Databases and ontologies

GlycoBase and autoGU: tools for HPLC-based glycan analysis

Matthew P. Campbell*,†, Louise Royle‡, Catherine M. Radcliffe‡, Raymond A. Dwek and Pauline M. Rudd†

Oxford Glycobiology Institute, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK

Received and revised on February 8, 2008; accepted on March 4, 2008

Advance Access publication March 14, 2008

ABSTRACT

Summary: The development of robust high-performance liquid chromatography (HPLC) technologies continues to improve the detailed analysis and sequencing of glycan structures released from glycoproteins. Here, we present a database (GlycoBase) and analytical tool (autoGU) to assist the interpretation and assignment of HPLC-glycan profiles. GlycoBase is a relational database which contains the HPLC elution positions for over 350 2-AB labelled N-glycan structures together with predicted products of exoglycosidase digestions. AutoGU assigns provisional structures to each integrated HPLC peak and, when used in combination with exoglycosidase digestions, progressively assigns each structure automatically based on the footprint data. These tools are potentially very promising and facilitate basic research as well as the quantitative high-throughput analysis of low concentrations of glycans released from glycoproteins.

Availability: http://glycobase.ucd.ie

Contact: matthew.campbell@nibrt.ie

1 INTRODUCTION

Glycobiology is a distinct field that seeks to determine the roles of sugars in biology; it includes the release, labelling, separation and sequencing of glycan structures that are covalently attached to proteins or lipids. The structure determination of glycans is a difficult task partly due to the large diversity of structures found in nature, the complex biosynthetic structural variations and the physical and chemical similarities of the monosaccharide. The next major challenge in the post-genomic era is to fully determine and understand the implications of post-translational modifications, such as glycosylation.

Bioinformatics solutions for glycobiology and glycomics are still in their infancy compared to those tools available in the proteomic and genomic fields. Several large scale initiatives have been established to develop technologies and resources to advance data handling of large and diverse datasets and to assist data interpretation these include the Consortium for Functional Glycomics (CFG) (Raman et al., 2006), Kyoto Encyclopedia of Genes and Genomes (KEGG) (Hashimoto et al., 2006), Glycosciences.de (Lutteke et al., 2006), CAZy (http://afmb.cnrs-mrs.fr/CAZY/), BCSDB (Toukach et al., 2007), GlycoSuiteDB (Cooper et al., 2003) and CarbBank (Doubet and Albersheim, 1992). The most recent of these projects is the EUROCarbDB initiative (www.eurocarbdb.org) which aims to establish a framework for glycan data deposition obtained by high-performance liquid chromatography (HPLC), MS and NMR techniques and the development of tools to assist data interpretation.

One of the main limiting factors restricting the development and application of glycobiology is the lack of a well established, rapid and automated high-throughput platform. We have recently developed and validated a HPLC technology based on a 96-well plate format for the detailed analysis of glycans at concentrations required for biomedical applications (low femtomoles of N-linked sugars released from micrograms of glycoproteins). However the interpretation, annotation and assignment of HPLC-glycan data is currently a manual and very time-consuming aspect of glycan analysis performed by experts (Anumula, 2000; Guile et al., 1996; Royle et al., 2006). Consequently, we have built a relational database (GlycoBase) and an analytical tool (autoGU) in conjunction with EUROCarbDB to assist data interpretation to bring glycan analysis within the reach of any well-established laboratory.

2 DESIGN AND IMPLEMENTATION

GlycoBase and autoGU are based on open-source technology, developed in Perl, Common Gateway Interface (CGI) and tested on SUSE 10.0 running Apache 2.0. The relational database was set-up in MySQL 5.0 and the schema is based on the EUROCarbDB-HPLC data model which can be easily modified and extended. The html-based web interface has been tested using IE6.0-7.0 and Firefox 1.0-2.0.

3 APPLICATIONS AND PERSPECTIVES

GlycoBase and autoGU are novel developments for the interpretation and annotation of glycan sequencing data.

3.1 HPLC strategy for glycan analysis

Normal–Phase (NP)-HPLC using amide-based columns is a robust and reproducible method for high-resolution separation
of N-linked glycans released from glycoproteins (Anumula, 2000, 2006; Guile et al., 1996; Tomiya et al., 1988). Released glycans labelled with a fluorophore e.g. 2-aminobenzamidazole enables detection at the femtomole level. The advantages of NP-HPLC analysis of 2-aminobenzamide (2AB)-labelled glycans are: (i) a total glycan pool including charged and neutral glycans can be analysed at once compared to capillary electrophoresis (CE) where samples are usually desialylated before analysis or matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) MS in which sialic acid linkages are unstable; (ii) the 1:1 stoichiometric labelling of the released glycans allows for accurate and quantitative measurement of the relative amounts of individual glycans; (iii) the addition or subtraction of monosaccharides shifts the retention time by predictable amounts (unlike RPHPLC, CE or HPAEC-PAD) enabling detailed structural assignments to be made when HPLC analysis is combined with exoglycosidase digestions and database matching and (iv) NP-HPLC can separate structures with the same composition on the basis of sequence and linkage type (mass-spectrometry analysis can distinguish differences when used in combination with chromatographic separation and/or fragmentation analysis).

Glycan profiles from NP-HPLC are calibrated against a 2AB-labelled dextran ladder (2AB-glucose homopolymer, Ludger Ltd.) and assigned glucose unit (GU) values by fitting a fifth order polynomial distribution curve to allocate GU values from retention times (using Empower GPC software from Waters Ltd). Glycan structures/composition are assigned using the database of GU values and confirmed by a series of exoglycosidase digestions (Royle et al., 2006). GU values can be used as reference standard values because the calibration minimizes day-to-day and/or system variations.

We have developed a high-throughput method for the efficient release and 2AB labelling of 96 glycoprotein samples in 2–3 days (Fig. 1). Preliminary HPLC analysis takes an additional 2 days (using 30 min run) followed by exoglycosidase sequencing steps and structural assignments of the glycan pool by database matching (GlycoBase and autoGU); a similar approach was recently used in a pilot scale to identify glycosylation changes in total serum of patients with advanced ovarian cancer (Saldova et al., 2007). The high-throughput methodology with supporting data has been recently published (Royle et al., 2007).

### 3.2 GlycoBase

GlycoBase contains more than 350 2AB-labelled N-linked glycan structures including 117 identified in the human serum glycome (Fig. 2). All structures were determined at the Oxford Glycobiology Institute by a combination of NP-HPLC with exoglycosidase sequencing and mass spectrometry (MALDI-MS, ESI-MS, ESI-MS/MS, LC-MS, LC-ESI-MS/MS). GlycoBase contains the HPLC elution positions (expressed as glucose unit values) for each individual glycan alongside the products of exoglycosidase digestions (Royle et al., 2006). The GU value for each glycan is related to the number and linkage types of its constituent monosaccharides. The SD of the GU value is ≤0.1 for 91% (≤0.2 for 97%) of neutral structures and ≤0.2 for 95% of charged structures, with more than one reference value. This is calculated for over 800 referenced GU values from the accumulated experimental data in GlycoBase. Each entry is comprehensively annotated and includes the following: a pictorial representation of the structure depicting monosaccharide sequence and linkages; NP-HPLC retention time expressed as an average GU value, with SD (calculated from all listed published data for that structure); the monosaccharide composition; related reference information; links to the identified exoglycosidase digest products and a list of the subgroups in which the glycan can be found. The subgroups eliminate structures which are not found/possible therefore assisting glycan searches; humans do not encode the glycosyltransferase responsible for adding core α1–3 fucose units unlike plants.

Many of the well-established glycan databases offer a vast amount of information including: glycan biosynthetic pathways; glyco- and microarray data and structural and chemical
information for natural and chemically synthesized glycans. We recognize that our dataset is comparatively small but none of these large databases provide standardized experimental information which can be used to interpret HPLC-glycan data with a high degree of accuracy. Consequently, GlycoBase is an important development in the field of glycobiology. The availability of high quality data generated by a highly sensitive, robust and quantitative method has recently been used to assist full glycan characterization (Domann et al., 2007; Saldova et al., 2007).

3.3 autoGU

The interpretation of a complete set of exoglycosidase digestions can be very time consuming and a difficult exercise for an inexperienced glycobiologists. We have developed database matching software (autoGU) which automatically assigns possible glycan structures to each HPLC peak. When used in combination with data from a series of exoglycosidases autoGU will create a refined list of structures based on the digest footprint i.e. shifts in GU values due to cleavage of terminal monosaccharides dependent on enzyme specificity.

Integrated HPLC data in the format of GU values and percentage areas is initially submitted to a search of entries in GlycoBase. Preliminary glycan structures are assigned to each peak where a GU match exists using the mean and standard deviation values for each structure or ±0.3 GU where only one reference source is reported.

Initially, several different glycan structures can be assigned to each peak (undigested sample) however each glycan structure has a unique exoglycosidase digest footprint (all digestion pathways are stored in GlycoBase). The sequential digestion of a glycan with exoglycosidases can be used to accurately and quantitatively sequence a glycan pool (Royle et al., 2006). The removal of monosaccharides from the non-reducing end changes the retention time by a predictable amount, therefore structural assignments can be made when HPLC analysis is combined with exoglycosidase digestions. When the results from a set of glycan exoglycosidase profiles are submitted, autoGU progressively analyses the data, to produce a refined list of final structural assignments that match the digestion data using the preliminary (undigested sample) assignments as the initial dataset (Fig. 3).

These glycan tools have been extensively evaluated and shown to assist and improve the accuracy of HPLC-glycan data interpretation (Domann et al., 2007; Saldova et al., 2007). GlycoBase will be updated on a regular basis and will include N- and O-linked glycans from natural sources e.g. immunoglobulins, human serum and modulation of glycosylation in cell culture. We anticipate that the continued growth will further enhance the application of GlycoBase to accurately annotate HPLC-glycan profiles. These tools will be incorporated into theEUROCarbDB framework for the community including the development of databases for additional labelling procedures and HPLC methods.

ACKNOWLEDGEMENTS

We are actively working in collaboration with the EUROCarbDB (http://www.eurocarbdb.org) RIDS contract number: 011952. This work was supported by the Oxford Glycobiology Institute endowment. We thank the Wellcome Trust and the Biotechnology and Biological Sciences Research Council for grants to purchase the MALDI-TOF and Q-Tof mass spectrometers that were used in this work.

Conflict of Interest: none declared.

REFERENCES


Hashimoto, K. et al. (2006) KEGG as a glycome informatics resource. Glycobiology, 16, 63R–70R.

Lutke, T. et al. (2006) GLYocosciences.de: an Internet portal to support glycomics and glycobiology research. Glycobiology, 16, 71R–81R.


Saldova, R. et al. (2007) Ovarian cancer is associated with changes in glycosylation in both acute-phase proteins and IgG. Glycobiology, 17, 1344–1356.
