

Taverna services for analysis of kinetic models of biological systems

Mikołaj Rybiński

Institute of Informatics, University of Warsaw
Mossakowski Medical Research Centre, Polish Academy of Sciences

Ustroń, October, 2nd, 2010
8th Workshop on Bioinformatics

Outline

- 1 Mathematical modelling of biochemical reactions
- 2 Taverna Workbench
- 3 tav4sb framework
- 4 Sample applications

Simple enzymatic reaction modelling example

1 Define model structure



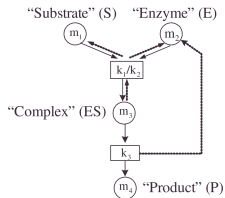
2 Choose kinetics (mathematical framework)

- deterministic framework: ordinary differential equations (ODEs) with mass action (polynomial) or Michaelis–Menten (Hill equation) kinetics.
- stochastic framework: continuous–time Markov process/chain (CTMC) with simple, combinatorial collision rates.

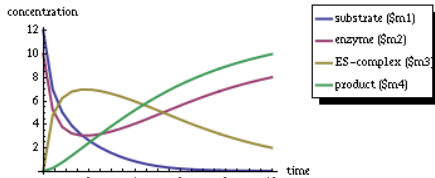
3 Parametrize model — identify the relevant parameters.

Simple enzymatic reaction ODE modelling in pictures

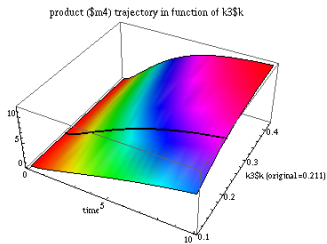
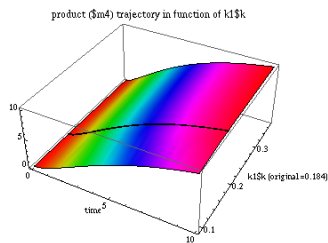
1 Model structure



2 Kinetics (ODE numerical simulations)



3 Model parametrization (OAT sensitivity)



The Taverna Workbench

Taverna: Computational biology tool for facilitating the design and execution of *in silico* experiments.

What is it for?

- Provides **easy access to various Web Services** available transparently via the Internet.
- Allows to model **experiments as workflows**.
- **Integrates** variety of **tools** (different programming languages, different platforms).



What are the benefits?

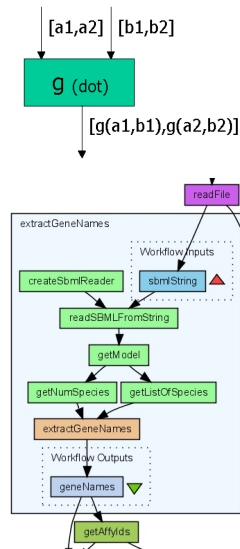
- Experiments are easily repeatable.
- Each run is fully documented.
- The provenance of data is recorded.

Taverna workflows

Taverna workflow: set of processors (local scripts or remote Web services) with connections between their inputs and outputs.

- Workflows are protocols for scientific experiments:
 - SCUFL/T2FLOW format;
 - ^{my}Experiment workflows repository.
- Large set of available WSs: EMBL–EBI services, NCBI Entrez Programming Utilities, DDBJ, PDBJ and KEGG APIs. . .

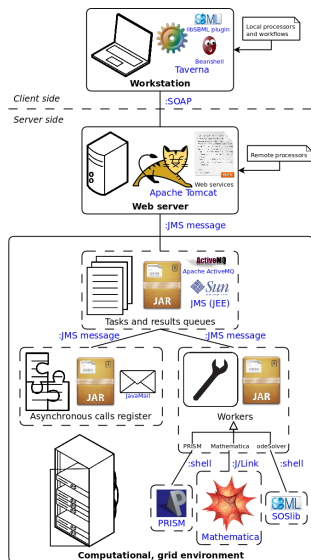
Vast majority of Web services available for use with Taverna are related to sequence analysis (bioinformatics).



What can it do?

- 1 **simulations** for the deterministic formulation of the network model (ODEs) with use of the `odeSolver` tool (the `simulate` service),
 - 2 **model checking** of Continuous Stochastic Logic (CSL) formula over a continuous-time Markov process/chain (CTMC) with use of the PRISM tool (the `prism` and `asyncPrism` services),
 - 3 **visualisation**, e.g. of trajectories or timed CSL properties with use of the Mathematica (the `mathPlot` service),
- + workflow-specific utilities like Mathematica scripts wrapped in Web services (e.g. Latin hypercube sampling) as well as local Beanshell scripts used in exemplary workflows.

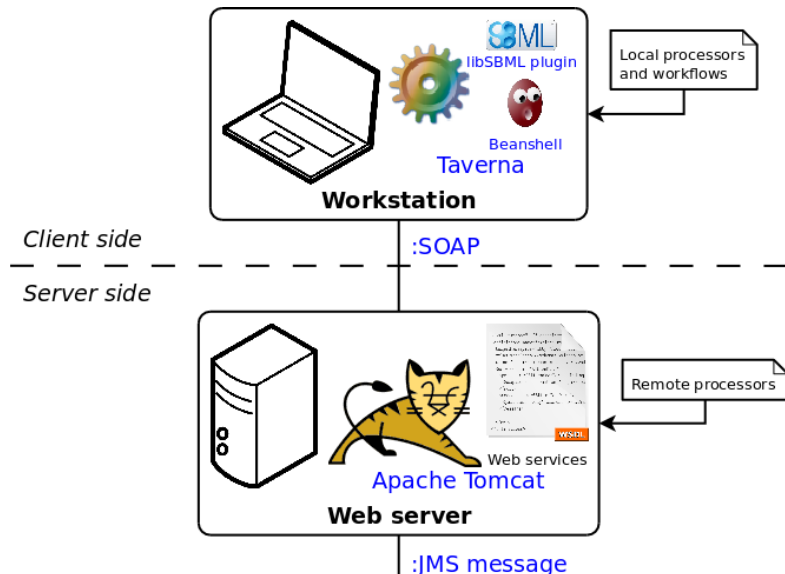
tav4sb framework architecture



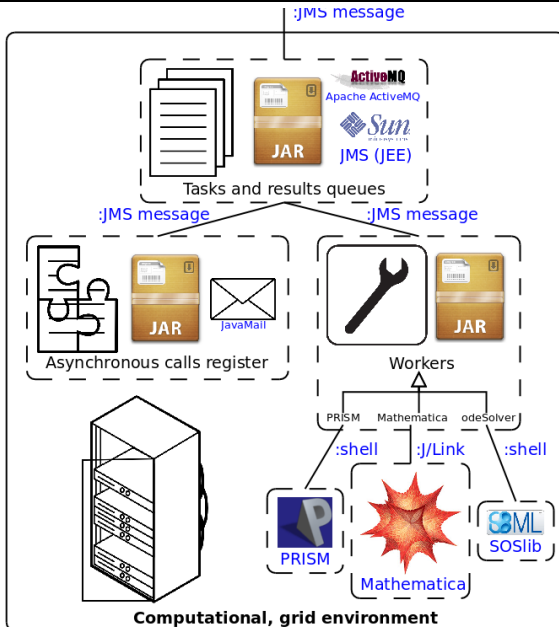
Web services utilization flow

- 1 User defines and runs workflows which include tav4sb Web services.
- 2 Taverna client communicates with Web server via Web services
- 3 Web server sends computational tasks to the grid environment.
- 4 Grid environment runs computational tasks.
- 5 Depending on service, response might be synchronous or asynchronous (sent via email).

tav4sb framework architecture



tav4sb framework architecture

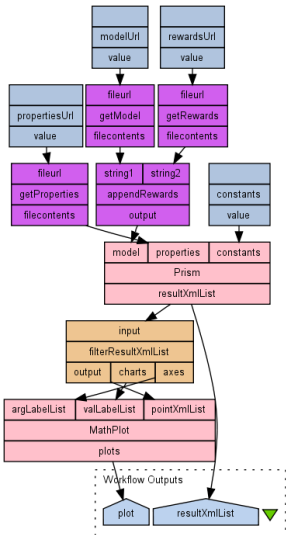


3

4

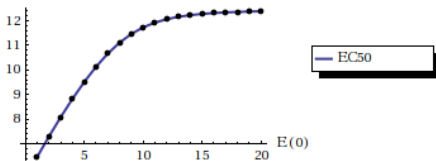
5

Enzymatic reaction probabilistic model checking (PMC)



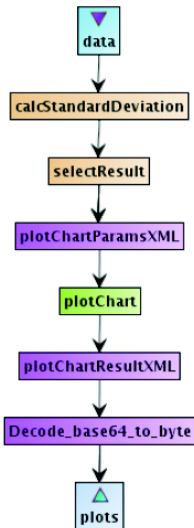
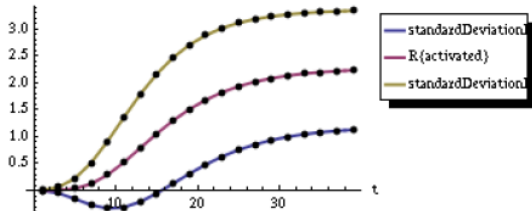
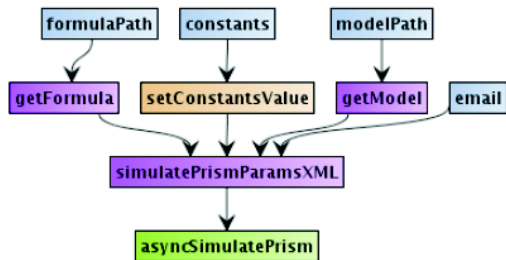
EC50 formula

$$R'' \# r_1'' = ? [F(P > 0.5 * \lim_{t \rightarrow \infty} P(t))]$$

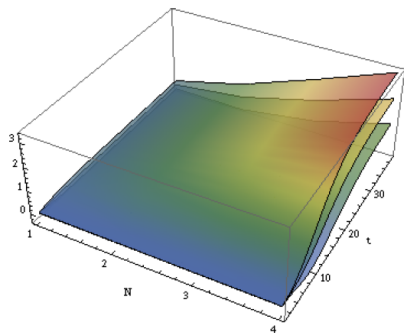


$$S(0) = 12, P(0) = ES(0) = 0$$

Asynchronous PRISM call



Grid time efficiency for MAPK model PMC



Setting

- Time-measurement experiment for $N = 1, 2, 3$, $t = 1, \dots, 50$.
- Single time point PMC with $N = 3$ runs in ≈ 9 minutes; for $N < 3$ it takes few seconds.
- Computational cluster of 16 PC with 2 dual-core CPUs each.

Time performance (in minutes)

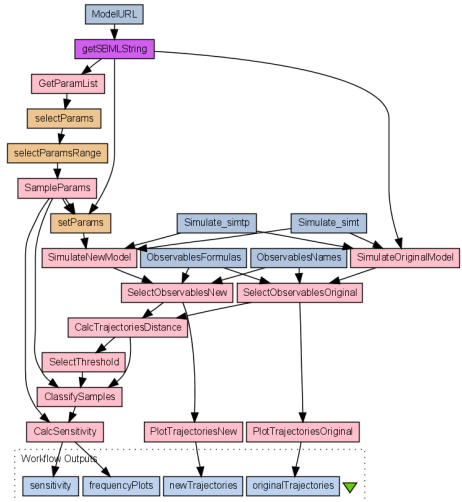
threads/machines	1	2	4	8	14
1	271,25	137,84	71,22	38,06	23,85
2	149,09	71,12	38,00	21,51	14,55
4	124,53	54,37	26,06	13,68	9,75
8	70,44	37,53	23,78	17,39	16,44

Multi-Parameter Sensitivity Analysis procedure

MPSA procedure

- Step 1.** Select parameters to assess.
- Step 2.** Set parameters range.
- Step 3.** Uniformly generate independent samples.
- Step 4.** Calculate samples errors.
- Step 5.** Classify samples as acceptable or unacceptable.
- Step 6.** Statistically compare samples sets.

Taverna MPSA workflow



Summary

- Our services **extend the functionality of Taverna** in the systems biology domain for which Web-based applications are still not widely available.
- Our grid environment runs on the computational cluster and servers, **extending the dedicated hardware base** for computationally expensive kinetic analysis tasks.
- Our grid facilitates **integration of the heterogeneous** tools, such as Mathematica, PRISM or SOSlib

IT motivational slogans supporting WSs & Taverna:

- data **standardization**,
- tools **integration**,
- services **accessibility**,
- computations **transparency**.

The grid itself ...

- ... is **deployed@bioexploratorium** computational cluster (alpha),
- ... will be also deployed@mimuw (beta) and @mrc (future);
- ... is in **clean-up and production tests phase** (release candidate);
- ... will go open source (future);
- ... needs to meet (OGF) grid systems standards and requirements.

The experiments and functionality

- MPSA workflow with PRISM
- Monte Carlo model checking of CSL formulas

Acknowledgments

mimuw

- Paweł Banasik
- Michał Lula
- Ania Gambin

bioexploratorium/mrc

- Paweł Daniluk
- Bogdan Lesyng

Thank you!