

Databases and ontologies

GlycoStore: a database of retention properties for glycan analysis

Sophie Zhao¹, Ian Walsh¹, Jodie L. Abrahams², Louise Royle³,
Terry Nguyen-Khuong¹, Daniel Spencer³, Daryl L. Fernandes³,
Nicolle H. Packer^{2,4}, Pauline M. Rudd¹ and Matthew P. Campbell^{2,*}

¹Bioprocessing Technology Institute, Agency for Science, Technology and Research (A*STAR), Singapore, ²Institute for Glycomics, Griffith University, Gold Coast, Australia, ³Ludger Ltd., Culham Science Centre, Oxfordshire, UK and ⁴Department of Chemistry and Biomolecular Sciences, Macquarie University, Sydney, Australia

*To whom correspondence should be addressed.

Associate Editor: Janet Kelso

Received on October 10, 2017; revised on March 28, 2018; editorial decision on April 18, 2018; accepted on April 19, 2018

Abstract

Summary: GlycoStore is a curated chromatographic, electrophoretic and mass-spectrometry composition database of *N*-, *O*-, glycosphingolipid (GSL) glycans and free oligosaccharides associated with a range of glycoproteins, glycolipids and biotherapeutics. The database is built on publicly available experimental datasets from GlycoBase developed in the Oxford Glycobiology Institute and then the National Institute for Bioprocessing Research and Training (NIBRT). It has now been extended to include recently published and in-house data collections from the Bioprocessing Technology Institute (BTI) A*STAR, Macquarie University and Ludger Ltd. GlycoStore provides access to approximately 850 unique glycan structure entries supported by over 8500 retention positions determined by: (i) hydrophilic interaction chromatography (HILIC) ultra-high performance liquid chromatography (U/HPLC) and reversed phase (RP)-U/HPLC with fluorescent detection; (ii) porous graphitized carbon (PGC) chromatography in combination with ESI-MS/MS detection; and (iii) capillary electrophoresis with laser induced fluorescence detection (CE-LIF). GlycoStore enhances many features previously available in GlycoBase while addressing the limitations of the data collections and model of this popular resource. GlycoStore aims to support detailed glycan analysis by providing a resource that underpins current workflows. It will be regularly updated by expert annotation of published data and data obtained from the project partners.

Availability and implementation: <http://www.glycostore.org>

Contact: m.campbell2@griffith.edu.au

Supplementary information: [Supplementary data](#) are available at *Bioinformatics* online.

1 Introduction

The majority of human proteins are glycosylated and alterations in glycosylation impact numerous physiological and pathological processes (Varki, 2017). Technology advances across a broad range of analytic approaches, including mass spectrometry (MS), liquid chromatography, capillary electrophoresis (CE) and other orthogonal strategies, are generating increasingly expansive glycomic datasets (Novotny and Alley, 2013). Robotised platforms are allowing the

facile and rapid preparation of thousands of fluorescently labelled released complex glycan samples using ultra-high performance liquid chromatography (UHPLC) (Stockmann *et al.*, 2015). Access to datasets provides an opportunity to expand our understanding of the role of glycosylation in biological processes and disease, and also highlights a growing requirement for suitable bioinformatics infrastructure. GlycoBase was released in 2008 (Campbell *et al.*, 2008)

and contained the HPLC elution positions for more than 350 2-AB labelled N-glycan structures (Royle *et al.*, 2008). Over the last decade GlycoBase has been enriched by CE, UPLC and reversed phase ultra-performance liquid chromatography elution positions for 330 additional glycan structures. However, despite its popularity and growth, maintaining and developing GlycoBase has become increasingly difficult due to reliance on the discontinued EUROCarbDB framework (von der Lieth *et al.*, 2011).

To preserve the information accumulated over the last few decades we have developed GlycoStore, which integrates all publicly accessible data from GlycoBase with new evidences such as glycosphingolipid glycan headgroups from human serum acquired by this multi-institutional program. In this manuscript, we describe the data coverage, search features, technical framework of GlycoStore, with supporting documentation available at <http://unicarbk.freshdesk.com>.

2 Design and implementation

The GlycoStore web application has been developed in Java and Scala using the Play Framework. All data is stored in an Apache Jena TDB triple store using a Resource Description Framework (RDF) format as defined by the GlycoRDF ontology (Ranzinger *et al.*, 2015). A Spring API has been developed to model SPARQL operations as Java objects that are rendered by the Play Framework. This platform has been developed to meet a demand to coordinate RDF activities and data cross-referencing in the glycosciences space. A simplified schema demonstrating connections between classes and properties is shown in [Supplementary Figure S1](#).

GlycoStore provides elution property information for over 850 unique structures including standardized retention times, expressed as glucose units (GU) and arabinose units (AU), for 2AB and procainamide labelled glycans determined by a combination of HPLC, UPLC and RP-UPLC techniques. Time based data for reduced N-glycans run on Porous Graphitized Carbon (PGC) is also available with supporting MS/MS spectra, along with a growing CE dataset of APTS labelled glycans. The database is organized into 12 collections, each associated with a set of samples analyzed using a described workflow that includes 1654 CE, 2429 HPLC, 5646 UPLC, 543 PGC and 450 RPUPLC evidence entries. These include technical repeats and/or sequential exoglycosidase digestion profiles. For each sample set we provide representative metadata describing (i) sample preparation procedures encompassing glycan release techniques and methods that alter glycan structure, including exoglycosidase treatment and derivatization, and (ii) general analytical approach. An overview of the data model is shown in [Supplementary Figure S2](#).

3 Search, browse and filter

GlycoStore offers a variety of search methods categorized by either: (i) experimental values such as GU, AU or time (min); (ii) monosaccharide composition; or (iii) metadata labels such as taxonomy, sample name and the Oxford linear notation. This functionality can be accessed from the 'Search' menu. 'Show Structures' is a quick method to view all glycan structures stored in the database. This new structure interface can be used to search for glycans matching particular properties e.g. mass, composition or specific motif. Each structure entry page summarizes all recorded experimental measurements (sourced from published or in-house analysis), the associated global profiles in which this glycan is present, as well as experimentally determined exoglycosidase digestion pathways and biological source ([Supplementary Fig. S3](#)). 'Show Collections' lists the 12 datasets described in [Supplementary Table S1](#). By following the appropriate

link, users can explore the glycan and experimental content associated with each collection. To expedite searches either by glycan name or by experimental values (GU, AU or time) a shortcut box is fixed to the navigation bar ([Supplementary Fig. S4](#)). An objective of GlycoStore is to provide public access to a growing, curated database of glycan structural information characterized by the above techniques. The 'References' link from 'Show All' lists all curated manuscripts providing similar contextual information described above.

3.1 Structure graph database

A structure RDF-graph database, based on an approach introduced by (Alloci *et al.*, 2015), has been implemented for (sub)structure searching based on the GlycoCT format (Herget *et al.*, 2008). This framework can be used to search related content available in the UniCarbKB database, and to facilitate data-sharing with other knowledgebases e.g. the structure repository GlyYouCan (Aoki-Kinoshita *et al.*, 2016). For an explanation of the Data endpoint refer to <https://bitbucket.org/glycostore>.

4 Conclusions and future work

In this application note we describe the first release of GlycoStore, an international effort to provide a centralized resource that combines glycan structure information with chromatographic separation and electrophoretic data. It contains the largest collection of curated and in-house LC and CE experimental data on glycan structures with associated research literature. We will continue to adapt its data gathering, processing and user interfaces to support on-going developments in separation-MS-based analytical workflows, especially integration with ion-mobility data collections (Struwe *et al.*, 2016). In the future GlycoStore will be linked with UniCarbKB and GlyGen projects to improve data interoperability.

Funding

This work was supported by the Institute for Glycomics, Macquarie University-Ludger Pilot Scheme, A*STAR's Joint Council Visiting Investigator Programme (HighGlycoART) and Biomedical Research Council Strategic Positioning Fund (GlycoSing).

Conflict of Interest: none declared.

References

- Alloci, D. *et al.* (2015) Property Graph vs RDF Triple Store: a Comparison on Glycan Substructure Search. *PLoS One*, **10**, e0144578.
- Aoki-Kinoshita, K. *et al.* (2016) GlyYouCan 1.0—The international glycan structure repository. *Nucleic Acids Res.*, **44**, D1237–D1242.
- Campbell, M.P. *et al.* (2008) GlycoBase and autoGU: tools for HPLC-based glycan analysis. *Bioinformatics*, **24**, 1214–1216.
- Herget, S. *et al.* (2008) GlycoCT—a unifying sequence format for carbohydrates. *Carbohydr. Res.*, **343**, 2162–2171.
- Novotny, M.V. and Alley, W.R. Jr. (2013) Recent trends in analytical and structural glycobiology. *Curr. Opin. Chem. Biol.*, **17**, 832–840.
- Ranzinger, R. *et al.* (2015) GlycoRDF: an ontology to standardize glycomics data in RDF. *Bioinformatics*, **31**, 919–925.
- Royle, L. *et al.* (2008) HPLC-based analysis of serum N-glycans on a 96-well plate platform with dedicated database software. *Anal. Biochem.*, **376**, 1–12.
- Stockmann, H. *et al.* (2015) Automated, high-throughput serum glycoproteomics platform. *Integr. Biol. (Camb.)*, **7**, 1026–1032.
- Struwe, W.B. *et al.* (2016) GlycoMob: an ion mobility-mass spectrometry collision cross section database for glycomics. *Glycoconj. J.*, **33**, 399–404.
- Varki, A. (2017) Biological roles of glycans. *Glycobiology*, **27**, 3–49.
- von der Lieth, C.W. *et al.* (2011) EUROCarbDB: an open-access platform for glycoinformatics. *Glycobiology*, **21**, 493–502.