BIOINFORMATICS APPLICATIONS NOTE Vol. 22 no. 16 2006, pages 2051–2052
doi:10.1093/bioinformatics/btl322

*Systems biology*

# JCell—a Java-based framework for inferring regulatory networks from time series data

C. Spieth*, J. Supper, F. Streichert, N. Speer and A. Zell

Centre for Bioinformatics Tübingen (ZBIT), University of Tübingen, Sand 1, 72076 Tübingen, Germany

Received on April 11, 2006; revised on June 5, 2006; accepted on June 7, 2006
Advance Access publication June 16, 2006
Associate Editor: Chris Stoeckert

**ABSTRACT**

**Motivation:** JCell is a Java-based application for reconstructing gene regulatory networks from experimental data. The framework provides several algorithms to identify genetic and metabolic dependencies based on experimental data conjoint with mathematical models to describe and simulate regulatory systems. Owing to the modular structure, researchers can easily implement new methods. JCell is a pure Java application with additional scripting capabilities and thus widely usable, e.g. on parallel or cluster computers.

**Availability:** The software is freely available for download at http://www-ra.informatik.uni-tuebingen.de/software/JCell

**Contact:** christian.spieth@uni-tuebingen.de

## 1 INTRODUCTION

Within the last few years, researchers obtained large amounts of datasets from experiments. For example, fermentation experiments yield data series of the metabolic system of yeast and DNA microarray technology allows the measurement of gene expression levels for a whole genome. These techniques provide new insights into activities of system components under different biochemical and physiological conditions and can therefore be used to extract time-dependent relationships of interacting molecules like RNA, or proteins.

Although an increasing number of such microarray and metabolic datasets exist, still only few software packages are available to determine regulatory networks from time series datasets. Existing software solutions that are freely available were developed for evaluation procedures addressing specific biological questions. Graphical tools like GenMAPP (Dahlquist *et al*., 2002) allow only for visualizing microarray data on pre-compiled pathways. Simulation frameworks such as CellDesigner (Kitano *et al*., 2005) or BioNetGen (Blinov *et al*., 2004) focus on the modeling of network topologies and their simulation, without the ability to learn network dependencies from experimental data. Tools for inferring such dependencies, such as GeneNetwork by Wu *et al*. (2004) or the SBToolbox (Schmidt and Jirstrand, 2006), on the other hand provide a limited number of mathematical models or focus only on a few optimization methods and rely on commercial software.

JCell is an application to address the problem of combining algorithms and models to infer regulatory systems from time series data. First, It aims to provide a framework structure that enables researchers to use these methods without deeper knowledge of the underlying mathematics. Second, one of the design goals is the independence of commercial software to ensure the spirit of free research. Further on, owing to the modular design it can be easily extended with new algorithms and strategies. For this purpose, we implemented non-mandatory interfaces to other packages such as MATLAB, Maple, JavaEvA, evolvica and libEA (Fig. 1).

## 2 CONCEPTS

The proposed application is aimed to support biological and medical researchers in generating hypotheses about topologies and kinetic parameters of regulatory systems. This is done by searching for dependencies in time series data. JCell was initially designed to handle gene expression data from microarray experiments and was later extended on metabolic data because they have similar properties with respect to the mathematical modeling of the underlying systems. To reconstruct regulatory systems, JCell tries to model time series data by finding a topology of the network, together with a mathematical description and the corresponding kinetic parameters of the given dynamic systems.

The actual computational search is performed mostly with Evolutionary Algorithms (Bäck *et al*., 1997) such as Genetic Algorithms, Evolution Strategies, Differential Evolution, Particle Swarm, Genetic Programming, Memetic Algorithms and several multi-objective optimizer together with direct heuristics, if applicable. Overall, the framework comprises more than 10 different optimization strategies including their parallel and distributed versions that allow for running the tool on high-performance clusters. Further on, the program provides several mathematical models for simulating regulatory systems, like Boolean Networks, linear and pseudolinear Systems, S-Systems, H-Systems and arbitrary differential equations. We also implemented models to simulate intracellular systems stochastically (e.g. Bayesian Networks), where a deterministic approach is unfavorable as it is in the case for immune relevant signaling cascades. For the inference process, multiple novel strategies have been developed at our institute and implemented in JCell, including iterative knock-outs, topology identification and graph minimization problems. The corresponding publications and a user guide/ tutorial can be found on the website. In addition, JCell enables the user to edit the network topology, thus adding biological information of the network into the inference process, enforcing the algorithm to find networks with specified properties. This can be done manually, using a built-in graph editor, or by importing topologies from external sources. For example,

*To whom correspondence should be addressed.

© The Author 2006. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org 

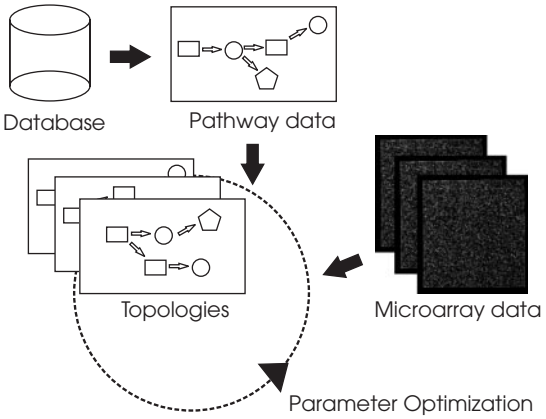


**Fig. 1.** JCell is able to import a priori knowledge of the pathway of interest. This imported pathway information is used as a template for the inference process.

pathways can be incorporated by querying public databases like KEGG (Kanehisa *et al*., 2000) and TransFAC (Wingender *et al*., 2001) to import topology graphs into JCell (Spieth *et al*., 2005a). The imported pathways are used as templates for the inference process. They can either be fixed throughout the inference, so that the tool only fits the kinetic parameters, or they can be used as an initial guidance structure that can be altered by the framework.

Another feature of JCell is the ability to simulate regulatory networks to examine effects of particular components *in silico* (Streichert *et al*., 2004). This enables researchers to study regulatory networks and the impact of drug treatment on a computer to reduce expensive 'wet lab' experiments. Publications on applications in Artificial Embryology are listed on the website.

One of the main goals was to develop a user-friendly framework that is able to run on most computer platforms without requirements of special hardware. To ensure this independence, JCell is written completely in Java. Furthermore, it can be run using Java WebStart in a standard web browser. JCell comprises a simple and self-explanatory user interface that presents only the most important parameter values. For most of the remaining hidden settings default values that showed good performance in preliminary experiments are used. Nevertheless, all algorithm settings are accessible via configuration files. These files can be either loaded from the interface or committed in the command line for batch mode. This enables experts to further tune the algorithms.

Beside the already implemented ability to import pathways in ASCII, XML and KGML (KEGG Markup Language), we plan comprehensive support of SBML (Systems Biology Markup Language) for the next release of JCell.

## ACKNOWLEDGEMENTS

This work was partially supported by the National Genome Research Network (NGFN) from the Federal Ministry of Education and Research in Germany under contract number 0313323. JCell was supported by the State of Baden-Württemberg with the doIT Software Research Award 2005.

*Conflict of Interest*: none declared.

## REFERENCES

Bäck,T., Fogel,D.B. and Michalewicz,Z. (1997) *Handbook on Evolutionary Computation*. Oxford University Press.

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.

Dahlquist,K. *et al*. (2002) GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways. *Nat. Genet*., **19**, 19–20.

Kanehisa,M. and Goto,S. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*., **28**, 27–30.

Kitano,H. *et al*. (2005) Using process diagrams for the graphical representation of biological networks. *Nat. Biotechnol*., **23**, 961–966.

Schmidt,H. and Jirstrand,M. (2006) Systems Biology Toolbox for MATLAB: a computational platform for research in systems biology. *Bioinformatics*, **22**, 514–515.

Spieth,C. *et al*. (2005a) Inferring regulatory systems with noisy pathway information. *German Conf. Bioinform*., LNI, **P-71**, 193–203.

Spieth,C. *et al*. (2005b) Predicting single genes related to immune-relevant processes. *Comput. Intell. Bioinform. Comput. Biol*., **2005**, 461–468.

Spieth,C. (2006) Modeling gene regulatory systems. PhD Thesis, University of Tübingen.

Streichert,F., Spieth,C., Ulmer,H. and Zell,A. (2004) How to evolve the head-tail pattern from Reaction–diffusion systems. In *Proceedings of Nasa/DoD Conference on Evolvable Hardware*, pp. 261–268.

Wingender,E. *et al*. (2001) The TRANSFAC system on gene expression regulation. *Nucleic Acids. Res*., **29**, 281–283.

Wu,C. *et al*. (2004) GeneNetwork: an interactive tool for reconstruction of genetic networks using microarray data. *Bioinformatics*, **20**, 3691–3693.