Mass Spectrometry
DOI: 10.5702/massspectrometry.S0016

Vol. 2 (2013), S0016

Is High Throughput Glycomics Possible?

Carlito B. Lebrilla

Department of Chemistry, University of California, Davis

The role of glycosylation and their biological functions whether as free oligosaccharides or glycoconjugates has been made possible by the recent advancements in the analyses of these compounds. The heterogeneity and the large structural diversity have made oligosaccharide analysis significantly more difficult than other biopolymers. The next stage of development is to achieve high throughput analysis. However, the structural elucidation of oligosaccharides remains an extremely difficult task. Recent reports reveal that the diversity of structures in a given biological system is finite and may not be large. It may be possible to create a database of structures that can be used to determine the identity of known compounds. This capability would therefore make high throughput glycomics possible. Achieving this task depends on the proper selection of chemical characteristics to identify the compound. In this presentation, nanoflow liquid chromatography retention times, accurate mass, and tandem MS is used to determine structure with a high degree of certainty. The method is used to determine the biological function of milk oligosaccharides as well as to discover glycan-based biomarkers for diseases.

Keywords: glycomics, nanoLC, glycans, disease markers

(Received September 12, 2012; Accepted November 13, 2012)

INTRODUCTION

The importance of glycosylation in areas as diverse as diseases and nutrition has led to the development of methods for rapid structural elucidation. However, while the analyses of biopolymers such as proteins have advanced significantly by providing rapid protein identification, glycomics has lagged considerably primarily because glycan structures do not follow a template. The major efforts in our research are focused in two areas: glycans as biomarkers for diseases and glycoconjugates as bioactive components of milk.

In disease markers, the glycomics approach has led to a new paradigm for biomarker discovery producing a host of new potential leads based on a common but structurally complicated posttranslational modification. Glycans constitute a new source of potential biomarkers for a host of diseases but specifically cancer. In the glycomics approach, glycans are harvested and enriched for structural analysis. Recent research in this laboratory and elsewhere has shown that glycans have significant potential value as markers for diseases. Glycans have distinct advantages for disease marker discovery as recently discussed by Packer *et al.*[1] and An *et al.*[2,3] Furthermore, there is over 50 years of glycobiology relating cancer and glycosylation.[3,4] In biomarker research glycan profiles of human serum using mass spectrometry (MS) were studied for breast,[5] ovarian,[6–8] pancreatic,[9] and prostate[10] cancer and demonstrated the potential of glycomic profiling for early disease diagnosis. Currently, most of these approaches have relied on compositional analysis employing MS of entire mixtures. The structures of the glycans obtained in these studies are only putative and based solely on the composition with the exact structure including linkages, branching, antennae, anomeric character, and specific glycan residues generally unknown.

To advance the field further, a method to identify precise structures in a high throughput manner must be developed. Although glycosylation has become an important target for biomarker discovery, a rapid method for structural identification remains a distant goal. Glycosylation is the most complicated posttranslational modification with an extremely high level of structural diversity.[11,12] Because glycosylation is a post-translational modification, there is no genomic template for rapid identification. Recent advances in glycomics analysis have increased speed, sensitivity and rapid compositional determination yielding some structural details, but they have not solved the problems of complete structural identification in a high or even rapid throughput manner.

STRATEGY FOR RAPID METHOD FOR STRUCTURAL IDENTIFICATION

A key component to this research is a method for the rapid identification of glycans that will be used to determine the exact structure of the cell surface components. The term *structure* in this discussion refers to the complete structure with all linkages elucidated. Rapid identification does not mean rapid structural elucidation because of the complexities of oligosaccharide structures. *De novo* structural elucidation of glycans is complicated and still requires considerable effort. Instead, the proposed method will be based on the construction of a database of annotated glycan structures that will be used for the rapid identification of oligosaccharide structures. It will further rely on the development of identifiers unique to each oligosaccharide structure. These identifiers will include chromatographic retention times, accurate mass, and tandem MS, although the relative weight of each still need to be evaluated as to its capacity to identify structures.

Glycan identification will rely primarily on LC/MS/MS methods. Three conditions are important to characterize structures: efficient chromatographic separation, high performance mass spectrometry, and unique tandem MS. The proposed method will employ an effective and highly

Correspondence to: Carlito B. Lebrilla, *Department of Chemistry, University of California, Davis*, e-mail: cblebrilla@ucdavis.edu

reproducible separation technique with high sensitivity. High performance mass spectrometry provides high mass accuracy and resolution. Tandem MS will provide characteristic fragment ions to distinguish further isomers. For this project, we have selected nanoflow liquid chromatography (nanoLC) MS employing the Agilent microfluidic-based Chip-quadrupole time-of-flight mass spectrometry (QTOF MS) system.

NanoLC MS is critical for the analysis. Nanoflow techniques have sample requirements more consistent with biological analysis. Nanoflow liquid chromatography is emerging as a valuable technique offering high sensitivity, shorter analysis time, high resolution, and effective separation.[13] The ability of nanoLC to separate linkage and other structural isomers makes it a valuable technique for oligosaccharide profiling, especially when handling small amounts of sample.[14–16] Shown in Fig. 1 is the nanoflow chromatographic separation of N-glycans from serum proteins. Integrating mass spectrometric detection with nanoLC improves sensitivity and provides an ideal platform for structure identification.[13] Recent studies from this laboratory have shown that N-glycans from serum samples and free oligosaccharides, which are similar to O-glycans, from human milk samples are readily separated using nanoLC yielding over 200 neutral and sialylated species in a single analysis while using porous graphitized carbon (PGC) as the stationary phase.[15,16] Coupling a high performance mass analyzer not only yields highly reproducible retention times but also high mass accuracy (1–6 ppm mass error). In this project, we separate N-glycans using a nanoflow device for liquid chromatography employing an integrated microchip-based device with graphitized carbon as the stationary phase. We have examined other methods capable of sensitive separation of oligosaccharides. The other separation method with the same sensitivity is capillary electrophoresis (CE). CE has shown early promise but has been not as well utilized for oligosaccharide analysis. We have published early articles on the separation of oligosaccharides with CE.[17] However, we found difficulty in coupling CE with MS and the inability of CE to separate large isomeric species.

Tandem MS (MS^2) spectra are used to fingerprint individual compounds and distinguish isomers. In the QTOF, we obtain extensive fragmentation throughout the mass range eliminating the need for MS^n. Shown in Fig. 2 are the tandem mass spectra of two isomeric species. The fragmentation patterns are distinct and uniquely identify each compound.

The utility of tandem MS for oligosaccharides differs from that of peptides. For peptides, peptide bond cleavages are typically the most common fragmentations and yield sequence and complete structure. The information is generally platform independent. For glycans, isomers are not structurally resolved by interpreting the tandem MS spectra. The most common fragment ions are typically due to glycosidic bond cleavages but they usually differ by the relative magnitude of the fragment intensities.[18] Furthermore, multiple energy regimes were used to vary the extent of fragmentation. However, the optimal conditions were found to yield the most structurally informative spectra.

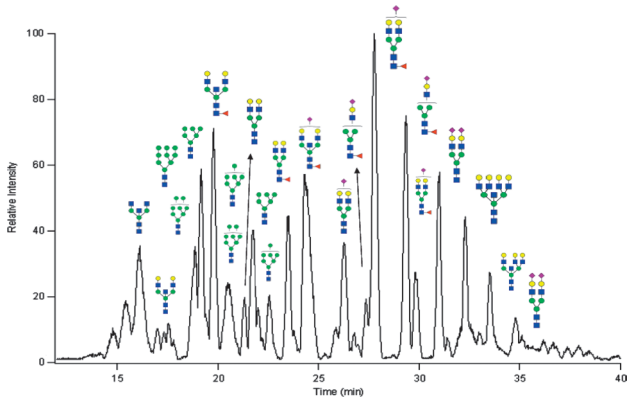


Fig. 1. NanoLC MS/MS chromatogram of N-glycans from serum proteins.

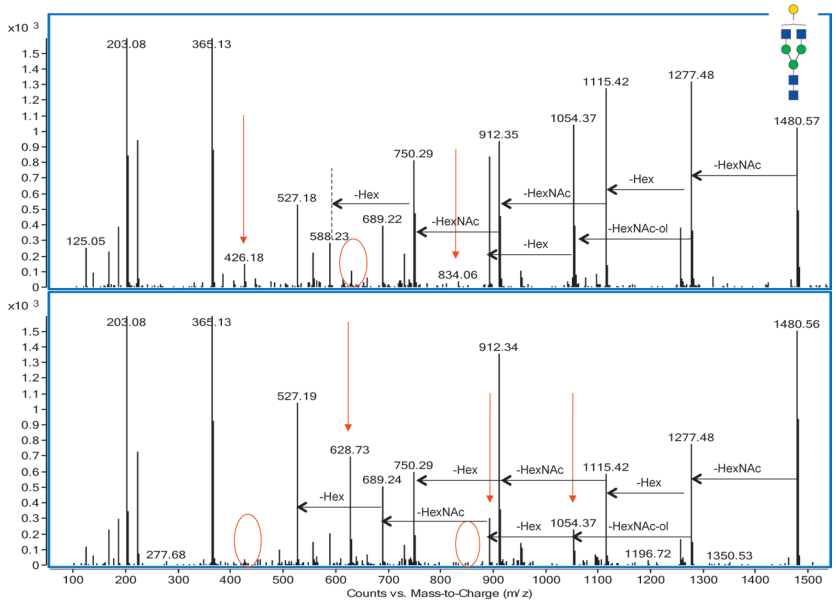


Fig. 2. Tandem mass spectra of two isomeric N-glycans. Each compound has distinct fragmentation pattern.

CREATE FUNCTIONAL DATABASE FOR GLYCOMICS

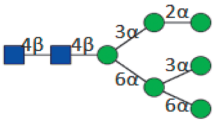
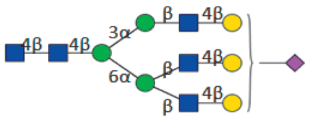
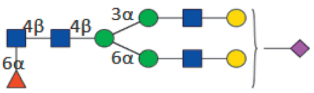
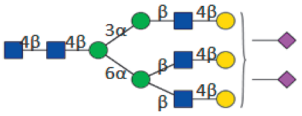
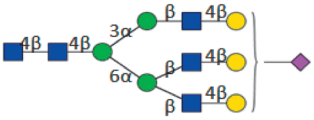
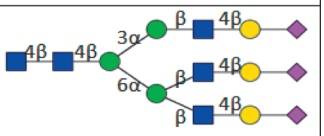
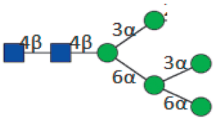
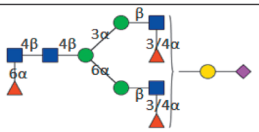
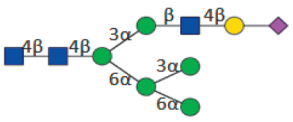
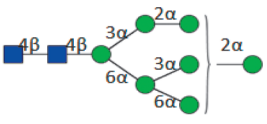
A glycan database is being constructed from serum glycans. Shown in Table 1 is a partial list, which includes retention time, accurate mass, tandem MS, and structure. Cells regularly transmit and shed glycoproteins into the blood. Blood therefore contains a suitable representation of every glycan in the body. We use primarily serum to build the glycan database because serum is readily available and is generally representative of blood at least in terms of glycosylation. Furthermore, while there are proteases that complicate proteomics analysis in serum, there are little or no glycosidases in serum that can complicate the analysis.

Structure elucidation is performed by a combination of tandem MS and exoglycosidase reactions. Accurate mass is used to determine the composition, while tandem MS is used to obtain the sequence. The tandem MS is used to guide the exoglycosidase reactions. Based on this strategy, we have over 100 complete and nearly complete N-glycan structures.

CONCLUSION

A functional and effective database is being constructed

Table 1. Partial list of N-glycans from serum glycoproteins.

Compound name	Mass	Relative abundance	Structure
N6200a	1398.5034	1.6062	
N6501a	2298.8383	1.5913	
N5411d	2079.7656	1.5666	
N6502c	2589.9265	1.4716	
N6501b	2298.8441	1.4576	
N6503a	2881.0274	1.3233	
N5200a	1236.4504	1.3108	
N4431a	2209.8401	1.3057	
N6301c	1892.7	1.2442	
N7200b	1560.5558	1.1995	

for the rapid identification of N-glycans for serum. The database is used to identify N-glycans based on their accurate mass, retention time, and tandem MS. We find that accurate mass and retention times are sufficient to identify nearly 90% of the glycan structures. Tandem MS can identify nearly all structures but cannot be performed with very low abundant species. The combination of three identifiers provides a comprehensive method for the rapid identification of oligosaccharide structures.

REFERENCES

1) N. H. Packer, C. W. von der Lieth, K. F. Aoki-Kinoshita, C. B. Lebrilla, J. C. Paulson, R. Raman, P. Rudd, R. Sasisekharan, N. Taniguchi, W. S. York. Frontiers in glycomics: Bioinformatics and biomarkers in disease. An NIH white paper prepared from discussions by the focus groups at a workshop on the NIH campus, Bethesda MD (September 11–13, 2006). *Proteomics* 8: 8–20, 2008.
2) H. J. An, S. R. Kronewitter, M. L. de Leoz, C. B. Lebrilla. Glycomics and disease markers. *Curr. Opin. Chem. Biol.* 13: 601–607, 2009.
3) C. B. Lebrilla, H. J. An. The prospects of glycan biomarkers for the diagnosis of diseases. *Mol. Biosyst.* 5: 17–20, 2009.
4) D. H. Dube, C. R. Bertozzi. Glycans in cancer and inflammation—Potential for therapeutics and diagnostics. *Nat. Rev. Drug Discov.* 4: 477–488, 2005.
5) C. Kirmiz, B. Li, H. J. An, B. H. Clowers, H. K. Chew, K. S. Lam, A. Ferrige, R. Alecio, A. D. Borowsky, S. Sulaimon, C. B. Lebrilla, S. Miyamoto. A serum glycomics approach to breast cancer biomarkers. *Mol. Cell. Proteomics* 6: 43–55, 2007.
6) R. Saldova, L. Royle, C. M. Radcliffe, U. M. Abd Hamid, R. Evans, J. N. Arnold, R. E. Banks, R. Hutson, D. J. Harvey, R. Antrobus, S. M. Petrescu, R. A. Dwek, P. M. Rudd. Ovarian cancer is associated with changes in glycosylation in both acute-phase proteins and IgG. *Glycobiology* 17: 1344–1356, 2007.
7) T. I. Williams, K. L. Toups, D. A. Saggese, K. R. Kalli, W. A. Cliby, D. C. Muddiman. Epithelial ovarian cancer: Disease etiology, treatment, detection, and investigational gene, metabolite, and protein biomarkers. *J. Proteome Res.* 6: 2936–2962, 2007.
8) H. J. An, S. Miyamoto, K. S. Lancaster, C. Kirmiz, B. Li, K. S. Lam, G. S. Leiserowitz, C. B. Lebrilla. Profiling of glycans in serum for the discovery of potential biomarkers for ovarian cancer. *J. Proteome Res.* 5: 1626–1635, 2006.
9) J. Zhao, W. Qiu, D. M. Simeone, D. M. Lubman. N-linked glycosylation profiling of pancreatic cancer serum using capillary liquid phase separation coupled with mass spectrometric analysis. *J. Proteome Res.* 6: 1126–1138, 2007.
10) Z. Kyselova, Y. Mechref, M. M. Al Bataineh, L. E. Dobrolecki, R. J. Hickey, J. Vinson, C. J. Sweeney, M. V. Novotny. Alterations in the serum glycome due to metastatic prostate cancer. *J. Proteome Res.* 6: 1822–1832, 2007.
11) R. G. Spiro. Protein glycosylation: Nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* 12: 43R–56R, 2002.
12) C. B. Lebrilla, L. K. Mahal. Post-translation modifications. *Curr. Opin. Chem. Biol.* 13: 373–374, 2009.
13) J. Hernández-Borges, Z. Aturki, A. Rocco, S. Fanali. Recent applications in nanoliquid chromatography. *J. Sep. Sci.* 30: 1589–1610, 2007.
14) Y. Mechref, M. V. Novotny. Miniaturized separation techniques in glycomic investigations. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 841: 65–78, 2006.
15) M. Niñonuevo, H. An, H. Yin, K. Killeen, R. Grimm, R. Ward, B. German, C. Lebrilla. Nanoliquid chromatography-mass spectrometry of oligosaccharides employing graphitized carbon chromatography on microchip with a high-accuracy mass analyzer. *Electrophoresis* 26: 3641–3649, 2005.
16) M. R. Ninonuevo, Y. Park, H. Yin, J. Zhang, R. E. Ward, B. H. Clowers, J. B. German, S. L. Freeman, K. Killeen, R. Grimm, C. B. Lebrilla. A strategy for annotating the human milk glycome. *J. Agric. Food Chem.* 54: 7471–7480, 2006.
17) H. J. An, A. H. Franz, C. B. Lebrilla. Improved capillary electrophoretic separation and mass spectrometric detection of oligosaccharides. *J. Chromatogr. A* 1004: 121–129, 2003.
18) B. Li, H. J. An, J. L. Hedrick, C. B. Lebrilla. Infrared multiphoton dissociation mass spectrometry for structural elucidation of oligosaccharides. *Methods Mol. Biol.* 534: 23–35, 2009.