

Glycoproteomics: New Technology Developments and Applications Provide Renewed Interest in Glycoproteins

Hui Zhang · Robert J. Cotter · Guest Editors

Published online: 1 November 2008
© Humana Press 2008

Glycoproteins are formed through covalent linkages of sugar residues to specific amino acids in proteins. Both amide and acyl linkages are possible, and the resultant N-glycosylated and O-glycosylated proteins carry remarkable complexity in their oligosaccharide chains. N- and O-glycosylations occur in proteins that remain in the endoplasmic reticulum, proteins in the plasma membrane, and proteins secreted into the extracellular surface. N-glycosylation is localized to Asn residues, in the consensus tripeptide sequence Asn-*X*-Ser/Thr (*X* is any amino acid except proline). O-glycosylation of proteins occurs commonly on both Ser and Thr residues, as well as posttranslational hydroxylation-created 5-hydroxylysine and 4-hydroxyproline side chains. There are reports of glycosylation to Cys residues. O-linked β -*N*-acetylglucosamine (O-GlcNAc) is a modification of serine or threonine side chains of nuclear and cytoplasmic proteins [1].

The study of glycoproteins began a very long time ago; progress in understanding their structure and function has awaited the development of methods for the isolation of glycoproteins and glycans and the analysis of the sequences of residues in both polysaccharides and polypeptides. The field of “glycoproteomics” has emerged recently from the global analyses of structures and functions of glycoproteins, similar to the approaches now used in “proteomics.” The

questions we ask in glycoproteomics include: (1) what glycoproteins are expressed in a cell type or tissue, (2) what particular glycan structures of a glycoprotein is modified, (3) which amino acid of a specific glycan structure is attached to the backbone of a glycoprotein, (4) how the function of a protein is modified by the specific glycosylation, (5) how abundant the glycoproteins and glycans are in different conditions, and (6) whether a specific glycosylation of a glycoprotein is associated with diseases that can be used for diagnosis, prognosis, or treatment purposes. Investigations of these questions in the glycoproteomics field are hampered by the fact that there are no simple and sensitive tools for efficient glycoprotein isolation, detection, and quantification and we lack methods to analyze the glycosylation sites of large number of glycoproteins.

The renewed interest in glycoproteomics has been initiated by the completion of the human genome project, new technological breakthroughs in mass spectrometry (MS) instrumentation, and proteomic methods that have been developed in the last decade. Recently, there has been considerable progress in the development of methodologies for elucidating the structures and functions of glycoproteins, which has provided the opportunities to advance glycoprotein analysis. The availability of affinity reagents for glycoproteins (lectins and antibodies) [2] and newly developed glycoprotein derivatization and enrichment approaches [2–4] have provided important breakthroughs. Application of these glycoproteomic technologies may facilitate our understanding of the perturbed glycosylation in disease progression and could be used to discover glycoproteins as biomarkers or therapeutic targets to improve disease diagnosis, prognosis, or treatment outcomes.

In this first special edition of *Glycoproteomics*, we have received excellent contributions from several experts in this field. These papers cover different areas of glycoproteomics,

H. Zhang (✉)
Department of Pathology,
Johns Hopkins University School of Medicine,
Baltimore, MD 21231, USA
e-mail: hzhang32@jhmi.edu

R. J. Cotter
Department of Pharmacology & Molecular Sciences,
Johns Hopkins University School of Medicine,
Baltimore, MD 21231, USA
e-mail: rcotter@jhmi.edu

including reviews of the existing methods of glycoproteomics, new technology developments, and applications. The articles are now published in two special issues due to the large numbers of excellent manuscripts received. The first issue includes six papers that describe innovative approaches to analyze glycoprotein sequences or glycan structures. The other issue includes five papers that describe the applications of glycoproteomic technologies.

New techniques for glycoproteomic analysis, including those for glycoprotein derivatization and enrichment, have provided significant opportunities in glycoproteomic analysis [2–4]. The *O*-GlcNAc modification of proteins was originally discovered in Gerald Hart's laboratory over two decades ago, and in this special issue of *Glycoproteomics*, Wang and Hart review the methods and the recently developed technologies for specific enrichment and detection of *O*-GlcNAc sites and their quantification [5]. Kaji and Isobe summarize MS-based glycoproteomic technologies, with a special focus on preproteomic analysis enrichment of glycopeptides from complex biological mixtures and their assignments and structural analysis by MS [6]. The methods include: (1) affinity chromatography for capturing glycopeptides using lectins, antibodies, or glycoprotein receptors that bind to certain glycans [2]; (2) chemical coupling to capture glycopeptides to solid phase [3, 4]; (3) carbohydrate tags via metabolic or chemo-enzymatic labeling [7, 8]; and (4) liquid chromatography (LC)-based enrichment of glycopeptides using hydrophilic interaction, size exclusion chromatography, boronic acid, strong cation exchange, and titanium dioxide [9–13]. Their review also describes the mass spectrometric methods for glycopeptide analysis and quantitative aspects of glycoproteomics. In addition to summarizing the existing methods for glycoproteomics in this special edition, Lewandrowski et al. describe a novel approach for glycopeptide enrichment by electrostatic repulsion hydrophilic interaction chromatography to separate glycopeptides from the bulk of nonmodified peptides. The eluted glycopeptides were then identified by tandem MS [14].

Given the number and diversity of glycans, there is a need for new technologies that can more rapidly provide differential carbohydrate–protein structural information on a large scale. These new technologies are expected to facilitate the rapid and sensitive identification of novel biomarkers based on carbohydrate changes in glycoproteins and lead to a better understanding of the connections of glycan function in cellular mechanisms of health and disease. In this special edition, Ralin et al. describe a new microarray platform based on a Label Free Internal Reflection Ellipsometry (LFIRE™) technique [15]. This microarray platform allows simultaneous detection of multiple glycoprotein–lectin interactions without the need for reporter labels, while still providing high-throughput

kinetic screening of carbohydrate–lectin recognition at much lower cost. The lectin immobilization format was also used to demonstrate improved selectivity, specificity, and increased functional avidity using multilectin affinity chromatography to capture glycoproteins from biological samples. Vakhrushev et al. describe a novel approach using ion mobility tandem MS for the analysis of human glycourinome to obtain carbohydrate pattern data of a patient with congenital disorders of glycosylation (CDG) [16].

Currently, 19 protein biomarkers are approved by the US Food and Drug Administration in serum, urine, and tissue and all are glycoproteins [17]. Prostate-specific antigen (PSA) is the only tumor marker that is approved as diagnostic marker, and others are intended for monitoring response to therapy due to their low specificity for screening purposes. To improve the specificity and obtain precise and consistent analytical results, Li et al. analyzed five clinically used tumor markers, including PSA, carcinoembryonic antigen, Her-2, human chorionic gonadotropin, and CA125, and provided a list of tryptic peptides and glycopeptides as proteotypic peptides [18]. Heavy-isotope-labeled-peptide standards could be synthesized based upon the peptide sequences identified here and used for the development of assays for these tumor markers using MS-based methods. Using the targeted MS-based approach, different glycoforms and isoforms of the glycoproteins can be distinguished and specifically analyzed from the interference of other proteins in a complex biological mixture. This will allow the investigation of the potential for glycosylation patterns of these glycoproteins to be used as markers for improved cancer detection.

Glycosylation changes in glycoproteins have been identified in various diseases, such as CDGs and cancers, and can be used as biomarkers for disease diagnosis or prognosis [19, 20]. There has been an increasing effort in applying glycoproteomic technologies to identify additional disease biomarkers from specific organs or detection of glycoproteins from a specific organ in body fluids. In the second issue of this special edition of *Glycoproteomics*, five papers were contributed to describe the application of glycoproteomic technologies to biomarker discovery. The paper by Kam and Poon provides an overview of the historical applications of glycan analyses to biomarker discovery [21]. In the past few years, attempts have been made to develop novel methodologies to identify potential glyco-biomarkers, and plasma has been the focus of testing different technologies for the identification of biomarkers due to its ready accessibility. However, it is not clear if direct proteomic analysis of plasma can be used to discover new marker proteins that are associated with tumor progression. Tian et al. reported that such tumor-specific proteins can be detected in plasma in a chemically induced

skin cancer mouse model [22]. They analyzed glycoproteins from normal skin, benign papillomas, and malignant carcinomas from mice using a recently developed platform, solid-phase extraction of glycopeptides (SPEG) and MS, and they identified glycoproteins preferentially expressed in skin cancer, which include most known extracellular proteins that have been reported to play roles in skin cancer development. They further investigated whether these tumor proteins could be detected in plasma from tumor-bearing mice using isotope labeling and 2D-LC–matrix-assisted laser desorption ionization–MS/MS. This result indicates that direct proteomic analysis of tumors and plasma using glycopeptide capture, isotopic labeling, and MS can be used to discover new cancer-derived proteins in plasma for early cancer detection. To identify changes in the glycoproteins from serum that could be used as biomarkers, Letarte et al. demonstrated a proof-of-concept for the use of label-free quantitative glycoproteomics for a biomarker discovery workflow [23]. Formerly, N-linked glycopeptides were isolated from plasma proteins of mouse model for skin cancer with a mutation in the p19ARF gene using SPEG, and the glycopeptides were used for LC-MS analyses. The computational framework, Corra, was then used to determine the most discriminating peptide features from cancer and control mice for their identification by tandem MS using inclusion lists. The identified proteins were assessed to determine the specific biological processes that correlate with the presence of disease.

In addition to plasma, other body fluids or tissues are also good sources for biomarker discovery of a specific disease. Saliva is a body fluid that holds promise for use as a diagnostic fluid. Ramachandran et al. employed SPEG to enrich for formerly N-linked glycopeptides for analysis using LC-tandem MS to identify the salivary N-glycoproteins [24]. N-glycoproteins from whole saliva, parotid fluid, submandibular fluid, and sublingual fluid were identified and compared in order to identify glycoproteins that are differentially found in submandibular and sublingual glandular secretions. The majority of the N-glycoproteins were annotated as extracellular proteins. The identifying glycoproteins in saliva may be valuable for future disease biomarker studies. Chaerkady et al. carried out tissue glycoproteomic profiling of tumor and adjacent non-cancer liver tissues from hepatocellular carcinoma (HCC) patients [25]. Glycoproteins enriched from liver samples using lectin affinity chromatography and subsequent $^{18}\text{O}/^{16}\text{O}$ labeling of peptides were used to obtain relative abundance levels of lectin-bound proteins. Western blotting and immunohistochemical staining were carried out for several proteins to confirm MS results. As a complementary approach, they also examined the relative expression of proteins in HCC without glycoprotein enrichment. Lectin affinity enrichment was found to be advantageous to

quantitate several interesting proteins, which were not detected in the whole proteome screening approach.

We are very grateful to the positive responses from researchers in the field of glycobiology and glycoproteomics for their suggestions, support, contributions, and reviewing of the manuscripts, which have made the first special edition of *Glycoproteomics* possible. We would also extend our gratitude to the editor-in-chief, reviewers, and staffs from *Clinical Proteomics* and Springer for their support of these special issues.

References

- Varki AEA. Essentials of glycobiology. Cold Spring Harbor: Cold Spring Harbor Laboratory Press; 1999.
- Kaji H, Saito H, Yamauchi Y, Shinkawa T, Taoka M, Hirabayashi J, et al. Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. *Nat Biotechnol* 2003;21(6):667–72.
- Zhang H, Li XJ, Martin DB, Aebersold R. Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat Biotechnol* 2003;21(6):660–6.
- Wells L, Vosseller K, Cole RN, Cronshaw JM, Matunis MJ, Hart GW. Mapping sites of O-GlcNAc modification using affinity tags for serine and threonine post-translational modifications. *Mol Cell Proteomics* 2002;1(10):791–804.
- Wang Z, Hart GW. Glycomic approaches to study GlcNAcylation: Protein identification, site-mapping, and site-specific O-GlcNAc quantitation. *Clin Proteomics* 2008;4(1).
- Kaji H, Isobe T. Liquid chromatography/mass spectrometry (LC/MS)-based glycoproteomics technologies for cancer biomarker discovery. *Clin Proteomics* 2008;4(1).
- Sprung R, Nandi A, Chen Y, Kim SC, Barma D, Falck JR, et al. Tagging-via-substrate strategy for probing O-GlcNAc modified proteins. *J Proteome Res* 2005;4(3):50–7.
- Khidkel N, Ficarro SB, Peters EC, Hsieh-Wilson LC. Exploring the O-GlcNAc proteome: direct identification of O-GlcNAc-modified proteins from the brain. *Proc Natl Acad Sci U S A* 2004;101(36):13132–7.
- Hagglund P, Bunkenborg J, Elortza F, Jensen ON, Roepstorff P. A new strategy for identification of N-glycosylated proteins and unambiguous assignment of their glycosylation sites using HILIC enrichment and partial deglycosylation. *J Proteome Res* 2004;3(3):556–66.
- Sparbier K, Koch S, Kessler I, Wenzel T, Kostrzewa M. Selective isolation of glycoproteins and glycopeptides for MALDI-TOF MS detection supported by magnetic particles. *J Biomol Tech* 2005;16(4):407–13.
- Alvarez-Manilla G, Atwood J 3rd, Guo Y, Warren NL, Orlando R, Pierce M. Tools for glycoproteomic analysis: size exclusion chromatography facilitates identification of tryptic glycopeptides with N-linked glycosylation sites. *J Proteome Res* 2006;5(3):701–8.
- Lewandrowski U, Zahedi RP, Moebius J, Walter U, Sickmann A. Enhanced N-glycosylation site analysis of sialoglycopeptides by strong cation exchange prefractionation applied to platelet plasma membranes. *Mol Cell Proteomics* 2007;6(11):1933–41.
- Larsen MR, Jensen SS, Jakobsen LA, Heegaard NH. Exploring the sialome using titanium dioxide chromatography and mass spectrometry. *Mol Cell Proteomics* 2007;6(10):1778–87.

14. Lewandrowski U, Lohrig K, Zahedi RP, Wolters D, Sickmann A. Glycosylation site analysis of human platelets by electrostatic repulsion hydrophilic interaction chromatography. *Clin Proteomics* 2008;4(1).
15. Ralin DW, Dultz SC, Silver JE, Travis JC, Kullolli M, Hancock WS, et al. Kinetic analysis of glycoprotein-lectin interactions by label free internal reflection ellipsometry. *Clin Proteomics* 2008;4(1).
16. Vakhrushev SY, Langridge J, Campuzano I, Hughes C, Peter-Katalinić J. Identification of monosialylated N-glycoforms in the CDG urine by ion mobility tandem mass spectrometry: the potential for clinical applications. *Clin Proteomics* 2008 4(1).
17. Sokoll LJ, Chan DW. Biomarkers for cancer diagnostics. In: Abeloff MD, Armitage JO, Niederhuber JE, Kastan MB, McKenna WG, editors. *Abeloff's clinical oncology*. 4th ed. Philadelphia: Elsevier; 2008.
18. Li Y, Sokoll LJ, Barker PE, Zhang H, Chan DW. Mass spectrometric identification of proteotypic peptides from clinically used tumor markers. *Clin Proteomics* 2008;4(1).
19. Roth J. Protein N-glycosylation along the secretory pathway: relationship to organelle topography and function, protein quality control, and cell interactions. *Chem Rev* 2002;102(2):285–303.
20. Durand G, Seta N. Protein glycosylation and diseases: blood and urinary oligosaccharides as markers for diagnosis and therapeutic monitoring. *Clin Chem* 2000;46(6 Pt 1):795–805.
21. Kam RK, Poon TW. The potentials of glycomics in biomarker discovery. *Clin Proteomics* 2008;4(2).
22. Tian Y, Kelly-Spratt KS, Kemp KS, Zhang H. Identification of glycoproteins from mouse skin tumors and plasma. *Clin Proteomics* 2008;4(2).
23. Letarte S, Brusniak MY, Campbell D, Eddes J, Kemp CJ, Lau H, et al. Differential plasma glycoproteome of p19ARF skin cancer mouse model using the corra label-free LC-MS proteomics platform. *Clin Proteomics* 2008;4(2).
24. Ramachandran P, Boonthueung P, Pang E, Yan W, Wong DT, Loo JT. Comparison of N-linked glycoproteins in human whole saliva, parotid, submandibular, and sublingual glandular secretions identified using hydrazide chemistry and mass spectrometry. *Clin Proteomics* 2008;4(2).
25. Chaerkady R, Thuluvath PJ, Kim MK, Nalli A, Vivekanandan P, Simmers J, et al. ¹⁸O Labeling for a quantitative proteomic analysis of glycoproteins in hepatocellular carcinoma. *Clin Proteomics* 2008;4(2).