Predicting the Retention Behavior of Specific O-Linked Glycopeptides

Majors J. Badgett,¹ Barry Boyes,^{1,2} and Ron Orlando^{1,*}

¹Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia 30602, USA; and ²Advanced Materials Technology, Wilmington, Delaware 19810, USA

O-Linked glycosylation is a common post-translational modification that can alter the overall structure, polarity, and function of proteins. Reverse-phase (RP) chromatography is the most common chromatographic approach to analyze O-glycosylated peptides and their unmodified counterparts, even though this approach often does not provide adequate separation of these two species. Hydrophilic interaction liquid chromatography (HILIC) can be a solution to this problem, as the polar glycan interacts with the polar stationary phase and potentially offers the ability to resolve the peptide from its modified form(s). In this paper, HILIC is used to separate peptides with O-N-acetylgalactosamine (O-GalNAc), O-N-acetylglucosamine (O-GlcNAc), and O-fucose additions from their native forms, and coefficients representing the extent of hydrophilicity were derived using linear regression analysis as a means to predict the retention times of peptides with these modifications.

KEY WORDS: Hydrophilic Interaction Chromatography, HILIC, LC-MS

INTRODUCTION

The attachment of a single monosaccharide to the side chain of serine and threonine residues has been found to play a wide range of biologic functions, including inflammation response, lubrication, or protein stability.^{1, 2} Two examples of glycans analyzed in the present study are O-GlcNAc and O-GalNAc. These isomeric monosaccharides are attached to serine or threonine residues in proteins and play important roles in numerous biologic processes. O-GlcNAc is involved in enzymatic activity, protein function, and disease-relevant signaling and has been found on over 1000 proteins.^{3–6} It can be added or removed rapidly from target proteins as a result of the presence of O-GlcNAc transferase and O-GlcNAcase, and the cycling of O-GlcNAc structures on proteins is a result of environmental changes, including many forms of stress.^{7–9} Abnormal O-GlcNAc modification is associated with numerous diseases, such as Alzheimer's disease, Huntington's disease, diabetes, systematic lupus erythematosus, and Parkinson's disease, as well as many types of cancer.^{3, 5, 6, 8, 10, 11} O-GalNAc modification is referred to as the Tn antigen and is often found on mucins or

mucin-like proteins produced in tumor cells. This modification is present in >10% of human proteins and over 85% of proteins passing through the Golgi apparatus.¹² In contrast to O-GlcNAc, O-GalNAc modification is initiated by 20 GalNAc transferases that control where the site of glycan attachment is in proteins.^{1, 12–15} However, similarly to O-GlcNAc, aberrant modification of O-GalNAc is associated with a wide range of diseases, such as acute coronary disease, congenital heart disease, various types of cancer, and many others.^{13, 15–17}

Another monosaccharide that is found attached to proteins through an O-linkage to a serine or threonine residue is fucose, commonly known as O-fucosylation. This modification was first discovered by Hallgren *et al.*¹⁸ in 1975 and is commonly found on epidermal growth factor-like repeats, as well as thrombospondin type-1 repeats.^{19–21} Similarly to O-GalNAcylation, other monosaccharides are frequently added to the core O-fucose, and this elongation can be critical for protein activity.²² The attachment of O-linked fucose to proteins is catalyzed by 11 different fucosyltransferases that have guanosine diphosphate-fucose as the donor substrate, and these fucosyltransferases most commonly reside in the Golgi apparatus but have also been found in the endoplasmic reticulum.^{19, 23–26} Although fucose is found to play an essential role in physiologic processes, such as blood type determination and Notch signaling, as well as other signal transduction processes, it is linked to many different cancers, rheumatoid arthritis, cystic fibrosis, and leukocyte adhesion deficiency type II, among others.^{20, 23, 25, 27–35}

The ability to resolve glycosylated peptides from their native forms would greatly facilitate glycoproteomics, as this

*ADDRESS CORRESPONDENCE TO: Ron Orlando, Complex Carbohydrate Research Center, University of Georgia, 315 Riverbend Rd., Athens, Georgia 30602, USA (Phone: 706-542-4412; Fax: 706-542-4412; E-mail: orlando@ccrc.uga.edu)

Abbreviations: GU=glucose unit, HILIC=hydrophilic interaction liquid chromatography, LC=liquid chromatography, MS=mass spectrometry, O-GalNAc=O-N-acetylgalactosamine, O-GlcNAc=O-N-acetylglucosamine, RP=reverse-phase

doi: 10.7171/jbt.17-2803-003

would enable glycan characterization while connected to a peptide tag that provides the glycan's location. RP-liquid chromatography (LC) is typically used in the analysis of peptide/glycopeptide mixtures, where hydrophobic interaction drives the retention. Hence, the hydrophilic glycans, which do not interact significantly with the RP stationary phase, play little to no role in the separation. This behavior makes it difficult to resolve glycopeptides from their unmodified counterparts by RP, which is particularly true when the glycan consists of a single monosaccharide. Alternatively, glycans interact extensively with the stationary phases used in HILIC. Therefore, HILIC should be capable of resolving these peptide/glycopeptide pairs.

We have previously devised a model that predicts peptide retention using HILIC that provides amino acid coefficients describing their hydrophilicity and are currently devising a model that predicts the retention of glycopeptides in HILIC separations.³⁶ Dextran is used as a retention calibrant, enabling peptide and glycopeptide retention to be expressed in glucose units (GUs) that permit a comparison of peptide retention across different LC-mass spectrometry (MS) systems. The combination of retention models of HILIC separations of peptides and glycans is directed to the prediction of retention of previously uninvestigated glycopeptides. Here, we show that glycopeptides can be resolved from the native peptide, and retention coefficients that represent the hydrophilicity of *O*-GlcNAc, *O*-GalNAc, and *O*-fucose are derived. This will permit a more straightforward determination of the extent and type of glycosylation for the glycans analyzed in this study.

MATERIALS AND METHODS

Glycopeptides

Synthetic peptide–glycopeptide pairs were obtained from GlycoScientific (Athens, GA, USA).

LC-MS/MS settings and instrumentation

Data were acquired using a Finnigan LTQ (Thermo Fisher Scientific, Waltham, MA, USA) in series with an 1100 Series Capillary LC system (Agilent Technologies, Santa Clara, CA, USA) with an electrospray ionization source that used spray tips made in-house. Samples were suspended in 25% H₂O, 75% acetonitrile, and 0.1% formic acid (Sigma-Aldrich, St. Louis, MO, USA) for direct injection into the LC system. Samples were separated using a 200 μ m $\times$ 150 mm HALO Penta-HILIC column packed with 2.7 μ m-diameter superficially porous particles (Advanced Materials Technology, Wilmington, DE, USA) at room temperature. The gradient elution conditions used a linear increase in aqueous solvent from 5 to 70%, over 90 min at a 2 μ l/min flow rate, using the column at room temperature. The

mobile phases contained 0.1% formic acid (Sigma-Aldrich) in acetonitrile, and the (strong) aqueous solvent contained 50 mM ammonium formate (Thermo Fisher Scientific). The settings for the mass spectrometer included taking the 5 most intense ions from each full mass spectrum for fragmentation using collision-induced dissociation, and the resulting MS/MS spectra were recorded.

Selection of peptides for prediction model and post-run data analysis

Retention times were determined manually from RAW files using the apex of the peaks displayed in Xcalibur software (Thermo Fisher Scientific), and resulting MS/MS data were visually inspected for fragmentation that was consistent with peptide assignments. Glycopeptide retention times in minutes were converted to GUs based on Dextran samples that were run immediately before. Linear regression analysis using StatPlus (AnalystSoft, Walnut, CA, USA) was used to find retention coefficients for the modifications.

RESULTS AND DISCUSSION

The separation of glycopeptides from their native forms

The ability to resolve the native peptide from the species with a single monosaccharide is shown by the HILIC chromatogram of the peptide GTTPSPVPTTSTTSAP, with the underlined amino acid residues representing the sites of *O*-GalNAcylation (**Fig. 1**). The unmodified version of the peptide elutes first, followed by the peptide with only 1 modification at the third threonine residue, followed by the peptide with both modifications at the third and 13th residues. All of the peaks are baseline resolved, demonstrating the ability of the HILIC column to separate the species with the hydrophilic modification from the native peptide.

FIGURE 1

The separation of the *O*-GalNAcyated peptide GTTPSPVPTTSTTSAP, where the earliest eluting peak is unmodified, the middle peak is modified at the third residue only, and the most retained peak has both modifications.

The difference between the unmodified peptide and singly modified peptide in this figure is 5.28 min, which is higher than the difference between the singly modified peptide and doubly modified peptide at 3.70 min. Taking into account the additional replicates, the average difference between the first 2 peaks is 1.58 min larger than the average difference between the last 2 peaks. This suggests that the addition of 1 sugar residue drives the interaction with the water-rich layer of the stationary phase and that the addition of the second moiety does not change the overall hydrophilicity of the peptide as much as the first. It is interesting to note that the position of the sugar does not have an effect on glycopeptide retention, as there was no retention time difference between the peptide glycosylated on the third residue and the species with the sugar on the 13th residue.

The separation of the *O*-GlcNAcylated peptide VPTTAASTPDAVDK is shown in **Fig. 2**. Similarly to the example discussed in the preceding paragraph, the modified form elutes later than the native form as a result of the hydrophilicity of the modification. These peaks are also baseline resolved. As a result of the near-identical structures of these 2 isomers (*O*-GlcNAc to *O*-GalNAc), the retention time difference between peptide and glycopeptide is analogous to the native and singly modified species discussed above.

O-Fucosylation is a deoxyhexose and thus is a slightly less hydrophilic monosaccharide compared with the 2 *N*-acetylated hexosamines discussed to this point. **Figure 3** shows the baseline separation of the native and modified forms of the peptide CQNGGTCHNTH, with the *O*-fucosylated form eluting later. However, the retention time difference of this peptide–glycopeptide pair is smaller than the differences observed with the peptides modified with *O*-GlcNAc or *O*-GalNAc. This is expected; as fucose lacks the amide and acetyl groups possessed by *O*-GlcNAc and *O*-GalNAc, these functional groups are both hydrophilic and thus increase the HILIC retention.

FIGURE 2

The separation of the *O*-GlcNAcylated peptide VPTTAASTPDAVDK, where the earlier peak is the native version of the peptide, and the later peak is the modified version.

FIGURE 3

The separation of the *O*-fucosylated peptide CQNGGTCHNTH, where the earlier peak is the native version of the peptide, and the later peak is the modified version.

These 3 examples show that HILIC is not only capable of separating *O*-glycosylated peptides from their native counterparts but that it can baseline resolve the peaks, allowing for quantitation and facilitating quicker and easier identifications.

Retention time prediction of *O*-glycans

The prediction of retention for peptides with these modifications can further the identification process. All of the retention times of the native and modified peptides over the course of 4 runs are shown in **Table 1**. Their retention times in minutes were converted to GUs from procainamide-labeled Dextran samples run immediately before and after the protein samples. This enables the comparison of retention across LC-MS systems and provides a retention time calibrant for these samples. The differences in retention times from run to run are miniscule, indicating the reproducibility and potential predictability of retention using the HILIC column. SDs from the 4 runs for each sample are also low, with the average SD at 0.094 GU or 0.357 min. This is extremely precise for a method that uses a 90 min long LC-MS gradient at a 2 μ L/min flow rate.

To analyze how the modifications are influencing the overall hydrophilicity of the peptides, linear regression analysis of the retention times of peptides with and without modification was used to derive coefficients (**Table 2**). These coefficients represent the hydrophilicity of the modification, which are almost as great as the most hydrophilic amino acids at a low pH using the same HILIC column: histidine, lysine, and arginine.³⁶ The coefficients are also displayed in GU, and their high coefficient values were expected as a result of the size and polar characteristics of the sugar additions. *O*-GlcNAc and *O*-GalNAc only differ by the position of some of their hydroxyl groups, so the isomeric glycans should have similar coefficients, and are consistent with the experimental observations. However, *O*-fucose is less hydrophilic than the 2 *N*-acetylated

TABLE 1

Peptide	Run 1 RT, GU	Run 2 RT, GU	Run 3 RT, GU	Run 4 RT, GU	Avg. RT, GU	SD
VPTTAASTPDAVDK	6.680	6.790	6.841	6.871	6.796	0.084
VPTTAASTPDAVDK	7.960	8.123	8.120	8.105	8.077	0.078
GTTPSPVPTTSTTSAP	5.111	5.026	5.174	5.174	5.121	0.070
GTTPSPVPTTSTTSAP	6.517	6.581	6.607	6.529	6.559	0.042
GTTPSPVPTTSTTSAP	6.517	6.492	6.563	6.483	6.514	0.036
GTTPSPVPTTSTTSAP	7.712	7.866	7.787	7.684	7.762	0.082
KRGRKESYSIYVYK	8.412	8.366	8.171	8.306	8.314	0.105
KRGRKESYSIYVYK	9.177	8.879	8.742	8.646	8.861	0.231
RGGVKRISGLIYEE	3.412	3.226	3.341	3.421	3.350	0.090
RGGVKRISGLIYEE	6.167	3.968	4.068	4.241	4.111	0.119
CKLLGRVTIAQGG	4.352	4.434	4.405	4.438	4.407	0.040
CKRGRKESYSIYVYK	7.877	7.956	7.867	7.916	7.904	0.041
CKRGRKESYSIYVYK	7.105	7.272	7.167	6.832	7.094	0.188
CELAHAVSEGTKA	6.821	6.982	6.852	7.011	6.916	0.094
CKSAPATGGVKK	7.191	7.312	7.275	7.360	7.284	0.072
CKSAPATGGVKK	6.480	6.614	6.584	6.661	6.585	0.076
CFNGGTCVDGIN	2.794	2.781	2.738	2.576	2.722	0.100
CFNGGTCVDGIN	3.322	3.130	3.314	3.386	3.288	0.110
CQNGGTCHNTH	6.073	6.175	6.124	6.180	6.138	0.050
CQNGGTCHNTH	6.821	7.022	6.940	6.898	6.920	0.084

Amino acids in bold represent the site(s) of modification.

hexosamines, which is reflected in this monosaccharide having the lowest coefficient for monosaccharide additions studied here.

It is important to note that the 3 peptides—KRGRKESYSIYVYK, RGGVKRISGLIYEE, and CKRGRKESYSIYVYK—that have *O*-GlcNAc modifications on a serine residue have an average retention time difference of 0.706 GU from their unmodified forms, whereas the 2 peptides—VPTTAASTPDAVDK and CKSAPATGGVKK—that have an *O*-GlcNAc modification on a threonine residue have an average retention time difference of 0.990 GU from their native forms. Whereas this could be attributed to the difference in amino acid structures, the theoretical difference between the coefficients for serine and threonine derived in our HILIC peptide retention prediction model (0.334 and 0.357, respectively) suggests that there should be a minimal change in retention. We are at a loss to explain the 0.2 GU difference between *O*-GlcNAc modifications on a serine or threonine residue. However, this difference is fairly small and allows the retention of these peptides to be predicted within ~30 s in a 90 min gradient, which we feel is acceptable.

SUMMARY

The attachment of a single monosaccharide to a peptide increased the overall hydrophilicity and causes the glycosylated

species to be more retained on an HILIC column. The shifts in HILIC retention caused by the addition of these monosaccharides were determined to be compared with amino acid coefficients that were the most hydrophilic in a previously made peptide-retention prediction model. This work showed the ability of the HILIC column to separate these peptide/glycopeptide pairs reproducibly and in a predictable manner, which allows easier identification, characterization, and quantification. Whereas there were no major differences between *O*-GalNAc and *O*-GlcNAc, *O*-fucosylated peptides were slightly less hydrophilic.

TABLE 2

Modification/amino acid residue	Coefficient, GU
<i>O</i> -GlcNAcylation	1.637
<i>O</i> -GalNAcylation	1.758
<i>O</i> -Fucosylation	1.438
Histidine (H)	1.564
Lysine (K)	2.121
Arginine (R)	1.828

ACKNOWLEDGMENTS

Support for this work comes from the U.S. National Institutes of Health (NIH) Grant R44GM093747 to B.B. and NIH Grants R41GM119879 and R41GM1040631 to R.O.

DISCLOSURE

The authors declare no conflicts of interest.

REFERENCES

1. Hang HC, Bertozzi CR. The chemistry and biology of mucin-type O-linked glycosylation. *Bioorg Med Chem* 2005;13:5021–5034.
2. Van den Steen P, Rudd PM, Dwek RA, Opdenakker G. Concepts and principles of O-linked glycosylation. *Crit Rev Biochem Mol Biol* 1998;33:151–208.
3. Bond MR, Hanover JA. A little sugar goes a long way: the cell biology of O-GlcNAc. *J Cell Biol* 2015;208:869–880.
4. Hart GW, Akimoto Y. The O-GlcNAc modification. In *Essentials of Glycobiology* (Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, Hart GW, Etzler ME, eds), 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1999.
5. Ma J, Hart GW. O-GlcNAc profiling: from proteins to proteomes. *Clin Proteomics* 2014;11:8.
6. Zhu Y, Shan X, Yuzwa SA, Vocadlo DJ. The emerging link between O-GlcNAc and Alzheimer disease. *J Biol Chem* 2014;289:34472–34481.
7. Kearse KP, Hart GW. Lymphocyte activation induces rapid changes in nuclear and cytoplasmic glycoproteins. *Proc Natl Acad Sci USA* 1991;88:1701–1705.
8. Slawson C, Copeland RJ, Hart GW. O-GlcNAc signaling: a metabolic link between diabetes and cancer? *Trends Biochem Sci* 2010;35:547–555.
9. Zeidan Q, Hart GW. The intersections between O-GlcNAcylation and phosphorylation: implications for multiple signaling pathways. *J Cell Sci* 2010;123:13–22.
10. Ngoh GA, Facundo HT, Zafir A, Jones SP. O-GlcNAc signaling in the cardiovascular system. *Circ Res* 2010;107:171–185.
11. Yuzwa SA, Shan X, Macauley MS, et al. Increasing O-GlcNAc slows neurodegeneration and stabilizes tau against aggregation. *Nat Chem Biol* 2012;8:393–399.
12. Gill DJ, Clausen H, Bard F. Location, location, location: new insights into O-GalNAc protein glycosylation. *Trends Cell Biol* 2011;21:149–158.
13. Bennett EP, Mandel U, Clausen H, Gerken TA, Fritz TA, Tabak LA. Control of mucin-type O-glycosylation: a classification of the polypeptide GalNAc-transferase gene family. *Glycobiology* 2012;22:736–756.
14. Joshi HJ, Steentoft C, Schjoldager K, Vakhrushev SY, Wandall HH, Clausen H. Protein O-GalNAc glycosylation: the most complex and differentially regulated PTM. In *Glycoscience: Biology and Medicine* (Taniguchi N, Endo T, Hart GW, Seeberger PH, Wong C-H, eds), Springer, Tokyo, Japan, 2015: 1049–1064.
15. Hanisch FG. O-Glycosylation of the mucin type. *Biol Chem* 2001;382:143–149.
16. O'Halloran AM, Patterson CC, Horan P, et al. Genetic polymorphisms in platelet-related proteins and coronary artery disease: investigation of candidate genes, including N-acetylgalactosaminyltransferase 4 (GALNT4) and sulphotransferase 1A1/2 (SULT1A1/2). *J Thromb Thrombolysis* 2009;27:175–184.
17. Fakhro KA, Choi M, Ware SM, et al. Rare copy number variations in congenital heart disease patients identify unique genes in left-right patterning. *Proc Natl Acad Sci USA* 2011;108:2915–2920.
18. Hallgren P, Lundblad A, Svensson S. A new type of carbohydrate-protein linkage in a glycopeptide from normal human urine. *J Biol Chem* 1975;250:5312–5314.
19. Luo Y, Koles K, Vorndam W, Haltiwanger RS, Panin VM. Protein O-fucosyltransferase 2 adds O-fucose to thrombospondin type 1 repeats. *J Biol Chem* 2006;281:9393–9399.
20. Mormann M, Maček B, De Peredo AG, Hofsteenge J, Peter-Katalinić J. Structural studies on protein O-fucosylation by electron capture dissociation. *Int J Mass Spectrom* 2004;234:11–21.
21. Wang Y, Lee GF, Kelley RF, Spellman MW. Identification of a GDP-L-fucose: polypeptide fucosyltransferase and enzymatic addition of O-linked fucose to EGF domains. *Glycobiology* 1996;6:837–842.
22. Serth K, Schuster-Gossler K, Kremmer E, Hansen B, Marohn-Köhn B, Gossler A. O-Fucosylation of DLL3 is required for its function during somitogenesis. *PLoS One* 2015;10:e0123776.
23. Luo Y, Haltiwanger RS. O-Fucosylation of Notch occurs in the endoplasmic reticulum. *J Biol Chem* 2005;280:11289–11294.
24. Moloney DJ, Lin AI, Haltiwanger RS. The O-linked fucose glycosylation pathway. Evidence for protein-specific elongation of O-linked fucose in Chinese hamster ovary cells. *J Biol Chem* 1997;272:19046–19050.
25. Luo Y, Nita-Lazar A, Haltiwanger RS. Two distinct pathways for O-fucosylation of epidermal growth factor-like or thrombospondin type 1 repeats. *J Biol Chem* 2006;281:9385–9392.
26. Okajima T, Xu A, Lei L, Irvine KD. Chaperone activity of protein O-fucosyltransferase 1 promotes Notch receptor folding. *Science* 2005;307:1599–1603.
27. Becker DJ, Lowe JB. Fucose: biosynthesis and biological function in mammals. *Glycobiology* 2003;13:41R–53R.
28. Ricketts LM, Dlugosz M, Luther KB, Haltiwanger RS, Majerus EM. O-Fucosylation is required for ADAMTS13 secretion. *J Biol Chem* 2007;282:17014–17023.
29. Müller J, Rana NA, Serth K, Kakuda S, Haltiwanger RS, Gossler A. O-Fucosylation of the Notch ligand mDLL1 by POFUT1 is dispensable for ligand function. *PLoS One* 2014;9:e88571.
30. Becker DJ, Lowe JB. Leukocyte adhesion deficiency type II. *Biochim Biophys Acta* 1999;1455:193–204.
31. Flögel M, Lauc G, Gornik I, Maček B. Fucosylation and galactosylation of IgG heavy chains differ between acute and remission phases of juvenile chronic arthritis. *Clin Chem Lab Med* 1998;36:99–102.
32. Gornik I, Maravić G, Dumić J, Flögel M, Lauc G. Fucosylation of IgG heavy chains is increased in rheumatoid arthritis. *Clin Biochem* 1999;32:605–608.
33. Kim ML, Chandrasekharan K, Glass M, et al. O-Fucosylation of muscle agrin determines its ability to cluster acetylcholine receptors. *Mol Cell Neurosci* 2008;39:452–464.
34. Orntoft TF, Vestergaard EM. Clinical aspects of altered glycosylation of glycoproteins in cancer. *Electrophoresis* 1999;20:362–371.
35. Scanlin TF, Glick MC. Terminal glycosylation in cystic fibrosis. *Biochim Biophys Acta* 1999;1455:241–253.
36. Badgett MJ, Boyes B, Orlando R. Prediction of peptide retention times in hydrophilic interaction liquid chromatography (HILIC) based on amino acid composition. *Chromatogr Today* 2015;8:39–42.