High-throughput Workflow for Glycan Profiling and Characterisation

Henning Stöckmann, Giorgio Carta, Ciara A. McManus, Mark Hilliard, and Pauline M. Rudd*

NIBRT GlycoScience Group, NIBRT – The National Institute for Bioprocessing Research and Training, Foster's Avenue, Mount Merrion, Blackrock, Co. Dublin, Ireland

E-Mail: *pauline.rudd@nibrt.ie

Received: 20th January 2014 / Published: 22nd December 2014

Abstract

Over the last 40 years, the understanding of glycosylation changes in health and disease has evolved significantly and glycans are now regarded as excellent biomarker candidates because of their high sensitivity to pathological changes. However, the discovery of clinical glycobiomarkers has been slow, mainly as a consequence of the lack of high-throughput glycoanalytical workflows that allow rapid glyco-profiling of large clinical sample sets. To generate high-quality quantitative glycomics data in a high-throughput fashion, we have developed a robotised platform for rapid *N*-glycan sample preparation and glycan characterisation. The sample preparation workflow features a fully automated, rapid glycoprotein affinity purification followed by sequential protein denaturation and enzymatic glycan release on a multiwell ultrafiltration device, thus greatly streamlining all required biochemical manipulations. After glycan purification on solid-supported hydrazide, glycans are fluorescently labelled to allow accurate quantification by ultra-high pressure liquid chromatography (ultra HPLC or UPLC). Subsequent peak assignment can be carried out utilising GlycoBase, a bespoke chromatographic data system developed to aid the analysis of glycans performed using different chromatographic techniques (UPLC, HPLC, Reverse Phase-UPLC, Capillary Electrophoresis).

This article is part of the Proceedings of the Beilstein Glyco-Bioinformatics Symposium 2013.
www.proceedings.beilstein-symposia.org

INTRODUCTION

Complex carbohydrates occur in a variety of forms and locations throughout the body and it is estimated that over half of eukaryotic proteins are in fact glycosylated [1]. Modern biopharmaceuticals are glycosylated, which can have a profound impact on pharmacodynamics and pharmacokinetics. *N*- and *O*-linked glycans are attached to both cell surfaces and secreted proteins. In turn, the diversity of possible monosaccharide combinations and conformations gives rise to oligosaccharides and protein site attachment variations also increase heterogeneity. Both the number and the structural diversity of glycans found in mammalian systems present significant analytical challenges for determining detailed glycan structural profiles in complex organisms. Glycosylation plays fundamental roles in many biological recognition events. Glycans undergo rapid structural changes in response to biological stimuli, providing a unique opportunity to identify and exploit glycans as clinical markers that can be indicative of specific disease states, disease progression, and/or therapy response. The concept of biomedical glycomics has gained considerable momentum and efforts are underway to rapidly identify glycans as disease biomarkers, to reveal the mechanisms that regulate glycan biosynthetic pathways, and to decode the functions of glycans in complex biological systems. In particular, large-scale glycoprofiling is being conducted in combination with human genome-wide association studies (GWAS) to understand the complex regulation of glycan expression in humans (Figure 1).

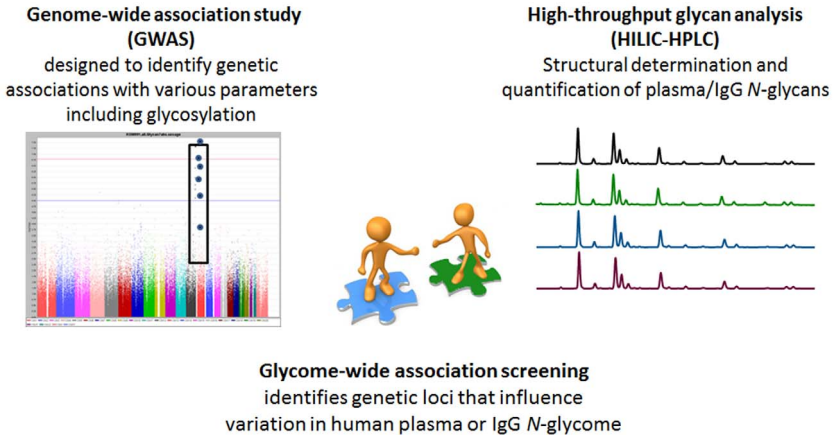

Figure 1. The 'Genomics meets Glycomics' concept. Genome-wide association studies (GWAS) are conducted to determine genetic variations that contribute to polygenic diseases, such as cancer and diabetes. Genetic associations can help develop improved strategies to detect, treat and prevent the disease. Glycome-screening studies help identify correlations between a disease phenotype and certain glycan features across glycomes. Linking GWAS and glycomics data unravels the regulation of glycan expression.

GWAS aim to identify genetic variations associated with particular diseases and involve the rapid screening of single-nucleotide polymorphism (SNP) markers across the genomes of thousands of individuals. When GWAS data are correlated with glycomics data, glycomic variations can be associated with genomic variations to elucidate the regulation of glycans, which can provide insights into pathophysiology. For example, linking GWAS data and glycomics data has been used to identify HNF1a as a master regulator of fucosylation [2].

Despite the great promise of GWAS-linked glycomics studies, low-cost technologies to rapidly obtain quantitative glycomics data from large sample numbers have not been developed. Although there is pressing need for high-throughput, reliable workflows in biomedical glycomics, biopharmaceutical development also require high-throughput glyco-profiling technologies, for example in the selection of clones that produce monoclonal antibodies or hormones with a particular glycosylation pattern.

High-Throughput Glycan Profiling Platform

We had previously developed a high performance liquid chromatography (HPLC)-based analysis for serum *N*-glycans in a 96-well format [3], and this has been successfully utilised in glycoprofiling studies in complex diseases such as cancer and diabetes [4 – 7]. However, while 'high-thoughput' can be carried out by intensive manual labour, the processing time can be up to three days for 96 samples and the method is difficult to automate. Thus, we have now developed the first fully automated, cost-effective multi-purpose glycomics platform, thereby considerably expanding the existing repertoire of glycomics workflows [8]. The main purpose of the platform is accurate glycan quantification in complex biological samples such as patient sera and cell culture supernatants. The platform is versatile and can be used to isolate individual glycoproteins or classes of glycoproteins (such as immunoglobulin G) and determine their glycosylation pattern. Glycoprotein samples are first prepared on a liquid handling robot (described in more detail below), and glycans are then separated with ultra-high performance liquid chromatography (UPLC), resulting in fluorescence chromatograms. Peak annotation is conducted using glycan sequencing and reference data stored in a dedicated database, termed GlycoBase (Figure 2). In addition, glycans can be labelled with different fluorophores, separated by a range of separation technologies both individual and coupled like LC/MS.

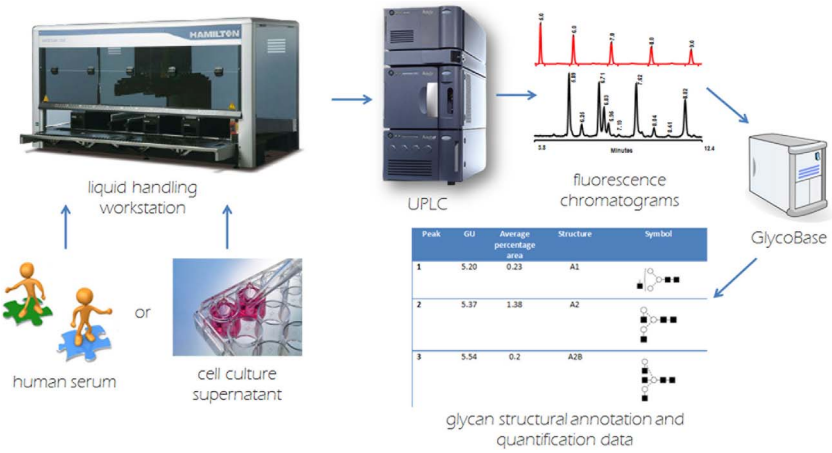

Figure 2. Glycomics workflow. Sera or cell culture supernatant samples are processed on a robotic workstation, resulting in fluorescently labelled glycans, which are subsequently separated and quantified by ultra-high pressure liquid chromatography (UPLC). Glycan peak annotations are performed based on enzymatic glycan sequencing and glucose unit (GU) data stored in the GlycoBase database.

Efficient and high-throughput glycan sample preparation is a key element of the glycomics platform. The workflow (Figure 3) is initiated by an automated affinity purification step of the glycoprotein of interest, e. g., immunoglobulin G (IgG). Subsequently, proteins are denatured and *N*-glycans released from the protein enzymatically by Peptide-*N*-Glycosidase F (PNGaseF). All biochemical operations are performed on an ultrafiltration plate to efficiently remove buffer and excess reagents and to separate the glycans from the protein.

Glyans are then captured on solid supported hydrazide to enable the removal of residual impurities which can interfere with fluorescent labelling. First, reducing-end glycans are reacted with solid-supported hydrazides to form hydrazones, thus, capturing glycans on the solid support. Contaminants such as excess reagents and buffer salts are removed by filtration. Next, glycans are released by acid catalysis in the presence of water. To equip glycans with stoichiometric amounts of fluorophore, they are labelled by reductive amination with 2-aminobenzamide (2-AB), a well-known and reliable reaction. Post-labelling clean-up is performed by solid-phase extraction.

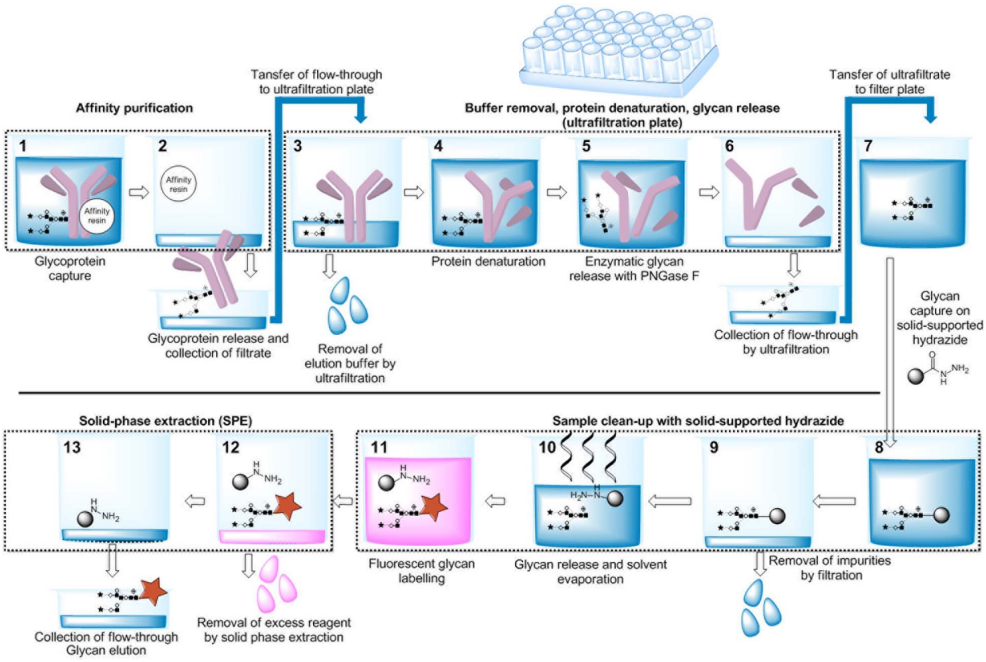

Figure 3. Sample preparation workflow on the NIBRT robotised platform. This schematic depicts the steps (1 – 13) for sample preparation, consisting of glycoprotein affinity purification (steps 1 – 2), buffer removal and protein denaturation (steps 3 – 4), enzymatic glycan release (step 5), glycan immobilisation on solid supports (step 8), removal of contaminants (step 9), glycan release (step 10), fluorescence labelling (step 11), and solid-phase extraction (steps 12 – 13). Reprinted and adapted with permission from reference [8]. Copyright 2013 American Chemical Society.

A robust robotic program with automatic error recovery was created to implement the workflow on a commercial liquid handling robot (Figure 4). The fully software-controlled workstation was equipped with eight robotic pipettes with individual liquid level and pressure sensors (Figure 4B), pipette tip racks, plate carriers, reagent reservoirs, a software-controlled vacuum manifold (Figure 4C), a temperature-controlled orbital shaker (Figure 4D), and a plate-transport tool to enable the movement of multi-well plates between positions and to operate the vacuum manifold.

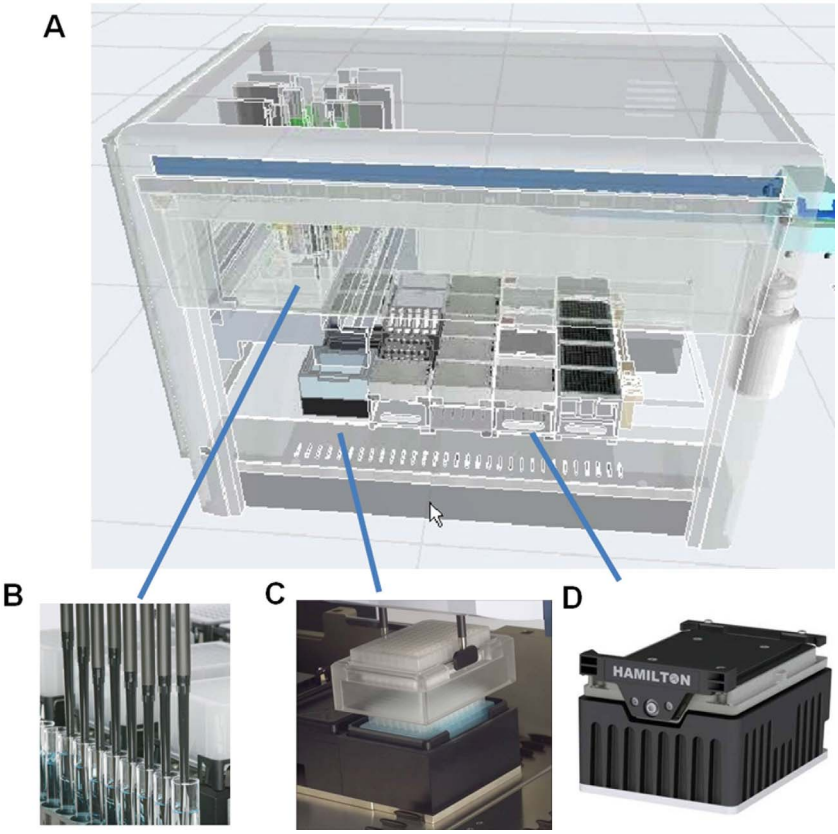

Figure 4. Schematic of the liquid-handling workstation utilised to perform high-throughput automated glycan sample preparation **(A)**. Key components include: pipetting channels with liquid-level detection and antidroplet control **(B)**, software-controlled robotic vacuum manifold and plate-transport tool **(C)**, and temperature-controlled orbital shaker **(D)**. Reprinted with permission from reference [8]. Copyright 2013 American Chemical Society.

The processing of up to 96 samples including glycoprotein affinity purification in a 96 well plate format typically takes around 22 h. The fluorescently labelled glycans are run on HPLC/UPLC instruments equipped with hydrophilic interaction chromatrography (HILIC) columns and the resulting peaks are correlated to a pre-run dextran ladder, thereby assigning a Glucose Unit (GU) value to each of the peaks. The use of standard glucose units makes these values independent of the running conditions; which allows for the direct comparison of chromatographic profile peaks and their relative glycan abundance (Figure 5). The coefficients of variation between samples prepared on different days with the automated robotised method for all major IgG peaks are typically below 10% (i. e., those peaks with a relative percentage area above 1%), indicating an excellent reproducibility.

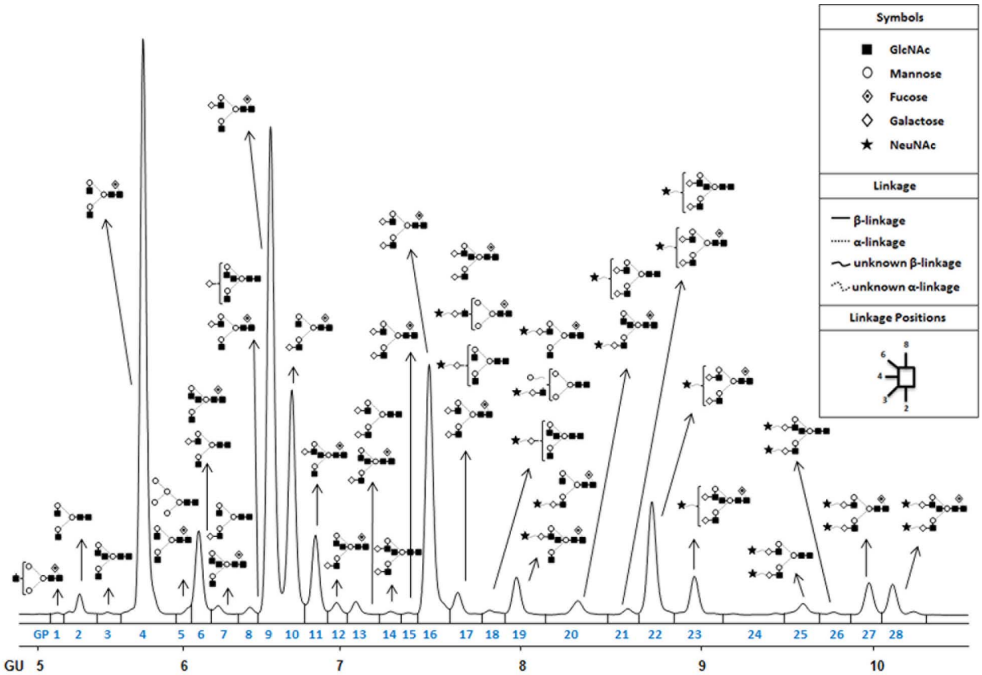

Figure 5. Human IgG N-glycome and peak assignments. IgG from native human serum was isolated and processed on the liquid-handling workstation followed by glycan analysis by ultra-high pressure liquid chromatography with fluorescence detection. GU: glucose units. Reprinted with permission from reference [8]. Copyright 2013 American Chemical Society.

Due to its robustness, high throughput and low cost, the platform is an ideal tool for efficient and accurate glycan profiling for GWAS and biopharmaceutical development and has been extensively used in these contexts (publications in preparation).

Glycan Structure Elucidation Using Glycan Sequencing and NIBRT's GlycoBase

Structure elucidation of glycan peaks requires reliable techniques and glycan reference data. Challenges in structural analysis include the large number of glycan classes and the efficient exploitation of analytical and bioinformatics tools that are available for structural interrogation. Data sources for glycan analytics encompass several orthogonal methodologies such as ultra-high pressure liquid chromatography (UPLC), capillary electrophoresis (CE) and mass spectrometry (MS), all of which have inherent difficulties in data interpretation. Assignment and characterisation of glycan structures in biotherapeutic products or high-throughput data from clinical profiling is a difficult and time-consuming process and is often a bottleneck in this type of research. It therefore requires automated

data-integration, data-mining and statistical analysis tools coupled with software engineering and database technology to advance this field of research. This would bring glycomic analysis in line with both the proteomic and genomic fields.

NIBRT's GlycoBase (www.glycobase.nibrt.ie, [9]) is an integrated solution for rapid and reproducible characterisation of glycan samples. Originally developed from the database EurocarbDB, GlycoBase is a resource for the storage, classification and reporting of glycan structures as well as their associated experimental values obtained using various chromatographic techniques such as HPLC, UPLC and CE. GlycoBase is a web-enabled, open-access resource that contains glycan data as normalised chromatographic retention time data, expressed as GU values, for more than 740 2-AB labelled *N*-linked glycan structures. These values were experimentally obtained by systematic analysis of released *N*-glycans from a diverse set of glycoproteins on the NIBRT glycan analytical platform utilising both Waters HPLC and UPLC analytical instruments. The database was built using data from many samples over the course of one decade. The UPLC data were obtained from many analytes including human serum. The Waters collection is a list of GU values pertaining to the analyses of a number of therapeutically interesting glycoproteins including erythropoietin and herceptin, haptoglobin, RNAse B and transferrin and is continually being expanded. Hydrophilic interaction liquid chromatography combined with fluorescence detection (HILIC-fluorescence), supplemented by exoglycosidase sequencing and mass spectrometric confirmation, was used to generate this high confidence glycan library. The resulting database has been made accessible through a customised web-application containing a simple and intuitive interface to assign and confirm glycan structures.

GlycoBase enables users to search for specific glycans using a variety of searching tools. These include searching by the regular expression name or by antennary composition (e. g., A1, A2 etc.). Alternatively searches can be carried out according to a GU value (± 0.3), or the user can search for a particular glycan feature, for example the presence or absence of sialic acid or core-fucose. The user also has the ability to carry out a stoichiometric search and thus search by, for example the number of hexoses or xyloses. All the searches can be performed on a global basis, thus searching the entire collection, a selected collection or a particular sample within a collection. GlycoBase provides users with access to a "summary report" which collates all the available data for a selected glycan. This includes information on general glycan properties such as the monoisotopic mass and the monosaccharide composition. Individual experimental records containing for example all the UPLC derived GU values recorded in the database are also shown on this summary page. Similarly, the user can view links to literature records, profile information as well as the instrument running conditions.

Reliable glycan peak assignments and structure elucidation are achieved through GU data from GlycoBase combined with glycan sequencing. Glycan sequencing is performed by exoglycosidase glycan digestion and is an ideal method for rapid oligosaccharide

characterisation including monosaccharide sequence and linkage information. Exoglycosidases remove carbohydrate residues from the non-reducing end of a glycan in a linkage-specific manner. For example, almond meal fucosidase (AMF) removes terminal α-fucose residues attached with a (1→3) or (1→4) linkage but not residues attached with a (1→6) linkage. In glycan sequencing, the glycan pool is analysed before and after sequential digestion with arrays of linkage-specific exoglycosidases. Glycan digestion results in peak shifts, the extent of which depends on the nature and the number of monosaccharides removed. The entire pool of glycans can be digested without separating individual peaks and aliquots of the pool can be digested simultaneously with panels of enzyme arrays. Figure 6 shows an example of a complete exoglycosidase digestion scheme for the structural analysis of a glycan pool obtained from Trastuzumab (trade name Herceptin), a monoclonal antibody used to treat certain types of breast cancer. Treatment of the glycan pool (i) with sialidase leads to the disappearance of two peaks at GU = 9.10 and GU = 8.33 and to a corresponding increase in the peak at GU = 7.60 (ii). A GU-shift of ca. 0.75 is characteristic for a sialic acid, so that the peaks at GU = 9.10 and GU = 8.33 must be glycans with two and one terminal sialic acids, respectively. The glycan pool obtained after sialidase digestion is then sequentially digested with fucosidase (iv), galactosidase (v) and hexosaminidase (vi), resulting in one single peak at GU = 4.30, which represents Man3GlcNAc2, the core structure of all *N*-linked glycans.

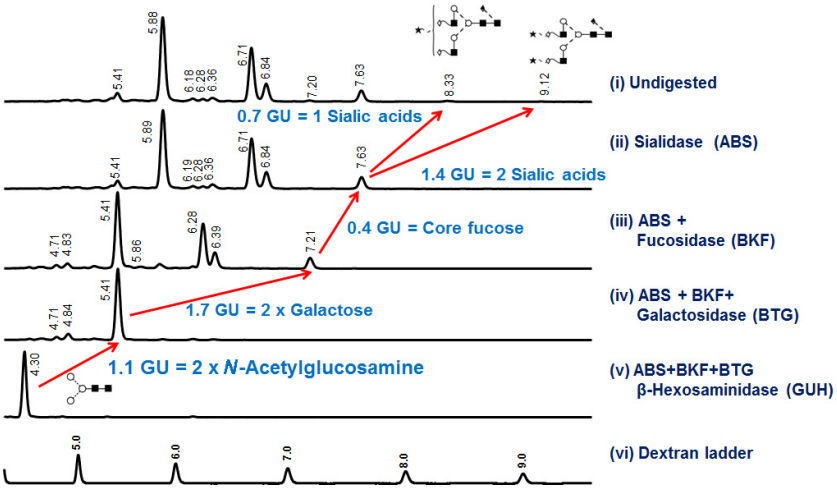

Figure 6. UPLC analysis and exoglycosidase array digestions of Trastuzumab (Herceptin) glycans analysed by UPLC with fluorescence detection. (i) Undigested glycan sample; (ii) ABS (*Athrobacter ureafaciens* sialidase) releases α(2 – 3,6,8)-linked sialic acids; (iii) ABS + BKF (Fucosidase from bovine kidney) releases α(1 – 2,6) linked fucose; (iv) ABS + BKF + BTG (Bovine testes β-galactosidase) releases β1 – 3 and 1 – 4 linkages, galactose and (v) ABS + BKF + BTG + GUH (hexosaminidase) releases β-GlcNAc but not GlcNAc linked to β(1 – 4) Man.

GlycoBase also enables the user to carry out *in silico* "GlycoBaseDigests". This predictive tool enables the generation of *in silico* exoglycosidase digests using various enzymes that are frequently used in the full characterisation of glycans. The user can select a particular glycan and perform a "virtual" digestion. GlycoBase returns the predicted digest product, and if experimental GU values associated with the queried and digested exist in the database, then these are also reported. GlycoBaseDigest currently provides *in silico* digestion for the following exoglycosidases: JBM, GUH, ABS, BKF, NAN1, AMF and SPG. (Figure 7).

Figure 7. Example of an *in silico* GlycoBase digest: The glycan A1G1 (GU = 5.7) is digested with SPG which releases β(1 – 4)-Galactosidase resulting in the glycan, A1 (GU = 4.8).

Additionally, GlycoBase allows users to conduct *in silico* "Extrapolated Profiling". GlycoBase stores glycan profiles using both the GU values as well as the area under the peak. Using this information and data, GlycoBase will then re-construct the original profile computing the Gaussian kernel density estimation. The relative percentage areas are used as a numeric vector of non-negative observation weights, while the GU values are the data from which the estimate is to be computed. The result is an *in silico* approximate profile (Figure 8). The computation is performed on the server side using the "R" statistical and graphic package.

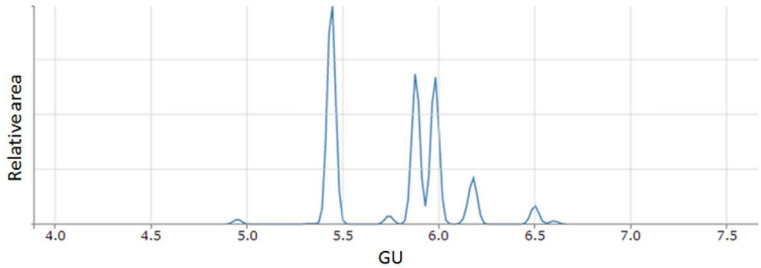

Figure 8. Example of an computationally calculated profile from human serum digested with a α(2 – 3,6,8) sialidase and a α(1 – 2,6) fucosidase.

If information regarding the exoglycosidase digest panels is available, GlycoBase users can then navigate from parent to child profiles. Thus, this allows the user to reconstruct the logic that was followed during the full characterisation of an undigested profile and enables the user to fully comprehend how each and every structure was identified.

Conclusion

The combined efficiencies of sample preparation, high resolution glycan analysis and computer assisted data processing can enable both the research community and the biopharmaceutical sector to perform released glycan analysis with greater confidence and speed than previously possible with the existing analysis workflows.

The robotic platform in combination with UPLC and GlycoBase as a robust database system, offer the basis for high-throughput glycoprofiling and characterisation of biological samples from biomarker discovery studies and clinical studies such as GWAS. While the robotic platform is an enabling technology to reduce processing time and cost associated with sample preparation, the UPLC method offers excellent peak resolution and performance for glycan analysis. The glycan sample preparation platform can be easily adapted and allows glycan labelling with a variety of labels so that it can be linked to complementary analytical technologies such as mass spectrometry and capillary electrophoresis.

Acknowledgments

The authors wish to acknowledge the following; GlycoBaseDigest: in collaboration with the Swiss Institute of Bioinformatics. GlycoBioM grant funding from the European Union Seventh Framework Programme FP7/2007 – 2013), Grant No 259869, HighGlycan grant funding from the European Union Seventh Framework Programme (FP7/2011 – 2013), Grant No 278535, HighGlycan

References

[1] Apweiler, R., Hermjakob, H., Sharon, N. (1999) On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochimica et Biophysica Acta – General Subjects* **1473**:4 – 8. doi: http://dx.doi.org/10.1016/S0304-4165(99)00165-8.

[2] Lauc, G., Essafi, A., Huffman, J. E., Hayward, C., Kneević, A., Kattla, J.J., Kattla, J.J., Polašek, O., Gornik, O., Vitart, V., Abrahams, J.L., Pučić, M., Novokmet, M., [...], Rudan, I. (2010) Genomics Meets Glycomics – The First GWAS Study of

Human *N*-Glycome Identifies HNF1α as a Master Regulator of Plasma Protein Fucosylation. *PLoS Genetics* **6**(12):e1001256.
doi: http://dx.doi.org/10.1371/journal.pgen.1001256.

[3] Royle, L., Campbell, M.P., Radcliffe, C.M., White, D.M., Harvey, D.J., Abrahams, J.L., Kim, Y.-G., Henry, G.W., Shadick, N.A., Weinblatt, M.E., Lee, D.M., Rudd, P.M., Dwek, R.A. (2008) HPLC-based analysis of serum *N*-glycans on a 96-well plate platform with dedicated database software. *Analytical Biochemistry* **376**(1):1 – 12.
doi: http://dx.doi.org/10.1016/j.ab.2007.12.012.

[4] Saldova, R., Royle, L., Radcliffe, C.M., Abd Hamid, U.M., Evans, R., Arnold, J.N., Banks, R.E., Hutson, R., Harvey, D.J., Antrobus, R., Petrescu, S.M., Dwek, R.A., Rudd, P.M. (2007) Ovarian cancer is associated with changes in glycosylation in both acute-phase proteins and IgG. *Glycobiology* **17**(12):1344 – 1356.
doi: http://dx.doi.org/10.1093/glycob/cwm100.

[5] Arnold, J.N., Saldova, R., Abd Hamid, U.M., Rudd, P.M. (2008) Evaluation of the serum *N*-linked glycome for the diagnosis of cancer and chronic inflammation. *Proteomics* **8**(16):3284 – 3293.
doi: http://dx.doi.org/10.1002/pmic.200800163.

[6] Abd Hamid, U.M., Royle, L., Saldova, R., Radcliffe, C.M., Harvey, D.J., Storr, S.J., Pardo, M., Antrobus, R., Chapman, C.J., Zitzmann, N., Robertson, J.F., Dwek, R.A., Rudd, P.M. (2008) A strategy to reveal potential glycan markers from serum glycoproteins associated with breast cancer progression. *Glycobiology* **18**(12):1105 – 1118.
doi: http://dx.doi.org/10.1093/glycob/cwn095.

[7] Arnold, J.N., Saldova, R., Galligan, M.C., Murphy, T.B., Mimura-Kimura, Y., Telford, J.E., Godwin, A.K. Rudd, P.M. (2011) Novel glycan biomarkers for the detection of lung cancer. *Journal of Proteome Research* **10**(4):1755 – 1764.
doi: http://dx.doi.org/10.1021/pr101034t.

[8] Stöckmann, H., Adamczyk, B., Hayes, J., Rudd, P.M. (2013) Automated, high-throughput IgG-antibody glycoprofiling platform. *Analytical Chemistry* **85**(18):8841 – 8849.
doi: http://dx.doi.org/10.1021/ac402068r.

[9] Campbell, M.P., Royle, L., Radcliffe, C.M., Dwek, R.A., Rudd, P.M. (2008) GlycoBase and autoGU: tools for HPLC-based glycan analysis. *Bioinformatics* **24**(9):1214 – 1216.
doi: http://dx.doi.org/10.1093/bioinformatics/btn090.