AN INTEGRATION OF THE DESCRIPTIONS OF GENE NETWORKS AND THEIR MODELS PRESENTED IN SIGMOID (CELLERATOR) AND GENENET

***Podkolodny N.L.*[*1, 2], *Podkolodnaya N.N.*[1], *Miginsky D.S.*[1], *Poplavsky A.S.*[1], *Likhoshvai V.A.*[1], *Compani B.*[3], *Mjolsness E.*[3]**

[1] Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia; [2] Institute of Computational Mathematics and Mathematical Geophysics, SB RAS, Novosibirsk, Russia; [3] Institute for Genomics and Bioinformatics, University of California, Irvine, USA

[*]Corresponding author: e-mail: pnl@bionet.nsc.ru

Key words: gene network, metabolic pathway, model, databases

SUMMARY

Motivation: The problems that arise when modeling complex molecular genetic systems at the cell level are so large-scale that they require integrated efforts of many research teams. Therefore, it is a topical problem to integrate the technologies applied to description of gene networks and their models developed at the ICG (Russia) and UCI (USA) with the aim to provide a tight cooperation in the field of systems biology.

Results: A software for conversion of gene networks and the models of gene network dynamics represented in the GeneNet system into a format suitable for loading to the SIGMOID database and Cellerator system was developed.

INTRODUCTION

An original software for supporting the technological chain of modeling was designed at the Institute of Cytology and Genetics, SB RAS, including creation of databases compiling the descriptions of various organization levels of genetic systems, formalization (specification) of the models of genetic systems, study of the models' behavior, search for the models' parameters according to the experimentally observed gene network behavior, accumulation of basic models, and solution of the target problems. A language for the specification, SiBML, was developed; it is oriented to the construction of mathematical models of molecular genetic systems taking into account the main specific features of their structure: a linear ordering and gene orientations within the genomes, polyvariance of genes (polyallelism), and polycompartment pattern of biological systems. An original software supporting SiBML for computing the direct problem was developed as well as for solving the problem of verification of mathematical models of gene networks. An original technology GeneNet (Ananko *et al*., 2002, 2005) that enables accumulation of data in a database with a consequent analysis of heterogeneous information on gene and metabolic networks was designed. A large number of the gene networks describing the vital molecular genetic processes were reconstructed and are presented in the GeneNet database.

A software for systems biology SIGMOID (Cheng *et al*., 2005; http://www.sigmoid.org/) that in turn calls Cellerator (Shapiro *et al*., 2003) was developed

at the University of California, Irvine; this software provides a wide range of options for description of biological processes and their mathematical models.

An integration of the gene networks and their models presented in GeneNet and SIGMOID will allow for an efficient combination of the technologies for modeling genetic systems developed at ICG and the technologies for distributed modeling that are developed at UCI.

THE REPRESENTATIONS OF GENE NETWORKS IN GENENET AND SIGMOID DATABASES

Functioning of a gene network is provided by complex relationships between different components, namely, genes, proteins, metabolites, signal molecules, energy-connected cell components, etc. Using an object-oriented approach, we recognize several following logical levels in the description of relationships between the gene network components. Ontological level, including general notions and relations between them. Here we describe as metaclasses the elementary structures, or Entities (genes, proteins and protein complexes, RNAs, and small molecules) and the elementary processes (reactions and regulatory events). A scheme of semantic relationships between the elementary structures and processes is given in Fig. 1.

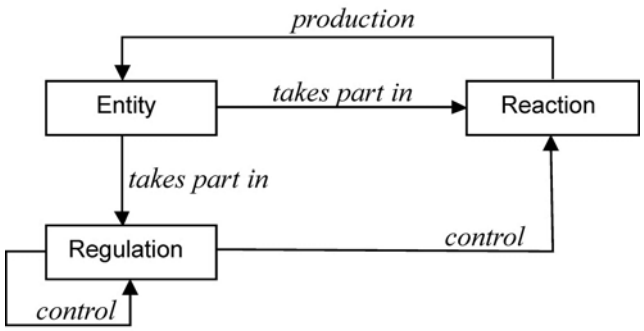

Figure 1. Semantic relations between elementary notions in a gene network.

The level of objects of study involves the descriptions of notions and classes of entities for particular objects of study. For example, description of some representatives of such classes as “genes”, “proteins and protein complexes”, “RNAs”, and “small molecules”, which are involved in functioning of particular objects studied (i.e., cells of *E. coli* K12).

The level of compartments, at which the entities described above are attached to a spatial compartment. At this level, the description of entities may include supplementary parameters such as, for example, the concentration of a given entity in a particular compartment. The whole bulk of information about elementary structures and functional relations in a gene network is represented at the three levels described above. The subsequent two levels are needed for describing the network as a whole.

The level of gene networks corresponds to the description of functional subsystems. At this level, the particular descriptions of a complex system (the object of study) from different viewpoints may be represented as well as its simplified description or a partial model representation. Many representations of a complex system may exist. Besides, there is a possibility of the gene network reconstruction via a query to the GeneNet database.

The level of representation provides the way for describing the pattern of visualization or graphical representation of a gene network. It is supposed that different ways of representation exist for each gene network. In particular, to this level we refer the representation of a gene network in a form of a tree or hierarchical relations, in a form of a hyper graph, and so on. At this level, it is possible to use different ways of layouts for the automated visualization of a gene network, obtained by querying the database.

GeneNet format is based on XML. Each separate XML file represents a single GeneNet diagram and consists of the following sections:

Header (<diagram> element). It includes the identifier of a diagram in GeneNet database, dates of its creation and modification, and the description of biological function of the gene network.

Nodes list (<objects> element). This section contains the descriptions of all nodes in graph representing diagram. Nodes are genes, proteins, substances, RNAs, as well as reactions and any regulatory events (reaction or regulation). Each description includes name and synonyms, links to articles, comments, and some service information. Moreover, the reactions and regulatory events contain information about their inputs and outputs. The reactions in an object representation of the SIGMOID (see the scheme of information representation in the Sigmoid database at are divided into two subclasses—biological and mathematical reactions. The subclass "Biological reactions" is designed for the description of various reaction types with a distinct biological interpretation (for example, replication, transcription, translation, allosteric interaction, enzymatic reaction, etc.). A mathematical representation of the model described in the subclass "Mathematical reactions", containing the hierarchy of methods for modeling biological reactions, is available to mathematically model each of the "Biological reaction" subclasses.

SiBML, A LANGUAGE FOR SPECIFICATION OF MODELS

The language SiBML is designed for an efficient (economic) specification of the models; it comprises three following description levels: (1) the level for description of elementary processes; (2) the level for description of "genetic maps" (G maps), which are constructed as the lists of ordered oriented objects named "genes", each "gene" carrying the information about the elementary processes related to it; and (3) the level of specification of a compartment structure of the object modeled. At the last level, a map of intercommunicating compartments (or C map) is specified. The C map is constructed as an ordered list of objects named "compartments". Each "compartment" as an object contains the information about its own name, the compartments whereto the substances flow from this compartment, the information about G maps localized to this compartment, and the information about the files where the description of elementary processes are stored. The corresponding descriptions are stored in the databases at each level and can be used multiply. The models are assembled by a specialized set of programs, named model constructor. The models that appear when describing the processes in terms of chemical kinetic reactions are formally belong to the class of autonomous systems of differential equations. However, in a general case, the models belong to a mixed type, since they may contain continuous, discrete, probabilistic, and other modules. The final constructed model is an ordered list of elementary processes. It is this pattern that is convertible into the CELLERATOR format without any losses.

IMPLEMENTATION AND RESULTS

1. A software providing the loading of the gene networks contained in the GeneNet system into the SIGMOID database was developed. This brings about the problems of matching the objects described in these databases and the attributes of these objects,

Object. The <gene>, <rna>, <protein>, and <substance> elements represented in GeneNet are mapped into SIGMOID as interfaces Gene, RNA, Protein, and Molecule, respectively. Using SIGMOID API, it is possible to map name, species, references to papers, and comments.

Reaction. This class of regulatory events is represented in GeneNet as <reaction> elements; it is mapped in SIGMOID using the CatalyticWithAllostericRegulation interface. However, the following data can be mapped directly: substrates, products, references to papers, and comments. Additional information about enzymes, activators, and inhibitors, necessary for specification of CatalyticWithAllostericRegulation interface is extracted from GeneNet through the analysis of regulatory relationships (see below).

Gene regulation. In GeneNet, this is represented by the <reaction> element, whose output (i.e., the object of regulation) is the reaction of transcription, translation, etc. or the indirect reactions that include all the stages involved in gene expression. The reaction of transcription has an input (single gene), output (RNAs), and a set of transcription factors (proteins or protein complexes). This reaction can be mapped to SIGMOID as the RegulatoryRelationship interface with the following fields: target gene, regulators, references to papers, and comments.

Regulatory event. In GeneNet, this is a special type of interaction (<reaction> element) where the inputs are objects and the output is another interaction. Regulatory events may be positive or negative as well as direct or indirect (which means the lack of precise information about the particular mechanism). Regulatory event could be organized in complex multilevel cascades (regulation of regulation of …). In SIGMOID regulatory elements participate in reactions as additional inputs. They are not consumed and therefore they exit the reaction as outputs. Since GeneNet and SIGMOID possess different perspectives on how regulation should be modeled, GeneNet regulatory events are mapped in SIGMOID as activators, inhibitors, enzymes, or regulators of reactions.

2. A converter was designed able to convert the mathematical models constructed using a limited version of the SiBML standard into the standard of CELLERATOR software package aiming to further loading into the package Mathematica 5.0.

The input format of the converter is specified by the SiBML standard (Likhoshvai *et al*., 2001). Three files are used as the input data; these files contain the parameters (constants) and their values, dynamic variables and their initial values, and the corresponding mathematical model in SiBML, respectively.

The output of the converter operation is the file in a Cellerator standard organized as a NoteBook for Mathematica 5.0. The output NoteBook consists of a unit switching of the cellerator.m module, the unit containing the list of initial concentrations, and the unit with the list of reactions. The presented version of converter supports the following set of SiBML blocks: B(1->1); B(1=>1); B(0->1); B(0->1) ; B(UNI1) (see Table 1).

Table 1. Interpretation of standard SiBML blocks in terms of Cellerator

Block	Interpretation in Cellerator
B(UNI1)	$\left\{\left\{ 0 \to \mathrm{A_i},\quad -\mathrm{a_i} * \mathrm{G}\right\},\quad \left\{ 0 \to \mathrm{B_j},\quad \mathrm{b_j} * \mathrm{G}\right\}\right\},\quad \mathrm{i}=1,...,\mathrm{m};\ \mathrm{j}=1,...,\mathrm{l}$
B(1->1)	$\{A \to B,\quad K\}$
B(1=>1)	$\{0 \xrightarrow{B} A.\quad K\}$
B(0->1)	$\{0 \to A,\quad K\}$
B(1->0)	$\{A \to 0,\quad K\}$

ACKNOWLEDGEMENTS

The work was supported by the grant of NSF "FIBR: developmental modeling and informatics" and innovation project of the Federal Agency for Science and Innovation IT-CP.5/001 "Development of software for computer modeling and design in postgenomic system biology (system biology *in silico*)".

REFERENCES

Ananko E.A. *et al*. (2002) GeneNet: a database on structure and functional organisation of gene networks. *Nucl. Acids Res*., **30**, 398–401.

Ananko E.A. *et al*. (2005) GeneNet in 2005. *Nucl. Acids Res*., **33**, Database issue D425–D427.

Cheng J. *et al*. (2005) Sigmoid: A software infrastructure for pathway bioinformatics and systems biology. *IEEE Intelligent Systems*, **20**(3), 68–75.

Likhoshvai V.A. *et al*. (2001) A generalized chemical-kinetic method for modeling gene networks. *Mol. Biol. (Mosk.)*, **35**(6), 1072–1079.

Shapiro B.E. *et al*. (2003) Cellerator: extending a computer algebra system to include biochemical arrows for signal transduction simulations. *Bioinformatics*, **19**(5), 677–678.