BIOINFORMATICS APPLICATIONS NOTE Vol. 29 no. 13 2013, pages 1706–1707
doi:10.1093/bioinformatics/btt190

Systems biology

Advance Access publication April 22, 2013

Automated annotation and quantification of glycans using liquid chromatography–mass spectrometry

Chuan-Yih Yu[1,*], Anoop Mayampurath[1], Yunli Hu[2], Shiyue Zhou[2], Yehia Mechref[2] and Haixu Tang[1,*]

[1]School of Informatics and Computing, Indiana University, Bloomington, IN 47408, USA and [2]Department of Chemistry and Biochemistry, Texas Tech University, Lubbock, TX 79401, USA

Associate Editor: Martin Bishop

ABSTRACT

Summary: As a common post-translational modification, protein glycosylation plays an important role in many biological processes, and it is known to be associated with human diseases. Mass spectrometry (MS)-based glycomic profiling techniques have been developed to measure the abundances of glycans in complex biological samples and applied to the discovery of putative glycan biomarkers. To automate the annotation of glycomic profiles in the liquid chromatography-MS (LC–MS) data, we present here a user-friendly software tool, MultiGlycan, implemented in C# on Windows systems. We tested MultiGlycan by using several glycomic profiling datasets acquired using LC–MS under different preparations and show that MultiGlycan executes fast and generates robust and reliable results.

Availability: MultiGlycan can be freely downloaded at http://darwin.informatics.indiana.edu/MultiGlycan/.

Contact: chuyu@indiana.edu or hatang@indiana.edu

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

Received on February 18, 2013; revised on April 16, 2013; accepted on April 18, 2013

1 INTRODUCTION

Glycosylation is one of the most common post-translational modifications of proteins, in which glycans are covalently linked to the side chain of Asn (i.e. N-linked glycosylations) or Ser/Thr (i.e. O-linked glycosylations) residues. The alteration of protein glycosylation is known to be associated with various biological processes and human diseases. Because of high sensitivity and throughput, mass spectrometry (MS) is routinely used to identify and quantify glycans derived from a complex biological sample. Several studies show that differences in glycomic profiles (in both N-linked and O-linked glycans) are observed between disease and healthy human individuals (de Leoz *et al.*, 2011).

To automate the application of glycomics to the discovery of putative glycan biomarkers, several software tools have been developed to annotate glycans in a mass spectrum by matching the measured mass-to-charge values (m/z) of glycans to theoretical values. Cartoonist (Goldberg *et al.*, 2005) can assign glycans from a list of ~2800 *N*-glycan candidates (derived from glycan synthetic pathways) to peaks in a matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) spectrum based on their masses that are calibrated using linear models. GlycoWorkbench (Ceroni *et al.*, 2008) provides an interface for users to draw glycan structures that can be subsequently used to match the peaks in MALDI-TOF spectra. Both Cartoonist and GlycoWorkbench can incorporate fragment ions from tandem mass spectra (MS/MS) to achieve higher confidence in glycan assignment. Recent advancement in glycomic profiling techniques incorporate electro-spray ionization (ESI), wherein native or derivatized glycans are chromatographically separated using hydrophilic interaction chromatography, reversed-phase chromatography (Hu and Mechref, 2012) or porous graphitized carbon column before online ESI–MS analyses. Enhanced ESI sensitivity is observed with derivatization, such as permethylation of glycans, which can be efficiently separated by reversed-phase chromatography. GlycReSoft (Maxwell *et al.*, 2012) is a recently published software tool for annotation of multiple charged glycan ions in liquid chromatography (LC)–ESI–MS data. However, GlycReSoft is limited to only one adduct type and requires previous data processing using deconvolution software, Decon2Ls (Jaitly *et al.*, 2009). Here, we present a new tool MultiGlycan that can annotate a user-defined list of glycans observed in MALDI-TOF or LC–ESI–MS data. For MALDI-TOF, MultiGlycan works by matching experimental isotopic envelopes with the theoretical envelopes of glycans. Using a mixture model, MultiGlycan can also account for overlapping glycan isotopic distributions. MultiGlycan-ESI, which we will focus on here, links together envelopes using the mass of a specified adduct, or multiple (including ammonia, sodium, potassium or other defined by the user) adducts simultaneously, leading to more confident glycan annotation.

Users can tackle specific glycan targets by providing a selected set of monosaccharide compositions, or conduct a blind search of glycans without specifying the compositions. The program will report abundances of annotated glycans based on the total ion intensities of the glycan carrying various charges and adducts within consecutive MS scans.

2 METHODS

MultiGlycan supports mzXML and Thermo Fisher .raw (provided vendor-specific Xcalibur software is installed) input file formats. For every MS spectrum, after peak detection, MultiGlycan uses the THRASH algorithm from Decon2LS code to de-isotope isotopic distributions. MultiGlycan contains a built-in glycan list representing N-linked

*To whom correspondence should be addressed.

Table 1. Performance of MultGlycan (M) and GlycReSoft (G)

Dataset	Running time (min)		Detected glycans		Manually annotated glycans
	M	G	M	G	
HBS	2.2	4.9[a]	79	45	79
Cancer cell line A(CL-A)[b]	1.5	3.9[a]	46	32	60
Cancer cell line B(CL-B)[b]	1.4	3.4[a]	74	40	73
Cancer cell line C(CL-C)[b]	1.4	3.6[a]	56	26	60

[a]Running time of GlycReSoft includes the required pre-processing using Decon2LS.
[b]Completed results are in Supplementary Figures S4–S6, respectively.

glycans that are generated by using the synthetic pathways of N-linked glycans in human (Krambeck and Betenbaugh, 2005). Alternatively, users can choose to build their own N- or O-linked glycan list in a comma separated values (CSV) file. For LC–ESI–MS data, after retrieving the ions from the input data, MultiGlycan matches these ions with a set of glycan masses and adducts from a built-in or user-defined list.

In the final report, MultiGlycan automatically merges the same glycan across consecutive MS scans. MultiGlycan also provides an intuitive quantitation measure. Each output associated with a glycan is the summation of peaks with and without adducts in the whole-isotope envelope in the spectrum, and then summed within an elution window. These results can be directly used in large-scale glycomics studies.

3 RESULTS

MultiGlycan was tested on six LC–MS datasets acquired from two N-glycan samples, released by using 60 U of PNGase F (*Flavobacterium meningosepticum*), from ribonuclease B (RiboB), human serum glycoproteins (HBS) and glycoproteins from three cancer cell lines, respectively (for details, see in Supplementary Table S1).

A Dionex 3000 Ultimate nano-LC system (Dionex, Sunnyvale, CA, USA) connected to a Velos LTQ Orbitrap hybrid mass spectrometer (Thermo Scientific, San Jose, CA, USA) was used for permethylated *N*-glycan analysis in this study. The mass spectrometer was operated in data-dependent acquisition mode with one scan event containing a full-scan MS along with a collision induced dissociation MS/MS, and another scan event containing a full-scan MS (m/z from 500 to 2000). In the analysis of native RiboB, only MS scans were acquired.

The complete list of glycan profiling results from MS spectra within datasets used in this analysis can be found in Supplementary Tables S2–S4. The program reports high-mannose *N*-glycans for both standard RiboB samples as expected: four of five known *N*-glycans for the permethylated and three of five for the native sample were reported (see Supplementary Tables S2 and S3, respectively). MultiGlycan annotated 79 *N*-glycans in the HBS sample, and 53 of them matched the manually annotated ones from an earlier study (Hu and Mechref, 2012) (Supplementary Table S4). Additionally, 26 *N*-glycans annotated by MultiGlycan were confirmed by manual annotation (Supplementary Fig. S1). The addition of adducts enables more confident assignment to glycan compositions that otherwise are difficult to assign simply using p.p.m. mass error. For example, in HBS experiment, an ion at 1090.572 m/z can be matched to two different glycan compositions within 10 p.p.m., viz., $[HexNAc_5Hex_8deHex_2+3H]^{3+}$ and $[HexNAc_5Hex_7deHex_1NeuAc_1+NH_4+2H]^{3+}$. However, the presence of isotopic distribution at 1084.893 m/z accounts for $[HexNAc_5Hex_7deHex_1NeuAc_1+3H]^{3+}$ (see the corresponding MS/MS spectrum in Supplementary Fig. S2) allowing the correct assignment of $[HexNAc_5Hex_7deHex_1NeuAc_1]$ to this glycan. Thus, clustering isotopic distributions that are linked by adducts in this manner enables accurate annotation for glycan profiling.

We compared MultiGlycan performance with GlycoReSoft on the HBS and three other cancer cell line samples. In HBS among the 45 glycans that GlycoReSoft found, 35 were confirmed manually that are all contained in the reported 79 glycans by MultiGlycan (Supplementary Table S5 and Supplementary Fig. S3). The remaining 10 were not reported in the study conducted by Hu and Mechref (2012). In three cancer cell line samples, MultiGlycan also identified more glycans than GlycoReSoft, most of which are consistent with manual annotation (Table 1, also see Supplementary Tables S6–S8 and Supplementary Figs S4–S6).

MultiGlycan allows users to run the analysis directly on the raw file. Moreover, it can handle multiple adducts, and any specified glycan list. All these features make MultiGlycan a robust tool ready for high-throughput glycomic data analysis.

ACKNOWLEDGEMENTS

The authors thank PNNL and http://omics.pnl.gov (along with affiliates NIH National Center for Research Resources, the W.R. Wiley Environmental Molecular Science Laboratory and the National Institute of Allergy and Infectious Diseases) for making Decon2LS open source.

Funding: National Science Foundation [DBI-0642897]; Persistent Systems (to A.M.).

Conflict of Interest: none declared.

REFERENCES

Ceroni,A. *et al.* (2008) GlycoWorkbench: a tool for the computer-assisted annotation of mass spectra of glycans. *J. Proteome Res.*, **7**, 1650–1659.

de Leoz,M.L. *et al.* (2011) High-mannose glycans are elevated during breast cancer progression. *Mol. Cell Proteomics*, **10**, M110.002717.

Goldberg,D. *et al.* (2005) Automatic annotation of matrix-assisted laser desorption/ionization N-glycan spectra. *Proteomics*, **5**, 865–875.

Hu,Y. and Mechref,Y. (2012) Comparing MALDI-MS, RP-LC-MALDI-MS and RP-LC-ESI-MS glycomic profiles of permethylated N-glycans derived from model glycoproteins and human blood serum. *Electrophoresis*, **33**, 1768–1777.

Jaitly,N. *et al.* (2009) Decon2LS: An open-source software package for automated processing and visualization of high resolution mass spectrometry data. *BMC Bioinformatics*, **10**, 87.

Krambeck,F.J. and Betenbaugh,M.J. (2005) A mathematical model of N-linked glycosylation. *Biotechnol. Bioeng.*, **92**, 711–728.

Maxwell,E. *et al.* (2012) GlycReSoft: a software package for automated recognition of Glycans from LC/MS Data. *PLoS One*, **7**, e45474.