OXFORD

Structural bioinformatics

GlycanAnalyzer: software for automated interpretation of *N*-glycan profiles after exoglycosidase digestions

Ian Walsh [1], Terry Nguyen-Khuong[1], Katherine Wongtrakul-Kish[1], Shi Jie Tay[1], Daniel Chew[1], Tasha José[2], Christopher H. Taron[2,†] and Pauline M. Rudd[1,*,†]

[1]Bioprocessing Technology Institute Agency for Science, Technology and Research (A*STAR), Singapore 138668 and [2]New England Biolabs, 240 County Road, Ipswich, MA 01938, USA

*To whom correspondence should be addressed.

[†]The authors wish it to be known that, in their opinion, the last two authors should be regarded as Joint Senior Authors.

Associate Editor: Alfonso Valencia

Abstract

Summary: Many eukaryotic proteins are modified by *N*-glycans. Liquid chromatography (ultra-performance –UPLC and high-performance–HPLC) coupled with mass spectrometry (MS) is conventionally used to characterize *N*-glycan structures. Software can automatically assign glycan structures by matching their observed retention times and masses with standardized values in reference databases. However, more precise confirmation of *N*-glycan structures can be derived using exoglycosidases, enzymes that remove specific monosaccharides from glycans. Exoglycosidase removal of monosaccharides results in signature peak shifts, in both UPLC and MS1, yielding an effective way to verify *N*-glycan structure with high detail (down to the position and isomeric linkage of each monosaccharide). Because manual interpretation of exoglycosidase data is complex and time consuming, we developed GlycanAnalyzer, a web application that pattern matches *N*-glycan peak shifts following exoglycosidase digestion and automates structure assignments. GlycanAnalyzer significantly improves assignment accuracy over other auto-assignment methods on tests with a monoclonal antibody and four glycan standards (100% versus 82% for the next best software). By automating data interpretation, GlycanAnalyzer enables the easier use of exoglycosidases to precisely define *N*-glycan structure.

Availability and implementation: http://glycananalyzer.neb.com. Datasets available online.

Contact: pauline.rudd@nibrt.ie

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

N-Glycans are carbohydrates that are covalently bonded to many glycoproteins. They play critical roles in numerous physiological and pathological processes. In addition, their characterization is imperative in biotherapeutic quality by design (Dalziel *et al.*, 2014). *N*-Glycans have diverse branching, topology and monosaccharide linkages (Cummings, 2009). Consequently, there are significant analytical challenges in their complete characterization. In recent years,

analytical techniques have advanced and are now able to detect and quantitate *N*-glycans in complex samples. Such techniques may involve various methodologies such as high/ultra-performance liquid chromatography (H/UPLC), capillary electrophoresis (CE), mass spectrometry (MS) or combinations thereof (Mariño *et al.*, 2010). In UPLC-MS two types of information are generated to help characterize *N*-glycans structures: (i) the retention time (RT) of *N*-glycans on an LC column, and (ii) the mass of each eluted *N*-glycan

(Supplementary Fig. S2). Typically, software approaches to assign structures to observed peaks in UPLC-MS involve matching observed RTs and masses to reference values in glycan repositories (Royle *et al.*, 2008). As RTs may vary across different experimental conditions, software often draws upon comparisons to the mobility of a dextran ladder standard [Glucose Units (GU)] (Mariño *et al.*, 2010). For example, the UNIFI Scientific Information System (Waters) utilizes both *N*-glycan GU and mass matching criteria at its core.

Exoglycosidases are enzymes that catalyze sequential removal of monosaccharides from the non-reducing end of glycans. They have specificity for a particular type of sugar, its stereochemistry (α/β anomer) and position of attachment to an adjacent sugar on the glycan (Supplementary Fig. S1). Consequently, monitoring the UPLC-MS peak movements after step-wise application of exoglycosidases provides a unique and effective way to annotate sugars with high detail down to the level of position and linkage isomers (Supplementary Fig. S3). This level of detail is not directly ascertained when only matching GU and/or mass values (Supplementary Fig. S2). Interpretation of exoglycosidase array data analysis is currently performed manually and can be complex and very time-consuming. Previously, the GlycoDigest tool (Gotz *et al.*, 2014) was developed to model the action of exoglycosidase(s) on particular glycans but it did not automate structure assignment. Additionally, GlycoProfileAssigner (Duffy and Rudd, 2015), sought to automate interpretation of exoglycosidase array digestion data (Supplementary Table S1 shows how it compares to GlycanAnalyzer). In our experiments, GlycoProfileAssigner handled simple scenarios well but failed on more complex samples.

Here, we present GlycanAnalyzer, a web application that automates the annotation of *N*-glycans in UPLC-MS chromatograms after treatment with one or more exoglycosidases. It presents data visualization and links to published reference databases (Aoki-Kinoshita *et al.*, 2016; Zhao *et al.*, 2018). Comprehensive help pages/tutorials are also included to facilitate user training and support the program's use (Supplementary Table S2). GlycanAnalyzer is the most accurate and comprehensive software for automated interpretation of *N*-glycan exoglycosidase digestion data described to date.

2 Usage and output

Input: GlycanAnalyzer uses GU mobility reference and mass data. All input is supplied through tab-separated text. H/UPLC peak data can be uploaded using tab-separated triplets (RT, GU, % Area) for each peak (Supplementary Fig. S4a). Three dimensional mass information can be supplied by providing the triplets (RT, m/z, intensity) in tab-separated form (Supplementary Fig. S4a). All variables (RT, m/z, intensity, GU and % area) can easily be extracted from H/UPLC-MS instruments, but users may wish to supply their own observed masses (Supplementary Fig. S4b). The software also executes when no mass information is supplied (i.e. only supplying UPLC peak GU and % area). However, accuracy is significantly diminished in this case (see 'NA' values in Supplementary Fig. S4b). Before input pre-processing proceeds, users must select the fluorescent label used in their experiment, or select 'label free' if no label was used (Supplementary Section S4.1.2). Additionally, users can select a type of glycoprotein to help refine the final glycan assignments.

Output: The *N*-glycans assigned by the software are sorted by a score assigned to each peak (the best score is zero). All glycans are drawn in accepted structural notation (Haltiwanger, 2016), and links to the databases GlycoStore (Zhao *et al.*, 2018) and

Table 1. Fraction of correctly assigned peaks measured against expert manual assignment

	Anti-Her2	Standards	Accuracy (%)
GlycanAnalyzer	18/18	4/4	100.0
GlycoProfileAssigner	4/18	1/4	22.7
UNIFI	17/18	1/4	81.8

GlyTouCan (Aoki-Kinoshita *et al.*, 2016) are provided. GlycanAnalyzer outputs three pieces of supporting evidence for each structure assignment: (i) GU similarity to reference values in GlycoStore, (ii) mass matching with known structures and (iii) interpretation of peak shifts after exoglycosidase digestion. This is the first software where all three pieces of evidence are available together. When all three lines of evidence are returned, the assignment has excellent support and a low score (close to zero) will be returned. In some cases, not all three lines of evidence will be found. Example output is shown in Supplementary Figures S5 and S6.

3 Materials and methods

GlycanAnalyzer extracts each glycan's mass from the 3-dimensional mass input (i.e. RT, m/z, intensity; Supplementary Fig. S4) or user-supplied values. Using detailed understanding of each exoglycosidase's specificity, it searches for peak shifts anticipated by the removal of individual monosaccharide masses. Supplementary Figure S7 shows example mass and UPLC peak shifts for a simple standard. However, in more complex samples, many peak movements are possible (Supplementary Fig. S9). A score is implemented to help predict and rank the most likely peak movements that result from enzyme digestions. An assigned glycan's score depends on a number of factors such as: optimized regression models that predict its expected GU shift, whether all masses are found in each profile of the exoglycosidase array, and the GU similarity to average values in GlycoStore (Supplementary Section S5.2). In cases where no mass is supplied, mass is estimated by linear regression and matching to GlycoStore (Supplementary Fig. S10). The algorithm and scoring technique are described in detail in the Supplementary Section S5.

4 Results

Correct assignments: Table 1 shows 100% assignment accuracy for GlycanAnalyzer compared to other software capable of automated assignment. Benchmarking was performed on *N*-glycans released from a monoclonal antibody and four glycan standards (one standard was used for illustration in Supplementary Fig. S7). Supplementary Table S3 shows examples where GlycanAnalyzer's score helps remove the GU and mass annotation ambiguity. Supplementary Table S4 shows GlycanAnalyzer's ability to detect co-elution.

Time savings: Depending on the complexity of a sample and the number of exoglycosidases used, manual annotation could involve interpreting hundreds of peak movements, potentially taking days/weeks to complete. Using GlycanAnalyzer all assignments were returned in 20 min when using the 'Assign all top 5' option for our monoclonal antibody.

Acknowledgement

Funding

Conflict of Interest: none declared.

References

Aoki-Kinoshita,K. *et al*. (2016) GlyTouCan 1.0 – The international glycan structure repository. *Nucleic Acids Res.*, **44**, D1237–D1242.

Cummings,R.D. (2009) The repertoire of glycan determinants in the human glycome. *Mol. BioSyst.*, **5**, 1087–1104.

Dalziel,M. *et al*. (2014) Emerging principles for the therapeutic exploitation of glycosylation. *Science*, **343**, 1235681.

Duffy,F.J. and Rudd,P.M. (2015) GlycoProfileAssigner: automated structural assignment with error estimation for glycan LC data. *Bioinformatics*, **31**, 2220–2221.

Gotz,L. *et al*. (2014) GlycoDigest: a tool for the targeted use of exoglycosidase digestions in glycan structure determination. *Bioinformatics*, **30**, 3131–3133.

Haltiwanger,R.S. (2016) Symbol nomenclature for glycans (SNFG). *Glycobiology*, **26**, 217–217.

Mariño,K. *et al*. (2010) A systematic approach to protein glycosylation analysis: a path through the maze. *Nat. Chem. Biol.*, **6**, 713–723.

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.

Zhao,S. *et al*. (2018) GlycoStore: a database of retention properties for glycan analysis. *Bioinformatics*, **1**.