Published in final edited form as:

Electrophoresis. 2017 September ; 38(17): 2193–2199. doi:10.1002/elps.201600547.

Study of structure-dependent chromatographic behavior of glycopeptides using reversed phase nanoLC

Petr Kozlik^{1,2}, Radoslav Goldman^{1,3}, and Miloslav Sanda^{1,*}

¹Department of Oncology, Lombardi Comprehensive Cancer Center PSB GF9, Georgetown University, 3800 Reservoir Road NW, Washington, DC 20057, United States

²Department of Analytical Chemistry, Faculty of Science, Charles University, Hlavova 8, 128 43 Prague 2, Czech Republic

³Department of Biochemistry and Molecular & Cellular Biology, Georgetown University, 3800 Reservoir Road NW, Washington, DC 20057, United States

Abstract

Analysis of glycosylation is challenging due to micro- and macro-heterogeneity of the protein attachment. A combination of liquid chromatography with tandem mass spectrometry is one of the most powerful tools for glycopeptide analysis. In this work, we show the effect of various monosaccharide units on the retention time of glycopeptides. Retention behavior of several glycoforms of six peptides obtained from tryptic digest of haptoglobin, hemopexin and sex hormone binding globulin was studied on a reversed phase chromatographic column. We observed reduction of the retention time with increasing number of monosaccharide units of glycans attached to the same peptide backbone. Fucosylation of larger glycans provides less significant retention time shift than for smaller ones. Retention times of glycopeptides were expressed as relative retention times (RRT). These relative retention times were used for calculation of upper and lower limits of glycopeptide retention time windows under the reversed phase conditions. We then demonstrated on the case of a glycopeptide of haptoglobin that the predicted retention time window boosts confidence of identification and minimizes false positive identification. Relative retention time, as a qualitative parameter, is expected to improve LC-MS/MS characterization of glycopeptides.

Keywords

Glycoform dependent retention prediction; glycopeptide chromatography; glycoproteomics; LC-MS/MS; reversed phase chromatography

*Corresponding author: Miloslav Sanda, Ph.D., Georgetown University, 3800 Reservoir Road NW, Washington, DC 20057, United States, ms2465@georgetown.edu.

Conflict of interest: The authors declare that there is no conflict of interest.

1 Introduction

Glycosylation of proteins is one of the most frequent posttranslational modifications which regulates protein expression and folding, intracellular signaling processes, or cell adhesion [1-5]. The importance of the analysis of glycosylation is highlighted by its involvement in disease-processes [6, 7] and glycoproteins are, therefore, considered relevant diagnostic biomarkers [8]. Glycoproteins are also wide spread therapeutic proteins for various cancer types, autoimmune diseases, and replacement therapies such as enzyme and hormone substitutes [9, 10].

Liquid chromatography coupled to mass spectrometry (LC-MS) is a standard technique of glycoproteomics analysis. High sensitivity, speed and resolution of current MS methods together with various modes of fragmentation such as collision induced dissociation (CID), electron capture or transfer dissociation (ECD or ETD), higher energy collisional C-trap dissociation (HCD), and infrared multi-photon dissociation (IRMPD) offer efficient glycoprotein analysis and provide some degree of structural information [11-13]. However, micro- and macro-heterogeneity of glycoproteins, lower ionization efficiency of glycopeptides compared to peptides, and incomplete fragmentation of either the glycan or peptide moieties of the glycopeptide present substantial analytical challenges [13, 14]. MS analysis alone does not always provide full linkage resolution due to the structural complexity and fragmentation behavior of glycopeptides [15] and chromatographic separation, based on the composition and structure of the attached glycan, can efficiently improve resolution of the analytes [16].

Separation of glycopeptides by reversed phase (RP) chromatography on C18 columns is driven mainly by physical-chemical properties of the peptide backbone with additional influence of the glycan moiety [14, 17-19]. In proteomics, retention and separation of peptides are an essential part of the general bottom up LC-MS/MS workflows [19]. Moreover, predictions of the retention time of peptides is considered in several data processing software tools [18] and the models for prediction of peptide retention times were reviewed recently [19]. A good correlation between predicted and observed retention times of modified peptides including acetylated, butyrylated, propionylated, methionine oxidized and sulfonated peptides were documented [20]. However, there is a lack of models of the chromatographic behavior of glycopeptides with only scattered reports of the influence of glycan moiety on the retention behavior of glycopeptides.

Wang et al. showed that the addition of hexose or N-Acetylhexosamine reduces the retention of glycopeptides under RP conditions while the presence of N-Acetylneuraminic acid or a phospho-group on the glycan increases retention [21]. Medzihradszky et al. showed the chromatographic effect of the different sialic acid compositions present on the Nucleobinding-1 glycopeptide [15]. These results show that the presence of each additional acetylation increased the retention of a glycopeptide and that incorporation of N-Glycolylneuraminic acid into the glycan resulted in slightly earlier elution compared to N-Acetylneuraminic acid. Ozohanics et al. showed in a nano-HPLC-MS-MS workflow for analysis of site-specific glycosylation that the addition of fucose and N-acetyllactosamine units decreased the retention time and the presence of sialic acid residues increased the

retention time of glycopeptides. This retention time shift were used by the authors to reduce the false positive rate of glycopeptide identification [22].

In this work, we study the influence of neutral monosaccharides on the retention of glycopeptides under reversed-phase (RP) nanoLC conditions. We chose hemopexin (HPX), haptoglobin (HP), and sex hormone binding globulin (SHBG) to describe the chromatographic behavior of their respective glycoforms and we derive models that should allow a general correlation of glycopeptide retention with the number of their monosaccharide units (glycan building blocks). The chromatographic information is complementary to mass spectrometric analysis and is useful for reliable assignments of the site specific glycoforms of proteins. We expect that the retention models can be implemented in glycoproteomics search engines as additional qualitative parameter for improved glycopeptide identification.

2 Materials and methods

2.1 Materials and Reagents

Acetonitrile (ACN, LC-MS grade), acetonitrile with 0.1% formic acid (LC-MS grade), water (LC-MS grade), water with 0.1% formic acid (LC-MS grade), ammonium bicarbonate (purity 99%, LC-MS grade), iodoacetamide (purity 99%) and endoglycosidase F2 and F3 from *Elizabethkingia miricola* with Endo F Reaction Buffer (250 mM sodium acetate, pH 4.5) were supplied by Sigma-Aldrich (St. Louis, MO). α (2-3,6,8)-neuraminidase with GlycoBuffer 1 (5 mM calcium chloride, 50 mM sodium acetate, pH 5.5), PNGase F with G7 Reaction Buffer (500 mM sodium phosphate, pH 7.5) and β 1-3,4 Galactosidase with GlycoBuffer 4 (50 mM sodium acetate, pH 4.5) were purchased from New England BioLabs, (Ipswich, MA). Glyko β N-Acetylhexosaminidase with Reaction Buffer (50 mM sodium phosphate, pH 5.5) was obtained from ProZyme (Hayward, CA,) and Trypsin Gold for MS from Promega (Madison, WI). Dithiothreitol (ultrapure grade) was purchased from Thermo-Fischer Scientific (Waltham, Massachusetts). HP and HPX from human plasma were supplied by Athens Research and Technology (Athens, Georgia,) and SHBG by Lee Biosolutions (Maryland Heights, MO). Stock solutions of the individual proteins at a concentration of 10 μ g/ μ L were prepared by dissolving the compounds in water.

2.2 Sample preparation

All glycopeptide standards were prepared by tryptic digest of 10 μ L (100 μ g) of each glycoprotein stock solution diluted in 190 μ L of 50 mM ammonium bicarbonate. Cysteine residues were reduced with 5 mM dithiothreitol at 60 °C for 60 min, alkylated with 15 mM iodoacetamide for 30 min in the dark, and residual iodoacetamide was quenched with 5mM dithiothreitol. Trypsin was added to the samples at an enzyme/protein ratio of 1:25 (w/w) and the glycoproteins were digested at 37 °C in Barocycler NEP2320 (Pressure BioSciences, South Easton, MA) for 1 h. The enzymatic digestions were stopped by heating at 99 °C for 10 min.

An aliquot (100 μ L) of the mixture was desialylated by the addition of 11 μ L of GlycoBuffer 1 and 4 μ L of neuraminidase overnight at 37 °C; another aliquot was deglycosylated by the

addition of 11 μ L of the G7 reaction buffer and 3 μ L of PNGase F overnight at 37 °C. The enzyme reactions were stopped by heating at 75 °C for 10 min. Prepared digests were desalted using solid phase extraction on a C18 cartridge 4 mm/1 mL (3M, St. Paul, MN). Desalting procedure was as follows: Condition step: 1 mL of 100% acetonitrile, 1.5 mL of 2% trifluoroacetic acid; loading step; washing step: 1 mL of 2% trifluoroacetic acid; elution step: 0.3 mL of 60% acetonitrile, 0.2 mL of 80% acetonitrile. The combined eluates were evaporated using a vacuum concentrator (Labconco, Kansas City, MO) and reconstituted in a solution of 0.1% formic acid in 2% acetonitrile.

20 μ L of the desialylated sample (1 μ g) was evaporated and reconstituted in 80 μ L of water, mixed with 20 μ L of reaction buffer, and digested with 2 μ L of each endoglycosidase F2 and F3 at 37 °C for two days. Another aliquot of 20 μ L of the desialylated sample was evaporated and reconstituted in 90 μ L of water, mixed with 10 μ L of reaction buffer, and digested with 3 μ L of β -Galactosidase at 37 °C overnight. 30 μ L of the degalactosylated sample (0.29 μ g) was mixed with 5 μ L of reaction buffer, 2 μ L of Glyco β -N-AcetylHexosaminidase was added, and the mixture was incubated at 37 °C overnight. All samples after enzymatic digestion were cleaned using solid phase extraction, reconstituted in a solution of 0.1% formic acid in 2% acetonitrile, and mixed in appropriate proportions to obtain an adequate MS signal of all analytes.

2.3 Glycopeptide identification

All chromatographic measurements were performed on a nanoAcquity UPLC system with a binary pump (Waters, Milford, MA) interfaced with 6600 TripleTOF (Sciex, Framingham, MA) mass spectrometer. The Analyst software (Sciex, Framingham, MA) was used for data acquisition. Glycopeptides were analyzed by RP chromatography on Acquity UPLC M-class Symmetry C18 Trap Column (100 Å, 5 μ m, 180 μ m $\times$ 20 mm) and Acquity UPLC Peptide BEH C18 nano column (300 Å, 1.7 μ m, 75 μ m $\times$ 150 mm) both from Waters (Milford, MA). Experimental conditions were as follows: solvent A, 0.1% formic acid in 2% acetonitrile; solvent B, 0.1% formic acid in 100% acetonitrile, flow rate 0.4 μ L/min, column temperature 40 °C, injection volume 1 μ L, gradient program (marked as gradient 3) included 5 min trapping step at 1% B at 15 μ L/min followed by the following: 0-60 min, 1-43% B; 60-65 min, 43-99% B; 65-75 min, 99% B; 75-77 min, 99-1% B; 77-97 min, 1% B. An information-dependent acquisition workflow was used for the assignment of *N*-glycopeptides. Full MS scan in the *m/z* range 400-1800 was followed by collision-induced dissociation of the 50 most intense ions. Collision energy was set automatically and the MS/MS spectra were recorded in the *m/z* range 100-1800. Identification of *N*-glycopeptides was performed manually based on precursor masses and characteristic fragments. Mass spectrometric experimental conditions were set to declustering potential, 80 V; curtain gas, 30; ion spray voltage, 2300 V; ion source gas 11; interface heater, 150 °C; exclusion time, 5 s.

2.4 Determination of retention time

To study the retention behavior under RP-LC conditions, we chose two occupied peptides from each of the three proteins: VVLHPNYSQVDIGLIK and MVSHHNLTTGATLINEQWLLTTAK for HP, ALPQPQNVTSLGCTH and

SWPAVGNCSALR for HPX, SHEIWITHSCPQSPGNGTDASH and LDVDQALNR for SHBG. For each peptide, we recorded retention times of seven glycoforms prepared by different enzymatic digestions: GlcNAc, GlcNAc₂Man₃, GlcNAc₂Man₃GlcNAc₂, GlcNAc₂Man₃GlcNAc₃, GlcNAc₂Man₃GlcNAc₂Gal₂, GlcNAc₂Man₃GlcNAc₂Gal₂Fuc₁, GlcNAc₂Man₃GlcNAc₃Gal₃, GlcNAc₂Man₃GlcNAc₃Gal₃Fuc₁, and additional two glycoforms GlcNAc₂Man₃GlcNAc₄Gal₄, GlcNAc₂Man₃GlcNAc₄Gal₄Fuc₁ for peptide VVLHPNYSQVDIGLIK.

All retention time measurements were performed in five replicates using full MS scan mode (Accumulation time – 100 ms; cycle time – 0.1250 s) from 400 to 2000 *m/z* to acquire more than 50 data points across the LC peak. LC experimental conditions were as described above. Three gradient programs with different gradient slope were tested. All gradient programs included 5 min trapping step at 1% B at a 15 μ L/min. The first gradient program (marked as gradient 1; gradient slope 0.40 %/min) was as follows: 0-60 min, 1-25% B; 60-65 min, 25-99% B; 65-75 min, 99% B; 75-77 min, 99-1% B; 77-97 min, 1% B. The second one (marked as gradient 2; gradient slope 0.55 %/min) was as follows: 0-60 min, 1-34% B; 60-65 min, 34-99% B; 65-75 min, 99% B; 75-77 min, 99-1% B; 77-97 min, 1% B. The third one (marked as gradient 3; gradient slope 0.70 %/min) was as described in section “Glycopeptide identification”.

PeakView software (Sciex, Framingham, MA) was used to process LC-MS data. Retention time of glycopeptides was evaluated from Gaussian smoothed extracted ion chromatograms (EICs) of the monoisotopic *m/z*. EICs were exported as text files to OriginPro 8.5.0 (OriginLab Corporation, Northampton, MA) for the visualization presented in Figure 1A; each EIC was normalized to the target glycopeptide and all target glycopeptide EICs of one peptide were overlaid and zoomed to the appropriate retention time window.

3. Results and discussion

In this study, we have used tryptic digests of three human serum glycoproteins (HP, HPX and SHBG) to describe the retention behavior of glycopeptides under RP chromatographic conditions. We chose glycoforms of two peptides of each protein to predict the retention windows of the glycoforms based on their monosaccharide composition. Retention of the glycoforms that share the same peptide sequence were expressed as relative retention times (RRTs) to GlcNAc, mannose core (GlcNAc₂Man₃), bi-antennary (GlcNAc₂Man₃GlcNAc₂Gal₂), and D forms of the peptides in order to derive the most accurate prediction of the retention windows.

3.1 Retention behavior of glycopeptides

It has previously been described that different neutral oligosaccharide units of glycopeptides influence the retention time under the RP chromatography [21-23]. We observed decreased retention times of glycoforms of tryptic glycopeptides of several proteins with increasing size of neutral glycan and we used a 1 minute retention window around the major glycoform as a criterion for identification of the peak as a related glycoform of the same peptide [23]. We used column with sub-2- μ m ethylene bridged hybrid particles, providing high resolving power, to achieve resolution of glycopeptides based on their monosaccharide units under the

C18 RP chromatography. We restrict our focus to neutral oligosaccharides and we removed negatively charged sialic acids with neuraminidase. We do this because sialic acid increases retention [15] and further studies would be needed to establish a retention model based on the sum of the contributions of the neutral and charged glycans.

The retention of glycopeptides in the RP mode is driven mainly by the peptide. Retention times of the studied peptides in their deglycosylated (D) form (conversion of asparagine to aspartic acid by PNGase F) under the gradient 3 condition (see methods) were as follows: 20.28 min for SHEIETHSCPQSPGDGTDASH, 26.19 min for LDVDQALDR, 29.26 min for SWPAVGDCSSALR, 34.65 min for VVLHPDYSQVDIGLIK, 37.57 min for ALPQPQDVTSLGCTH and 38.35 min for MVSHDLTTGATLINEQWLLTTAK. All RRTs and retention times under the gradient 3 condition are presented in Supporting Information Table S1. A chromatogram of the overlaid normalized EICs of the glycoforms of VVLHPNYSQVDIGLIK is depicted in Figure 1A. This figure shows how the polar character of neutral glycans causes a glycopeptide to elute earlier. In general, increasing the number of monosaccharide units of a glycan decreases retention progressively so that GlcNAc affects retention of the peptide less than the larger glycans. We showed that six complex glycoforms of the VVLHPNYSQVDIGLIK tryptic peptide of haptoglobin aligned in retention time window: 32.65 and 33.07 min [23]. Presence of fucose provides a less pronounced retention shift for larger glycans, which is seen in differences of RRTs expressed as $t_{R_x}/t_{R_{GlcNAc \text{ form}}}$ for bi-, tri-, and tetra-antennary forms with and without fucose. The RRT difference introduced by fucose for the bi-antennary glycoform of VVLHPNYSQVDIGLIK is 0.00204, for tri-antennary 0.00096, and for tetra-antennary 0.00035. A larger glycan has a greater hydrophilicity and glycopeptide retention is less affected by the addition of the fucose monosaccharide unit (lower partial hydrophilicity of the monosaccharide unit). The same trends in the retention behavior were observed for the other glycopeptides (Supporting Information Table S1). Our data showed that average decrease of retention time caused by GlcNAcGal and Fuc is 0.4% (standard deviation 0.3%) and 0.2% (standard deviation 0.2%), respectively. Similar trends were described by Ozohanics et al; the authors showed that the presence of GlcNAcGal decreased the retention time average by 0.8% (standard deviation 0.4%) and addition of Fuc reduced the retention time of glycopeptides by 0.4% (standard deviation 0.4%) [22].

Figure 1B shows the dependence of RRT ($t_{R_x}/t_{R_{GlcNAc \text{ form}}}$) of the glycoforms of VVLHPNYSQVDIGLIK on the number of monosaccharide units. The dependence is described by second order polynomial regression curve with correlation coefficient 0.9968. Very similar results were obtained for all studied compounds (see Supporting Information Figure S2).

When RRTs ($t_{R_x}/t_{R_{GlcNAc \text{ form}}}$) of all the studied glycoforms of VVLHPNYSQVDIGLIK were plotted against the number of monosaccharide units, the second order polynomial fit is less accurate ($R^2 = 0.9151$) and shows a visible gap between galactosylated and degalactosylated structures. Elimination of galactose units by β -Galactosidase resulted in significant shift in the retention time (Figure 1A). This behavior could be related to different polarities of galactose (log $P = -2.6$) and N-acetylglucosamine (log $P = -1.7$) (<https://www.ncbi.nlm.nih.gov/pccompound>) and/or to adsorption and steric effects of the

glycopeptides [24]. Similar retention behavior was also observed for the glycoforms of MVSHNLTTGATLINEQWLLTTAK and SWPAVGNCSSALR and, to a lesser extent, for ALPQPQNVTSLGCTH (Supporting Information Figure S2-C, E, and G). In the case of the glycoforms of SHBG, we did not observe significant shift in the retention time after elimination of galactose (Supporting Information Figure S2-I, K). This could be related to the overall higher polarity of SHBG glycopeptides whose retention time is therefore less affected by different polarities (log *P*) of galactose and N-acetylglucosamine compared to the peptides of haptoglobin. Based on these observations, the retention of glycopeptides in RP chromatography does not depend only on the polarity of an analyte, but it seems that partitioning and adsorption contribute to the retention mechanism [25].

In order to increase the precision of presented fits, we tried to fit our data into different multiparametric forms. We found the best correlation between observed RRT and fitted equation when we used two parameter fit. First parameter, $N(\text{sacch-gal})$, represents the total number of glycan units minus the number of galactose units and the second $N(\text{gal})$ is the number of galactose units. This partially reflects main structural feature of studied saccharides (different polarities of galactose and N-acetylglucosamine). We used quadratic terms to describe the retention behavior which resulted in five terms to be fitted: $\text{RRT} = a \cdot N(\text{sacch-gal})^2 + b \cdot N(\text{sacch-gal}) + c \cdot N(\text{gal})^2 + d \cdot N(\text{gal}) + e$. A second order polynomial fit of all the studied glycoforms of VVLHPNYSQVDIGLIK with respect to the total number of saccharide units showed correlation $R^2 = 0.9151$ (Figure S2A); the improved quadratic model, achieved $R^2 = 0.9959$. We present a generalized fit, valid for whole spectrum of all studied glycopeptides, in Table 1 and we show regression coefficients of the fitted multiparametric functions for all the studied glycopeptides. Fitted values offer good correlation with the measured RRT. This further supports our hypothesis that RRT is primarily influenced by the number and identity of the saccharide units and less by the peptide sequence. The main factor which describes the RRT of the studied glycopeptides is the length and composition of saccharide chain. We also investigated the effect of three different gradient slopes (gradient 1 0.4 %/min, gradient 2 0.55 %/min, and gradient 3 0.70 %/min) on the RRTs of glycopeptides (Figure 2). The lowest differences of RRT ($\bar{t}_{\text{R}}/\bar{t}_{\text{GlcNAc form}}$) were observed for the VVLHPNYSQVDIGLIK glycoforms (Figure 2A) where the differences of RRT ($\bar{t}_{\text{R}}/\bar{t}_{\text{GlcNAc form}}$) ranged from 0.10 % (GlcNAc₂Man₃) to 0.29 % (GlcNAc₂Man₃GlcNAc₂Gal₂). The highest differences of RRT ($\bar{t}_{\text{R}}/\bar{t}_{\text{GlcNAc form}}$) were observed for the SHEIWITHSCPQSPGNGTDASH glycoforms (Figure 2B) where the differences of RRT ($\bar{t}_{\text{R}}/\bar{t}_{\text{GlcNAc form}}$) ranged from 0.32 % (GlcNAc₂Man₃GlcNAc₃) to 0.74 % (GlcNAc₂Man₃GlcNAc₃Gal₃Fuc₁). Differences of RRT ($\bar{t}_{\text{R}}/\bar{t}_{\text{GlcNAc form}}$) for all studied glycopeptides are shown in the Supporting Information Figure S3. In summary, the change of the gradient slope affected the retention but had negligible impact on the RRTs of glycopeptides.

3.2 Prediction of the retention time windows

Peptides and glycopeptides are usually identified based on their intact mass and fragmentation spectra [26]. CID (HCD) fragmentation mass spectra of N-glycopeptides are characterized by the high yield of glycan fragments (oxonium ions) and marginal yields of peptide fragment ions (y and b). Several methods and bioinformatics approaches based on

additional qualitative information have been developed to improve the identification of glycopeptides [26]. The most common is the use of parallel analysis of glycopeptides trimmed or deglycosylated by enzymatic treatment such as PNGase F [27-29]. Chen et al. reported an approach based on the reduction of the glycan structures to a mannose core which resulted in an increased confidence of the glycopeptide assignments [27].

We focused in our study on the analysis of retention behavior of complex glycans which provide the most difficult CID fragmentation spectra for automated glycopeptide identification. We have compared retention of bi-, tri-, and tetra-antennary glycoforms. RRTs related to the deglycosylated (D) form, trimmed GlcNAc and mannose core forms and the bi-antennary form as the major serum glycoform were used to calculate the upper and lower limits of the retention windows. Central value, upper and lower limits, and retention time width (expressed in %) are shown in Table 2. Limits were calculated for gradient 3 and for all gradients together. Calculation of upper and lower limits was based on a maximum and a minimum value of RRTs for all studied peptides. For example, the maximum and the minimum value of RRT ($\alpha_x/\alpha_{\text{GlcNAc form}}$) for gradient 3 was obtained for GlcNAc₂Man₃GlcNAc₂Gal₂ of ALPQPQNVTSLLGCTH peptide (0.97237) and for GlcNAc₂Man₃GlcNAc₃Gal₃Fuc₁ of LDVDQALNR peptide (0.94367), respectively (Supporting Information Table S1). Upper and lower limits were corrected to average peak width which corresponded approximately to 0.01% RRT. Limits calculated for bi-antennary form provide the narrowest retention time window and the widest retention time window was observed for the RRTs related to the D form. Calculation for all gradients compared to gradient 3 slightly increases the retention time window.

To illustrate how prediction of retention time windows work, we present results for VVLHPNYSQVDIGLIK glycoforms of the HP protein. Calculated retention time windows (based on limits for all gradients) are shown in Figure 3. Retention time windows were 2.41 min, 1.70 min, 1.41 min and 1.13 min compared to the D form, GlcNAc form, mannose core and bi-antennary form, respectively, which provides the narrowest predicted retention window. This significantly strengthens the confidence of the glycopeptide identification. Zacharias et al. used retention time window of 4 min relative to the D form for identification of glycopeptides in six cell lines [28]. Chen et al. used retention time tolerance of 5 min as qualitative parameter for identification in the case of early-eluting glycopeptides and 3 min for the late-eluting glycopeptides relative to the glycopeptides with mannose core [27].

Advantage of the retention time window based on the D form and GlcNAcylated peptides is the similarity of the fragmentation to the naked peptides; interpretation of such spectra is compatible with standard proteomics search engines. Combination of several enzyme digests (neuraminidase, β -Galactosidase and β -N-AcetylHexosaminidase) is needed to obtain the glycoforms with a mannose core which is time consuming and provides lower yields. The bi-antennary forms are typically common and present in tryptic digests without additional enzymatic digests. We expect that the retention time windows can be incorporated to the glycoproteomics search engines to boost confidence of glycopeptides identification and minimize false positive identification.

4. Concluding remarks

Tandem mass spectrometry is a powerful analytical tool of glycoproteomics analysis efficiently assisted by the chromatographic separation of glycopeptides. In this work, we describe the influence of monosaccharide units of the neutral glycans on the retention behavior of glycopeptides in RP chromatography. The retention behavior was studied on multiple glycoforms of six peptides obtained by enzymatic digestions of haptoglobin, hemopexin and sex hormone binding globulin. We have documented decreasing retention of glycopeptides with increasing number of monosaccharide units of the glycan moiety. Retention times of glycopeptides were expressed relative to different glycoforms generated *in vitro* (GlcNAc, mannose core, or D form) or relative to the naturally occurring bi-antennary glycoform. The retention time windows were predicted and the narrowest windows were obtained in the case of normalization to the bi-antennary form. In contrast, the widest RRTs and windows are observed with normalization to the D form. We propose that the predicted retention time windows of glycopeptides based on RRTs can be incorporated to the glycoproteomics search engines which is expected to improve the reliability of glycopeptide identification.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors gratefully acknowledge the financial support by the National Institutes of Health Grants UO1 CA168926, UO1 CA171146, and RO1 CA135069 (to R.G.) and CCSG Grant P30 CA51008 (to Lombardi Comprehensive Cancer Center supporting the Proteomics and Metabolomics Shared Resource).

References

1. Schnaar RL. J Leukocyte Biol. 2016; 99:825–838. [PubMed: 27004978]
2. Woronowicz A, Amith SR, De Vusser K, Laroy W, Contreras R, Basta S, Szewczuk MR. Glycobiology. 2007; 17:10–24. [PubMed: 16971381]
3. Krasnova L, Wong CH, Kornberg RD. Annu Rev Biochem. 2016; 85:599–630. [PubMed: 27145845]
4. Varki A. Cold Spring Harbor Perspect Biol. 2011; 3:1–14.
5. Varki, A.Cummings, RD.Esko, JD.Stanley, P.Hart, G.Aebi, M.Darvill, A.Kinoshita, T.Packer, NH.Prestegard, JJ.Schnaar, RL., Seeberger, PH., editors. Essentials of Glycobiology. Cold Spring Harbor Laboratory Press; Cold Spring Harbor, NY; 2015.
6. Song E, Mechref Y. Biomarkers Med. 2015; 9:835–844.
7. Peracaula R, Sarrats A, Rudd PM. Proteomics: Clin Appl. 2010; 4:426–431. [PubMed: 21137061]
8. Zhao JF, Song EW, Zhu R, Mechref Y. Electrophoresis. 2016; 37:1420–1430. [PubMed: 26957414]
9. Ghaderi D, Zhang M, Hurtado-Ziola N, Varki A. Biotechnol Genet Eng Rev. 2012; 28:147–175. [PubMed: 22616486]
10. Bork K, Horstkorte R, Weidemann W. J Pharm Sci. 2009; 98:3499–3508. [PubMed: 19199295]
11. Novotny MV, Alley WR. Curr Opin Chem Biol. 2013; 17:832–840. [PubMed: 23790311]
12. Zhao J, Li SY, Li C, Wu SL, Xu W, Chen YT, Shameem M, Richardson D, Li HJ. Anal Chem. 2016; 88:7049–7059. [PubMed: 27291648]
13. Marino K, Bones J, Kattla JJ, Rudd PM. Nat Chem Biol. 2010; 6:713–723. [PubMed: 20852609]

14. Sanda M, Pompach P, Benicky J, Goldman R. Electrophoresis. 2013; 34:2342–2349. [PubMed: 23765987]
15. Medzihradsky KF, Kaasik K, Chalkley RJ. Anal Chem. 2015; 87:3064–3071. [PubMed: 25654559]
16. Alley WR, Mann BF, Novotny MV. Chem Rev. 2013; 113:2668–2732. [PubMed: 23531120]
17. Benicky J, Sanda M, Pompach P, Wu J, Goldman R. Anal Chem. 2014; 86:10716–10723. [PubMed: 25302577]
18. Bruderer R, Bernhardt OM, Gandhi T, Reiter L. Proteomics. 2016; 16:2246–2256. [PubMed: 27213465]
19. Tarasova IA, Masselon CD, Gorshkov AV, Gorshkov MV. Analyst. 2016; 141:4816–4832. [PubMed: 27419248]
20. Moruz L, Staes A, Foster JM, Hatzou M, Timmerman E, Martens L, Kall L. Proteomics. 2012; 12:1151–1159. [PubMed: 22577017]
21. Wang BL, Tsybovsky Y, Palczewski K, Chance MR. J Am Soc Mass Spectrom. 2014; 25:729–741. [PubMed: 24549892]
22. Ozohanics O, Turiak L, Puerta A, Vekey K, Drahos L. J Chromatogr A. 2012; 1259:200–212. [PubMed: 22677411]
23. Sanda M, Zhang L, Edwards NJ, Goldman R. Anal Bioanal Chem. 2016; 406:619–627.
24. Geng XD, Regnier FE. J Chromatogr. 1984; 296:15–30. [PubMed: 6480740]
25. Dorsey JG, Dill KA. Chem Rev. 1989; 89:331–346.
26. Chalkley RJ, Baker PR. Anal Bioanal Chem. 2016; doi: 10.1007/s00216-016-9981-2
27. Chen R, Cheng K, Ning Z, Figeys D. Anal Chem. 2016; doi: 10.1021/acs.analchem.6b03531
28. Wohlgemuth J, Karas M, Jiang W, Hendriks R, Andrecht S. J Sep Sci. 2010; 33:880–890. [PubMed: 20222078]
29. Zacharias LG, Hartmann AK, Song E, Zhao J, Zhu R, Mirzaei P, Mechref Y. J Proteome Res. 2016; 15:3624–3634. [PubMed: 27533485]

A list of abbreviations

ECD	Electron capture dissociation
EIC	Extracted ion chromatogram
ETD	Electron transfer dissociation
HCD	Higher energy collisional C-trap dissociation
HP	Haptoglobin
HPX	Hemopexin
IRMPD	Infrared multi-photon dissociation
RRT	Relative retention time
SHBG	Sex hormone binding globulin

Figure 1.

Analysis of the VVLHPNYSQVDIGLIK peptide of HP: A. Overlay of the normalized extracted ion chromatograms of all the analyzed glycoforms under the gradient 3 nanoLC conditions; B. Polynomial fit of the dependence of RRT ($t_x/t_{\text{GlcNAc form}}$) on the number of monosaccharide units of the glycoforms. Symbols: ■, N-acetylglucosamine (GlcNAc); ●, Galactose (Gal); ▲, Fucose (Fuc); ●, Mannose (Man). N form – peptide backbone of *N*-glycopeptides containing asparagines (the glycan is attached to this asparagine). D form – deglycosylated glycopeptide (conversion of asparagine to aspartic acid by PNGase F).

Figure 2.

The influence of the three gradient slopes (0.4 %/min; 0.55 %/min; and 0.70 %/min) on the RRT (tr_X/tr_{GlcNAc} form) of glycopeptides: A. Glycoforms of the VVLHPNYSQVDIGLIK peptide of HP; B. Glycoforms of the SHEIWTHTSCPQSPGNGTDASH peptide of SHBG.

Figure 3.

Prediction of the retention time windows (calculated based on upper and lower limit for all gradients) for VVLHPNYSQVDIGLIKbi-, tri-, and tetra-antennary forms. Wine color lines create the retention time window calculated for RRT ($t_x/t_{D \text{ form}}$), red color lines for RRT ($t_x/t_{\text{GlcNAc form}}$), orange color lines for RRT ($t_x/t_{\text{mannose core}}$) and pink color lines for RRT ($t_x/t_{\text{bi-antennary form}}$).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Regression coefficients of the fitted RRT dependence: $RRT = a \cdot N(\text{sacch} - \text{gal})^2 + b \cdot N(\text{sacch} - \text{gal}) + c \cdot N(\text{gal})^2 + d \cdot N(\text{gal}) + e$, where $N(\text{sacch} - \text{gal})$ is total number of monosaccharide units minus the number of galactose units and $N(\text{gal})$ is the number of galactose units

Table 1

RRT	a	b	c	d	e	R ²
r_X/r_{H_2O} -antennary	0.0002	-0.0066	0.0013	-0.0087	1.0064	0.9581
$r_X/r_{\text{mannose core}}$	0.0002	-0.0068	0.0007	-0.0071	1.0274	0.9230
r_X/r_{GlcNAc}	0.0002	-0.0066	0.0013	-0.0087	1.0064	0.9201
$r_X/r_{\text{D form}}$	0.0002	-0.0061	0.0029	-0.0125	0.9659	0.8841

Comparison of upper, lower limits and retention time window (%) for bi-, tri-, and tetra-antennary glycoforms relative to the D, GlcNAc, mannose core, and bi-antennary forms.

Table 2

	RRT ($t_{RX}/t_{D form}$)		RRT (t_{RX}/t_{GlcNAc})		RRT ($t_{RX}/t_{mannose core}$)		RRT ($t_{RX}/t_{bi-antennary}$)	
	Gradient 3	All gradients	Gradient 3	All gradients	Gradient 3	All gradients	Gradient 3	All gradients
Central value	0.91661	0.91164	0.95802	0.95677	0.97480	0.97465	0.99236	0.99236
Upper limit	0.94669	0.94669	0.98209	0.98209	0.99525	0.99599	1.01000	1.01000
Lower limit	0.88693	0.87710	0.93423	0.93177	0.95455	0.95354	0.97487	0.97487
%	6.52	7.63	4.99	5.26	4.18	4.36	3.54	3.54

$$\text{Central value} = (RRT_{\max} + RRT_{\min})/2$$

$$RRT = RRT_{\max} - RRT_{\min}$$

$$\text{Upper limit} = \text{central value} + 0.5 * RRT + 0.01 * RRT_{\max}$$

$$\text{Lower limit} = \text{central value} - 0.5 * RRT - 0.01 * RRT_{\min}$$

$$\% = 100 * (\text{Upper limit} - \text{lower limit} / \text{central value})$$