**Supplementary Materials for ‘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 Packer2,4, Pauline M Rudd1 and Matthew P Campbell2,\*

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, United Kingdom, 4Department of Chemistry and Biomolecular Sciences, Macquarie University, Sydney, Australia.

Table S1: A summary of the data collections available in GlycoStore including a description of the analytical technique and sample sets (http://www.glycostore.org/collections). New data collections released for GlycoStore 1.0 include: i) MQ Porous Graphitic Carbon (PGC), ii) BTI N-glycans, iii) BTI-GSL Standards, and iv) Ludger.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Collection Name | Description | Samples in Collection | Technique | Reference |
| BTI GSL Standards | 60 glycosphingolipids standards labelled with 2-AB and procainamide | Ganglio-, Lacto-, Neolacto-, Globo- and Iso-globo | HILIC-UPLC |  |
| CE Database | *N-*glycans analysed by capillary electrophoresis | Haptoglobin, IgG, Purified Standards, RNase B, Transferrin | CE-LIF | (Mittermayr, et al., 2011) |
| Human Serum *N-*glycans  NIBRT | Glycans characterised in house | Human Serum | HILIC-UPLC | (Saldova, et al., 2014) |
| Milk Oligosaccharides | A comprehensive overview of milk oligosaccharides from a variety of domestic animal species including cow, goat, sheep, pig, horse, and dromedary camel. | Cow, Dromedary Camel, Goat, Horse, Pig, Sheep | HILIC-UPLC | (Albrecht, et al., 2014) |
| MQ Porous Graphitic Carbon (PGC) | 90 *N-*glycan structures released from a number of standard glycoproteins. | Commercially available glycoprotein standards | PGC-LC-ESI-MS/MS | (Abrahams, et al., 2017) |
| NIBRT GSL | 28 glycosphingolipids from human serum and mammalian cell surfaces | Mammalian cell surfaces,  blood serum | UPLC-HILIC-FLD | (Albrecht, et al., 2016) |
| O-Glycans | 68 *O-*glycans | Various *O-*Glycans | HILIC-HPLC | (Royle, et al., 2003) |
| Original Release of GlycoBase | Original collection of *N-*glycans analysed at the Oxford Glycobiology Institute. | Human Serum, Human Immunoglobulins, Monoclonal Antibodies, Follicle Stimulating Hormone, and cell surface glycoproteins including  human CD’s | HILIC-HPLC | (Campbell, et al., 2008) |
| Royle (2008) Paper | HPLC-based analysis of serum *N-*glycans on a 96-well plate  platform with dedicated database software. | Human Serum | HILIC-HPLC | (Royle, et al., 2008) |
| RP IgG glycans | *N-*glycans characterised by reversed-phase UPLC | Human IgG | RP-UPLC |  |
| UPLC BTI | *N-*glycan structures analysed from a variety of samples | Human IgG  Alpha-1 antitrypsin (AAT),  Human Serum,  Mouse Serum | HILIC-UPLC |  |
| UPLC Ludger | A collection of *N-* and *O-*glycan structures characterised from Ludger standards and a variety of biological samples | Ludger commercial glycoprotein and glycan standards, Saliva, and Ocular mucin | HILIC-HPLC, HILIC-UPLC |  |

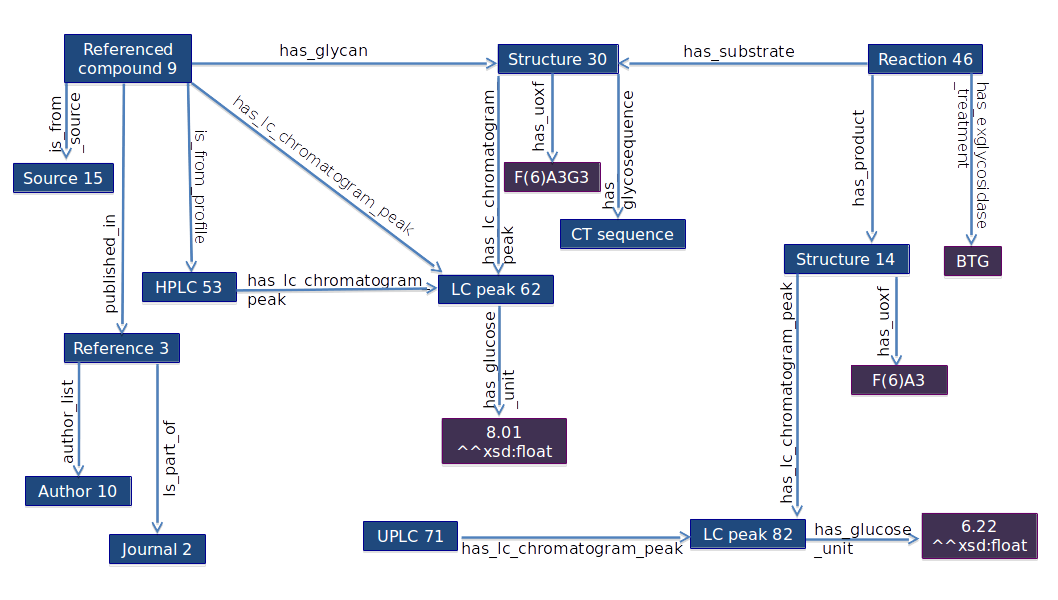


Figure S1: An overview of the GlycoStore data model. The RDF schema is based on the GlycoRDF syntax, which has been extended to support chromatography and electrophoretic data collections. Important inclusions are the experimental classes glycan:evidence\_hplc, glycan:evidence\_uplc and glycan:evidence\_rpuplc, which differentiate modes of chromatography with glycan:has\_glucose\_unit and glycan:has\_arabinose\_unit describing the evidence type. The ontology supports annotation of exoglycosidase array digestions with predicates describing commonly used enzymes, which can be applied to the *Reaction* class. The *ReferencedCompound* class connects information such as source, reference and evidence from one experiment profile to a single glycan structure. Finally, the *Source* class describes the origin of a glycan structure e.g. specification of the species, tissue, fluid, cell type, cell line, strains, life stage and related diseases. For simplicity, the figure focuses on the experimental data and omits the *Saccharide* and *Reference* classes that are described by the GlycoRDF ontology. GlycoRDF release 1.2.1 is available at https://github.com/ReneRanzinger/GlycoRDF (Ranzinger, et al., 2015).

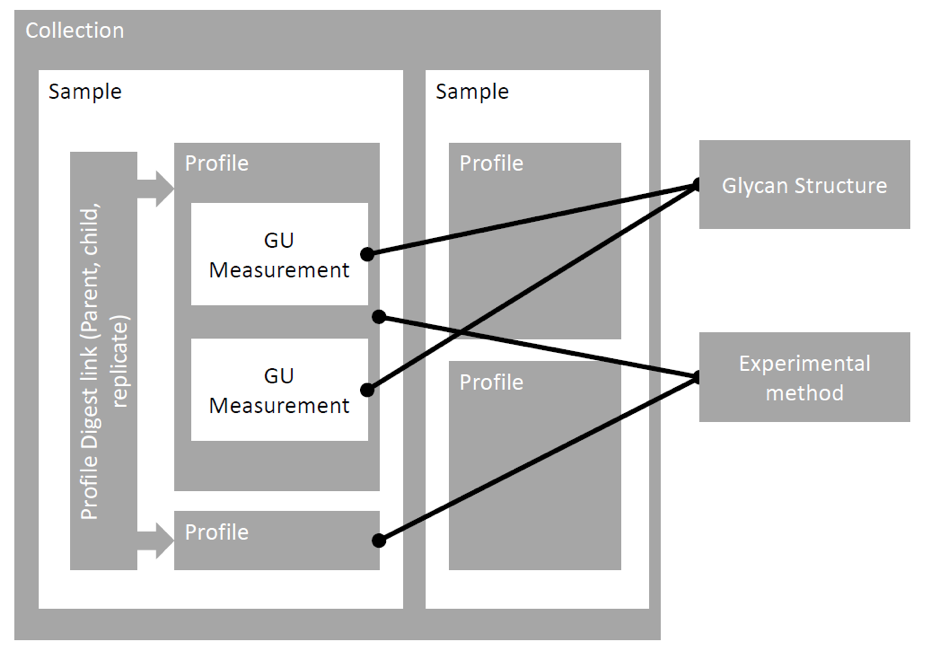


Figure S2: Data organisation and relationships. Glycan structures and experimental methods are stored outside the context of ‘Collections’ and ‘Samples’, but attached to GU/AU/Time measurements and ‘Profiles’ respectively. Individual glycan structures can be linked to multiple elution positions, across several reports and collections.



Figure S3: Glycan summary page for the tri-antennary fully galactosylated structure A3G(4)3. (a) For each glycan structure, all retention values and supporting literature records are summarised, and a complete description of the data source is shown in (b), which links to (c) a list of experimental profiles and (d) all recorded exoglycosidase evidence.

****

Figure S4: Example of a GlycoStore quick search. The glycan name search retrieves entries corresponding to the specified structure based on the Oxford shorthand nomenclature (Harvey, et al., 2009). The retention value search will return glycan entries that match the queried value with a tolerance of ± 0.5.

**References**

Abrahams, J.L., Campbell, M.P. and Packer, N.H. Building a PGC-LC-MS N-glycan retention library and elution mapping resource. *Glycoconj J* 2017.

Albrecht, S.*, et al.* A comparative study of free oligosaccharides in the milk of domestic animals. *Br J Nutr* 2014;111(7):1313-1328.

Albrecht, S.*, et al.* Comprehensive Profiling of Glycosphingolipid Glycans Using a Novel Broad Specificity Endoglycoceramidase in a High-Throughput Workflow. *Anal Chem* 2016;88(9):4795-4802.

Campbell, M.P.*, et al.* GlycoBase and autoGU: tools for HPLC-based glycan analysis. *Bioinformatics* 2008;24(9):1214-1216.

Harvey, D.J.*, et al.* Proposal for a standard system for drawing structural diagrams of N- and O-linked carbohydrates and related compounds. *Proteomics* 2009;9(15):3796-3801.

Mittermayr, S.*, et al.* Multiplexed analytical glycomics: rapid and confident IgG N-glycan structural elucidation. *J Proteome Res* 2011;10(8):3820-3829.

Ranzinger, R.*, et al.* GlycoRDF: an ontology to standardize glycomics data in RDF. *Bioinformatics* 2015;31(6):919-925.

Royle, L.*, et al.* HPLC-based analysis of serum N-glycans on a 96-well plate platform with dedicated database software. *Anal Biochem* 2008;376(1):1-12.

Royle, L.*, et al.* Secretory IgA N- and O-glycans provide a link between the innate and adaptive immune systems. *J Biol Chem* 2003;278(22):20140-20153.

Saldova, R.*, et al.* Association of N-glycosylation with breast carcinoma and systemic features using high-resolution quantitative UPLC. *J Proteome Res* 2014;13(5):2314-2327.