Available online at www.sciencedirect.com

ScienceDirect

Electronic Notes in Theoretical Computer Science 299 (2013) 61–67

Electronic Notes in Theoretical Computer Science

www.elsevier.com/locate/entcs

A Comprehensive Web-based Platform For Domain-Specific Biological Models[1]

M. Klement, D. Šafránek, T. Děd, A. Pejznoch, L. Nedbal

Faculty of Informatics, Masaryk University
Brno, Czech Republic

R. Steuer, J. Červený, S. Müller

Global Change Research Centre AS CR, v. v. i.
Brno, Czech Republic

Abstract

A *Comprehensive Modeling Platform*, that is, a general framework for public sharing, annotation, and visualization of domain-specific biological models, is presented. For a selected organism, the framework is instantiated as a web-based application which allows to capture several aspects of biological models represented as biochemical reaction networks or ordinary differential equations. The key feature of the instantiation for a given organism relies on mapping kinetic models to a precise textual and a schematic graphical representation of the related biological knowledge, thereby supporting the systems biological view of the modeled organism. Besides model repository and annotation, the platform includes basic model analysis features such as simulation and static analysis.

Keywords: biological models, model annotation, systems biology, simulation, database

1 Introduction

In the last decade, many different platforms aiming to speed up and facilitate propagation of systems biology findings were revealed. They provide sharing of dynamical models. These "online" models allow us to simulate behaviour of a living organism or give us information about a single part of a living organism or its integration in large complex units. It must be said that no such functionality would be widely available without the Internet and web.

Despite the large number of model repositories and annotation databases, there is no platform which combines the models with analysis techniques and detailed biological annotation together and, thus, trade on such a symbiosis. For this reason,

[1] The work has been supported by the EC OP project No. CZ.1.07/2.3.00/20.0256.

http://dx.doi.org/10.1016/j.entcs.2013.11.006

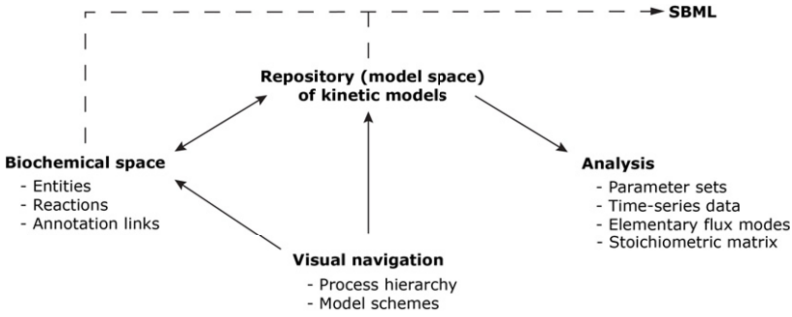

Fig. 1. Comprehensive Modeling Platform Architecture.

the activity e-cyanobacterium/e-photosynthesis [1] [13] has been initiated. It aims at handling the domain-specific problems profoundly. It also intends to trade on advantages arising from repository and database synthesis, which all in all allow to create complex relationships between models and ontological knowledge bases, different types of simulations and analysis, and integration of models in the domain-specific context.

We employ the idea of mapping biological models to a unified space of biological objects making the core components and building blocks for models. This is inspired by the so-called *Comprehensive Modeling Space* (CMS) that has been first introduced in [11] as a concept for formal representation of internally consistent reduced models of oxygenic photosynthesis. In general, CMS should simplify model-building tasks by providing simple and clear way of notation easily understandable by modelers and biologists. Wide model sharing can be provided only by Internet in case those models are translated to unified format with uniformly annotated components so they can be used in their original form, compared with each other and with wet-lab experiments. All those demands shape basics of CMS and its main features.

To support the idea of CMS, we introduce the *Comprehensive Modeling Platform* that combines model building, model analysis and annotation tasks in a single public site. Central part of CMP is made by the model repository (*model space*). Biological background of models is captured in *biochemical space*, which formally represents reaction networks linked to existing ontologies. Since models primarily target mechanisms behind biological processes, process-based hierarchy is used to navigate through the biochemical space. To this end, for each level of the process hierarchy, CMP provides visual representation of relevant biochemical mechanisms allowing to understand the non-trivial biological context of models. Finally, CMP also supports common model analysis tasks in terms of a module that provides a gate to existing tools and algorithms. General overview of the CMP architecture is depicted in Fig. 1.

In order to work reasonably with models in CMP, multi-level structure of user rights is employed. This structure allows collaborative development of a model in the team.

[1] http://www.e-photosynthesis.org, http://www.e-cyanobacterium.org

Recently, the platform is used for modeling of processes related to cyanobacteria *Synechocystis* sp. PCC6803. The main aim is to establish a consortium-approved biochemical space that unifies the state-of-the-art knowledge of cyanobacteria processes. On the other hand, CMP provides an environment shared by modelers focusing on different parts of cyanobacteria. This is the preliminary step on the way to obtain an integrated model of the selected organism.

It is important to note that there exist several tools that provide simulation, management, development, annotating, sharing of models, and comparing of model simulations against experimental data. Tools which are tightly related to CMP are, e.g., Biomodels.net [9], VCell [10] or JWSonline [12]. However, each of these tools is focused on general dynamic (mathematical) models and does not provide facilities for implementation of the domain-specific view. Especially, none of these tools is coupled with systematically organized biological background. Models captured in these tools have a wide scope in terms of biological organisms and phenomena modeled.

2 Background

2.1 Computational Models and SBML

ODE (Ordinary differential equation) models are one of the primary ways to research the behavior of biological systems. To model the behavior of the system it is necessary to create a mathematical model that describes the progress of reactions for certain concentrations. To share this type of models of systems, the format SBML [6] has been developed. It is derived from XML and provides a way to prepare the models for computational processing. Commonly known repositories support this approach, probably the best known of these is Biomodels.org [9].

2.2 Model Annotation

Annotations are essential for easy and fast understanding of ODE models. Their non-uniformity and incompleteness are a contemporary problem, which should not be underestimated. The main function of annotation is to put models into a well-organized biological and biochemical context provided by existing ontologies (e.g., KEGG [7], ChEBI [3], or GeneOntology [1]).

Guidelines for consistent annotation and curation of computational models are captured by the MIRIAM standard [8] (minimum information required in the annotation of models). It defines the necessary minimal information that must be assigned to a computational model. Besides precise model description enhancing the model quality there is also a requirement to annotate model components (i.e., organism structures, biochemical species and processes, ...).

2.3 Model Analysis

Models can be analysed in several different ways. The most common method is time-course simulation. Another area of model analysis is static analysis that al-

lows to analyse models without the need to specify model parameters. For each specific model analysis task, there are several tools which provide features needed for calculation of results (e.g., Copasi [5], CopasiWS, Octave [4], etc.). Since there is no universal tool for model analysis (which is due to a wide range of tasks) it is beneficial to support cooperation with a number of tools and enable their suitable combination. To process the results in a unified way, there has been developed the SBRML format [2] which is based on XML and provides universal way of storing results of common analysis tasks.

3 Comprehensive Modeling Platform

3.1 Model Repository Module

Our framework provides a repository of biological (kinetic) models represented as systems of ODEs. We employ the data model as defined in SBML level 2 [6]. There are two basic styles of model representation. In particular, the model can be represented directly as a purely *mathematical model*, given as a system of ODEs possibly combined with algebraic assignments, or as a *biochemical model*, given as a set of chemical reactions associated with generic (e.g., mass action, Michaelis-Menten) or user-defined kinetic functions (kinetic laws). Both styles can be suitably combined to fit the optimal level of presentation detail targeting easy understanding of the given model. In the current version we do not support differential algebraic equations.

There is the following functionality provided by the repository: biochemical and mathematical representation arbitrarily combined, user-defined kinetic functions for biochemical reactions, export to SBML level 2.

3.2 Biochemical Space Module

The goal of biochemical space module is to provide the functionality for formalization, annotation and maintenance of the biochemical space. Components of the biochemical space are divided into two groups – reactions and entities. Reactions represent elemental chemical reactions behind biological processes. Entities are cellular components and chemical species. An important function is linking of model components to items in existing annotation sources through entities and reactions consistently defined and annotated under the biochemical space.

The module allows annotators to manage the biochemical space. Import of new items is carried out by uploading the text file containing the respective annotations represented in the specific syntax. Similarly to KEGG [7], the format of annotation files is organized in block records (here separated by empty lines). There are two kinds of records – entity and reaction records. In both cases, common attributes are present: `ID`, `NAME`, `DESCRIPTION`. These are complemented by `NOTES` (internal communication among annotators), `LINKS` (references to ontologies and other sources), `CLASSIFICATION` (domain-specific classification, e.g., reaction type, organization of metabolic reactions into pathways, complexity of chemical species, ...).

A specific emphasis is given to organization of biochemical space. Entities can be arbitrarily nested provided that a particular entity can be understood both as a *species* or as a *location* for other entities. E.g., an enzyme can be understood as a location where simple molecules can be placed and transformed by reactions catalyzed by the enzyme. An entity can be assigned multiple locations. Moreover, entities may appear in different modifications, so-called *states*. To this end, a list of allowed `STATES` can be specified for any entity. E.g., a molecule of chlorophyll d, denoted ChlD, can appear in states ChlD{∗},ChlD{+},ChlD{−}, denoting the excited, oxidized and reduced state, respectively.

Reactions are formalized using a precise reaction `EQUATION` where substrates, products and modifiers are referred by an entity identifier of the general form "`entity_id{state_id}::location_id`". The location identifier is compulsory. In particular, the reaction is uniquely determined by specifying the related entities at concrete locations (and in a concrete state whenever the entity has a non-empty list of its states). Moreover, reactions can be understood as rules allowing compact representation of combinatorial reactions switching among individual modification states of large protein complexes.

3.3 Visualization Module

Undoubtedly, biochemical space has a significant informative potential. Since this space is complex, its representation is not easy to understand. We propose a separate module allowing graphical visualisation of the biochemical space. In this module, a process-based hierarchy of biochemical data is specified first, and then graphical visualisation traversing several levels of abstraction and biological structures is maintained. As a result, we obtain navigation through different levels of biochemical space based on clickable parts.

Additionally, the visualisation module also allows to attach a specific graphical scheme to each particular model in the repository. In the current version, the scheme is only flat without the possibility of zooming to specific details.

3.4 Analysis Module

Model evaluation is provided by the analysis module. It is possible to make both dynamic and static analysis. Dynamic analysis is presented by simulation of model dynamics in time. This task is extended by possibility to change parameters of the model. For simulation, a computation kernel can be chosen with respect to the type of connection. Currently, it is allowed to use Octave and Copasi solvers on our server side or to use CopasiWS remotely. The output data are displayed in the form of a graph or exported to a file. Static analysis consists of the following set of tasks: elementary flux modes, mass conservation, stoichiometric matrix construction. Static analysis is entirely performed by the Copasi tool. Output data of static analysis are displayed in the table or they can be exported to a file in SBRML format [2].

4 Current State and Future Work

First applications of CMP are e-photosynthesis.org and e-cyanobacterium.org [1] websites. The former is a completed site focusing on models of photosynthetic processes [13]. It has a simplified version of biochemical space targeting biological entities only (not reactions). The latter site implements all of the features presented in this paper. It focuses on models explaining cellular processes of *Synechocystis* sp. PCC6803. The procedures of top-down and bottom-up modeling are expected to meet at the complex biochemical space traversing from the level of a bioreactor to the level of individual components of the cell (photosynthesis and respiration, metabolism, carbon fixation mechanism, circadian clock). The functional part (the software) is completed while the biochemical space and visualisation of cyanobacterium is currently under development in the CyanoNetwork consortium [2].

There are currently two models presented on the e-cyanobacterium.org site: (*i*) a high-level bioreactor model of the gas exchange (Müller et al.) and (*ii*) a detailed model of light reactions and respiration (Plyusnina et al.). The model (*i*) is based on elemental physics and simple chemistry, its presentation in the biochemical space is thus direct. The model (*ii*) is more complicated and fully uses advanced features of the platform, i.e., projecting lumped dynamics of femtosecond-scale light reactions to elemental reaction rules defined in the biochemical space.

To sum up our contribution, the platform in the current state as presented in this paper, has several novel features. Important original features include multiple locations for entities, locations considered as regular entities allowing unlimited nesting, a formalism for entity states and elemental reaction rules preventing combinatorial explosion of the number of individual chemical reactions. These features are captured in the formal syntax of reaction equations. In this aspect the biochemical space of CMP goes beyond SBML level 2 format. At the model repository module, the compatibility with SBML level 2 is fully preserved.

Our next aims are an offline annotation tool which will relieve hard work from annotators hands and a user-friendly online model editor supporting non-expert users. The platform is currently being tested by the consortium focusing on modeling cyanobacteria. Application under this setting will help us to establish the platform as a robust technology that can be easily reused for other organism-specific modeling projects highly relevant in systems biology.

References

[1] Ashburner, M. et al., *Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.*, Nat Genet **25** (2000), pp. 25–29.

[2] Dada, J. O., I. Spasić, N. W. Paton and P. Mendes, *Sbrml: a markup language for associating systems biology data with models*, Bioinformatics **26** (2010), pp. 932–938.

[3] de Matos, P., R. Alcántara, A. Dekker, M. Ennis, J. Hastings, K. Haug, I. Spiteri, S. Turner and C. Steinbeck, *Chemical entities of biological interest: an update*, Nucleic Acids Research **38** (2010), pp. 249–254.

[2] http://www.cyanoteam.org

[4] Eaton, J. W., "Gnu Octave Manual," Network Theory Ltd., 2002.

[5] Hoops, S., S. Sahle, R. Gauges, C. Lee, J. Pahle, N. Simus, M. Singhal, L. Xu, P. Mendes and U. Kummer, *Copasia complex pathway simulator*, Bioinformatics **22** (2006), pp. 3067–3074.

[6] Hucka, M. et al., *The systems biology markup language (sbml): a medium for representation and exchange of biochemical network models*, Bioinformatics **19** (2003), pp. 524–531.

[7] Kanehisa, M. and S. Goto, *Kegg: Kyoto encyclopedia of genes and genomes*, Nucleic Acids Research **28** (2000), pp. 27–30.

[8] Le Novere, N. et al., *Minimum information requested in the annotation of biochemical models (MIRIAM)*, Nat Biotech **23** (2005), pp. 1509–1515.

[9] Li, C., M. Donizelli, N. Rodriguez, H. Dharuri, L. Endler, V. Chelliah, L. Li, E. He, A. Henry, M. I. Stefan, J. L. Snoep, M. Hucka, N. Le Novère and C. Laibe, *BioModels Database: An enhanced, curated and annotated resource for published quantitative kinetic models.*, BMC Systems Biology **4** (2010), p. 92.

[10] Moraru, I. I., J. C. Schaff, B. M. Slepchenko, M. L. Blinov, F. Morgan, A. Lakshminarayana, F. Gao, Y. Li and L. M. Loew, *Virtual Cell modelling and simulation software environment*, IET Systems Biology **2** (2008), pp. 352–362.

[11] Nedbal, L., J. Červený and H. Schmidt, *Scaling and integration of kinetic models of photosynthesis: Towards comprehensive e-photosynthesis*, in: A. Laisk, L. Nedbal and Govindjee, editors, *Photosynthesis in silico*, Advances in Photosynthesis and Respiration **29**, Springer, 2009 pp. 17–29.

[12] Olivier, B. G. and J. L. Snoep, *Web-based kinetic modelling using jws online*, Bioinformatics **20** (2004), pp. 2143–2144.

[13] Šafránek, D., J. Červený, M. Klement, J. Pospíšilová, L. Brim, D. Lazár and L. Nedbal, *E-photosynthesis: Web-based platform for modeling of complex photosynthetic processes*, Biosystems **103** (2011), pp. 115–124.