Journal of Biological Physics 32: 173–176, 2006.
DOI: 10.1007/s10867-006-9005-0

Short Note

Systematic Reduction of a Stochastic Signalling Cascade Model

COLIN GUANGQIANG DONG, LUKE JAKOBOWSKI
and DAVID R. McMILLEN
Institute for Optical Sciences and Dept. of Chemical and Physical Sciences, University of Toronto at Mississauga, 3359 Mississauga Rd N, Mississauga, ON L5L 1C6 CANADA

Abstract. Biochemical systems involve chemical reactions occurring in low-number regimes, wherein fluctuations are not negligible and thus stochastic models are required to capture the system behaviour. The resulting models are often 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 propose a method that reduces the complexity of stochastic biochemical network models, and apply this method to the reduction of a mammalian signalling cascade. Our results indicate that the simplified model gives an accurate representation for not only the average number of all species, but also for the associated fluctuations and statistical parameters.

Key words: stochastic biochemical modelling (modeling), model reduction, signalling (signaling, signal) cascade

The goal of achieving a quantitative, systematic understanding of biological phenomena has driven a recent surge of interest in the formulation of mathematical models in biology [1, 2]. The resulting models tend to be complex, exhibiting both nonlinear and stochastic behaviours; further, detailed models quickly grow to include large numbers of interacting species and their associated chemical reactions. Signalling cascades are a classic example of this complexity, involving many participating species interacting in highly branched networks [3]. Models of such systems are computationally expensive and difficult to understand and analyze, and thus any reduction in their complexity is welcome, provided it can be achieved without substantially altering the system's behaviour. Here, we propose a systematic method for reducing the complexity of stochastic biochemical models while keeping their statistical properties unchanged. We apply the method to a mammalian receptor tyrosine kinase signalling cascade, reducing it to substantially fewer reactions and species while maintaining the same overall behaviour.

In time-scale analysis [4], models are reduced by identifying slowly-varying species, and the key problem is the determination of the lower-dimensional space on which the slow species' nonlinear differential equations are constrained to evolve. Several different methods have been proposed to determine this lower dimensional space [6], but none are designed for stochastic models, and difficulties can arise in this case. Based on an approach for deterministic nonisothermal systems [7], we have developed a method specifically to reduce stochastic reaction systems:

1. Convert the reaction system into differential equations $\frac{dx}{dt} = v(x) \cdot r(x)$, where $v(x)$ is the stoichiometric matrix and $r(x)$ is the reaction rate vector. Identify fast and slow reactions, separate the $r(x)$ into fast ones $r_f(x)$and slow ones $r_s(x)$, and pick out the corresponding $v_f(x)$ and $v_s(x)$. Then we get $\frac{dx}{dt} = f(x) + v_f(x) \cdot r_f(x)$, where $f(x) = v_s(x) \cdot r_s(x)$.
2. Calculate the column rank p' of $v_f(x)$. If $v_f(x)$ has full column rank, ${v_f}'(x) = v_f(x)$, ${r_f}'(x) = r_f(x)$. Otherwise select the independent columns of $v_f(x)$ as ${v_f}'(x)$, and set ${r_f}'(x) = \{[{v_f}'(x)]^T[{v_f}'(x)]\}^{-1}[{v_f}'(x)]^T \cdot [v_f(x)]r_f(x)$. We get $\frac{dx}{dt} = f(x) + {v_f}'(x) \cdot {r_f}'(x)$.
3. Calculate the Jacobian of the vector ${r_f}'(x) : J = \frac{\partial {r_f}'(x)}{\partial x}$ and the row rank $p^* = \text{rank}(J)$. If $p^* = p', v_f^*(x) = {v_f}'(x)$ and $r_f^*(x) = {r_f}'(x)$. Otherwise construct the nonsingular matrix $E(x)$ such that $r_f^*(x) = E(x) \cdot {r_f}'(x)$ has the first p^* rows with independent scalar functions, and the last $p' - p^*$ rows identically equal to 0. $v_f^*(x) = {v_f}'(x) \cdot E(x)^{-1}$ and the differential equations become $\frac{dx}{dt} = f(x) + v_f^*(x) \cdot r_f^*(x)$.
4. Solve $r_f^*(x) = 0$, and put the solutions into these differential equations: $\frac{dx}{dt} = f(x) + v_f^*(x) \cdot (L_{vf} r_f^*(x))^{-1}(L_f r_f^*(x))$, where $(L_f r_f^*(x))_i = \sum_{j=1}^{N_s}(\frac{\partial (r_f^*(x))_i}{\partial x_j}) \cdot (f(x))_j (i = 1 \ldots p^*)$, $(L_{vf} r_f^*(x))_{ij} = \sum_{k=1}^{N_s}(\frac{\partial (r_f^*(x))_i}{\partial x_k}) \cdot (v_f^*(x))_{kj}$ $(i, j = 1 \ldots p^*)$. Then we get the reduced differential equations.
5. Translate the differential equations back into a set of reactions whose time evolution may then be simulated stochastically [8].

The mammalian signalling cascade used here as a case study begins at the cell surface with the binding of epidermal growth factor (EGF) to its associated receptor. This binding induces a series of protein binding and phoshorylation events that culminate in the activation of extracellular signal-regulated kinases (ERK) [5]. The subsequent activities of activated ERK (also known as ERKPP), which include translocation into the nucleus of the cell where it activates transcription factors, are not included in the model. The original deterministic model of the signalling cascade contains 41 species and 63 reactions [5], of which a few species are of particular biological interest, including activated ERK. A stochastic model reduction requires the conversion of the chemical reaction system into differential equations; this conversion is then reversed following the reduction process. Due to the different time scales present in the original signalling cascade model, time scale analysis provides a suitable method for reducing the intermediate differential equations.

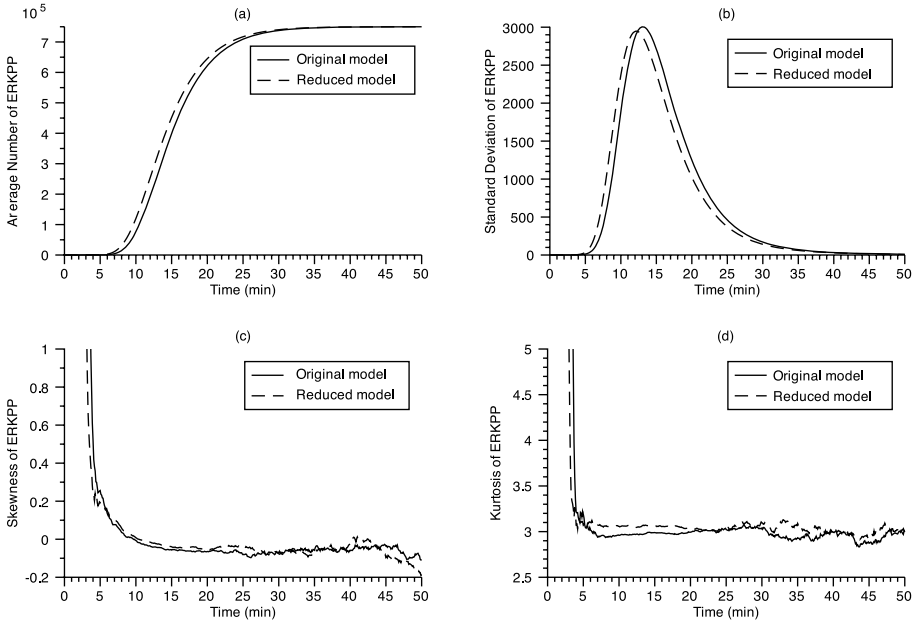

Figure 1. Comparisons of statistical parameters as a function of time for the ERKPP species, from the original and reduced models. In each case, statistics were accumulated over 10000 runs. (a) Mean number of ERKPP. (b) Standard deviation. (c) Skewness. (d) Kurtosis.

The original and reduced signalling cascades were simulated using the Biochemical Network Stochastic simulator (BioNetS), a software package developed for the purpose of efficiently and accurately simulating stochastic models of biochemical networks [8]. The simplification work was carried out using Mathematica. The reaction model is read directly from a BioNetS file and transferred into Mathematica, making it simple to apply to other biochemical network systems.

The reduced model contains 27 species and 34 reactions with an overall reduction in computational time of 50% as compared to the original model. To compare the fluctuations in the original and reduced models, both were run 10000 times with varying random number generator seeds, with a simulated extracellular signal applied to the network at time zero. This yields 10000 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 Figure 1. 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. Some species that have been removed in the model reduction existed only in numbers fewer than 10; the statistical information present in the original model

is satisfactorily preserved despite these excisions. This result points out that small particle numbers in intermediate species do not necessarily translate to significant fluctuations in the "output" of a biochemical cascade. A few species in the reduced model show deviations in their skewness and kurtosis, but the remainder match as well as ERKPP (results not shown).

The method used to reduce the stochastic model gives an accurate representation of the original model while significantly reducing the computational time and presenting the information in a manner that is simpler and easier to manipulate. Extension of this approach to other biochemical networks is straightforward, and offers the prospect of a systematic means of stripping away unnecessary detail when examining biological reaction systems.

This work was supported by NSERC Canada. We thank Tony Pawson and Bruce Seet of the Samuel Lunenfeld Research Institute for their insights into signalling cascades.

References

1. Kaern, M., Elston, T.C., Blake, W.J., and Collins, J.J.: Stochasticity in Gene Expression: From Theories to Phenotypes. *Nat. Rev. Genet.* **6** (2005), 451–464.
2. Hasty, J., McMillen, D., Isaacs, F., and Collins, J.J.: Computational Studies of Gene Regulatory Networks: *in numero* Molecular Biology. *Nat. Rev. Genet.* **2** (2001), 268–279.
3. Weng, G.Z., Bhalla, U.S., and Iyengar, R.: Complexity in Biological Signaling Systems. *Science* **284** (1999), 92–96.
4. Okino, M.S., and Mavrovouniotis, M.L.: Simplification of Mathematical Models of Chemical Reaction Systems. *Chem. Rev.* **98** (1998), 391–408.
5. Babu, S.C.V., Yoon, S., Nam, H.-S., and Yoo, Y.S.: Simulation and Sensitivity Analysis of Phosphorylation of EGFR Signal Transduction Pathway in PC12 Cell Model *Systems Biol.* **1** (2004), 213–221.
6. Gorban, A.N., Karlin, I.V., and Zinovyev, A.Y.: Constructive Methods of Invariant Manifolds for kinetic Problems. *Phys. Rep.* **396** (2004), 197–403.
7. Vora, N., and Daoutidis, P.: Nonlinear Model Reduction of Chemical Reaction Systems, *AICHE J.* **47** (2001), 2320–2332.
8. Adalsteinsson, D., McMillen, D., and Elston, T.C.: Biochemical Network Stochastic Simulator (BioNetS): Software for Stochastic Modelling of Biochemical Networks, *BMC Bioinformatics* **5** (2004), 24.