BIOINFORMATICS APPLICATIONS NOTE Vol. 24 no. 8 2008, pages 1109–1111
doi:10.1093/bioinformatics/btn061

Systems biology

In silico Biochemical Reaction Network Analysis (IBRENA): a package for simulation and analysis of reaction networks

Gang Liu[1] and Sriram Neelamegham[1,2,*]

[1]Department of Chemical and Biological Engineering, and [2]NY State Center for Excellence in Life Science and Bioinformatics, State University of New York, Buffalo, NY 14260, USA

Received on December 7, 2007; revised on February 11, 2008; accepted on February 12, 2008

Advance Access publication February 28, 2008

Associate Editor: Limsoon Wong

ABSTRACT

Summary: We present *In silico* Biochemical Reaction Network Analysis (IBRENA), a software package which facilitates multiple functions including cellular reaction network simulation and sensitivity analysis (both forward and adjoint methods), coupled with principal component analysis, singular-value decomposition and model reduction. The software features a graphical user interface that aids simulation and plotting of *in silico* results. While the primary focus is to aid formulation, testing and reduction of theoretical biochemical reaction networks, the program can also be used for analysis of high-throughput genomic and proteomic data.

Availability: The software package, manual and examples are available at http://www.eng.buffalo.edu/~neel/ibrena

Contact: neel@eng.buffalo.edu

1 INTRODUCTION

Cellular reaction networks, for example models of signal transduction, can be simulated when biochemical information about the nature of protein interactions, and kinetic information about individual rate constants/enzyme kinetics is available. While the above information may be inferred using independent, isolated biochemical assays, *in silico* simulations combined with cell-based assays help test the degree to which *ex vivo* measurements translate to the real cellular environment.

Computational tools for simulation of biochemical reaction networks have been developed, including Gepasi (Mendes, 1993), E-cell (Tomita *et al*., 1999), Bionetgen (Blinov *et al*., 2004) and others that are listed at SBML.org. Most programs are either visualization tools for reaction networks, or they are stochastic/deterministic simulators. They place less emphasis on post-simulation analysis of data. To address this limitation, we developed a new software package called '*In silico* Biochemical Reaction Network Analysis (IBRENA)'. The coupling of forward sensitivity analysis (Liu *et al*., 2005) with multivariate analysis techniques, especially principal component analysis (PCA) and singular value decomposition (SVD), is a unique feature of this software that is not implemented in previous programs. In addition, this program features adjoint sensitivity analysis and model reduction methods. Overall, the compilation of individual routines into a suite with focus on post-simulation analysis is a useful tool for the identification of rate-limiting reactions/species, automated determination of connectivity between the reactions/species, and model reduction.

*To whom correspondence should be addressed.

2 FEATURES AND IMPLEMENTATION

IBRENA is a mixed-language program using MATLAB and FORTRAN. While simple matrix computations and the Graphical User Interface (GUI) interface are coded in MATLAB, FORTRAN in some instances is used to solve ordinary differential equations for enhanced computational efficiency. The compiling of FORTRAN codes into MATLAB Executable (MEX) file seamlessly links the code. This software is available in two versions: (a) a stand-alone application which accepts input from Excel Comma Separated Value files (.csv). This version implements most, but not all, of the program's functions and (b) a full-version that runs in the MATLAB environment. This uses Systems Biology Markup Language (SBML) format input. Detailed differences between the two versions including installation instructions are outlined in Chapter 1 of the software manual.

The software features a GUI with built-in plot panels for rapid visualization and interpretation of results. Simulation results can also be stored in local files for plotting using other software. IBRENA has been tested using two *in silico* biochemical reaction networks that examine Epidermal Growth Factor (EGF) (Liu *et al*., 2005) and Tumor Necrosis Factor (TNF) reaction pathways. Microarray data (Alter *et al*., 2000) have been used to test components of the program that relate to multivariate analysis. Detailed instructions including mathematical derivations and examples are provided with the software. IBRENA has been tested using Windows 2000 and XP operating systems.

Figure 1A shows an overview of the software. Here, the input for the model includes files that contain information about: (1) reaction rate expression, (2) initial concentration and (3) rate constant. These can be provided either in .csv or SBML format. Conversion from .csv to SBML format files is possible using IBRENA.

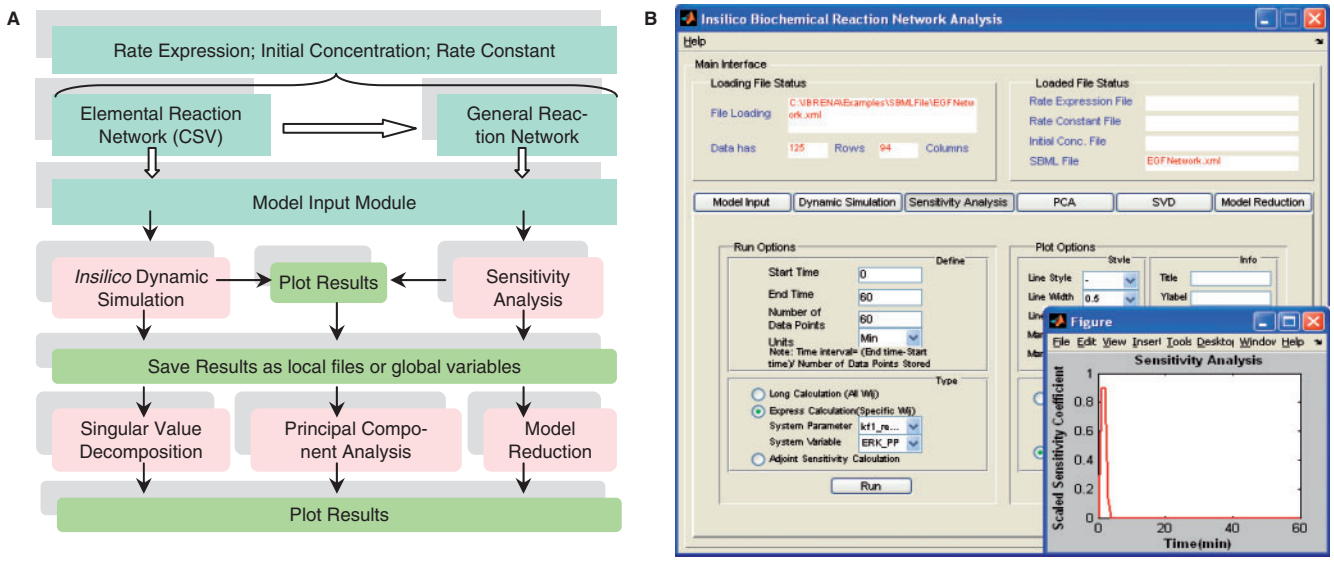

Fig. 1. (**A**) Software implementation flow chart. (**B**) Screenshot of IBRENA GUI highlighting the Sensitivity Analysis option.

Once these data are loaded, the user can run the 'Dynamic Simulation' function, which monitors time-dependent changes of various species concentrations, **C**, using:

$$\frac{d\mathbf{C}}{dt} = f(\mathbf{k}, \mathbf{C}) = \alpha^{T} \cdot \mathbf{v}$$
$$\mathbf{C}(t = 0) = \mathbf{C_0} \quad (1)$$

Here, **k**, **v** and $\mathbf{C_0}$ are matrices containing the rate constants, velocities and initial species concentrations, respectively. α^{T} is the transpose of the stoichiometric coefficient matrix. t denotes time.

'Sensitivity Analysis' can also be performed on the reaction network using forward sensitivity analysis (Liu *et al.*, 2005) or using backward/adjoint sensitivity analysis methods (Cao *et al.*, 2003). The later are enabled using functions outlined in SUNDIALS (Hindmarsh *et al.*, 2005). In such analysis, the effect of infinitesimal changes in system parameters (individual reaction rate constants, k_j) on system variables (species concentration, C_i) is monitored in terms of the scaled sensitivity coefficient, W_{ij}:

$$W_{ij} \equiv \frac{\partial C_i k_j}{\partial k_j C_i} \quad (2)$$

Since large W_{ij} correspond to those system parameters the perturbation of which can drastically affect the system variable/output of interest, sensitivity analysis helps identify rate-limiting reactions in the reaction network. Such analysis also helps identify the time-scales in which particular system parameters significantly affect system variable, and it aids the design of knock-in/-out experiments. Following 'Dynamic Simulation' and 'Sensitivity Analysis' global variables are either plotted or they are stored locally.

A sample plot showing the time dependent evolution of a single W_{ij} is shown in Figure 1B. Based on this, it is noted that the volume of W_{ij} data increases with the size of the biochemical network and experiment time-scale. Furthermore, consideration of a wide range of model input (i.e. different $\mathbf{C_0}$ values) can result in further expansion of W_{ij} output. Similar to the above, dynamic *in silico* simulations also yield vast amounts of concentration (C_i) data that are difficult to analyze individually.

In order to reduce the dimensionality of the dataset and to facilitate grouping of reactions into functionally related components, PCA, SVD and model reduction algorithms are applied to both dynamic simulation and sensitivity analysis results. For these modules, previously generated local files using IBRENA may be used as input data. Alternatively, multivariate data generated using experiments or other simulation packages can be loaded.

PCA has been applied to proteomic data (Janes *et al.*, 2004). It has also be applied to analyze the response of a biochemical reaction network to perturbation (Liu *et al.*, 2005). In the later context, the response of a biochemical reaction to perturbation of the kth rate constant, $Q(k)$, is shown to vary in proportional to the square, symmetric matrix $\mathbf{S^T S}$. Eigenvalue-eigenvector analysis results in:

$$\mathbf{S^T S} = \mathbf{U \Lambda U^T} \quad (3)$$

with $\mathbf{\Lambda}$ representing a diagonal matrix containing the eigenvalues and **U** containing the uncorrelated and orthogonal eigenvectors or principal components (PC). The magnitude of individual elements of $\mathbf{\Lambda}$ quantify the relative importance of the corresponding eigenvector. The elements of the eigenvector represent closely coupled combinations of network elements that contribute independently to biological system response.

SVD has been applied to analyze gene expression and simulation data. This technique involves the decomposition of the data matrix **X** (containing either *in silico* or experimental results) using:

$$\mathbf{X} = \mathbf{U \Sigma V^T} \quad (4)$$

Here, the columns of **U** and **V** are orthogonal, and are termed the left and right singular vector, respectively. Σ is a diagonal matrix with singular values arranged in descending order along the main diagonal. The singular values are a measure of the importance of the corresponding left/right singular vectors and they help determine fundamental patterns/modes, the linear combination of which can explain the complex output of biological processes.

In a final aspect, we implemented one possible strategy to automate model reduction. This routine enables deletion of redundant reactions with the goal of reducing model complexity. This feature is based on the observation that reactions with minimal contribution to system output exhibit two features: (i) They have low W_{ij}, i.e. perturbation of these reactions results in little changes to system output. (ii) Their reaction flux is low, i.e. they are not very active at any time during the course of the simulation (Liu *et al.*, 2005).

ACKNOWLEDGEMENTS

The authors acknowledge grant support from NIH HL63014.

Conflict of Interest: none declared.

REFERENCES

Alter,O. *et al.* (2000) Singular value decomposition for genome-wide expression data processing and modeling. *Proc. Natl Acad. Sci. USA*, **97**, 10101–10106.

Blinov,M.L. *et al.* (2004) BioNetGen: software for rule-based modeling of signal transduction based on the interactions of molecular domains. *Bioinformatics*, **20**, 3289–3291.

Cao,Y. *et al.* (2003) Adjoint sensitivity analysis or differential-algebraic equations: the adjoint DAE system and its numerical solution. *Siam Journal on Scientific Computing*, **24**, 1076–1089.

Hindmarsh,A.C. *et al.* (2005) SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers. *Acm Transactions on Mathematical Software*, **31**, 363–396.

Janes,K.A. *et al.* (2004) Cue-signal-response analysis of TNF-induced apoptosis by partial least squares regression of dynamic multivariate data. *J. Computat. Biol.*, **11**, 544–561.

Liu,G. *et al.* (2005) Sensitivity, principal component and flux analysis applied to signal transduction: the case of epidermal growth factor mediated signaling. *Bioinformatics*, **21**, 1194–1202.

Mendes,P. (1993) Gepasi - a software package for modeling the dynamics, steady-states and control of biochemical and other systems. *Comput. Appl. Biosci.*, **9**, 563–571.

Tomita,M. *et al.* (1999) E-CELL: software environment for whole-cell simulation. *Bioinformatics*, **15**, 72–84.