

Model-Reduction by Simultaneous Determination of Network Topology and Parameters: Application to Modules in Biochemical Networks

Mano Ram Maurya^a, Scott J. Bornheimer^b, Venkat Venkatasubramanian^c and Shankar Subramaniam^{a, b, d, 1}

^aLaboratory for Bioinformatics and Computational Biology, San Diego Supercomputer Center

^bDepartments of Chemistry & Biochemistry and Cellular & Molecular Medicine

^dDepartment of Bioengineering, University of California, San Diego
9500 Gilman Dr La Jolla, CA 92093, USA

^cLaboratory for Intelligent Process Systems, School of Chemical Engineering, Purdue University, West Lafayette, IN 47907, USA

Abstract

Development and analysis of detailed quantitative models of biological systems is computationally prohibitive due to inherent complexity and nonlinearities. Even if a mathematical model can be developed, lack of large-scale good-quality data makes accurate estimation of a large number of parameters impossible. Hence, reduced-order models (ROMs) consisting of essential biochemical mechanisms are more suitable for computational analysis and for studying important systemic functions. In this paper, a mixed-integer nonlinear-optimization-based approach to model-reduction is presented in which, starting with a detailed biochemical model with concomitant computational details (reaction network and mathematical description), the structure and the parameters of a reduced-order model are determined simultaneously. A genetic-algorithm is used to solve the optimization problem. The methodology is demonstrated by developing a ROM for the GTPase cycle module of M1 muscarinic acetylcholine receptor, Gq, and regulator of G-protein signaling 4 (RGS4, a GTPase activating protein or GAP) starting from a detailed model of 48 reactions. The resulting ROM has only 14 reactions, and the ROM fits experimental data well and predicts 4 limiting signaling regimes corresponding to the extremes of receptor and GAP concentration. The method is about an order of magnitude faster than a previous method for reduced-order modeling that used a multiparametric variability analysis approach.

Keywords

Reduced-order modeling, Computational complexity, Mixed-integer nonlinear optimization.

Introduction

The nonlinearity and complexity of biochemical systems is attributed to the presence of numerous chemical species and their complexes, which together participate in reactions and interactions spanning multiple timescales and spatial domains. Computational analysis of these systems is impractical due to unavailability of good-quality data and the computational complexity of simulation. An example of this complexity can be seen in a detailed model for the activation of the MAP kinase pathway by platelet-derived growth factor (PDGF) (Bhalla et al., 2002). This model consists of about 100 nonlinear ordinary differential equations (ODEs) and

¹ Author to whom all correspondences should be addressed: E-mail: shankar@sdsc.edu, Phone: (858) 822 0986, Fax: (858) 822 5722.

algebraic equations and about 200 parameters even without enumerating all possible complexes. Conzelmann et al. (2004), studying the similar epidermal growth factor (EGF) pathway, reported about 40,000 states in a complete model. The complexity becomes even more appreciable when considering the signaling networks of an entire cell. To simplify these complex networks, they can be broken down into distinct modules based upon the underlying sub-processes (functional decomposition), subcellular-location and principle of retroactivity (Asthaigiri and Lauffenburger, 2000; Conzelmann et al., 2004; Hartwell et al., 1999; Neves and Iyengar, 2002). However, the modules themselves can be quite complex. Hence, there is a need to develop methods to reduce the size and complexity of computational models of biochemical networks while retaining functional features and predictive accuracy. With a library of reduced-order models (ROMs) for the modules, a ROM for an entire system can be developed by combining the ROMs for the modules.

The generation of ROMs for linear systems is well studied (Green and Limebeer, 1995); however, for nonlinear systems including most biological systems, model-reduction is not well studied and is not as straight-forward as for linear systems (Conzelmann et al., 2004; Petzold and Zhu, 1999). A model-reduction method for biological models should take into consideration that a) the parameters of the system are seldom accurately measured, and b) the available experimental data may provide a number of context-specific constraints that must be incorporated into the optimization to guide parameter estimation. A number of methods have been proposed to reduce models of chemical systems. These include lumping, sensitivity analysis and time-scale analysis (Okino and Mavrovouniotis, 1998), state-space reduction methods (Obinata and Andersson, 2001) and optimization-based parameter and/or state elimination approaches using genetic algorithm (Edwards et al., 1998) or integer programming (Androulakis, 2000; Bhattacharjee et al., 2003). Conzelmann et al. (2004) have argued that most of these approaches are not suitable for model-reduction of biological systems. In all these methods, it is assumed that the values of the parameters are known. Recently, Maurya et al. (2005) presented a multiparametric variability analysis (MPVA)-based approach in which unknown parameters could be estimated using experimental data. However, the method is recursive and hence time consuming in that the size of the model is reduced in several rounds by eliminating few parameters in each round. In addition, the parameter-elimination space is not searched well even in a pseudo-global sense.

In this article a mixed-integer nonlinear-optimization-based approach to model-reduction is presented in which the structure (e.g., reaction network) and the parameters of the ROM are determined simultaneously by solving the optimization problem using a genetic-algorithm (GA) (Goldberg, 1989). In principle this is a parameter-elimination approach. The method is used to develop a reduced-order model for the GTPase cycle signaling module of m1 muscarinic acetylcholine receptor, Gq, and the GTPase activating protein (GAP) named regulator of G protein signaling 4 (RGS4) starting with the detailed model recently developed by Bornheimer et al. (2004).

Methods: Simultaneous Determination of Network Topology and Estimation of Parameters

In a general parameter-estimation problem, model parameters (p_i) are estimated by minimizing the fit-error between experimental data and model predictions while satisfying appropriate constraints. For model-reduction, the parameter-estimation problem is extended by including binary variables (u_i) to indicate whether or not a parameter is retained in the ROM. The key idea is to substitute each parameter, say, p_i (that can be possibly eliminated) by the expression $u_i^* p_i$, and then to minimize a suitable objective function with respect to both p_i and u_i . $u_i = 1$ or 0 imply that the parameter is retained or eliminated, respectively. Complex expressions in which some parameters should be retained or be eliminated simultaneously can be handled by introducing appropriate constraints. As an example, in a Michaelis-Menten flux-expression, both V_{\max} and K_M should be either retained or eliminated. The objective function is composed of two terms: (1) the number of retained parameters, and (2) an expression to reflect the fit-error so as to differentiate between ROMs with an equal number of retained parameters but different structures. The procedure is as follows:

1. Given the model equations, substitute each parameter p_i by $u_i^* p_i$.
2. Transform the constraints (if any) appropriately.
3. Develop the mixed-integer nonlinear program (MINLP):

$$\begin{aligned}
 \min_{(\{p_i\}, \{u_j\})} \text{obj} &= \sum_j u_j + \alpha * e(\{p_i\}, \{u_j\}) \\
 s/t: e(\{p_i\}, \{u_j\}, \Omega) &\leq e_{th} \\
 h_k(\{p_i\}, \{u_j\}) &= 0, \quad k = 1, \dots, m_1 \\
 g_l(\{p_i\}, \{u_j\}) &\leq 0, \quad l = 1, \dots, m_2 \\
 p_{i, LB} \leq p_i \leq p_{i, UB}, i &= 1, \dots, n_1; u_j = 0/1, j = 1, \dots, n_2
 \end{aligned} \tag{1}$$

where n_1 is the number of parameters to be optimized, n_2 is the total number of parameters that can be eliminated, m_1 is the number of equality constraints and m_2 is the number of inequality constraints. α is a factor used to adjust relative weight of the fit-error e . Different indices, i and j on p and u , respectively, are used to indicate that some of the fixed parameters can also be possibly eliminated and that some of the estimated parameters can be specifically retained if appropriate. The fit-error, e , is usually a weighted sum of squared-errors between model prediction and experimental data. e_{th} is the threshold on fit-error and is usually decided on the basis of acceptable fit-error for the detailed model and error or noise in experimental data.

4. Solve the resulting constrained MINLP using GA (Maurya et al., 2005).

The values of the retained parameters are used for computational purposes. The estimates of values of other parameters may not be used without appropriate justification since they did not play a role in fitting the experimental data even though they satisfied all the constraints on the parameters. In a general case, where all parameters are unknown and they can be potentially eliminated (i.e., $n_1 = n_2$), the computational complexity of the model-reduction process is about twice of the complexity of only parameter estimation. It can be noted that similar to any methodology for model development, the amount of experimental data used and its quality strongly affect the reduced-order model developed.

Test Case Study: A Reduced-Order Model for GTPase-Cycle Module

The GTPase-cycle module controls signal transduction in heterotrimeric G protein signaling networks by regulating the activity of heterotrimeric G proteins, and serves as a key control point for many cellular processes such as gene-transcription and cell-cycle regulation (Neves et al., 2002). In the module, G protein coupled receptors (GPCRs) activate G proteins by accelerating the exchange of GDP for GTP, and GTPase activating proteins (GAPs) deactivate G proteins by accelerating hydrolysis of GTP to GDP. The detailed model of GTPase-cycle module presented by Bornheimer et al. (2004) is shown in Figure 1A (and includes the reactions shown both in black and grey). A_i ($i = 1, 2, 3$, etc.) denotes exchange (association or dissociation) of GAP, R_i denotes exchange of GPCR, T_i denotes exchange of GTP, P_i denotes exchange of a phosphate (GTP hydrolysis reaction) and D_i denotes exchange of GDP. The detailed model consists of 17 ODEs and 48 reaction rate parameters. The key model outputs are the fraction of active G protein (Z), and GTP turnover rate (v) in steady state (Bornheimer et al., 2004). This model fitted the experimental data well (Figure 1B) and predicted that local concentrations of active receptor and GAP regulate Z and v within 4 *limiting signaling regimes* (LSRs) (Z : similar to Figure 1C; v : not shown). In the LSRs, active R and GAP are either absent or are present in saturating concentrations, such that only one of the 4 *extreme pathways* (i.e., $G \rightarrow G^*T \rightarrow GD \rightarrow G$, $GA \rightarrow G^*AT \rightarrow GAD \rightarrow GA$, etc. (Fig. 1A)) is active. In LSR G low peak response (Z) and slow signaling dynamics (v) are achieved, whereas low peak and rapid signaling is achieved in LSR GA. LSR RG is specialized for slow (sustained) signaling with large peak and LSR RGA for both large peak and rapid signaling. The Physical significance of the LSRs is that by controlling the local concentration of active receptor and GAP, cells can exhibit a variety of response.

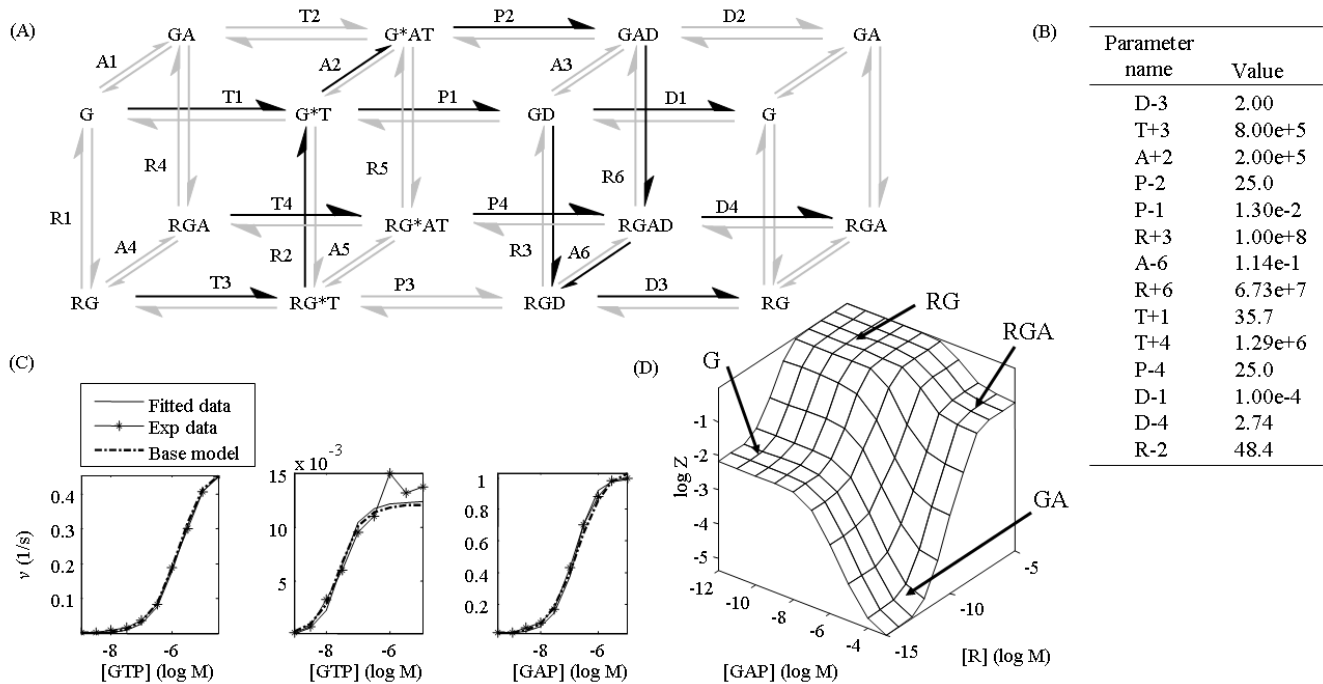


Figure 1. (A) Reaction network for the detailed model (all reactions) (Bornheimer et al., 2004) and the 14 parameter ROM (reactions in dark-color) of the GTPase-cycle module. G (G*): G-protein (active form), R: active G-protein coupled receptor (GPCR), A: GAP, T: GTP, D: GDP.

Free GTP, GDP and phosphate (P_i) are not shown for simplicity. (B) The values of the retained parameters: the units of association rate constants (e.g., T+3, A+2) are $1/(s.M)$ and the units of dissociation rate constants (e.g., D-3, P-2) are $1/s$. (C) Fit of the data predicted by ROM and comparison with the detailed (base) model. (D) The ROM predicts 4 limiting signaling regimes.

A reduced-order model of the GTPase-cycle module with 14 parameters was generated using the method described in the previous section. The data used for optimization consisted of the experimental data on GTP turnover rate under different conditions (Figure 1B), 2 data points on Z (Bornheimer et al., 2004), and three data points on Z from the region between LSR G and LSR GA. Several constraints based on experimental data were described for the detailed model (Bornheimer et al., 2004); all were imposed on the network during model-reduction; in addition, to ensure the connectivity of each species in the reduced-network, a constraint was imposed to retain at least one of the reactions A+1, A+2 or A+3 (binding of GAP to G-protein). As shown in Figure 1B and 1C, the ROM fitted the experimental data well and predicts the existence of 4 LSRs. The ROM retained the ternary complexes RGA, RG^*AT and RGAD, thus validating the ternary complex hypothesis proposed by Biddlecome et al. (1996) and recently used by Bornheimer et al. (2004). An estimate of the values of the parameters retained is listed in Figure 1D. GA-based optimization identified several other solutions with the same network topology (same parameters were retained) but somewhat different values of the estimated parameters as well as solutions with different network topology and different values of the estimated parameters. The 14 parameter ROM is the smallest ROM from 5 runs of GA.

As is true about any model, new data may necessitate model-revision, potentially leading to expansion of a reduced-order model or a detail model. Indeed, for the GTPase case study, when additional data on the LSRs (v) predicted by the detailed model is added and it is required that the reactions A+5 and A-6 be retained, a model with 17 parameters is obtained (not shown). Since this new reduced-order model (17-parameter ROM) captures the LSRs in both the Z -space and v -space, its range of validity is expected to be much better than that of the 14-parameter ROM.

Discussion and Conclusions

Restricting the size of computational models is a major problem in systems biology. We developed a MINLP-based method for model-reduction and tested it on a detailed model of the GTPase-cycle module. The ROM thus obtained is comparable to (good fit to data, prediction of the 4 LSRs and retaining the active ternary complex RG^*AT) or is better (slightly smaller size) than a ROM developed earlier by using a MPVA approach (Maurya et al., 2005). The predictions of the ROM qualitatively agree with other experiments on regulation of G-protein activity (e.g., dose-response curve and decrease in half-saturation concentration of pheromone α -factor in RGS vs. RGS-knockout strains of yeast (Yi et al., 2003)). This method is particularly suitable for model-reduction of biological systems which are nonlinear and for which little is known about parameter values. In general, the methodology is also expected to result in substantial reduction of models of highly connected modules.

Acknowledgements

This work was supported by National Institute of Health Grant U54 GM62114 (SS) and a grant from Hilblom Foundation. We thank Prof. Elliot M. Ross (The University of Texas Southwestern Medical Center) for the experimental data used in this work.

References

- Androulakis, I. P. (2000). Kinetic mechanism reduction based on an integer programming approach. *AIChE Journal*, 46, 361-371.
- Asthagiri, A. R., Lauffenburger, D. A. (2000). Bioengineering models of cell signaling. *Annual Review of Biomedical Engineering*, 2, 31-53.
- Bhalla, U. S., Ram, P. T., Iyengar, R. (2002). MAP kinase phosphatase as a locus of flexibility in a mitogen-activated protein kinase signaling network. *Science*, 297, 1018-1023.
- Bhattacharjee, B., Schwer, D. A., Barton, P. I., Green, W. H. (2003). Optimally-reduced kinetic models: reaction elimination in large-scale kinetic mechanisms. *Combustion and Flame*, 135, 191-208.
- Biddlecome, G. H., Berstein, G., Ross, E. M. (1996). Regulation of phospholipase C-beta 1 by G(q) and m1 muscarinic cholinergic receptor - Steady-state balance of receptor-mediated activation and GTPase-activating protein-promoted deactivation. *Journal of Biological Chemistry*, 271, 7999-8007.
- Bornheimer, S. J., Maurya, M. R., Farquhar, M. G., Subramaniam, S. (2004). Computational modeling reveals how interplay between components of a GTPase-cycle module regulates signal transduction. *Proc Natl Acad Sci USA*, 101, 15899-15904.
- Conzelmann, H., Saez-Rodriguez, J., Sauter, T., Bullinger, E., Allgower, F., Gilles, E. D. (2004). Reduction of mathematical models of signal transduction networks: Simulation-based approach applied to EGF receptor signaling. *Systems Biology*, 1, 159-169.
- Edwards, K., Edgar, T. F., Manousiouthakis, V. I. (1998). Kinetic model reduction using genetic algorithms. *Computers & Chemical Engineering*, 22, 239-246.
- Goldberg, D. E. (1989). *Genetic algorithms in search, optimization and machine learning*, Addison-Wesley, Reading, MA.
- Green, M., Limebeer, D. J. N. (1995). *Linear robust control*, Prentice Hall.
- Hartwell, L. H., Hopfield, J. J., Leibler, S., Murray, A. W. (1999). From molecular to modular cell biology. *Nature*, 402, C47-C52.
- Maurya, M. R., Bornheimer, S. J., Venkatasubramanian, V., Subramaniam, S. (2005). Reduced-order modeling of biochemical networks: Application to the GTPase-cycle signaling module. *Systems Biology*, accepted.
- Neves, S. R., Iyengar, R. (2002). Modeling of signaling networks. *Bioessays*, 24, 1110-1117.
- Neves, S. R., Ram, P. T., Iyengar, R. (2002). G protein pathways. *Science*, 296, 1636-1639.
- Obinata, G., Andersson, B. D. O. (2001). *Model reduction for control system design*, Springer, London.
- Okino, M. S., Mavrovouniotis, M. L. (1998). Simplification of mathematical models of chemical reaction systems. *Chemical Reviews*, 98, 391-408.
- Petzold, L., Zhu, W. J. (1999). Model reduction for chemical kinetics: An optimization approach. *AIChE Journal*, 45, 869-886.
- Yi, T. M., Kitano, H., Simon, M. I. (2003). A quantitative characterization of the yeast heterotrimeric G protein cycle. *Proc Natl Acad Sci USA*, 100, 10764-10769.