An Object-Oriented Data Model for Signal Transduction

Yves Deville [1], David Gilbert [2], Christian Lemer [3], Jacques van Helden [3], Shoshana J. Wodak [3]

[1] Computing Science and Engineering Department, Université catholique de Louvain,
Place Saint-Barbe 2, 1348 Louvain-la-Neuve, Belgium , deville@info.ucl.ac.be

[2] Bioinformatics Research Centre, Department of Computing Science, University of Glasgow,
17 Lilybank Gardens, Glasgow G12 8QQ, Scotland, UK, drg@brc.dcs.gla.ac.uk

[3] Unité de Conformation des Macromolécules Biologiques, Université Libre de Bruxelles,
50 av. F.D. Roosevelt, B-1050 Bruxelles, Belgium {chris,jvanheld,shosh}@ucmb.ulb.ac.be

Introduction

Metabolic, regulatory and signal transduction pathways refer to cellular processes, in which the interactions between the different components form intricate networks, often referred to as biochemical networks. The size of available data on such networks is growing rapidly. Information on such networks and their components is usually stored in various generic and more specialised databases. When available, existing information in a database is limited to one type of biochemical network, and is usually poorly structured. It is therefore difficult to perform elaborated analysis on relationships between different types of biochemical networks.

We propose an object-oriented data model for signal transduction. This model is an extension of the aMAZE data representation [1,2]. Integrated with the latter, it is capable of handling metabolic pathways as well as regulatory and signal transduction pathways, thus covering different types of biochemical networks in a single data representation. The signal transduction data model has been designed to fit a set of about 500 interactions on signal transduction, curated in the aMAZE project. It also integrates protein-protein interactions into their cellular context.

A Generic Model of Interactions

In Figure 1, we present the generic model of interactions, which is at the basis of our data representation. All the considered interactions are instances of this schema. For example, a reaction (one of the subclasses of Transformation) has several inputs (substrates), and several outputs (products). The Transf_Output and Transf_Input objects are used to store information on the particular role of the BioEntities in this reaction (e.g. stochiometry). In a reaction, the cellular location of the different substrates and products can be different (especially in signal transduction), location is therefore not an attribute of Reaction, but an attribute of Transf_Output and Transf_Input. Transformation can be controlled by another BioEntity. The classical example is the catalytic reaction. In this model, the same reaction can be catalyzed by different catalyses; each catalysis can be carried out by several BioEntities. Finally, BioEntities can control another control. For instance, Proline can inhibit the ReactionCatalysis (a Control) of a reaction

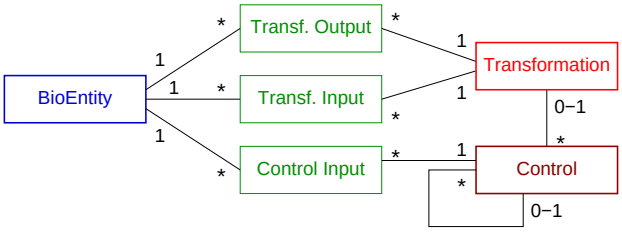

Fig. 1 UML generic model of interactions

Data Model for Signal Transduction

In Figure 2, we propose an object-oriented data model for signal transduction. It is an extension of the aMAZE data model, which already deals with metabolic and regulatory pathways following the principles illustrated by the generic model in Fig. 1. By a signal transduction pathway, we mean a biochemical pathway that carries signals from one cellular location (e.g. receptors on the cell-surface) to another (e.g. nucleus). In signal

transduction, the following elements are particularly important : the cellular location of the BioEntities, the formation and dissociation of complexes and the molecular states of the proteins (e.g. phosphorylation, acetylation, ubiquitylation.). Translocation is modeled as a subclass of Transformation. The input and output BioEntities are the same; but the initial location is a property of Transf_Input, and the final location a property of Transf_Output. A translocation can also be controlled by some TransportFacilitator (permeases). The Assembly and Dissasembly of complexes is a subclass of Transformation, with its associated AssemblyCatalysis control. Complex is a subclass of BioEntity, but there is an aggregation relation between the complex and all its components. This allows an easy access from a complex to its components. Some reactions are very common in signal transduction (phosphorylation, ubiquitylation). No subclass has been defined, but a classification is provided for reactions. This allows a more flexible and extendable structure for existing reactions. Polypeptides in a state are modeled as the Polypeptide subclass of BioEntity. A Polypeptide object only contains the information on its states; it is related to its polypeptide reference. Finally, pathways are represented as a directed graph of transformations; arcs relating different pathways are useful to model interactions between pathways as well as subpathways.

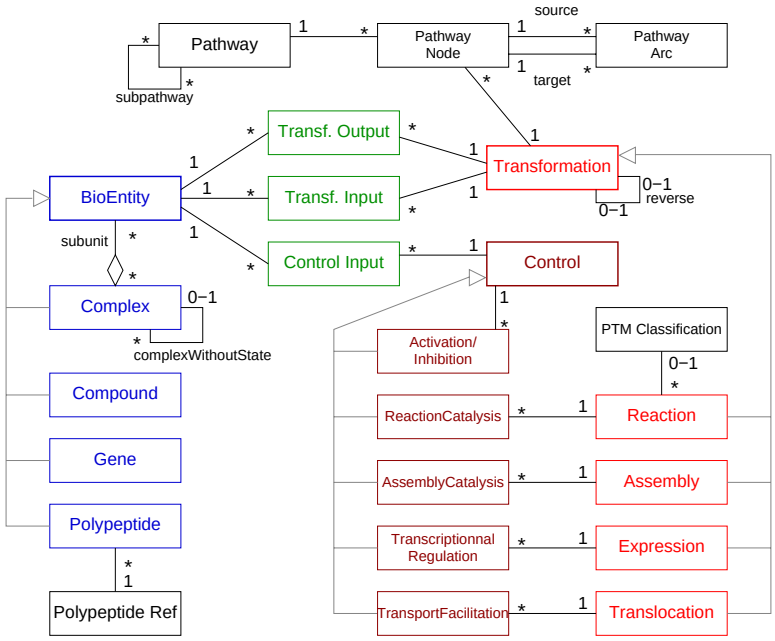

Fig. 2 UML data model for signal transduction

Prototype

A signal transduction prototype database has been developed; it will be integrated into aMAZE upon validation. The objective of this prototype is to validate the model on real data, and to allow the development of analysis methods, tools and techniques. The object-oriented model (with hierarchy) has been translated into a relational model comprising 30 tables; it is implemented in MySQL and in Oracle; the DB contains around 700 BioEntity and about 500 Transfomation and Control.

References

[1] van Helden, J., A. Naim, R. Mancuso, M. Eldridge, L. Wernisch, D. Gilbert, and S.J. Wodak: Representing and analysing molecular and cellular function using the computer. Biol Chem. 381(9-10), 921-35, 2000.

[2] van Helden, J., A. Naim, C. Lemer, R. Mancuso, M. Eldridge, and S.J. Wodak. From molecular activities and processes to biological function. Briefings in Bioinformatics. 2(1), 81-93, 2001.