Vol. 7, No. 3, 2016
ISSN 2233-4203/ e-ISSN 2093-8950
LETTER www.msletters.org | Mass Spectrometry Letters

Differentiation of Glycan Diversity with Serial Affinity Column Set (SACS)

Jihoon Shin and Wonryeon Cho*

Department of Chemistry, Wonkwang University, 460 Iksandae-ro, Iksan, Jeonbuk 54538, Republic of Korea

Received September 06, 2016; Revised September 22, 2016; Accepted September 26, 2016
First published on the web September 30, 2016; DOI: 10.5478/MSL.2016.7.3.74

Abstract : Targeted glycoproteomics is an effective way to discover disease-associated glycoproteins in proteomics and serial affinity chromatography (SAC) using lectin and glycan-targeting antibodies shows glycan diversity on the captured glycoproteins. This study suggests a way to determine glycan heterogeneity and structural analysis on the post-translationally modified proteins through serial affinity column set (SACS) using four *Lycopersicon esculentum* lectin (LEL) columns. The great advantage of this method is that it differentiates between glycoproteins on the basis of their binding affinity. Through this study, some proteins were identified to have glycoforms with different affinity on a single glycoprotein. It will be particularly useful in determining biomarkers in which the disease-specific feature is a unique glycan, or a group of glycans.

Keywords : Targeted glycoproteomics, Serial Affinity Chromatography (SAC), Serial Affinity Column Set (SACS), lectins, glycan-targeting antibody, glycan diversity, disease associated glycoprotein

Introduction

Proteomics, the large-scale analysis of proteins, contributes greatly to our understanding of gene function in the post-genomic era.[1-3] Mass spectrometry (MS) is a significant methodology for proteomics in harmony with a wide variety of separation methods. Tandem mass spectrometry (MS/MS) is a major tool in protein identification. Mass spectrometers measure the mass to charge ratio of analytes; which include intact proteins and protein complexes,[3,4] fragment ions of protein ions,[5-7] peptides produced by enzymatically digested proteins, and fragment ions of selected peptide ions.[8-10] The application of mass spectrometry and MS/MS to proteomics takes advantage of the huge amount of genomic and proteomic data stored in databases.[11]

Proteome complexity usually exceeds the analytical capacity of instrumentation. One solution to this problem is to use higher resolution mass spectrometers, i.e. to "put everything in the mass spectrometer, analyze it, and let acquired data sorted out". However, this method is very costly and also ignores the fact that the difficulty of getting peptides into the gas phase increases as sample complexity increases due to different ionization efficiencies, which is known in that relatively non-polar peptides in mixtures suppress the ionization of polar peptides in electrospray ionization (ESI) mass spectrometry.[12] In matrix assisted laser desorption ionization (MALDI) mass spectrometry, positively charged species suppress the ionization of neutral and negatively charged peptides.[13] Ultra-high resolution instruments are useless if the requisite proteins and peptides do not enter the mass spectrometer. Another solution is that complex components should be fractionated with much higher resolution before entering the mass spectrometer.[14] Even though this strategy would diminish matrix effects, it still needs ultra-sophisticated separation systems with huge numbers of theoretical plates and an accompanying long analysis time in sorting huge spectral data from 10^5 to 10^6 components to find a small number of relevant peptides.[15]

In both of these approaches, one needs to examine an entire proteome to answer biological questions, but this will not be necessary in many cases. Structure specific affinity selection can be used as a way to look at a portion of a particular proteome. This is a targeted strategy. A targeted strategy has been used in several different ways. One is to derivatize a specific functional group in proteins with an affinity selectable reagent that can be targeted with an affinity chromatography column. Avidin selection of proteins or peptides in which sulfhydryl or carbonyl groups have been derivatized with biotin is such an example.[16,17] A biotin residue (attached through a linker) permits the selective capture of the cysteine-containing peptides out of

*Reprint requests to Wonryeon Cho
E-mail: wcho@wku.ac.kr

the digest resulting in a significant reduction of the complexity of the sample through an avidin-biotin interaction.[18] Another approach is to directly select proteins or peptides based on shared structural features; for instance, direct selection of histidine containing peptides with copper loaded immobilized metal affinity chromatography (Cu^{2+}-IMAC) columns.[19] Selection of phosphorylated peptides with either Fe^{3+}-IMAC or Ga^{3+}-IMAC columns is another example.[18] An additional powerful approach is to select a subset of proteins which have a common post-translational modification related to the diseases being examined, such as lectins and glycan targeting antibodies for specific protein glycosylation.[20-26] If combined together, there could be more than a dozen ways to target structural features in proteomics.[27]

Protein structure plays an important role in disease progression and has been reported with post translational modifications.[21-24,28,29] In case of glycosylation associated with diseases, targeting disease-associated glycans would be an effective way to select and identify disease-related glycoproteins in complex matrices such as plasma.[30] Based on this idea, lectins and glycan-targeting antibodies have been used as tools in capturing glycosylated proteins.[20-25,31-32] This study shows a way to differentiate glycan diversity in the same protein in a serial affinity column set (SACS) using four *Lycopersicon esculentum* lectin (LEL) columns.

Materials and Methods

Materials and Chemicals

Agarose-bound *Lycopersicon esculentum* lectin (LEL) sorbent was purchased from Vector Laboratories (Burlingame, CA, U.S.A.). Normal pooled plasma from 100 human subjects was generously supplied by the National Institute of Standards and Technology (NIST, Gaithersburg, MD, U.S.A.). Acetic acid, sodium hydroxide, formic acid, calcium chloride, magnesium chloride, and HPLC grade acetonitrile (ACN) were purchased from Mallinckrodt Chemicals (Phillipsburg, NJ, U.S.A.). Ammonium bicarbonate, glycine, manganese chloride, proteomics grade N-*p*-tosyl-phenylalanine chloromethyl ketone (TPCK)-treated trypsin, 4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid (HEPES), iodoacetic acid (IAA), and L-cysteine were obtained from Sigma-Aldrich (St. Louis, MO, U.S.A.). Dithiothreitol (DTT) and urea were provided by Bio-Rad Laboratories (Hercules, CA, U.S.A.). PNGase F (glycerol free) was purchased from New England BioLabs (Ipswich, MA, U.S.A.). The C18 microspin column was obtained from The Nest Group, Inc. (Southborough, MA, U.S.A.). HLB Oasis SPE cartridges were provided by Waters (Milford, MA, U.S.A.). The DI water system was purchased from Millipore (Boston, MA, U.S.A.). Centrivap Concentrator was purchased from Labconco, Corp. (Kansas City, MO, U.S.A.).

Serial Affinity Column Set (SACS)

Agarose-bound LEL sorbent was individually self-packed in four 4.6 mm × 50 mm columns. Protein concentration in healthy human plasma was estimated using the Bradford assay to prepare samples with equal amounts of total proteins. For the SACS experiment, the four LEL columns were connected in series. Healthy human plasma was loaded directly onto serially connected soft-gel affinity columns with mobile phase A (0.10 M HEPES buffer, pH 7.5 containing 1 mM $CaCl_2$ and 1 mM $MgCl_2$) at a flow rate of 0.3 mL/min. Following extensive washing with mobile phase A to remove nonspecifically and weakly bound proteins, the serial LEL affinity columns were disassembled and affinity-selected proteins were eluted from each column individually with solution B (0.5 M acetic acid–HCl solution, pH 2.5). All the elution curves were obtained with an absorbance detector operating at 280 nm using 20 AD LC systems from Shimadzu Scientific Instruments, Inc. (Kyoto, Japan).

Saturation Test of Columns

A saturation test was done with an LEL column based on previous works.[24,26-30] The amount of injected sample was increased from 25 μg to 250 μg by 25 μg increments. The saturation amount of healthy human plasma samples was 200 μg of human plasma in the LEL column (data not shown). Therefore, the maximum injection amount was decided as 200 μg of plasma below the saturation amount in order to prevent overloading.

Proteolysis

Captured proteins were desalted using HLB Oasis SPE cartridges, dried with a centrivap concentrator after being adjusted to pH 7.5 with a 0.5 M NH_4HCO_3 buffer, and were then reconstituted with 8 M urea in a 50 mM HEPES buffer containing 10 mM $CaCl_2$. The denatured proteins were reduced with 10 mM DTT. After two hours of incubation at 50°C, iodoacetic acid was added to a final concentration of 20 mM for alkylation and incubated in darkness for additional two hours. L-cysteine was then added to the reaction mixture to a final concentration of 40 mM, and the mixture was incubated for 30 minutes at room temperature. After dilution with a 50 mM HEPES buffer to a final urea concentration of 1 M, proteomics grade trypsin (2%, w/w, enzyme to protein) was added and incubated overnight at 37°C. The proteolysis reaction was stopped by the addition of TLCK (trypsin/TLCK ratio of 1:1 (w/w)). The resulting peptide mixture was desalted with HLB Oasis SPE cartridges and dried into 50-100 μL with the centrivap concentrator after being adjusted to pH 7.5 with the 0.5 M NH_4HCO_3 buffer for PNGase F digestion.

PNGase F Digestion

N-Linked glycopeptides in a digested peptide mixture were deglycosylated by treatment with PNGase F. A

50 mM ammonium bicarbonate buffer was added to the desalted tryptic peptides to adjust to pH 7.0-8.0. Five microliters (2500 U) of PNGase F was added to the pH-adjusted tryptic peptides and then the mixture was incubated overnight at 37°C. Following deglycosylation, samples were desalted using HLB Oasis SPE cartridges and concentrated using a C18 microspin column. The PNGase F-treated peptide mixtures were reconstituted in 0.1% formic acid solution and stored at -80°C until analysis with an LTQ-Orbitrap instrument.

LC-MS/MS-Based Protein Identification

Proteins were identified with an LTQ-Orbitrap hybrid mass spectrometer (Thermo Fisher, Bremen, Germany). The peptide mixtures resulting from PNGase F treatment were separated on an Agilent 1100 HPLC system using a 75 μm × 120 mm C18 Reversed-phase Chromatography (RPC) column packed with 5 μm C18 Magic beads. RPC separations were achieved using a 60 min linear mobile phase gradient from 98% solvent A with 2% solvent B to 60% solvent A with 40% solvent B at a flow rate of 300 nL/min. Solvent A was composed of DI water to which formic acid had been added to a concentration of 0.1%. Solvent B was prepared with ACN to which formic acid had been added to a concentration of 0.1%. The electrospray ionization emitter tip was generated on the prepacked column with a laser puller (model P-2000, Sutter Instrument Co.). The HPLC system was coupled directly to the LTQ-Orbitrap hybrid mass spectrometer equipped with a nanoelectrospray ion source (Proxeon Biosystems, Odense, Denmark). The MS was operated in the data-dependent mode, in which a survey full scan MS spectrum (from m/z 300 to 1600) was acquired in the Orbitrap with a resolution of 60,000 at m/z 400. This was then followed by MS/MS scans of the three most abundant ions with +2 and +3 charge states. Target ions already selected for MS/MS were dynamically excluded for 180s. The resulting fragment ions were recorded in the linear ion trap.

Automated MS/MS data analysis was performed utilizing Protein Pilot software 5.0 using the Pro Group™ algorithm (SCIEX, Concord, Ontario, Canada) for protein identification. The minimum acceptance criterion for peptide identification was a 99% confidence level. Most of the proteins were identified based on the presence of at least two peptides from a protein identified by the Pro Group™ algorithm at the 99% confidence level. An unused score cutoff of 4 was the minimum value for identifying proteins with the Protein Pilot 5.0 Software. Proteins are listed according to their Swiss-Prot entry names and accession numbers.

Results and Discussion

Glycoproteomics enables comprehensive analysis of glycoproteins using proteomics tools. There are two approaches to recognize disease-associated aberrations of glycans in glycoproteomics. One is to do affinity selection of glycopeptides from trypsin digests of plasma. The selected peptides are identified by RPC-MS/MS after removal of the glycans from glycopeptides with N-glycosidase F (PNGase F). An attractive characteristic of this approach is that proteins and glycosylation sites are identified simultaneously.[33]

The other strategy is to carry out affinity selection of glycoproteins from plasma at the protein level. Then the selected glycoproteins are identified in two ways (Figure 1). At first, the glycoproteins are trypsin digested immediately and the produced peptides are analyzed directly by MALDI-MS/MS or by ESI-MS/MS after RPC for the identification of proteins. The other method is to fractionate glycoproteins further by RPC or SDS-PAGE before proteolysis and then analyze with RPC-MS/MS for protein identification.[34-35]

Affinity selection of glycoproteins at the protein level has a great advantage. In general affinity selection of glycopeptides at peptide level requires deglycosylation for identification. Although that is relatively easy with N-glycosylation proteins, it is difficult to deglycosylate and identify O-glycosylated peptides. In contrast, in affinity selection at the protein level, affinity selected N- and O-glycosylated proteins are easily identified by their non-glycosylated peptides.[20,25] The number of non-glycosylated peptides are usually about 50 times more than that of glycosylated peptides.

An interesting feature of glycoproteins is that it may contain multiple lectin- or antibody-targetable structures within an oligosaccharide at a single or several sites. When lectin or glycan targeting antibody affinity columns are connected in series, loaded, and then eluted separately, this method is referred to as serial affinity chromatography (SAC). It can determine whether individual glycan structures appear alone or together with other glycans in

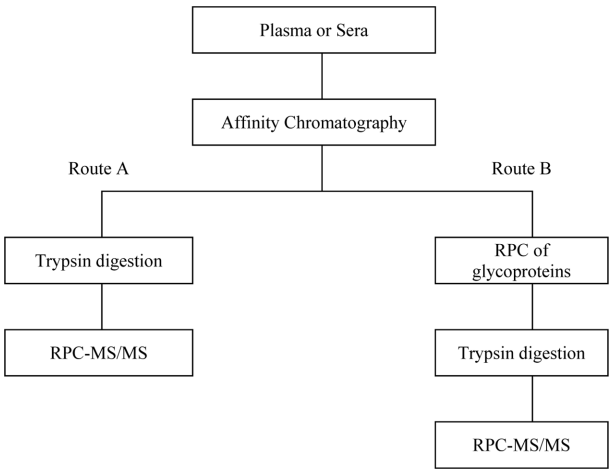

Figure 1. Two strategies of protein identification in protein level affinity selection.

Table 1. Glycoproteins captured from two non-serial LEL columns in the LEL1→LEL2→LEL3→LEL4 serial affinity column set (SACS).

No.	LEL 1	LEL 2	LEL 3	LEL 4	Swiss-Prot Accession no. and entry name	Names	Type of Glycosylation
1	O			O	P01877 |IGHA2_HUMAN	Ig alpha-2 chain C region	N
2	O			O	P04220 |MUCB_HUMAN	Ig mu heavy chain disease protein	N
3		O		O	P22792 |CPN2_HUMAN	Carboxypeptidase N subunit 2	N
4		O		O	P18428 |LBP_HUMAN	Lipopolysaccharide-binding protein	N

specific glycoproteins using different types of affinity columns in different orders under the protein level selection. It also elucidates that there is independent occurrence of different affinity targetable glycan features in the same glycoprotein, and multiple targetable glycan features are co-resident in the same glycoprotein, i.e. the heterogeneity of the glycan structures at different glycosylation sites. SAC is able to assess whether serial lectin and glycan-targeting antibody affinity chromatography is of any value in assessing glycan occurrence and heterogeneity among glycoproteins in blood. The major advantage of this approach is that it can determine whether a targeted glycan appearing at one site in an affinity selected protein will be present with other glycans at other sites in the same protein.[20,25] This was achieved using LEL, a lectin widely used in monitoring cancer. LEL binds specifically to β-(1 → 4)-linked galactosyl residues in poly-N-acetyllactosamine and GlcNAc residues in N-acetylglucosamine $(GlcNAc)_n$, where n ranges from 1 to 4. The structure and number of residues in a glycan can play a role in affinity selection. For example, in the series $(GlcNAc\ \beta_{1-4})_1$, $(GlcNAc\ \beta_{1-4})_2$, $(GlcNAc\ \beta_{1-4})_3$, to $(GlcNAc\ \beta_{1-4})_4$, the binding affinity of LEL from tomato increases with the size and number of the N-acetylglucosamine oligomers bound to a glycan.[22] This allows LEL to differentiate between glycoforms.[21]

In this study, four small identical LEL columns were coupled in series to form a single affinity chromatography system with the last column in the series connected to an absorbance detector. The serial affinity column set (SACS)[25] was then loaded with human plasma proteins. At the completion of loading, the column set was disassembled, the four columns were eluted individually, the captured proteins were trypsin digested, the peptides were deglycosylated with PNGase F, and the parent proteins were identified through mass spectral analyses.

Two proteins were found to be bound to LEL1 and also to LEL4, while other two proteins bound to LEL2 and LEL4 in Table 1. This is very important. Failure to bind sequentially suggests that proteins bound to (LEL1 and LEL4) or (LEL2 and LEL4) probably carry glycans that differ in their structure and LEL binding affinity. These proteins would lead to glycoproteins being bound non-sequentially in the LEL train.

In this way, SAC can determine glycan diversity of a protein; the types of glycan structures which the protein has. For example, there might be glycan diversity such as a single glycan (A, or B) or A-B containing protein isoforms such as:

In peptide-affinity SAC, it is not possible to differentiate these three glycoprotein isoforms, because after digestion, the protein sequence is destroyed. As a result, even though both of AB and BA sequence of glycan targeting serial affinity selectors are used separately, there is no difference in peptide captures between two serial selectors. However, in protein-affinity SAC, when AB sequence of glycan targeting serial affinity selectors are used, the B targeting affinity selector cannot capture the following A-B glycan structure,  because this A-B glycan structure has already been stripped off by the A targeting affinity selector. On the other hand, when BA sequence of glycan targeting serial affinity selector is used, the A targeting affinity selector cannot capture this A-B glycan structure because this A-B glycan structure has been already stripped off by the B targeting affinity selector. By comparing these two results, a type of protein heterogeneity can be elucidated by this SAC.

Conclusion

In conclusion, targeted proteomics is an effective way in proteomics to save time and effort enormously and SAC and SACS suggest ways to differentiate glycan diversity and structural analysis on post-translationally modified proteins using various affinity selectors.

Acknowledgments

Authors acknowledge the support of Wonkwang University in 2012.

References

1. Gajadhar, A. S.; White, F. M. *Curr. Opin. Biotechnol.* **2014**, 28, 83.
2. Pandey, A.; Mann, M. *Nature* **2000**, 405, 837.
3. Breuker, K.; Jin, M.; Han, X.; Jiang, H.; McLafferty, F. W. *J. Am. Soc. Mass Spectrom.* **2008**, 19, 1045.
4. Collier, T. S.; Hawkridge, A. M.; Georgianna, D. R.; Payne, G. A.; Muddiman, D. C. *Anal. Chem.* **2008**, 80, 4994.
5. Armirotti, A.; Benatti, U.; Damonte, G. *Rapid Commun. Mass Spectrom.* **2009**, 23, 661.
6. McLafferty, F. W.; Breuker, K.; Jin, M.; Han, X.; Infusini, G.; Jiang, H.; Kong, X.; Begley, T. P. *FEBS J.* **2007**, 274, 6256.
7. Macek, B.; Waanders, L. F.; Olsen, J. V.; Mann, M. *Mol. Cell. Proteomics* **2006**, 5, 949.
8. Dai, S.; Jia, Y.; Wu, S.-L.; Isenberg, J. S.; Ridnour, L. A.; Bandle, R. W.; Wink, D. A.; Roberts, D. D.; Karger, B. L. *J. Proteome Res.* **2008**, 7, 4384.
9. Benton, H. P.; Wong, D. M.; Trauger, S. A.; Siuzdak, G. *Anal. Chem.* **2008**, 80, 6382.
10. Getie-Kebtie, M.; Franke, P.; Aksamit, R.; Alterman, M. A. *J. Proteome Res.* **2008**, 7, 3697.
11. Goudenege, S.; Dargelos, E.; Claverol, S.; Bonneu, M.; Cottin, P.; Poussard, S. *Proteomics* **2007**, 7, 3289.
12. Cech, N. B.; Enke, C. G. *Mass Spectrom. Rev.* **2001**, 20, 362.
13. Krause, E.; Wenschuh, H.; Jungblut, P. R. *Anal. Chem.* **1999**, 71, 4160.
14. Smith, R. D. *Int. J. Mass. Spectrom.* **2000**, 200, 509.
15. Mellors, J. S.; Gorbounov, V.; Ramsey, R. S.; Ramsey, J. M. *Anal. Chem.* **2008**, 80, 6881.
16. Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. *Nat. Biotech.* **1999**, 17, 994.
17. Yoo, B.-S.; Regnier, F. E. *Electrophoresis* **2004**, 25, 1334.
18. Thiellement, H.; Zivy, M.; Damerval, C.; Méchin, V. *Plant Proteomics*, Humana Press, **2007**.
19. Ren, D.; Penner, N. A.; Slentz, B. E.; Mirzaei, H.; Regnier, F. *J. Proteome Res.* **2003**, 2, 321.
20. Jung, K.; Cho, W. *Anal. Chem.* **2013**, 85, 7125.
21. Cho, W.; Jung, K.; Regnier, F. E. *Anal. Chem.* **2008**, 80, 5286.
22. Cho, W.; Jung, K.; Regnier, F. E. *J. Proteome Res.* **2010**, 9, 5960.
23. Drake, P. M.; Schilling, B.; Niles, R. K.; Prakobphol, A.; Li, B.; Jung, K.; Cho, W.; Braten, M.; Inerowicz, H. D.; Williams, K.; Albertolle, M.; Held, J. M.; Iacovides, D.; Sorensen, D. J.; Griffith, O. L.; Johansen, E.; Zawadzka, A. M.; Cusack, M. P.; Allen, S.; Gormley, M.; Hall, S. C.; Witkowska, H. E.; Gray, J. W.; Regnier, F.; Gibson, B. W.; Fisher, S. J. *J. Proteome Res.* **2012**, 11, 2508.
24. Jung, K.; Cho, W.; Regnier, F. E. *J. Proteome Res.* **2008**, 8, 643.
25. Cho, W. *Anal. Chem.* **2015**, 87, 9612.
26. Dwek, M. V.; Ross, H. A.; Leathem, A. J. C. *Proteomics* **2001**, 1, 756.
27. Mirzaei, H.; Regnier, F. *J. Chromatogr. B* **2005**, 817, 23.
28. Gong, C. X.; Liu, F.; Grundke-Iqbal, I.; Iqbal, K. *J. Neural Transm.* **2005**, 112, 813.
29. Rademacher, T. W.; Parekh, R. B.; Dwek, R. A. *Annu. Rev. Biochem.* **1988**, 57, 785..
30. Nilsson, C. L. *Lectins: Analytical Technologies*, Elsevier, **2007**.
31. Julka, S.; Regnier, F. *J. Proteome Res.* **2004**, 3, 350.
32. Ueda, K.; Fukase, Y.; Katagiri, T.; Ishikawa, N.; Irie, S.; Sato, T.-A.; Ito, H.; Nakayama, H.; Miyagi, Y.; Tsuchiya, E.; Kohno, N.; Shiwa, M.; Nakamura, Y.; Daigo, Y. *Proteomics* **2009**, 9, 2182.
33. Heo, S.-H.; Lee, S.-J.; Ryoo, H.-M.; Park, J.-Y.; Cho, J.-Y. *Proteomics* **2007**, 7, 4292.
34. Larsen, M. R.; Højrup, P.; Roepstorff, P. *Mol. Cell. Proteomics* **2005**, 4, 107.
35. Issaq, H. J.; Veenstra, T. D. *Proteomic and Metabolomic Approaches to Biomarker Discovery*, Elsevier, **2013**.