BIOINFORMATICS APPLICATIONS NOTE Vol. 25 no. 15 2009, pages 1987–1988
doi:10.1093/bioinformatics/btp268

Systems biology

CycSim—an online tool for exploring and experimenting with genome-scale metabolic models

F. Le Fèvre[1], S. Smidtas[1], C. Combe[1,2], M. Durot[1], Florence d'Alché-Buc[2] and V. Schachter[1,*]

[1]CEA, DSV, IG, Genoscope, UMR 8030, Evry, F-91057 and [2]IBISC, FRE 3190 CNRS, Université d'Evry, Evry, France

Received on October 7, 2008; revised on March 20, 2009; accepted on April 17, 2009

Advance Access publication May 6, 2009

Associate Editor: Thomas Lengauer

ABSTRACT

Summary: *CycSim* is a web application dedicated to *in silico* experiments with genome-scale metabolic models coupled to the exploration of knowledge from BioCyc and KEGG. Specifically, CycSim supports the design of knockout experiments: simulation of growth phenotypes of single or multiple gene deletions mutants on specified media, comparison of these predictions with experimental phenotypes and direct visualization of both on metabolic maps. The web interface is designed for simplicity, putting constraint-based modelling techniques within easier reach of biologists. CycSim also functions as an online repository of genome-scale metabolic models.

Availability: http://www.genoscope.cns.fr/cycsim

Contact: cycsim@genoscope.cns.fr

1 INTRODUCTION

Constraint-based modelling (Price *et al.*, 2004) is a framework, simple and abstract enough to allow tractable modelling of metabolism at genome-scale, providing direct insights into the genotype–phenotype relationship. Constraint-based models (CBM) consist of a stoichiometric representation of the whole-cell metabolism together with a set of constraints on reaction fluxes. A wide variety of computational methods have been developed for this framework to characterize metabolic capabilities, help to discover new reactions, simulate scenarios of metabolic evolution or design experimental strategies to investigate metabolic behaviours (Feist and Palsson, 2008).

A few simulation tools (Becker *et al.*, 2007; Beste *et al.*, 2007; Lee *et al.,* 2003; Sympheny, www.genomatica.com) and model repositories (Le Novère *et al*., 2006; BiGG, unpublished data, http://bigg.ucsd.edu) have been proposed to the growing community of CBM users. These software tools have been limited in their usefulness to biologists for several reasons. First, most are either commercial, or add-ons to commercial platforms (e.g. MATLAB, http://www.mathworks.com). Next, they are typically directed at users with a background in modelling. Lastly, these tools are not designed to explore the biochemical and genomic knowledge underlying the metabolic models. Currently, the most convenient tools to reconstruct metabolic networks from genome annotation are databases of reference pathways such as BioCyc (Karp *et al.,* 2005) and KEGG (Kanehisa *et al.,* 2008). These databases provide descriptive and queriable views of the genetic and biochemical components of metabolism, but do not support modelling, simulation or prediction.

To address these shortcomings, we introduce *CycSim*, a web platform which supports *in silico* experiments with a variety of metabolic models, puts both the design and the results of these experiments in the visual context of reference pathways databases and allows confrontation with experimental data.

2 FUNCTIONALITIES

Predictions: CycSim supports *in silico* experiments with metabolic models. Each experiment consists in selecting a wild-type strain, choosing one or several genetic perturbations (e.g. knockout), and picking a set of growth media. Growth phenotype predictions are then generated for all (mutant, medium) pairs. These predictions can be compared against experimental growth phenotypes when available (Fig. 1). Two prediction methods are implemented: flux balance analysis and metabolites producibility check (Feist and Palsson, 2008). For any given (mutant, medium) pair, CycSim can also compute a flux distribution that is compatible with the model constraints and the objective function.

Visualisation: reactions, pathways and genes can be visualized in their context through a tight coupling of the CycSim core with the pathway display layers of BioCyc and KEGG. For instance, clicking on a reaction in the simulation panel will show the corresponding BioCyc reaction page augmented with information from the active model (i.e. balanced reaction equations or the Boolean gene-reaction correspondence). Conversely, a gene can be deleted from the current model by selecting it from a pathway map. Predictions and experimental results can be directly visualized and compared on pathways.

Model and data repository: the online CycSim repository stores information relative to three organisms: *Escherichia coli* (Feist *et al.,* 2007), *Saccharomyces cerevisae* (Duarte *et al.,* 2004) and *Acinetobacter baylyi ADP1* (Durot *et al.*, 2008). For each, CycSim includes (i) a genome-scale metabolic model; (ii) a detailed correspondence between that model and relevant data of that organism [EcoCyc, (Karp *et al.*, 2007); YeastCyc (Christie *et al.*, 2004); and AcinetoCyc (Durot *et al.*, 2008)]; (iii) a set of media definitions; and (iv) experimental growth phenotype datasets. Altogether, CycSim includes 2800 genes, 3700 reactions, 1400 metabolites, 190 media, 20 000 experimental phenotypes and 550

*To whom correspondence should be addressed.

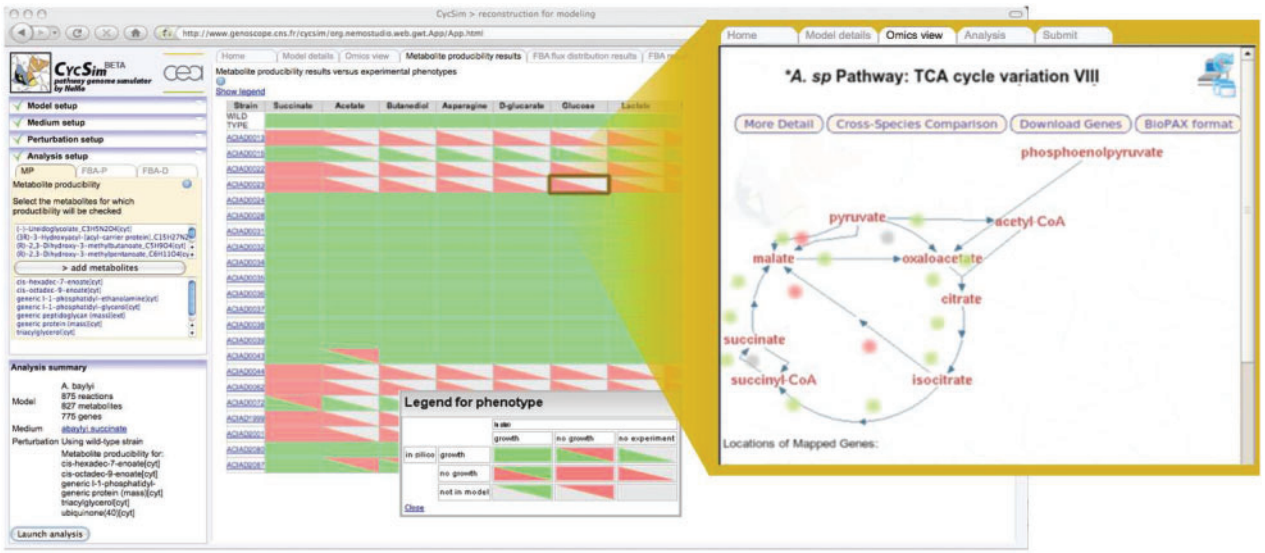

Fig. 1. CycSim screenshots. From the analysis of growth phenotypes of multiple mutants on multiple media (left), a flux distribution can be computed and visualized directly on relevant pathways (right).

pathways. Any of these four data types can be submitted online, using for models the SBML format, enhanced with MIRIAM annotations (Finney and Hucka, 2003; Le Novère *et al.,* 2005).

3 ARCHITECTURE AND TECHNOLOGIES

In order to facilitate operations from any computer, CycSim was developed as a web application using the AndroMDA framework (http://www.andromda.org) deployed on a Java application server (JBoss, http://www.jboss.org) with a MySQL backend (http://www.mysql.com). CycSim uses the AJAX technology (GWT, http://code.google.com/webtoolkit). In order to ensure the availability of sufficient computational resources, computations are performed on the server. A simple mechanism ensures some persistence of user sessions: the settings of each analysis are saved on the server and can be retrieved through a unique identifier.

In order to foster extensions by its developers or by the bioinformatics community, CycSim is based on a comprehensive UML model, which covers biochemical information (reactions and phenotype experiments) and information specific to CBM (fluxes and perturbations). Furthermore, web services are provided to programmatically access the models contained in CycSim (http://www.genoscope.cns.fr/cycsim/webservices.html).

4 CONCLUSIONS

CycSim is a simple online tool capable of handling several genome-scale metabolic models from a central repository in order to perform phenotype predictions, confronted to experimental data, and interpreted in the context of biological knowledge. CycSim facilitates the identification of inconsistencies, the design of new experiments and the iterative refinement of models using experimental data. We expect that the value of the biochemical insights obtained using CycSim will rise as more metabolic models are added to the repository, facilitating comparative analyses.

Funding: European FP6 Networks of Excellence BioSapiens (LSHG-CT-2003-503265); ENFIN (LSHG-CT-2005-518254).

Conflict of Interest: none declared.

REFERENCES

Becker,S. *et al.* (2007) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nat. Protocols*, **2**, 727–738.

Beste,D.J. *et al.* (2007) GSMN-TB: a web-based genome-scale network model of Mycobacterium tuberculosis metabolism. *Genome Biol.*, **8**, R89.

Christie,K.R. *et al.* (2004) Saccharomyces Genome Database (SGD) provides tools to identify and analyze sequences from Saccharomyces cerevisiae and related sequences from other organisms. *Nucleic Acids Res.*, **32**, D311–D314.

Duarte,N.C. *et al.* (2004) Reconstruction and validation of Saccharomyces cerevisiae iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res.*, **14**, 1298–1309.

Durot,M. *et al.* (2008) Iterative reconstruction of a global metabolic model of Acinetobacter baylyi ADP1 using high-throughput growth phenotype and gene essentiality data. *BMC Syst. Biol.*, **2**, 85.

Feist,A.M. and Palsson,B.Ø. (2008) The growing scope of applications of genome-scale metabolic reconstructions using Escherichia coli. *Nat. Biotechnol.*, **26**, 659–667.

Feist,A.M. *et al.* (2007) A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol. Syst. Biol.*, **3**, 121.

Finney,A. and Hucka,M. (2003) Systems biology markup language: level 2 and beyond. *Biochem. Soc. Trans.*, **31**, 1472–1473.

Kanehisa,M. *et al.* (2008) KEGG for linking genomes to life and the environment. *Nucleic Acids Res.,* **36**, D480–D484.

Karp,P.D. *et al.* (2005) Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. *Nucleic Acids Res.,* **33**, 6083–6089.

Karp,P.D. *et al.* (2007) Multidimensional annotation of the Escherichia coli K-12 genome. *Nucleic Acids Res.*, **35**, 7577–7590.

Le Novère,N. *et al.* (2005) Minimum information requested in the annotation of biochemical models (MIRIAM). *Nat. Biotechnol.*, **23**, 1509–1515.

Le Novère,N. *et al.* (2006) BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Res.*, **34**, D689–D691.

Lee,S.Y. *et al.* (2003) MetaFluxNet, a program package for metabolic pathway construction and analysis, and its use in large-scale metabolic flux analysis of Escherichia coli. *Genome Inform.*, **14**, 23–33.

Price,N.D. *et al.* (2004) Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat. Rev. Microbiol.*, **2**, 886–897.