The prospects of glycan biomarkers for the diagnosis of diseases

Carlito B. Lebrilla* and Hyun Joo An

DOI: 10.1039/b811781k

Over 40 years of literature shows that glycosylation is greatly affected by diseases such as cancer. This *opinion* article argues the intrinsic advantages of using glycans as disease markers over other biomolecules and the potential of glycan profiling for diagnosing and determining the progression of disease.

Recent interest in profiling the glycome stems from the potential of glycans as disease markers. Over 40 years of literature are available that clearly show glycosylation is greatly affected by diseases such as cancer.[1,2] With glycans as disease markers there are several intrinsic advantages compared to other biomolecules, specifically proteins: (1) glycan biosynthesis is more significantly affected by disease states than protein production. Glycans are essentially metabolic products that are amplified with even small changes in the associated protein (glycosyltransferase) expression. (2) Glycans are produced and modified in primarily two places in the cell. Aberrant glycosylation can therefore potentially affect nearly every glycoprotein produced in the diseased cell. (3) Given the current technology, it is far simpler to quantitate oligosaccharide expression than protein expression. Oligosaccharides have similar chromophoric activities and can be readily profiled by chromatographic detectors. As glycans have similar sizes and structures (within specific types), their ionization efficiencies in mass spectrometry are relatively similar so that relative intensities permit the calculation of the relative oligosaccharide abundances.[3]

Only recently has the global glycome been used as a source for disease markers. Progress has been hindered by the complexities of oligosaccharides and the lack of analytical methods for elucidating structures. This condition is, however, changing rapidly. Oligosaccharides are highly amenable to mass spectrometric analyses and in recent years, considerable efforts have been expanded in developing mass spectrometry as the central method for oligosaccharide analysis. Mass spectrometry provides a rapid and fine tool for component analysis. It is also a precise tool for structural elucidation. As new methods for profiling oligosaccharide composition and structure are becoming available, the search for biomarkers is now intensifying.

The glycome is defined as the glycan constituent of all glycoconjugates from a single specific biological source. Glycans may be either free oligosaccharides or glycoconjugates such as glycoproteins and glycolipids. Protein-bound glycans include N-linked glycans, linked to an asparagine on the protein backbone, O-linked glycans, linked to either threonine or serine, or glycosaminoglycans (GAGs), highly anionic and

Department of Chemistry and Department of Biochemistry and Molecular Medicine, University of California, Davis, USA. E-mail: cblebrilla@ucdavis.edu; Fax: +1–530–752–8995; Tel: +1–530–752–6364

Carlito B. Lebrilla

Dr Carlito B. Lebrilla is a professor at University of California, Davis in the Departments of Chemistry and Biochemistry and Molecular Medicine at the School of Medicine. He received his BS degree from the University of California, Irvine, and PhD from the University of California, Berkeley. He was an Alexander von Humboldt Fellow and an NSF-NATO Fellow at the Technical University in Berlin. He returned to the UC Irvine as a President's Fellow and has been at UC Davis since 1989. His research is in analytical chemistry, primarily mass spectrometry with applications to clinical glycomics and bioanalytical chemistry.

Hyun Joo An

Hyun Joo An received her BS and MS degrees (1997) from Inha University, Incheon, South Korea. She received her PhD in the Department of Chemistry at University of California, Davis. She has been working with Prof. Lebrilla as a postdoctoral researcher since 2004. She is currently working on the application of Fourier transform mass spectrometry in bioanalysis. Her specific research interests include oligosaccharide analysis (glycomics), biomarker discovery for infectious diseases and cancer and glycoprotein analysis (glycoproteomics) to determine site specific glycosylation and heterogeneity.

polymeric oligosaccharides linked to serine or threonine.

Glycosylation is one of the most common post-translational modifications of proteins and often occurs on cell surfaces and on secreted proteins. It is estimated that 50% of eukaryotic proteins are glycosylated.[4] For human proteins, it is estimated to be as high as 70%.[4] Glycosylation is the only modification with significant structural complexity. Unlike proteins, the primary structures of oligosaccharides are often nonlinear, with branching common. Furthermore, the structure of oligosaccharides cannot be predicted from genomic data, as with proteins. Glycosylation is also highly heterogeneous and diverse. A specific glycosylation site is rarely associated with a single glycan structure. Over a hundred different structures can be associated with a single glycan site. From the glycoproteomics point of view, this makes it difficult to determine both the proteome and the glycome simultaneously. A protein with three sites of glycosylation and ten different glycans on each site can have as many as 1000 different glycoforms (neglecting unoccupied variants). For this reason, glycoproteomics is an area that remains largely uncharted.

Glycans are important determinants of protein function in disease biology. Changes in glycosylation have the potential to be more pronounced than protein expression during the transformation to the diseased condition.[2] The glycosylation machinery involves a set of competing glycosyltransferases that add to the nascent oligosaccharide chain. When a transferase is either up or down regulated, as for example with the conditions in several diseases including cancer, its effect is amplified. Examples of aberrant glycosylation are rampant in the literature and include increases in the amount of fucose, a monosaccharide residue, and sialic acids (a nine-carbon monosaccharide containing a carboxylic acid). The glycans expressed by diseased cells are different as compared to those of normal cells.[1,5] For example, the glycosylation of prostate specific antigen (PSA) differs between normal individuals and patients with prostate cancer.[6] Glycans from major serum proteins such as α-1 acid glycoprotein, haptoglobin, and α-fetoprotein have also been known to vary between control and cancer patients.[7,8] There have been several potential glycan markers discovered by employing standard procedures such as lectin affinity, immuno-affinity electrophoresis, and blotting methods.[9] Lectin-based glycan detection methods have also been used for studying the roles of glycans in diseases.[10]

The majority of glycosylation processing occurs in two major loci of the cell, in the Golgi and in the endoplasmic reticulum. Aberration in the glycosylation process would mean that not only a few low abundance proteins, but potentially every single glycoprotein from a diseased cell could be mis-glycosylated. This glycan biosynthesis points to another benefit of the glycome as a source for disease markers, that is, the effect of the disease is significantly amplified compared to protein expression. The aberrant glycans need not necessarily come from diseased cells as glycosylation changes are known to occur as a response to the disease.[1,11] For example, glycosylation of immunoglobulin (Ig) is known to change between control and cancer patients.[11] In general, incomplete or truncated glycan structures on Ig are expressed persistently in cancer patients. In ovarian cancer the level of core fucosylated, agalactosyl biantennary glycans on IgG is similarly increased. Mannose-type sugars or agalactosyl oligosaccharides on IgG in the serum of patients with gastric cancer are increased. In lung cancer agalactosyl glycans are significantly increased while mono- and digalactosyl IgG glycans are decreased. Changes in *N*-glycosylation of IgG and incomplete galactosylation of IgA are also known to be associated with rheumatoid arthritis. Though perhaps not as useful from the therapeutics perspective, glycosylation changes of Ig will still have significant value as sources for cancer markers.

While characterizing the entire glycome is desirable, it is still methodologically difficult to study it completely and simultaneously. The most common differentiation is between N- and O-linked glycans. The majority of the reported studies have been on N-linked glycans, which are significantly more abundant in serum and plasma. N-Linked glycans are often released with an enzyme whereas O-linked glycans are released chemically. It is possible to release both types simultaneously but not with the desired completeness. In addition, O-linked glycans are often smaller than N-linked glycans. These size differences are often sufficient to require different conditions for analysis. Other glycans such as GAGs are so anionic and have such different chemical properties that analyzing them with the N- and O-linked is not possible. For disease markers therefore, it is best to focus on either N-linked, O-linked, or the GAGs separately.

Analysis of protein glycosylation involves several levels of complexity. The pyramid in Fig. 1 represents each of the different levels beginning with a simple compositional profile and moving downwards (with increasing complexity) to complete molecular profile, protein-specific glycosylation, and the site-specific glycosylation. The compositional profile is primarily the number of hexoses (Hex), *N*-acetylhexosamines

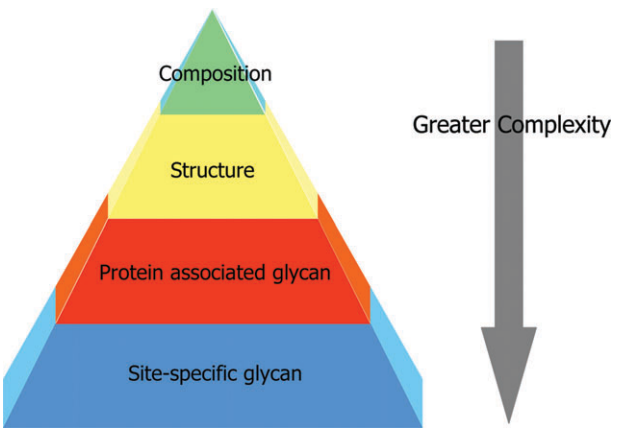

Fig. 1

(HexNAc), deoxyhexoses (usually fucose for humans), and sialic acids. As we proceed to the base of the pyramid, the complexity caused by oligosaccharide heterogeneity increases significantly. However, as we argue below, there are significant opportunities at each level not only for disease markers but also for therapeutic targets.

It is commonly believed that the glycan and its protein attachment are necessary to discover disease markers. However, an assay of the glycan, even a compositional profile of the glycans with little or no knowledge of the proteins, may in many cases be sufficient. The glycomics approach relies extensively on mass spectrometry. Changes in glycosylation are readily monitored by simple glycan compositional analysis, which is obtained from accurate mass analysis. High mass accuracy and high resolution are advantages in profiling oligosaccharides. The capability of a high resolution mass analyzer, such as by Fourier transform ion cyclotron resonance (FT ICR), to provide accurate mass has been crucial for obtaining basic glycan structure information. For example, an oligosaccharide with a quasi-molecular ion ($[M + Na]^+$) at m/z 2201.819 has three possible compositions within a tolerance of ± 0.1 mass units. Only with a tolerance of 0.01 mass units is the correct composition of two Fuc, four Hex, and six HexNAc obtained. Furthermore, putative structures of N-linked glycans having a trimannosyl core ($Man_3GlcNAc_2$) can be derived primarily from the composition. A simple approach to disease marker discovery might be to release the glycans from all glycoproteins and examine only the glycans based on their masses. Indeed, the first publications that have employed a glycomics approach examined only oligosaccharides released from the proteins and yielded only composition.[12–16]

Glycan mass profiling that provides compositions has already yielded several potential markers. However, individual oligosaccharides (specific isomeric structures) may provide more robust glycan markers with higher specificity than composition alone. There have been several glycan markers proposed for diseases such as gastrointestinal cancer and cystic fibrosis with lung disease.[17,18] The glycomic approaches require rapid and sensitive methods for separating and analyzing the glycome. In this regard, analytical methods for separating components into isomers, such as nanoflow liquid chromatography (LC) and capillary electrophoresis (CE), coupled with high performance mass spectrometry for providing composition and structural information have proved highly useful.[19,20] High reproducibility, particularly with methods such as nanoLC, will eventually allow rapid identification of known oligosaccharides.

Protein specific glycosylation requires a higher level of complexity yet offers a richer source of glycan markers. Analyzing glycosylation of proteins between disease states may yield more sensitive and specific markers in addition to some important targets for therapeutics. Glycoproteins can be separated and glycans released from even, for example, gel spots. In this way glycoproteins, such as the prostate specific antigen (PSA) clinical marker, have been shown to change in glycosylation between cancer and non-cancer states.[6]

More recently, reasonably fast and comprehensive analysis of site-specific glycosylation is now possible using a digestion method with non-specific proteases.[3,10,21] However, a "true" glycoproteomics approach still eludes us, although there have been considerable efforts in this area that may soon come to fruition. There have been attempts to analyze glycoproteins employing proteomics-type technology, however, the information regarding the glycan is often discarded.[22] These methods should not therefore be considered glycoproteomics in the complete sense.

Glycan markers hold considerable opportunities and challenges for disease diagnosis. The analytical tools for glycomic profiling have advanced considerably to the point that they are more than adequate for glycomics profiling and disease marker discovery. Further developments are needed in bioinformatic tools and in methods of glycoproteomic analyses where both glycan and protein information are obtained. Glycomics has the potential to provide a single platform to monitor several diseases simultaneously with both high sensitivity and high specificity. While this discussion has focused on mass spectrometry, there are other technologies, such as glycan arrays, that hold similar promise.[23] Glycans are abundant in most biological fluids, including serum, plasma, saliva, and tears. The marker may be based simply on the glycan profile at the composition level or perhaps on more involved analysis involving site-specific glycosylation. However, protein identification is not a necessity for solely glycomics-based markers. We predict that, in the future, it will be possible to not only diagnose diseases, but perhaps even determine disease progression and specific strain based solely on glycan profiling.

References

1 D. H. Dube and C. R. Bertozzi, *Nat. Rev. Drug Discovery*, 2005, **4**, 477–488.
2 M. M. Fuster and J. D. Esko, *Nat. Rev. Cancer*, 2005, **5**, 526–542.
3 H. J. An, T. R. Peavy, J. L. Hedrick and C. B. Lebrilla, *Anal. Chem.*, 2003, **75**, 5628–5637.
4 R. Apweiler, H. Hermjakob and N. Sharon, *Biochim. Biophys. Acta*, 1999, **1473**, 4–8.
5 I. Brockhausen, *Biochim. Biophys. Acta*, 1999, **1473**, 67–95.
6 R. Peracaula, G. Tabares, L. Royle, D. J. Harvey, R. A. Dwek, P. M. Rudd and R. de Llorens, *Glycobiology*, 2003, **13**, 457–470.
7 I. L. Ang, T. C. W. Poon, P. B. S. Lai, A. T. C. Chan, S. M. Ngai, A. Y. Hui, P. J. Johnson and J. J. Y. Sung, *J. Proteome Res.*, 2006, **5**, 2691–2700.
8 J. Zhao, T. H. Patwa, W. L. Qiu, K. Shedden, R. Hinderer, D. E. Misek, M. A. Anderson, D. M. Simeone and D. M. Lubman, *J. Proteome Res.*, 2007, **6**, 1864–1874.
9 A. Kuno, N. Uchiyama, S. Koseki-Kuno, Y. Ebe, S. Takashima, M. Yamada and J. Hirabayashi, *Nat. Methods*, 2005, **2**, 851–856.
10 T. H. Patwa, J. Zhao, M. A. Anderson, D. M. Simeone and D. M. Lubman, *Anal. Chem.*, 2006, **78**, 6411–6421.
11 J. N. Arnold, M. R. Wormald, R. B. Sim, P. M. Rudd and R. A. Dwek, *Annu. Rev. Immunol.*, 2007, **25**, 21–50.
12 G. S. Leiserowitz, C. Lebrilla, S. Miyamoto, H. J. An, H. Duong, C. Kirmiz, B. Ll, H. Liu and K. S. Lam, *Int. J. Gynecol. Cancer*, 2008, **18**, 470–475.
13 D. Isailovic, R. T. Kurulugama, M. D. Plasencia, S. T. Stokes, Z. Kyselova, R. Goldman, Y. Mechref, M. V. Novotny and D. E. Clemmer, *J. Proteome Res.*, 2008, **7**, 1109–1117.
14 Z. Kyselova, Y. Mechref, M. M. Al Bataineh, L. E. Dobrolecki, R. J. Hickey, J. Vinson, C. J. Sweeney and M. V. Novotny, *J. Proteome Res.*, 2007, **6**, 1822–1832.

15 C. Kirmiz, B. S. 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 and S. Miyamoto, *Mol. Cell. Proteomics*, 2007, **6**, 43–55.
16 H. J. An, S. Miyamoto, K. S. Lancaster, C. Kirmiz, B. S. Li, K. S. Lam, G. S. Leiserowitz and C. B. Lebrilla, *J. Proteome Res.*, 2006, **5**, 1626–1635.
17 J. Fernandez-Rodriguez, O. Dwir, R. Alon and G. C. Hansson, *Glycoconjugate J.*, 2001, **18**, 925–930.
18 B. L. Schulz, A. J. Sloane, L. J. Robinson, L. T. Sebastian, A. R. Glanville, Y. L. Song, A. S. Verkman, J. L. Harry, N. H. Packer and N. G. Karlsson, *Biochem. J.*, 2005, **387**, 911–919.
19 M. Ninonuevo, H. J. An, H. F. Yin, K. Killeen, R. Grimm, R. Ward, B. German and C. Lebrilla, *Electrophoresis*, 2005, **26**, 3641–3649.
20 H. J. An, A. H. Franz and C. B. Lebrilla, *J. Chromatogr., A*, 2003, **1004**, 121–129.
21 Y. Zhang, E. P. Go and H. Desaire, *Anal. Chem.*, 2008, **80**, 3144–3158.
22 P. Hagglund, J. Bunkenborg, F. Elortza, O. N. Jensen and P. Roepstorff, *J. Proteome Res.*, 2004, **3**, 556–566.
23 T. Horlacher and P. H. Seeberger, *OMICS*, 2006, **10**, 490–498.

