

## Structural bioinformatics

# GlycoProfileAssigner: automated structural assignment with error estimation for glycan LC data

Fergal J. Duffy and Pauline M. Rudd\*

National Institute for Bioprocessing Research and Training, Fosters Avenue, Mount Merrion, Co. Dublin, Ireland

\*To whom correspondence should be addressed.

Associate Editor: Anna Tramontano

Received on December 17, 2014; revised on February 17, 2015; accepted on February 21, 2015

### Abstract

**Motivation:** Sequencing glycan structures is a difficult problem that requires the use of multiple experimental approaches. One powerful approach to glycan sequencing is the combination of liquid chromatography with sequential exoglycosidase digestions; however, interpreting this can be difficult and time-consuming. To aid this process, we introduce GlycoProfileAssigner, software for automated structural assignment of glycan profile data from liquid chromatography experiments.

**Results:** GlycoProfileAssigner has been tested on human IgG data, and can retrieve the correct structure in 14 out of 16 peaks tested.

**Availability and Implementation:** The programme and its source code is available at <https://bitbucket.org/fergaljd/glycoprofileassigner>

**Contact:** pauline.rudd@nibrt.ie

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

## 1 Introduction

The study of glycans and glycoproteins is central to understanding a wide range of biological problems. The majority of human proteins are glycosylated (Wong, 2005) by oligosaccharides attached to asparagine (N-glycans) or serine/threonine (O-glycans). Glycosylation is also the single most important post-translational modification observed on biotherapeutic proteins (Walsh and Jefferis, 2006). Glycans affect the stability, safety and efficacy of biotherapeutics and understanding which glycans are present on a glycoprotein is vital to understanding its function.

Glycan structures are challenging to analyze, because a large variety of possible monosaccharide linkages and complex isomeric branched structures may exist for any given monosaccharide composition. A large range of analytical tools have been developed to elucidate glycan structures, including liquid chromatography, mass spectrometry, and capillary electrophoresis (Mittermayr and Guttman, 2012). These approaches are often supplemented with exoglycosidase digestions to reduce glycan structures by sequentially removing defined monosaccharides to a well-defined sugar core (Royle *et al.*, 2006).

GlycoProfileAssigner is a software tool that assists in the interpretation of HILIC-HPLC and UPLC data. HILIC (Buszewski and Noga, 2012; Hydrophobic interaction chromatography) is a technique that separates structures on the basis of their hydrophilicity. For glycan data, this is roughly equivalent to glycan size. HPLC and UPLC are High and Ultra performance liquid chromatography respectively. For glycan LC experiments, chromatographic retention times can be expressed in normalised Glucose Units (GU) for comparability across different experimental setups.

Combining LC-derived GU data with sequential enzyme digestions is a powerful method for analyzing the glycan structures and their relative abundances in a biological sample. The main difficulty with the current U/HPLC approach is that assignment of glycan structures is difficult and time consuming as it can require manually ascertaining the correct digestion pathways for dozens of glycans between five or more digestion panels. GlycoProfileAssigner aims to aid this process by offering a fast and accurate method of enumerating and scoring glycan structure assignments (Royle *et al.*, 2008).

## 2 Design and implementation

GlycoProfileAssigner calculates the correct structural assignments using GU and enzymatic digestion information. [Supplementary Fig. S1](#) outlines the process GlycoProfileAssigner uses to assign and estimate assignment error for glycan structures. For the assignment, glycan digestion profiles are ordered from least to most digested, and a rough GU-based assignment step is used to determine a starting list of potential glycan structures. This step uses Glycobase ([Campbell et al., 2008](#)) data to assign structures with a matching GU value  $\pm 0.30$  to each peak. Following this step code derived from GlycoDigest ([Gotz et al., 2014](#)) is used to identify assigned structures which would be degraded by the applied glycosidases, which are then removed. Finally profiles are made consistent by removing glycan structures that do not correspond to structures in a more digested profile. For example, the most digested profile in [Supplementary Fig. S1](#) has M3 as the only assigned glycan after digestion with ABS, BKF, BTG and GUH. The next most digested profile has the enzymes ABS, BTG and BKF applied ([Supplementary Table S3](#) describes the specificities of the abbreviated enzymes). Since these profiles come from the same undigested sample, all glycans in the second most digested profile must digest to M3 on the application of GUH. Therefore, any glycans that do not match this criterion are removed.

Finally, glycan error values are calculated. Error values are calculated as:  $\text{err} = \text{abs}(\Delta\text{GU}_{\text{expected}} - \Delta\text{GU}_{\text{observed}})$

To calculate error values, peaks are matched across digestions by simulated digestion of structures in peaks, e.g. A1 and A2 in the ABS, BKF, BTG digestion panel will digest to M3 in the ABS, BKF, BTG, GUH digestion panel. An observed GU shift is then calculated as the difference in GU values between the peaks. The expected GU shift is calculated by observing the monosaccharide units removed in the simulated digestion, and using [Supplementary Table S1](#) values to calculate an expected GU shift. Errors are added to each other cumulatively over the entire set of digest panels. The assignments and error values can then be exported to a CSV file from the GlycoProfileAssigner GUI.

## 3 Results

### 3.1 Structural assignment of human IgG glycans

To validate and demonstrate GlycoProfileAssigner, we compared GlycoProfileAssigner's scored assignments to a manually assigned human serum immunoglobulin G glycan profile, analyzed using UPLC.

The output of the manual assignment, compared to GlycoProfileAssigner is shown in [Supplementary Table S2](#). The table shows the GU values of the UPLC profile peaks, with their manually assigned peaks in the middle column and the GlycoProfileAssigner scored output on the right. In 14 out of 16 peaks, the GlycoProfileAssigner predictions include the correct glycan structure in its predictions. Glycobase assumes that peak GU will vary by no more than 0.05 of a GU for the same peak in a different digest. In the case of the peak at GU 6.75 in this data, the variation was slightly higher, leading to an incorrect structural assignment. In the case of the peak at GU 6.38, the recorded GU values in Glycobase differed too much from the observed value seen here, so FA2[6]G1 was not assigned.

Looking at GlycoProfileAssigner scores in [Supplementary Table S2](#), it can be seen that the structure with the lowest estimated error is not always the correct structure, particularly in the case of bisected and sialylated glycans, but, in general, correct assignments will have an error of under 0.5.

## 4 Discussion

GlycoProfileAssigner is not the first time automated structural assignment software has been developed. AutoGU ([Campbell et al., 2008](#)) was the original work on this topic; however, this software was limited to experimentally observed glycan digestions. GlycoProfileAssigner represents a major advance on this software by being able to predict the activity of any common glycosidase on any N- or O-glycan structure, and by including an assignment score to estimate the likelihood of a correct assignment.

GlycoProfileAssigner does have some limitations: it assumes a single glycan structure per chromatographic peak, and that all enzymatic digestion steps run to full completion. Applying GlycoProfileAssigner to under-digested profile data, or profile data where glycan structures are not separated into unique peaks will result in poor quality assignments.

GlycoProfileAssigner can also be a powerful tool in conjunction with mass spectrometry data. The GlycoProfileAssigner GUI displays the theoretical mass of all the predicted assignments, allowing them to be easily compared with MS experimental data.

## Funding

This work was supported by funding from the European Union's Seventh Framework Programme under the GlycoBioM grant agreement FP7-HEALTH-259869.

*Conflict of interest:* none declared.

## References

- Buszewski, B. and Noga, S. (2012) Hydrophilic interaction liquid chromatography (HILIC)—a powerful separation technique. *Anal. Bioanal. Chem.*, **402**, 231–247.
- Campbell, M.P. et al. (2008) GlycoBase and autoGU: tools for HPLC-based glycan analysis. *Bioinformatics*, **24**, 1214–1216.
- Gotz, L. et al. (2014) GlycoDigest: a tool for the targeted use of exoglycosidase digestions in glycan structure determination. *Bioinformatics*, **30**, 3131–3133.
- Mittermayr, S. and Guttman, A. (2012) Influence of molecular configuration and conformation on the electromigration of oligosaccharides in narrow bore capillaries. *Electrophoresis*, **33**, 1000–1007.
- Royle, L. et al. (2006) Detailed structural analysis of N-glycans released from glycoproteins in SDS-PAGE gel bands using HPLC combined with exoglycosidase array digestions. *Methods Mol. Biol.*, **347**, 125–143.
- 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.
- Walsh, G. and Jefferis, R. (2006) Post-translational modifications in the context of therapeutic proteins. *Nat. Biotechnol.*, **24**, 1241–1252.
- Wong, C.-H. (2005) Protein glycosylation: new challenges and opportunities. *J. Org. Chem.*, **70**, 4219–4225.