Tav4SB: grid environment for analysis of kinetic models of biological systems*

Mikołaj Rybiński[1,2], Michał Lula[1], Sławomir Lasota[1] and Anna Gambin[1,2]

[1] Institute of Informatics, University of Warsaw
[2] Mossakowski Medical Research Centre, Polish Academy of Sciences

Abstract. Taverna Workbench eases integration of software tools for life science research in experiments expressed as workflows. The Taverna services for Systems Biology (Tav4SB) project provides a set of new Web service operations which extend the functionality of Taverna Workbench in the systems biology domain. Tav4SB operations allow to perform numerical simulations or model checking of, respectively, deterministic or stochastic semantics of biological models. To visualize the results of model analysis a flexible plotting operation is provided as well. Tav4SB operations are executed in a grid environment, integrating heterogeneous software such as Mathematica, PRISM and SBML ODE Solver. User guide, contact information and full documentation of available Web service operations, exemplary workflows and other, additional resources can be found at the Tav4SB project's Web page: http://bioputer.mimuw.edu.pl/tav4sb/.

Introduction. The Taverna Workbench [11] is a tool which facilitates the design and execution of the *in silico* experiments. Experiments are constructed as workflows which can be stored and executed when needed. The building blocks of a workflow are services, called processors. Technically, workflow is a set of processors, together with connections between their inputs and outputs. The remote processors are implemented as Web service (WS) operations. Scattered physically throughout computational resources of numerous scientific facilities, combined WSs allow to perform highly complex analyses, surpassing power of a standard workstation.

Taverna services come from a diverse set of life sciences domains. In the field of computational biology, Taverna mainly provides services related to sequence annotation and analysis. Here, we present remote processors that extend Taverna's functionality in the systems biology domain, specifically, in the analysis of kinetic models of biological systems. Our hardware base offers computational resources sufficient for computationally demanding experiments, such as multiple invocations of the model-checking procedure. Essentially, Taverna Workbench provides a convenient user interface for our WS operations. Analysis of the behavior of cellular systems under various conditions can be conducted without the need of programming own WS client.

Main features of Tav4SB. Mathematical framework determines the structure of the kinetic formulation for a given biochemical network model. The most common repre-

* This work was partially supported by the Polish government grant N N206 356036 and by the Biocentrum Ochota project (POIG.02.03.00-00-003/09). The first author is a scholar within the Human Capital Operational Programme financed by European Social Fund and state budget.

sentations are ordinary differential equations (ODEs) for the deterministic framework and continuous–time Markov chain (CTMC) for the stochastic framework [1,14].

Operations provided by our Web server allow to perform: (1) numerical simulations for the deterministic formulation of the network model with use of the SBML ODE Solver library[13], (2) probabilistic model checking of Continuous Stochastic Logic (CSL) [2] formula over a CTMC with use of the PRISM tool[9], and (3) visualization of data series such as ODEs trajectories, or values of parametrized CSL properties, with use of the Mathematica tool (Wolfram Research, Inc., 2008,Version 7.0).

The SBML ODE Solver library enables numerical analysis of models encoded directly in the Systems Biology Markup Language (SBML) [10], the standard of our choice. The library employs libSBML [4] to automatically derive ODEs plus their Jacobian and higher derivatives as well as CVODES package — state of the art numerical integration library from SUNDIALS [8].

PRISM is one of the leading tools implementing probabilistic model checking, a technique of formal verification of systems that exhibit a stochastic behavior. A system to be analyzed is modeled as a Markov chain, and a correctness property is expressed in a suitable probabilistic temporal logic. Some recent works, see e.g. [7,12], demonstrate applicability of PRISM to analysis of models of biological systems. Case studies include models of cell cycle control, fibroblast growth factor signaling, and MAPK cascade. For biological applications, a CTMC (continuous-time Markov chain) is typically chosen as an underlying model, and the properties are specified in a continuous time logic, for instance in CSL. The approach seems promising as it often can yield a better understanding of the dynamics of systems to be analyzed.

Wolfram's Mathematica has one of the most advanced graphics capabilities among computer mathematics tools. Tav4SB provides Mathematica's two– and three–dimensional list plots together with a versatile set of options for customizing their display.

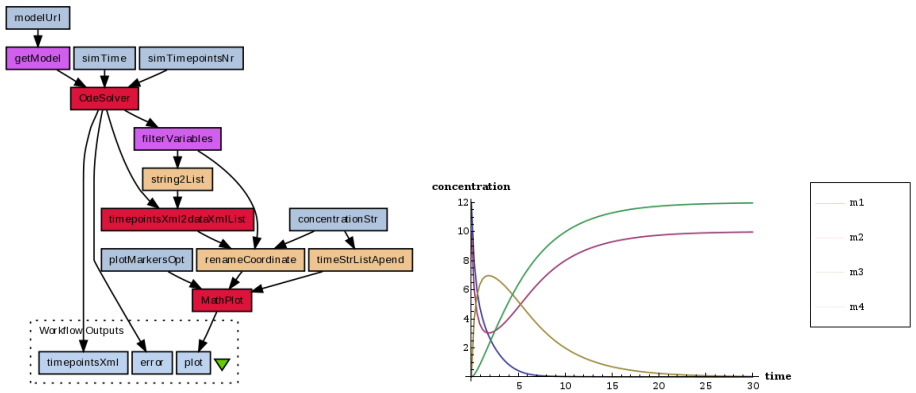

Fig. 1. The "Simulate SBML–derived ODEs" workflow and resulting trajectories plot for the enzymatic reaction model of [5]. The red boxes represent nested workflows, corresponding to Tav4SB WS operation wrappers and a helper. See text for more details.

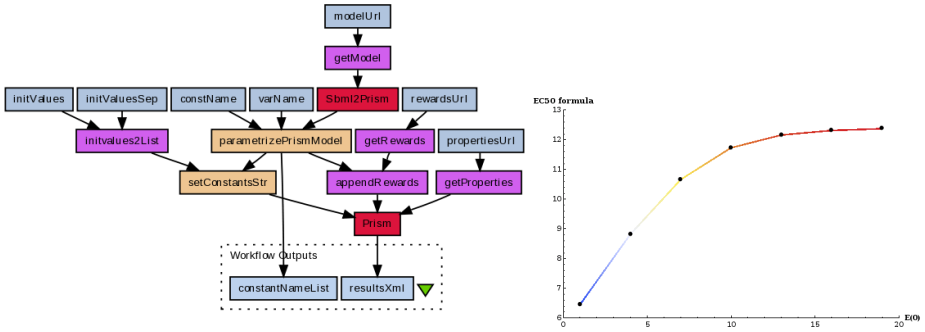

Fig. 2. The computational part of the "Probabilistic model checking of the SBML stochastic model" workflow and the resulting plot for the stochastic version of the enzymatic reaction model. The red boxes represent nested workflows, corresponding to Tav4SB WS operation wrappers. See text for more details.

Exemplary use cases. We constructed a set of exemplary workflows. Their main purpose is to demonstrate usage of the Tav4SB WS operations from the Taverna Workbench client. There are two kinds of exemplary workflows: Tav4SB WS operation wrappers and exemplary *in silico* experiments.

The wrapper workflows illustrate the direct usage of Tav4SB operations in Taverna. Their purpose is to be re–used as a nested workflows, as demonstrated in two exemplary experiments described below. Additionally, we built a number of helper Taverna processors, used for interacting with XML–formatted inputs and outputs of the WS operations. Those helpers are standard Taverna's XML splitters and local services as well as additional BeanShell scripts.

In our two exemplary experiments we have used an enzymatic reaction model with species names and parameters values from the [5]. The first workflow numerically simulates models' ODEs and plots the results. ODEs are derived automatically from the SBML model file, based on the rate laws for described reactions. The enzymatic reaction deterministic model contains mass–action kinetics rates. As a result of running this simple experiment one gets the trajectories of the species (ODEs) variables, together with their plot. Figure 1 depicts the simulation workflow and the resulting plot for all model's species variables over a 30 seconds time period.

The second experiment runs a probabilistic model checking for the stochastic version of the enzymatic reaction model of [5], also encoded in the SBML format. The reward–based CSL formula, which is being checked, is

$$\mathrm{R}_{\text{"\#}r1\text{"}=?}[\mathrm{F}(p > 0.5 * \lim_{t\to\infty} p(t))].$$

Roughly speaking, the formula answers the following question: how many times, on average, the enzyme-substrate complex association reaction $r1$ has to occur before the amount of product p reaches 50% of its maximum? This corresponds to the half maximal effective concentration (EC_{50} coefficient). The formula is evaluated for different

enzyme initial amounts to find its optimal efficiency. As this is a time consuming task, and plotting usually requires many runs to fine–tune the plot parameters, the experiment is divided into two separate parts: a computational part and a plotting part. Figure 2 depicts the computational part of the workflow and the resulting plot.

The plot can be read as: if $E(0)$ is equal to 1 then on average, before product reaches half of its maximum, each enzyme has to convert slightly more than 6 substrates. To no surprise, when $E(0)$ is equal to 12 — the initial amount of substrate, each enzyme has to convert at most one substrate. The total, parallel enzymatic reaction system's efficiency doesn't improve significantly from that point as not much more than 12 complex formation reactions $r1$ are needed to achieve half of the maximum product amount.

Availability. The definition of the operations provided by the Tav4SB WS plus exemplary workflows files, together with installation and execution instructions are available from the project's Web page: http://bioputer.mimuw.edu.pl/tav4sb/. Documentation of the Tav4SB WS can be found in the BioCatalogue [3], a curated catalogue of life sciences Web services. Wrappers and experiments workflows are also available from the myExperiment repository [6], together with the workflow figures.

References

1. Aldridge, B. B. et al.: Physicochemical modelling of cell signalling pathways. Nature Cell Biology 8(11), 1195–1203 (2006)
2. Aziz, A. et al.: Verifying continuous time markov chains. In: Proc. 8th International Conference on Computer Aided Verification (CAV'96). LNCS, vol. 1102, pp. 269–276 (1996)
3. Bhagat, J. et al.: BioCatalogue: a universal catalogue of web services for the life sciences. Nucleic Acids Research 38(Web Server issue), W689–694 (2010)
4. Bornstein, B. J. et al.: LibSBML: an API library for SBML. Bioinformatics (Oxford, England) 24(6), 880–1 (2008)
5. Cho, K.-H. et al.: Experimental design in systems biology, based on parameter sensitivity analysis using a monte carlo method: A case study for the TNFα–mediated NF–κB signal transduction pathway. SIMULATION 79, 726–739 (2003)
6. Goble, C. A. et al.: myExperiment: a repository and social network for the sharing of bioinformatics workflows. Nucleic Acids Research 38(Web Server issue), W677–682 (2010)
7. Heath, J. et al.: Probabilistic model checking of complex biological pathways. Theor Comput Sci 391(3), 239–257 (2008)
8. Hindmarsh, A. C. et al.: SUNDIALS: suite of nonlinear and differential/algebraic equation solvers. ACM Transactions on Mathematical Software 31(3), 363–396 (2005)
9. Hinton, A. et al.: PRISM: A tool for automatic verification of probabilistic systems. Lect Notes Comput Sc 3920, 441–444 (2006)
10. Hucka, M. et al.: The systems biology markup language (sbml): a medium for representation and exchange of biochemical network models. Bioinformatics 19, 524–531 (2003)
11. Hull, D. et al.: Taverna: a tool for building and running workflows of services. Nucleic Acids Res 34(Web Server issue), W729–732 (2006)
12. Kwiatkowska, M. et al.: Using probabilistic model checking in systems biology. ACM SIGMETRICS Performance Evaluation Review 35(4), 14–21 (2008)
13. Machné, R. et al.: The SBML ODE solver library: a native API. for symbolic and fast numerical analysis of reaction networks. Bioinformatics 22, 1406–1407 (2006)
14. Wolkenhauer, O. et al.: Modeling and simulation of intracellular dynamics: choosing an appropriate framework. IEEE Transactions on Nanobioscience 3(3), 200–207 (2004)