and full models is shown in Fig. 11. After fitting the reduced model to the original model, the statistical parameters of the protein species show a good match to those in the original model. However, in a more extreme case with much stronger feedback (a binding constant of *protein_small* to *operon_ns* of 0.01), the reduced model is no longer be able to capture the exact stochastic behavior of the original model (results not shown). Further investigation

Left	Right	Forward	Backward
operon_ns+Rpoly	operon_ns+incom_mRNA_1 + Rpoly_operon_ns_1	0.00036322/v	
Rpoly_operon_ns_1	Rpoly_operon_ns_2+incom_mRNA_2+mRNA	0.0578999	
Rpoly_operon_ns_2	Rpoly_operon_ns_3+incom_mRNA_3+mRNA	0.0578999	
Rpoly_operon_ns_3	Rpoly_operon_ns_4+incom_mRNA_4+mRNA	0.0578999	
Rpoly_operon_ns_4	Rpoly_operon_ns_5+incom_mRNA_5+mRNA	0.0578999	
v	2v	k_cell_div	
	operon_ns	k_rep_operon_ns	
	operon_s	k_rep_operon_s	
	Rpoly	k_rep_Rpoly	
	ribo	k_rep_ribo	
Rpoly_operon_ns_5	Rpoly_operon_ns_6+incom_mRNA_6+mRNA	0.0578999	
ribo + incom_mRNA_1	incom_mRNA_1+ribo_incom_mRNA_1	0.00984106/v	
ribo_incom_mRNA_1	ribo+protein	0.0663217	
ribo + incom_mRNA_2	incom_mRNA_1+ribo_incom_mRNA_2	0.00984106/v	
ribo + incom_mRNA_3	incom_mRNA_1+ribo_incom_mRNA_3	0.00984106/v	
ribo + incom_mRNA_4	incom_mRNA_1+ribo_incom_mRNA_4	0.00984106/v	
ribo + incom_mRNA_5	incom_mRNA_1+ribo_incom_mRNA_5	0.00984106/v	
ribo + incom_mRNA_6	incom_mRNA_1+ribo_incom_mRNA_6	0.00984106/v	
Rpoly_operon_ns_6	mRNA_small+Rpoly+mRNA	0.0578999	
ribo_incom_mRNA_2	ribo+protein	0.0663217	
ribo_incom_mRNA_3	ribo+protein	0.0663217	
ribo_incom_mRNA_4	ribo+protein	0.0663217	
ribo_incom_mRNA_5	ribo+protein	0.0663217	
ribo_incom_mRNA_6	ribo+protein	0.0663217	
ribo+ mRNA	ribo_mRNA + mRNA	0.00984106/v	
ribo_mRNA	ribo + protein	0.0663217	
operon_s+Rpoly	operon_s+Rpoly + stable_RNA	0.0537928/v	
mRNA_small + ribo	mRNA_small + ribo + protein_small	0.00984106/v	

Fig. 15 First reduction of the gene expression model (28 reactions, 29 species), shown in the BioNetS user interface, with columns labelled as in Fig. 12

will be required to elucidate the details of the conditions under which the method can yield satisfactory results for systems incorporating feedback.

4 Discussion

We have described a method of reducing the complexity of stochastic models of sets of biochemical reactions, and demonstrated that complex systems may be reduced significantly while retaining a good match with the statistical features of the original model. Such reductions offer not only faster computational times, but have the advantage of simplifying the system, making it easier to understand and manipulate. Depending on the requirements of the particular modelling effort, varying degrees

Left	Right	Forward	Backward
v	2v	k_cell_div	
	operon_ns	k_rep_operon_ns	
	operon_s	k_rep_operon_s	
	Rpoly	k_rep_Rpoly	
	ribo	k_rep_ribo	
operon_s+Rpoly	operon_s+Rpoly+ stable_RNA	0.0537928/v	
operon_ns+Rpoly	operon_ns+Rpoly+ mRNA	0.00217932/v	
ribo + mRNA	Protein+ribo + mRNA	0.00984106/v	
operon_ns+Rpoly	operon_ns + Rpoly + mRNA_small	0.00036322/v	
ribo + mRNA_small	Protein_small + ribo + mRNA_small	0.00984106/v	

Fig. 16 Second reduction of the gene expression model (10 reactions, 10 species), again given in the form of the BioNetS interface screen, with columns labelled as in Fig. 12

of model reduction may be done, preserving the structure of the original to varying degrees; for example, we have seen that a highly reduced model may preserve only steady-state behaviour rather than transients, while a less-reduced version of the same system can capture both aspects of the behaviour. Application of this approach to other biochemical networks is relatively straightforward, and offers the prospect of a systematic means of stripping away unnecessary detail when examining biological reaction systems.

Acknowledgements This work was supported by NSERC Canada. We thank Tony Pawson and Bruce Seet of the Samuel Lunenfeld Research Institute (Mount Sinai Hospital, Toronto) for their insights into signalling cascades.

Appendix

Important information that was thought to disrupt the flow of the main body of the text is provided here, including all rate constants, species names, and initial values used in the models.

Online supplementary material. The scripts used in this paper are available online at www.utm.utoronto.ca/~mcmillen/JOBP2007-ModelReduction (please see the file Readme.txt for a full description of the contents of the site). This includes a Mathematica script to carry out the algebraic manipulations in the reduction process, MATLAB scripts to carry out the simulated annealing process on each of the models, and BioNetS scripts to perform stochastic simulation of the original and reduced versions of each of the models discussed in Section 3. Information on obtaining and installing BioNetS [25] is available at x.amath.unc.edu/BioNetS.

Rate constants for the signal cascade model. The same rate constants are used in both the full and reduced versions of the signalling cascade model, and their values are (all in units of min^{-1}): $k_1 = 2.1 \times 10^{-5}$, $k_2 = 7.8 \times 10^{-5}$, $k_3 = 163$, $k_4 = k_{20} = 2.1 \times 10^{-4}$, $k_5 = 446$, $k_6 = 8$, $k_7 = k_{24} = 7.1 \times 10^{-4}$, $k_8 = 1200$, $k_9 = k_{10} = k_{16} = k_{22} = 1.1 \times 10^{-5}$, $k_{11} = .066$, $k_{12} = 42$, $k_{13} = 3.6 \times 10^{-5}$, $k_{14} = 9$, $k_{15} = .012$, $k_{17} = 7.1 \times 10^{-5}$, $k_{18} = .17$, $k_{19} = 7.1 \times 10^{-6}$, $k_{21} = 9$, $k_{22} = .006$, $k_{23} = 291$, $k_{25} = 10^{-4}$, $k_{26} = k_{38} = k_{24} = 30$, $k_{27} = 7 \times 10^{-6}$, $k_{28} = 60$, $k_{29} = k_{31} = 8 \times 10^{-5}$, $k_{30} = 210$, $k_{32} = 174$, $k_{33} = k_{35} = 8 \times 10^{-7}$, $k_{34} = 960$, $k_{36} = 342$, $k_{37} = k_{18} = .011$, $k_1 = k_{17} = k_{20} = 3.6$, $k_2 = 49$, $k_4 = 355$, $k_7 = 60$, $k_9 = 3$, $k_{10} = .18$, $k_{11} = .0036$, $k_{12} = 3.6 \times 10^{-6}$, $k_{13} = 5.7$, $k_{16} = 12$, $k_{19} = 6$, $k_{21} = 2.5 \times 10^{-4}$, $k_{23} = 4.6 \times 10^{-6}$, $k_{25} = k_{37} = 78$, $k_{27} = .32$, and $k_{29} = k_{31} = 1.1$.

Rate constants for the full gene expression model. Complete details of the model shown in Fig. 14 have been reported elsewhere [20]. The rate constants are: $k_{\text{RBS_clearance}} = .5$, $k_{\text{cell_div}} = 4.8 \times 10^{-4}$, $k_{\text{isomerization}} = 1$, $k_{\text{off_Rpoly}} = 10$, $k_{\text{off_ribo}} = 2.25$, $k_{\text{on_Rpoly_prom_ns}} = 3.9 \times 10^{-6}$, $k_{\text{on_Rpoly_prom_s}} = 5.9 \times 10^{-4}$, $k_{\text{on_ribo_RBS}} = 5.4 \times 10^{-5}$, $k_{\text{prom_clearance}} = 1$, $k_{\text{rep_Rpoly}} = 5.2$, $k_{\text{rep_operon_ns}} = 1.1$, $k_{\text{rep_operon_s}} = .07$, $k_{\text{rep_ribo}} = 33.3$, $k_{\text{ribo_conversion}} = .05$, $k_{\text{transcription_ns}} = .05$, $k_{\text{transcription_s}} = .09$, and $k_{\text{translation}} = .06$. All

Brief description of species in signal cascade model

Name of species	Description of species	Initial number
IR	Inactive EGFR (epidermal growth factor receptor)	15,000
L	Ligand which in this case is EGF (epidermal growth factor)	100,000
iRL	= epidermal growth factor bound to epidermal growth factor receptor	0
iRLRL	Two epidermal growth factor receptors, each with their own bound epidermal growth factor, bound together forming a dimer (still inactive)	0
RP	Phosphorylated (now active) epidermal growth factor dimer	0
PT1	Tyrosine phosphotases 1	15,000
PT2	Tyrosine phosphotases 2	30,000
SH	Src homology and collagen domain protein A (ShcA)	0
SHIP	Src homology and collagen domain protein A phosphorylated	30,000
GR	Growth factor receptor-binding protein 2 (Grb2)	17,000
SO	Son of sevenless homolog protein	20,000
RD	Ras in its GDP (guanosine diphosphate) bound (inactive) state	20,000
RasGTP	Ras in its GTP (guanosine triphosphate) bound (active) state	0
iRaf	Inactive inactive Raf	0
aRaf	Active active Raf	0
MEK	Mitogen-activated protein kinase	360,000
ERK	Extracellular signal-regulated kinase	750,000

All complexes not shown in this table have a initial value of zero.

The following standardized notation is used for all species:

- - refers to a bound complex (A_B implies a complex where A is bound to B); and
- P - represents a phosphorylated tyrosine residue (EP and EPP represent species E phosphorylated at one and two residues, respectively).

Brief description of species in gene expression model

Name of species	Description of species	Initial number
Rpoly	RNA polymerase	7,600
Rpoly_operon_ns	RNA polymerase elongating an average mRNA transcript from a template operon	0
Rpoly_operon_s	RNA polymerase elongating an average RNA transcript from a template operon	0
closed_Rpoly_prom_ns	RNA polymerase in a closed-complex with an mRNA operon promoter	0
closed_Rpoly_prom_s	RNA polymerase in a closed-complex with an mRNA operon promoter	0
mRNA	Average mRNA (gene length)	0
mRNA_small	Approximately 90% of the average mRNA; all species names that include OsmallO refer to this shorter species and its products/complexes	0
open_Rpoly_prom_ns	RNA polymerase in an open-complex with an mRNA operon promoter	0
open_Rpoly_prom_s	RNA polymerase in an open-complex with a stable RNA operon promoter	0
operon_ns	Average mRNA operon	1,541
operon_s	Average stable RNA operon	104
stable_RNA	Average stable RNA (full operon length)	0
incom_mRNA	Nascent (incomplete) mRNA	0
incom_mRNA_small	Nascent small mRNA	0
open_ribo_incom_RBS	A ribosome bound to the RBS of a nascent mRNA	0
open_ribo_incom_RBS_small	A ribosome bound to the RBS of a small nascent mRNA	0
open_ribo_RBS	A ribosome bound to the RBS of an average mRNA	0
open_ribo_RBS_small	A ribosome bound to the RBS of a small mRNA	0
protein	Average polypeptide	0
protein_small	Polypeptide that is approximately 90% the size of an average polypeptide	0
ribo	Ribosome	48,000
ribo_incom_mRNA	Ribosome currently elongating on a nascent mRNA	0
ribo_incom_mRNA_small	Ribosome currently elongating on a small nascent mRNA	0
ribo_mRNA	Ribosome currently elongating on an average mRNA	0
ribo_mRNA_small	Ribosome currently elongating on a small mRNA	0
v	Volume of the system	1,000

zeroth and first order rate constants are given in units of sec^{-1} , while second order rate constants are given in $\mu\text{m}^3 \times \text{sec}^{-1}$.

Rate constants for first reduction of the gene expression model. The first reduction of the gene expression model (to 28 reactions and 29 species, as shown in Fig. 15) used simulated annealing to adjust the rate constants to obtain a better fit to the original model. The fitted rate constants (with the values from the original model noted when a direct correspondence exists) are: $k_{\text{cell_div}} = 4.8 \times 10^{-4}$ (unchanged), $k_{\text{rep_Rpoly}} = 5.2778$ (vs. 5.2), $k_{\text{rep_operon_ns}} = 1.0978$ (vs. 1.1), $k_{\text{rep_operon_s}} = .0721$ (vs. 0.07), $k_{\text{rep_ribo}} = 33.3$ (unchanged), $k_{\text{ribo_on}} = .0095$, $k_{\text{transcription_initial}} = 3 \times 10^{-4}$, $k_{\text{transcription_ns}} = .071$ (vs. .05), $k_{\text{transcription_stable}} = .0285$ (vs. .09), $k_{\text{translation}} = .051$ (vs. .06), and $k_{\text{translation_small}} = .0094$. As above, all zeroth and first order rate constants are given in units of sec^{-1} , while second order rate constants are given in $\mu\text{m}^3 \times \text{sec}^{-1}$.

Rate constants for second reduction of the gene expression model. The second reduction of the gene expression model (to 10 reactions and 10 species, as shown in Fig. 16) again used simulated annealing to adjust the rate constants to improve the match between the ODE solutions of the reduced and full models. The fitted rate constants (with the values from the original model noted when a direct correspondence exists) are: $k_{\text{cell_div}} = 4.8 \times 10^{-4}$ (unchanged), $k_{\text{rep_Rpoly}} = 5.2778$ (vs. 5.2), $k_{\text{rep_operon_ns}} = 1.0978$ (vs. 1.1), $k_{\text{rep_operon_s}} = .0721$ (vs. 0.07), $k_{\text{rep_ribo}} = 33.3$ (unchanged), $k_{\text{transcription_ns}} = .00167$ (vs. .05), $k_{\text{transcription_s}} = .074$ (vs. .09), $k_{\text{transcription_small}} = 1.1 \times 10^{-4}$, $k_{\text{translation}} = .001$ (vs. .06), and $k_{\text{translation_small}} = .001$. As above, all zeroth and first order rate constants are given in units of sec^{-1} , while second order rate constants are given in $\mu\text{m}^3 \times \text{sec}^{-1}$.

Rate constants for the reduction of the gene expression model with negative feedback. The reduction of the gene expression model with negative feedback (to 12 reactions and 11 species, as shown in Fig. 17) again used simulated annealing to adjust the rate constants to improve the match between the ODE solutions of the reduced and full models. The fitted rate constants (with the values from the original model noted when a direct correspondence exists)

Left	Right	Forward	Backward
v	2v	$k_{\text{cell_div}}$	
	operon_ns	$k_{\text{rep_operon_ns}}$	
	operon_s	$k_{\text{rep_operon_s}}$	
	Rpoly	$k_{\text{rep_Rpoly}}$	
	ribo	$k_{\text{rep_ribo}}$	
operon_s+Rpoly	operon_s+Rpoly+ stable_RNA	$k_{\text{transcription_s/v}}$	
operon_ns+Rpoly	operon_ns+Rpoly+ mRNA	$k_{\text{transcription_ns/v}}$	
ribo + mRNA	Protein+ribo + mRNA	$k_{\text{translation/v}}$	
operon_ns+Rpoly	operon_ns + Rpoly + mRNA_small	$k_{\text{transcription_small/v}}$	
ribo + mRNA_small	Protein_small + ribo + mRNA_small	$k_{\text{translation_small/v}}$	
Protein_small + operon_ns	operon_ns_protein_small	$k_{\text{negative/(v/1000)}}$	$k_{\text{off_Negative}}$

Fig. 17 Reduction of the gene expression model with negative feedback (12 reactions, 11 species), again given in the form of the BioNetS interface screen, with columns labelled as in Fig. 12

are: $k_{\text{cell_div}} = 4.8 \times 10^{-4}$ (unchanged), $k_{\text{rep_Rpoly}} = 5.2778$ (vs. 5.2), $k_{\text{rep_operon_ns}} = 1.0978$ (vs. 1.1), $k_{\text{rep_operon_s}} = .0721$ (vs. 0.07), $k_{\text{rep_ribo}} = 33.3$ (unchanged), $k_{\text{transcription_ns}} = .0017$ (vs. .05), $k_{\text{transcription_s}} = .074$ (vs. .09), $k_{\text{transcription_small}} = 1.69 \times 10^{-4}$, $k_{\text{translation}} = .0109$ (vs. .06), $k_{\text{translation_small}} = .0182$, $k_{\text{negative}} = 0.0004$ (unchanged), and $k_{\text{off_Negative}} = 0.001$ (unchanged). As above, all zeroth and first order rate constants are given in units of sec^{-1} , while second order rate constants are given in $\mu\text{m}^3 \times \text{sec}^{-1}$.

References

1. Kaern, M., Elston, T.C., Blake, W.J., Collins, J.J.: Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev. Genet.* **6**, 451–464 (2005)
2. Hasty, J., McMillen, D., Isaacs, F., Collins, J.J.: Computational studies of gene regulatory networks: *in numero* molecular biology. *Nat. Rev. Genet.* **2**, 268–279 (2001)
3. Ventura, B.D., Lemerle, C., Michalodimitrakis K., Serrano, L.: From in vivo to in silico biology and back. *Nature* **443**, 527–533 (2006)
4. Andrews, S.S., Arkin, A.R.: Simulating cell biology. *Curr. Biol.* **6**, R523–R527 (2006)
5. van Zon, J.S., ten Wolde, P.R.: Simulation biochemical networks at the particle level in time and space: Green's function reaction dynamics. *Phys. Rev. Lett.* **94**, 128103 (2005)
6. Andrews, S., Bray, D.: Stochastic simulation of chemical reactions with spatial resolution and single molecule detail. *Phys. Biol.* **1**, 131–151 (2004)
7. Meng, T.C., Somani, S., Dwar, P.: Modeling and simulation of biological systems with stochasticity. In *Silico Biol.* **4**, 293–309 (2004)
8. Fedoroff, N., Fontana, W.: Genetic networks: small numbers of big molecules. *Science* **297**, 1129–1131 (2002)
9. Ozbudak, E.M., Thattai, M., Kurtser, I., Grossman, A.D., van Oudenaarden, A.: Regulation of noise in the expression of a single gene. *Nat. Genet.* **1**, 69–73 (2002)
10. Elowitz, M.B., Levine, A., Siggia, E., Swain, P.: Stochastic gene expression in a single cell. *Science* **297**, 1183–1186 (2002)
11. Paulsson, J.: Summing up the noise in gene networks. *Nature* **427**, 415–418 (2004)
12. Weng, G.Z., Bhalla, U.S., Iyengar, R.: Complexity in biological signaling systems. *Science* **284**, 92–96 (1999)
13. Okino, M.S., Mavrouniotis, M.L.: Simplification of mathematical models of chemical reaction systems. *Chem. Rev.* **98**, 391–408 (1998)
14. Djouad, R., Sportisse, B.: Partitioning techniques and lumping computation for reducing chemical kinetics. APLA: an automatic partitioning and lumping algorithm. *Appl. Numer. Math.* **43**, 383–398 (2002)
15. Petzold, L., Zhu W.J.: Model reduction for chemical kinetics: an optimization approach. *AIChE J.* **45**, 869–886 (1999)
16. Gorban, A.N., Karlin, I.V., Zinovyev, A.Y.: Constructive methods of invariant manifolds for kinetic problems. *Phys. Rep.* **396**, 197–403 (2004)
17. Valorani, M., Creta, F., Goussis, D.A., Lee, J.C., Najm, H.N.: An automatic procedure for the simplification of chemical kinetic mechanisms based on CSP. *Comb. Flam.* **146**, 29–51 (2006)
18. Vora, N., Daoutidis, P.: Nonlinear model reduction of chemical reaction systems. *AIChE J.* **47**, 2320–2332 (2001)
19. Dong, C.G., Jakobowski, L., McMillen D.R.: Systematic reduction of a stochastic signalling cascade model. *J. Biol. Phys.* **32**, 173–176 (2006)
20. Iafolla, M., McMillen, D.R.: Extracting biochemical parameters for cellular modeling: a mean-field approach. *J. Phys. Chem. B* **110**, 22019–22028 (2006)
21. Kumar, A., Daoutidis, P.: Daoutidis, Control of Nonlinear Differential Alge-braic Equation Systems. *Research Notes in Mathematics*. Chapman and Hall, CRC, London (1999)
22. Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P.: Optimization by simulated annealing. *Science* **220**, 671–680 (1983)

23. Vandekerckhove, J.: anneal.m, obtained from <http://www.mathworks.com/matlabcentral/fileexchange>, accessed on May 31, 2007; Category "Optimization," file ID 10548. Matlab Central File Exchange (2006)
24. Alves, R., Antunes, F., Salvador, A.: Tools for kinetic modeling of biochemical networks. *Nat. Biotechnol.* **24**, 667–672 (2006)
25. Adalsteinsson, D., McMillen, D.R., Elston, T.C.: Biochemical Network Stochastic Simulator (BioNetS): software for stochastic modelling of biochemical networks. *BMC Bioinformatics* **5**:24, (2004)
26. Gillespie, D.: Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**, 2340–2361 (1977)
27. Gibson, M., Bruck, J.: Efficient exact stochastic simulation of chemical systems with many species and many channels. *J. Phys. Chem.* **104**, 1876–1889 (2000)
28. Francisco, C.: Signal transmission by epidermal growth factor receptor: coincidence of activation and dimerization. *Biochem.* **31**, 4493–501 (1992)
29. Venus, L.K., Pawson, T.: The ShcA phosphotyrosine docking protein sensitizes cardiovascular signaling in the mouse embryo. *Genes Dev.* **14**, 1132–45 (2000)
30. Markevich, N.I.: Signal processing at the ras circuit: what shapes ras activation patterns? *Syst. Biol.* **1**, 104 (2004)
31. Shuichi, O., Yamauchi, K., Pessin, J.E.: Shc isoform-specific tyrosine phosphorylation by the insulin and epidermal growth factor receptors. *J. Biol. Chem.* **270**, 20737–20741 (1995)
32. Porter, A.C., Vaillancourt, R.R.: Tyrosine kinase receptor-activated signal transduction pathways which lead to oncogenesis. *Oncogene* **17**, 1343–52 (1998)
33. Schlessinger, J.: Cell signaling by receptor tyrosine kinases. *Cell* **103**, 211–225 (2000)
34. Babu, S.C.V., Yoon, S., Nam, H.-S., Yoo, Y.S.: Simulation and sensitivity analysis of phosphorylation of EGFR signal transduction pathway in PC12 cell model. *Systems Biol.* **1**, 213–221 (2004)