

**EXPANDING PROTEOMICS TO GLYCOBIOLOGY:
BIOCOMPUTING APPROACHES UNDERSTANDING THE
FUNCTION OF SUGAR**

C.-W. VON DER LIETH

DKFZ (German Cancer Research Centre), Molecular Modelling, INF 280, 69120

Heidelberg, Germany, {a.bohne,w.vonderlieth}@dkfz.de

The recognition of complex carbohydrates and glycoconjugates as mediators of important biological processes has stimulated investigations into the understanding of the underlying principles. Unfortunately, the rate of generating new information has been slow during the last decade. Carbohydrates differ from the two other classes of biological macromolecules (proteins and DNA/RNA) in two important characteristics: their residues can be connected by many different linkage types and they can form highly branched molecules. As a consequence, carbohydrate chains contain an evolutionary potential of information content, which is several order of magnitude higher in a short sequence than any other oligomer formed by nucleotides or amino acids. This structural variance allows oligosaccharides to encode information for specific molecular recognition and to serve as determinants of protein folding and stability. However, structural complexity is also one of the major barriers for a rapid progress in the field of glycobiology because carbohydrates are laborious to analyze and extremely difficult to synthesize. Knowing that glycosylation of proteins and lipids are the most ubiquitous forms of posttranslational modification, the unexpectedly small number of genes identified in the initial analysis of the human genome sequence provides even more efforts for understanding the biological roles of oligosaccharides.

Whereas the conformational behavior of oligosaccharides has been investigated intensively both by experimental techniques (mainly NMR) as well as various theoretical methods their study (function and structure) has been widely neglected so far in the area of bioinformatics. The utilization of proteomics databases has become indispensable for the daily work of the molecular biologist, but this situation has not been reached for carbohydrate applications yet. Several new and useful applications for the glycosciences (see http://www.dkfz.de/spec/links/glyco_list.html) have appeared on the web during the last few years. Unfortunately, existing data collections are only rarely annotated and cross-linked to other resources. The need to develop and maintain Internet based, databases for carbohydrate structures, carbohydrate binding proteins and carbohydrate active enzymes has recently been emphasized by the Consortium for Functional Glycomics (see <http://glycomics.scripps.edu/>)

The paper of Cooper, Harrison, Webster, Wilkins and Packer points out the need for the standardization of the entries in glyco-databases. GlycoSuiteDB is an annotated

database of glycans structures designed to provide rapid access to information on protein glycosylation. The paper describes how the glycan structure representation is normalized to provide consistency and how to enable different searching criteria. Currently, no generally accepted linear, canonical description for carbohydrates exists. Such a code, which can be processed by computers with ease, will enable efficient automatic cross-linking of distributed carbohydrate data collections by serving as a unique and unambiguous database access key. One of the major points to bring up for discussion during this session will probably be the question how to come to an agreement about a generally accepted linear, canonical description for carbohydrates.

Glycosylation may influence considerably the physicochemical properties and function of a protein. Since the experimental determination of glycosylated sites is difficult to archive - the percentage of annotated glycoprotein entries in SWISS-PROT is still low - the need for developing theoretical approaches to predict glycosylation potential of sequons is obvious. The group of S. Brunak has developed methods to predict glycosylation sites using artificial neural networks that examine correlations in the local amino acid sequence context and their surface accessibility. Here Gupta and Brunak analyze the data available in SWISS-PROT of human proteins with respect to certain functional categories. The proteins are clearly classified highlighting that the N-glycosylation occurs mainly in the 'transport and binding' category of proteins. The authors show that the glycosylation is one of the most important determinants for functional classification of proteins and has definitively taken into account when deciphering protein function and characterizing complete proteomes.

The question if glycans attached to proteins and lipids do show certain secondary or tertiary structural motifs still remains to be answered due to the lack of sufficient crystallographic data. The paper of Bohne and von der Lieth claims that the spatial structure of glycans provides the driving force for many intermolecular interactions and thus predetermines their functions. Moreover, the authors emphasize that flexibility and dynamics of glycans may play a key role in their biological activity and must also be taken into account. The paper describes a new computational method to explore the conformational space of N-glycans. Since the approach is very fast it is well suited for a web-based application.

As already mentioned the treatment of complex carbohydrates and glycoconjugates has so far been no major topic in the area of bioinformatics. The small number of only five papers submitted for this session reflects this situation. Moreover, it turned out to be rather difficult to encourage glycoscientists, who are not used to submit a full paper six months before the conference takes place, to contribute. Nevertheless, I am convinced that the subject is hot and that it will attract more attention in the near future. Since the three selected contributions address scientific questions with emerging interest where much research and development remains to be carried out, they provide an excellent basis for stimulating discussions.

