

## Summer School Bioinformatics — Information Systems and Network Analysis of Gene Regulation and Metabolism

RALF HOFESTÄDT\*

*Department for Computer Science, University Magdeburg*

Development in molecular biology have resulted in very significant increase in the volume of data available to scientists. Handling and modelling there data requires electronic tools, for which world-wide access via the internet is essential. This paper summarises contributions on this area to a summer school organiser in magdelling in September 1999.

*Keywords:* internet, molecular modeling

Between September 5<sup>th</sup> and September 10<sup>th</sup> 1999 *The First International Summer School of Information Systems and Network Analysis of Gene Regulation and Metabolism* was organized in Magdeburg.

Post-genomic methods in molecular biology generate large amounts of expression data, which require computational methods for their storage, management, and analysis. Knowledge databases, containing information on genes not only as genome constituents but also as therefore an elements of regulatory, signaling, and metabolic pathways, are essential resource for “reverse engineering” regulatory and metabolic networks in the cell. An updated illustrative list of different methodologies available on the internet can be obtained at (see [http://www.iti.cs.uni-magdeburg.de/iti\\_bm/dagstuhl/demos.html](http://www.iti.cs.uni-magdeburg.de/iti_bm/dagstuhl/demos.html)). The conclusion when looking at such a variety of approaches is that in order to support the analysis of gene regulation and metabolic network control we

need distributed data access integrated with visualization, simulation, and analysis tools. Based on the experience of two international Dagstuhl Seminars on this topic (1995 and 1998 Schloss Dagstuhl with URL: <http://www.dag.uni-sb.de>) we invited scientists to prepare specific courses address through issues.

Sequencing technology has produced an exponential growth in the amount of data on DNA and protein sequences. However, in the case of the human genome project ca.  $64 \cdot 10^9$  bases will be available at the end of the year 2001. That means that to publish the human genome will require approximately 1000 books of 1000 pages. This example highlights the imports electronic representation of molecular data. The internet is a powerful medium which allows worldwide access to molecular information systems. Therefore, the internet could be the main catalyst for the development of molecular database and information systems. Today the backbone of

\*Corresponding Author: E-mail: [hofestaed@iti.cs.uni-magdeburg.de](mailto:hofestaed@iti.cs.uni-magdeburg.de)

the electronic infrastructure for biotechnology and molecular biology can be characterized by distributed molecular database systems, for which have different analysis tools already been developed and implemented. Most of these systems are also available via the internet. Fundamental to the analysis of cell functions is an understanding of gene regulation. Genes can be activated under specific biochemical situations. Thus, metabolites have activate or deactivate specific genes themselves provided as a consequence of gene directs, protein synthesis, activate biochemical reactions and influence signal induction pathways.

The idea of the summer school was to present lectures for postgraduate students, on the electronic infrastructure for biotechnology. It opened with a mini symposium, which presented actual topics as metabolic pathways and metabolic engineering through invited talks. V. Ratner (Russia Academy of Science) discussed the metabolic processes using formal languages. He presented a summary of his 30 years' work and discussed DNA as a programming language. Rolf Apweiler (EBI Cambridge) presented the basic molecular database systems and introduced the protein database system SWISSPROT. Finally, actual Bioinformatics research projects supported by the EBI were discussed. The information of promotor sequences and the transcription factors were collected by E. Wingerder (GBF Braunschweig). He developed and implemented the database system, which represents the fundamental data of gene regulation. In his course he discussed the phenomena of gene regulation and the contents of the TRANSFAC system. However, the identification of Promotor sequences is fundamental to the discussion of gene regulation and metabolic pathways. Different tools are available for the detection of promotor sequences. Manson McGuire (Harvard Medical School) presented an overview of these methods in her lecture. Genes code for proteins such as Enzymes which catalyse biochemical reactions. Therefore, genes control metabolism indirectly. More than 30 years ago G. Michal (Tutzingen) started collecting the knowledge about biochemical reactions. He is the

editor of the Boehringer pathway chart, which represents the first collection and visualisation of this biochemical data. He presented the key ideas of his new book — the atlas of metabolic pathways. Based on that data the KEGG group implemented the first electronic system for metabolic processes. The leader of this group M. Kanehisa (University Kyoto) presented the KEGG system and its important applications. Moreover, he presented new ideas about the simulation of metabolic pathways using graph theoretical approaches. The next topic of the summer school was the electronic representation of signal induction pathways. R. Brent (Molecular Science Institute, Berkeley) discussed the protein/protein interactions and N. Kolchanov (Russia Academy of Science, Novosibirsk) discussed the biology of cell signal pathways and the electronic representation. The fundamental processes of cell differentiation was presented by J. Reinitz (Mt. Sinai Med. School, New York). Behind the molecular data and its systematic representation using electronic tools and database systems, one application is the modeling and simulation of these processes. Different models and simulation environments have been developed. All these models can be classified into two classes, quantitative and qualitative models. Simulation of metabolic processes was discussed by M. Tomita (University Keio). He presented his object oriented simulation environment E-Cell, which is able to simulate virtual cells and parts of biochemical networks. The advantage of E-Cell is its user interface, which supports experiments in hypothetic worlds. E-Cell or any other simulation environment in the Internet currently to the disadvantage of no automatic access to the molecular database systems. In the case of E-Cell the user has to open for example the KEGG system and to create E-Cell models by hand. For the support of the ideas of metabolic engineering this is not sufficient. The idea is to implement software tools for the analysis and synthesis of metabolic processes. M. Reuss (University Stuttgart) presented a basic course of the methods and goals of this new research field. The key idea is that metabolic engineering is a multi-disciplinary subject and that

we have to develop virtual labor, which can be used as a simulator of complex metabolic processes. The integration of molecular databases is fundamental for this vision and was the last topic of the summer school. Therefore, T. Mück (University Vienna) presented the basic course of database systems and integration methods. Based on this course R. Hofestädt, U. Scholz and M. Lange (University Magdeburg) presented the prototype of a Information Fusion System for the simulation of metabolic processes.

Biotechnology belongs to the most important subject of the next century. During the last years ca. 2000 companies were founded in western Europe and we are just at the beginning of this process. An important tool is the electronic representation and analysis of these data. The problem is a shortage of bioinformatic scientists. For example, in Germany only three universities have provide education in this new subject. To close

this gap was one reason to organize and establish this international summer school. However, the next five years the German universities will not produce interdisciplinary scientists in the field of bioinformatics. To close this gap we have to organize more interdisciplinary summer schools or special courses.

This summer school was supported by the Volkswagen-Stiftung. Based on this grant we invited the lectures and selected 28 graduate students of Biology, Medicine, Computer Science, Biophysics, Biomathematics, and Biochemical Engineering. One idea of this Summer school was to develop and implement electronic courses. These courses will be available at the end of October and can be accessed via internet.

[http://www.witi.cs.uni-magdeburg.de/iti\\_bm/ibss/](http://www.witi.cs.uni-magdeburg.de/iti_bm/ibss/)  
Furthermore, a book will appear at the beginning of the next year.