Electronic Biochemical Pathways[*]

Carl-Christian Kanne[1], Falk Schreiber[2], and Dietrich Trümbach[3]

[1] Lehrstuhl für Praktische Informatik III, Prof. Guido Moerkotte,
Universität Mannheim, D7, 27, 68131 Mannheim, Germany,
[2] Lehrstuhl für Informatik, Prof. Franz J. Brandenburg,
Universität Passau, 94030 Passau, Germany,
[3] Institut für Organische Chemie, Prof. Johann Gasteiger,
Universität Erlangen-Nürnberg, Nägelsbachstr. 25, 91052 Erlangen, Germany,

The biochemistry of living beings is a complex network of reactants, products and enzymes with multiple interconnections representing reactions and regulation. Examples are given by the Boehringer poster [5] and the atlas "Biochemical Pathways" [6]. In the *Electronic Biochemical Pathways* project we intend to build a platform for a convenient electronic access to the information covered by the poster and the atlas.

The *Electronic Biochemical Pathways* has three major parts. First, chemical information is modeled accurately and in detail. This data is stored in a database which also contains explicit connections between the different subjects (for example, showing the diseases related to a certain reaction pathway, see also [4]). Using this database and advanced automatic visualization of the involved reaction graphs, the resulting software system will provide users with new ways of considering and analyzing metabolic pathways.

The project is divided into three parts: Acquiring and modeling the data, the suitable database environment, and the graph visualization system.

Acquiring and modeling the data

This part deals specifically with the molecule- and reaction-attributes:

- Code all chemical structures from the Boehringer poster (part 1) on the atomic level. Thereafter, we can implement a full structure and a substructure search [8].
- Registration of the compound-names in German and English both in HTML format.
- Assignment of the compounds to (co-)reactant and (co-)product and the classification of these compounds into chemical groups.
- Specification of the conditions for a reaction (enzymes, regulation, environment). Detailed information about the reaction-mechanisms by specifying a mapping of product atoms to reactant atoms.

This core data from the poster is enriched by structural information, like grouping of relation reactions into standard reaction chains (pathways like glycolysis or the citrate cycle), layout constraints for reaction graphs, amino acid and gene sequences for proteins, and links into specific web databases.

Building a suitable database environment

The above-mentioned concepts belong to a wide range of different abstraction levels. They were mapped into a complex, object-oriented, conceptual model using UML [7]. The data model was implemented using a commercial database management system. It is the base component of the *Electronic Biochemical Pathways.*

Special consideration was given to the addition, export and interaction with external data. A common exchange format for chemical and biochemical data is XML (eXtensible Markup Language) [1]. Instead of multiple conversion programs, a specialized XML database system is being developed which stores all the relevant project's information directly in XML and will replace the commercial database server in the deployment version of our system.

Creating a graph visualization system

Existing graph drawing algorithms [2,3] are not sufficient to represent pathways according to the conventions of biochemistry. As a consequence, in this part of the project we are developing algorithms tailored to biochemical pathways.

The capabilities of automatic drawing go far beyond the ones shown by the poster: diagrams of pathways can be combined with other pieces of information (like cell compartments or membrane transport processes) or similar pathways of different organisms can be compared with each other by overlying identical parts and emphasizing different parts. One can use graphic features such as highlighting and animation. The automatic generation of pathway diagrams offers new chances for the navigation through pathways like different resolution steps from general pathway maps to diagrams of all details of a single pathway. Pathways can be expanded dynamically and drawings of new parts can be produced on demand. It is possible to connect diagrams with each other, for example pathways and the tree of the enzyme classification of the corresponding enzymes.

Forthcoming information is available under http://www.biochemical-pathways.de.

References

1. T. Bray, J. Paoli, and C. M. Sperberg-McQueen. Extensible Markup Language (XML) 1.0 W3C Recommendation 10-Feb.-98. http://www.w3.org/TR/REC-xml.
2. G. Di Battista, P. Eades, R. Tamassia, and I. G. Tollis. Algorithms for Drawing Graphs: An Annotated Bibliography. *Comput. Geom.*, 4(5):235–282, 1994.
3. G. Di Battista, P. Eades, R. Tamassia, and I. G. Tollis. *Graph drawing: algorithms for the visualization of graphs.* Prentice Hall, New Jersey, 1999.
4. P. W. Erhardt. *Drug Metabolism - Databases and High-Throughput Testing During Drug Design and Development.* Blackwell Science, Oxford, 1999.
5. G. Michal. Biochemical Pathways (Poster), 1993. Boehringer Mannheim GmbH.
6. G. Michal. *Biochemical Pathways.* Spektrum Akadem. Verlag, Heidelberg, 1999.
7. J. Rumbaugh, I. Jacobson, and G. Booch. *The Unified Modeling Language Reference Manual.* Addison-Wesley, 1998.
8. R. E. Stobaugh. Chemical Substructure Searching. *J. Chem. Inf. Comput. Sci.*, 25:271–275, 1985.