Full text provided by www.sciencedirect.com
SCIENCE DIRECT®

ELSEVIER

Prospects for glycoinformatics

Serge Pérez[1] and Barbara Mulloy[2]

High-throughput and automated techniques (mass spectrometry in particular) allow increasingly rapid structural analysis of complex glycans. Information concerning the primary structure (composition, sequence and linkages), three-dimensional structure (including dynamics) and interactions of glycans is now available in sufficient quantity to justify the maintenance of databases and search facilities. Several such resources (both commercial and open access) are now available as web tools. To derive the full value of glycan databases, it will be necessary to develop a universally accepted machine-readable structural representation of glycans.

Addresses
[1] Centres de Recherches sur les Macromolécules Végétales, CNRS, BP 53, 38041 Grenoble, France
[2] Laboratory for Molecular Structure, NIBSC, Blanche Lane, South Mimms, Potters Bar, Herts EN6 3QG, UK

Corresponding author: Pérez, Serge (serge.perez@cermav.cnrs.fr)

Current Opinion in Structural Biology 2005, **15**:517–524

This review comes from a themed issue on
Carbohydrates and glycoconjugates
Edited by Ten Feizi and Barbara Mulloy

Available online 6th September 2005

0959-440X/$ – see front matter
© 2005 Elsevier Ltd. All rights reserved.

DOI 10.1016/j.sbi.2005.08.005

Introduction

Over the past few years, there has been an increase in the application of systematic methods to the study of glycans and their interactions. In parallel with this trend, a revival in the number and scope of related databases, web sites, and both real and virtual glycan libraries has begun to address the information needs of glycobiologists and glycochemists. We intend the term 'glycoinformatics' to refer to informatics tools available for assessing 'primary data' (covalent and three-dimensional structures of glycans and glycoconjugates), and organizing these primary data into databases that can be used for speeding up the production of primary data, predicting new features and characterising structure/activity or structure/function relationships (Figure 1). There are several levels of glycan structural data for which this approach may be useful, from sequence through three-dimensional structure to interactions (Figure 2).

This review gives an overview of the status, methods, requirements and perspectives on the application of bioinformatics to the field of glycosciences. It covers the integration of computational methods in the elucidation of the several levels of glycan structural organisation, dynamics and interactions, and provides a list of web tools and databases relevant to the field.

From primary sequence to three-dimensional structures, dynamics and interactions

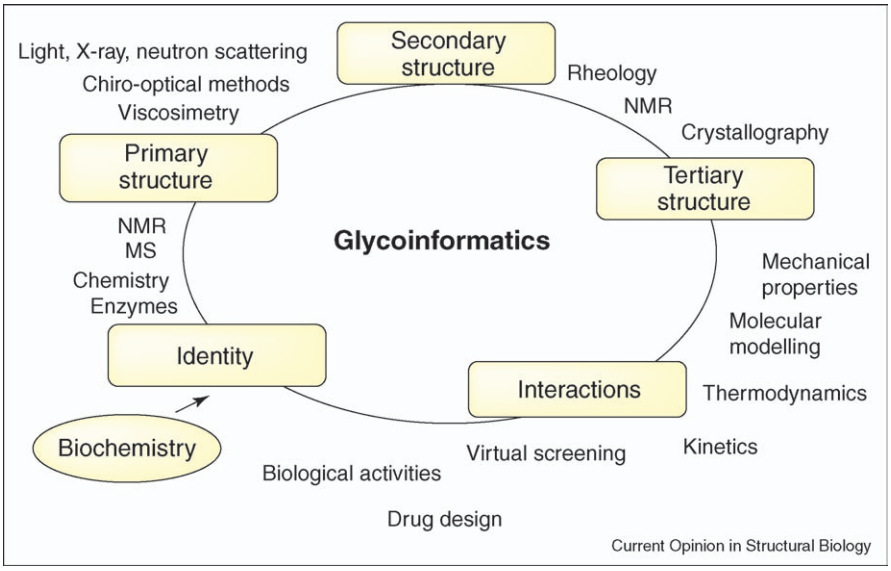
Representation in text of the primary structure, or sequence, of complex carbohydrates is best described following the IUPAC-IUBMB (International Union for Pure and Applied Chemistry and International Union for Biochemistry and Molecular Biology) terminology [1,2], in its extended, abbreviated and condensed forms. These forms are widely used within the carbohydrate community and are adequate for describing the complete sugar sequence, including monosaccharide stereochemistry, anomeric configuration and linkage information. Recommendations also apply to the description of polysaccharides [3] and glycoconjugates [4]. Glycoprotein glycans are also often represented graphically, using coloured shapes to indicate specific monosaccharides (e.g. http://www.functionalglycomics.org/static/gt/gtdb.shtml).

Sequencing complex carbohydrates

Despite recent progress in spectroscopy, the sequencing of structurally complex and diverse carbohydrates remains a challenging task. Mass spectrometry (MS) is the method of choice for sensitive identification and characterization, particularly for glycans released from glycoproteins, and automated high-throughput methodologies are currently being developed [5,6], some of which are also applicable to glycosaminoglycans (GAGs) [7•]. Automatic MS spectrum interpretation is still a matter of active exploration [8•,9] and the success of developed algorithms depends on the family of complex carbohydrate investigated. Significant successes can be obtained with screening specific classes, such as *N*-linked glycoprotein glycans, for which databases of existing structural data are particularly complete and for which we have solid knowledge of biosynthesis pathways [10,11•]. The combination of MS techniques with the use of accessible databases can also be used in the early stages of glycan analysis [12].

Structure determination often also involves the use of NMR spectroscopy, with recent improvements in sensitivity; NMR provides complementary information to MS and the combination is powerful [13]. Assignment of ^{1}H and ^{13}C NMR spectra of complex carbohydrates can be achieved with an algorithm using an additivity scheme, such as that developed in CASPER (Computer-Aided

Figure 1



Data types, techniques and the generation of knowledge concerning complex glycans are inter-related. At all stages in this cycle of knowledge generation, comprehensive collections of reliable data on standard and new compounds can accelerate the process, avoiding the costly replication of effort involved in individual searches through the primary literature.

Spectrum Evaluation of Regular Polysaccharides) [14]. This allows both simulation of spectra of known structures and sequence determination of unknown structures using information from chemical analysis and unassigned NMR spectra.

Three-dimensional features

Both NMR and X-ray diffraction can be used to assess three-dimensional features of complex carbohydrates. In solution, the method of choice to study conformation is NMR, through parameters such as chemical shifts, coupling constants, nuclear Overhauser effects (NOEs) and also relaxation time measurements [15]. However, a major difficulty arises from the flexibility of carbohydrates, especially the glycosidic links. When multiple conformations are present in solution, NMR data will represent a time-averaged conformation. As the geometrical parameters are usually related in a non-linear way to the experimental data, these data can be misleading.

Almost all the available experimental data on three-dimensional structures come from X-ray diffraction. Molecular and crystal structures of small and medium-sized oligosaccharides can be found in the Cambridge Structural Database (CSD) [16], which does not have a web interface. Out of the 4000 entries classified as carbohydrates, many are monosaccharides. Compounds of either structural or biological interest have been described in a recent review [17]. Among the crystal structures, no more than ten disaccharidic fragments of glycoconjugates and three trisaccharides are reported. The largest oligosaccharide crystal structure so far reported is that of tricolorin, a glycolipid extracted from a Mexican plant used in traditional medicine [18].

For glycoproteins and protein–carbohydrate complexes, an increasing number of crystal structures have been reported, as the result of significant and rapid progress in the use of synchrotron radiation; sufficient data are available for useful statistical analyses [19]. In the case of glycan macromolecular structures, intrinsic conformational flexibility is mainly responsible for their difficulty to crystallize. Therefore, relatively few crystallographically solved structures are available from the Protein Data Bank (PDB) [20]. A recent study identified 1562 entries in the PDB, with a total of 5397 carbohydrate chains. The majority of the chains correspond to *N*-glycosidically bound glycans, whereas *O*-glycans are a minority. Non-covalently bound carbohydrate ligands are also found; for example, more than 250 crystal structures of lectins have been solved, most of them as complexes with carbohydrate ligands. Many of these complexes involve biologically important oligosaccharides for which no structural information was available [21].

Polysaccharides form the most abundant family of biopolymers, offering a diversity of structures ranging from simple linear homopolymers to branched heteropolymers, with repeat units that consist of up to octasaccharides. In contrast to other macromolecules, the three-dimensional structural data that can be obtained on polysaccharides come from X-ray fibre diffraction and are not sufficient to permit crystal structure determination. Modelling techniques must be used that allow the calculation of diffrac-

Download English Version:

https://daneshyari.com/en/article/10822747

Download Persian Version:

https://daneshyari.com/article/10822747

Daneshyari.com