Editorial

Petri Nets and the Simulation of Metabolic Networks

Ralf Hofestädt

*Bioinformatics / Medical Informatics, Technische Fakultät, Universität Bielefeld,
Postfach 10 01 31, D-33501 Bielefeld, Germany
E-mail: hofestae@techfak.uni-bielefeld.de*

Based on the Human Genome Project, the new interdisciplinary subject of Bioinformatics has become an important research topic during the last decade. An important catalytic element of this process is that methods of molecular biology (DNA-sequencing, proteomics etc.) allow the automatic data generation of cellular components. Based on this technology roboter systems allow to sequence small genomes in a few weeks. Moreover, the semi-automatic assembly and annotation of the sequence data can only be done using methods of computer science. The molecular data is stored in database systems available via the Internet. Based on that data, different questions can be solved using specific analysis tools. Regarding the DNA sequences we are looking for powerful software tools which will predict functional units in the DNA. Today this topic is called "From Sequence to Function" or "Post-Genomics".

The common definition of Bioinformatics addresses the application of methods and concepts of computer science in the field of biology. Bioinformatics currently stresses three main topics. The first major topic is sequence analysis or genome informatics. Its basic tasks are: assembling sequence fragments, automatic annotation, pattern matching and implementation of database systems, like EMBL, TRANSFAC, PIR, RAMEDIS, KEGG etc. The sequence alignment problem is still representing the kernel of sequence analysis tools. Nevertheless, sequence analysis is not a new topic. It was, and still is, a topic of Theoretical Biology and Computational Biology. Protein Design is the second current major research topic of Bioinformatics. The first task was to implement information systems that represent knowledge about the proteins. Today many different systems, like PIR or PDB, are available. The main goal of this research topic is to develop useful models which will allow the automatic calculation of 3D structures, including the prediction of the molecular behavior of this protein. Until now, molecular modeling failed. Protein design is also not a new research topic. Its roots are coming from Biophysics, Pharmaco Kinetics and Theoretical Biology. The third major topic is Metabolic Engineering. Its goal is the analysis and synthesis of metabolic processes. The basic molecular information of metabolic pathways is stored in database systems, like KEGG, WIT, etc. Models and specific analysis algorithms, based on the molecular knowledge represented by these database and information systems, allow the implementation of analysis tools.

The idea of Metabolic Engineering represents the basic idea of the Virtual Cell. Using molecular data and knowledge, the implementation of specific models allows the implementation of simulation tools for cellular processes. Behind the algorithmic analysis of molecular data, modelling and simulation methods and concepts allow the analysis and synthesis of complex gene-controlled metabolic networks. The actual data and knowledge of the structure and function of molecular systems is still rudimentary. Furthermore, the experimental data available in molecular databases have a high error rate, while biological knowledge has a high rate of uncertainty. Therefore, only modelling and simulation and methods of artificial intelligence will suffice to discuss arising important questions. Such formal description can be used to specify a simulation environment. Therefore, modelling and simulation can be interpreted as the basic step for implementing virtual worlds that allow virtual experiments.

As already mentioned more than 500 database and information systems are available, which represent molecular knowledge today. Furthermore, a lot of analysis tools and simulation environments are available. That means that basic components of the electronical infrastructure for the implementation of a virtual cell are present. The concepts and tools which are available in the literature and the Internet are based on specific questions, such as the gene regulation process phenomena, or the biochemical process control. To solve current questions, we have to implement integrative tools (Integrative Bioinformatics) which can be used finally to implement a virtual cell. If we take a look at the Internet, we can see that only online representations of cellular illustrations, taken directly from books, are available today (http://www.life.uiuc.edu/plantbio/cell/). One of the first implementations is the E-Cell project of M. Tomita (www.e-cell.org). Many new virtual cell projects are following the E-Cell project. Regarding the different methods for modelling and simulation of metabolic pathways we can divide these tools into two classes. The classical methods are members of the so called analytical class. All these tools are based on the theory of differential equations and try to realize the exact molecular simulation. The main argument against this class is that we do not have the dynamic molecular data. This was the main argument for a lot of different scientists coming from different research areas to develop discrete models. These models are based on the theory of formal languages, automata, objects, rules, expert systems etc. However, a few of these models are also hybrid models. Until now it is not clear which kind of model will be the best to help to implement the virtual cell. Exactly 10 years ago M. Mavrovouniotis and colleagues presented the first paper using Petri nets for this important application [Reddy *et al.*, 1993]. In this paper they used only simple case condition systems for simulation of simple biochemical processes. During the last decade a lot of deeper papers were published using this method of simulation of metabolic networks. The advantages of this method are:
1. Deep theory and results are available,
2. Powerful simulation shells are available and
3. Petri nets can combine both classes in a simple way.

Point three is very important because using higher Petri nets we are able to place differential equations to the arcs of the model. That means that we are able to expand a discrete to an analytical model at any time.

This special issue represents four invited papers of a DFG workshop, which was organized 2001 at the University of Magdeburg. As one of the organizers I would like to thank the DFG for the support of this workshop. Moreover, I would like to thank the authors for presenting their talks and papers.

Bielefeld, 14. September 2003

REFERENCES

- Reddy, V. N., Mavrovouniotis, M. L. and Liebman, M. N. (1993). Petri net representations in metabolic pathways. Proc. Int. Conf. Intell. Syst. Mol. Biol. **1**, 328-336.