Separation and detection of minimal length isomeric glycopeptide neoantigen epitopes centering GSTA region of MUC1 by LC-MS

Dapeng Zhou[1*], Kaijie Xiao[2], Zhixin Tian[2*]

1, School of Medicine, Tongji University, Shanghai 200092, China

2, School of Chemical Science and Engineering & Shanghai Key Laboratory of Chemical Assessment and Sustainability, Tongji University, Shanghai 200092, China

Correspondence: dapengzhoulab@tongji.edu.cn;

zhixintian@tongji.edu.cn

Abbreviations list:

TR, tandem repeat;

Tn, GalNAc;

mAb, monoclonal antibody;

MS, mass spectrometry;

ESI, electro-spray ionization;

ETD, electron transfer dissociation;

HCD, higher-energy collisional dissociation;

IPACO, isotopic peak abundance cutoff;

IPMD, isotopic peak *m/z* deviation;

IPAD, isotopic peak abundance deviation.

Abstract

MUC1 ranks No.2 on the list of targets for cancer immunotherapy. We previously reported monoclonal antibodies binding to glycopeptide neoantigen epitopes centering GSTA sequence of the highly glycosylated tandem repeat region of MUC1. Epitopes centering GSTA sequence are also predicted by NetMHC programs to bind to MHC molecules, although empirical data are lacking. Detecting isomeric MUC1 glycopeptide epitopes by mass spectrometry (MS) remains a technical challenge since antigenic epitopes are often shorter than 10 amino acids. MUC1 digests by Arg-C-specific endopeptidase clostripain could generate heterogenous icosapeptides, but isomeric 20-residue glycopetides could not be separated by liquid chromatography. In this study, we used pronase from *Streptomyces griseus,* which has no amino acid sequence preference for enzymatic cleavage sites, to digest a pair of synthetic glycopeptide isomers RPAPGST(Tn)APPAHG and RPAPGS(Tn)TAPPAHG, and analyzed the digests by LC-MS using electron transfer dissociation (ETD) and higher-energy collisional dissociation (HCD) methods. The results showed that glycopeptide isomers containing 8 to 11 amino acids could be efficiently generated by pronase digestion. Such glycopeptide isomers of minimal epitope lengths were clearly distinguished by characteristic MS/MS ion patterns and elution profiles of liquid chromatography. A glycopeptide library was generated which may serve as standards for measuring neoantigen epitopes centering GSTA sequence.

Key words: MUC1; glycopeptide; mass spectrometry; electron transfer dissociation; neoantigen; epitope

Introduction

MUC1 is a type I membrane glycoprotein currently ranked as number 2 on the list of targets for cancer immunotherapy (1-2). Its extracellular portion contains 20 to 120 or more tandem repeats (TR) of 20-residue sequence (HGVTSAPDTRPAPGSTAPPA) with five potential O-linked glycosylation sites through N-acetylgalactosamine to serine and/or threonine residues. It is expressed at low levels at the apical surface of most healthy glandular epithelial cells, with normal pattern of O-glycosylation. In tumor settings MUC1 loses its polarity and normal pattern of glycosylation to expose GalNAc (Tn) residue.

We and others have dissected the abnormally glycosylated TR region as three antigenic motifs: PDTR, GSTA, and GVTS (2). The glycosylation of the three aforementioned peptide motifs may influence their binding to mAbs by several mechanisms. Identification of glycosylation site on above motifs is critical for dissecting the exact MUC1 antigenic epitopes for cancer diagnosis and therapy. For example, we have reported the expression of abnormally glycosylated GSTA neoantigen motif in non-small cell carcinoma cells (3-4).

MUC1 glycopeptides may also be processed by MHC Class I and/or MHC class II pathway, and serve as neoantigen epitopes for T cells. Barnea et al. reported the elution of MUC1-derived NLTISDVSV sequence from HLA-A2 molecules in breast cancer cell line MCF-7, although glycosylated NLTISDVSV has not been reported (5-6). Hanisch et al. reported the processing of a recombinant glycoprotein containing six MUC1 tandem repeats by mouse dendritic cells, that glycopeptides centering all PDTR, GSTA, and GVTS motifs could be generated by lysosomal protease digestion (7). Ninkovic et al. reported the loading of SAPDT(GalNAc)RPAPG by

3

HLA-A2 molecule, that stimulated CTL cells (8). However, the direct detection of endogenous

glycopeptides eluted from MHC molecules have not been reported.

Significant progresses have been made for glycopeptide analysis in past decade. Databases (9-

10) and methods (11-31) for analyzing N-glycopeptides and O-glycopepitdes are accumulating,

especially for site-specific glycosylation. Obviously, data containing both identities of glycan

structures and glycosylation sites are most valuable for their functional analysis, compared to

data containing glycan structures or glycosylation sites alone.

The accumulation of glycopeptide data benefit from most widely used MS ionization and ion

activation techniques. Especially, electro-spray ionization (ESI) and electron transfer

dissociation (ETD) greatly paved studies on O-linked glycosylation, allowing in-depth analysis

of both O-glycan structures and the occupancy of O-glycosylation sites. Levery and Clausen

group published systemic analysis of O-glycopeptides in trypsin-digested proteins from CHO

cells (21). Zhang group recently published site-specific extraction of O-linked glycopeptides in

trypsin-digested tissues (31). However, no data has been available for the TR region of mucin-1

protein yet due to the resistance of TR region to trypsin digestion. Muller et al. used Arg-C-

specific endopeptidase clostripain to digest human milk MUC1, which generated heterogenous

icosapeptides starting with the PAP sequence. O-glycosylation sites were localized by post-

source decay matrix-assisted laser desorption ionization mass spectrometry or by solid phase

Edman degradation (32). Sihlbom et al. used LC-MS with electron-capture dissociation (ECD)

fragmentation method to analyze clostripain-digested icosapeptides, and could localize the O-

glycoylation sites in heterogenous icosapeptides (33). However, the separation of isomeric icosapeptides could not be achieved by liquid chromatography.

In this study, we tested the efficacy of pronase from *Streptomyces griseus* in digesting synthetic MUC1 glycopeptides, and generated short glycopeptides library containing 8 to 11 amino acids.

Methods

Synthesis of RPAPGST(Tn)APPAHG and RPAPGS(Tn)TAPPAHG. Antigenic epitopes for antibody binding are mostly shorter than 10 amino acids. We chose to use two synthetic 13-residue glycopeptides, RPAPGST(Tn)APPAHG and RPAPGS(Tn)TAPPAHG, as the starting material for generating glycopeptides shorter than 10 amino acids. The chemical synthesis of glycopeptides was as described (3) using fluorenylmethyloxycarbonyl (Fmoc)-protected amino acids as the building blocks. The purity of synthetic glycopeptides were examined by reversed-phase HPLC and MS determination of molecular masses.

Pronase digestion of MUC1 glycopeptides. Glycopeptides were digested by pronase from *Streptomyces griseus* (Roche Diagnostics, Germany) according to the manufacturer's protocol. In brief, 5 µg glycopeptides were digested in 100 µL Tris buffer (50 mM, pH 7.6) with 5 mM $CaCl_2$ and 10 mg/mL pronase at 50 °C for 12 hours. The pronase was inactivated by heating to 100 °C for 5 minutes. The digests were desalted by a home-made C18 reverse chromatography column, before analyzed by mass spectrometry.

LC-MS/MS analysis of pronase digests of RPAPGST(Tn)APPAHG and RPAPGS(Tn)TAPPAHG

The pronase digests of both RPAPGST(Tn)APPAHG and RPAPGS(Tn)TAPPAHG were analyzed on an Orbitrap Fusion Lumos MS (Thermo Scientific, San Jose, CA, USA) coupled with a nano-ESI source and a Dionex Ultimate 3000 RSLCnano HPLC system.

A C18 (Agilent ZORBAX 300SB, 5 μm, 300 Å) pre-column (360 μm o.d. × 200 μm i.d., 7 cm long) was used for sample loading. Chromatographic separation was performed on a 75-cm-long analytical column (360 μm o.d. × 75 μm i.d.) packed with the same C18 particles with the pre-column; buffer A is composed of 99.9% H_2O and 0.1% FA, and buffer B is composed of 99.9% ACN and 0.1% FA. The flow rate of the mobile phase was 300 nL/min with a multi-step gradient starting with 4% B: 8% B, 4 min; 30% B, 49 min; 100% B, 4 min; 100% B, 3 min.

MS spectra were acquired as follows: mass resolution 60 k; m/z range 375-1800, max ion injection time 50 ms, automatic gain control (AGC) target 4e5, microscans 1, RF lens 40%. MS/MS spectra were acquired with the following settings: data-dependent mode, cycle time 3 s, isolation width 1.2 Th, first mass 100, ETD activation time 200 ms, ETD reagent target 2e5, ETD reagent injection time 200 ms, supplemental activation 35%, mass resolution 30 k, max ion injection time 200 ms, AGC target 5e4, microscans 1, dynamic exclusion 30 s, included charge states 2-7. The ESI conditions were as follows: spray voltage 2.2 kV, capillary temperature 320 °C.

Database search and peptide identification

Database search and peptide identification from the LC-MS/MS datasets of the full-length synthetic peptides and their pronase digests were carried out using ProteinGoggle; the detailed

6

interpretation steps have been published elsewhere (34-39). Briefly, a customized peptide database was created for the full-length RPAPGS(Tn)TAPPAHG and RPAPGST(Tn)APPAHG, as well as their pronase digests with GalNAc as a dynamic modification on either S or T, containing theoretical isotopic envelopes information of both the precursor and fragment ions (a, b, c, x, y, and z). Matched precursor and fragment ions were searched with the following IPACO/IPMD/IPAD parameters: 40%/15ppm/100%, 20%/15 ppm/50%.

Results

Synthesis and detection of GalNAc-modified MUC1 antigen epitopes centering GSTA motif.
The shortest glycopeptides of MUC1 tandem repeat region that have been detected in the past were icosapeptides (20-residue glycopeptides) prepared from MUC1 protein digests by Arg-C-specific endopeptidase clostripain (32-33). However, antibody-binding epitopes are often shorter than 10 amino acids. We previously tested a 13-residue glycopeptide, RPAPGS(Tn)TAPPAHG, which contains antibody binding sites for 16A (3), and synthesized its isomer RPAPGST(Tn)APPAHG.

The glycopeptide RPAPGS(Tn)TAPPAHG was confidently detected as shown in the MS spectrum (Figure 1A) with IPAD value within 5% and IPMD value within 5 ppm. 18 matched fragment ions were observed in the MS/MS spectrum (Figure 1B, *=GalNAc), among which 8 contain GalNAc (as shown in the red box). We detected c6* (*m/z* 786.40936) and z7 (*m/z* 634.30573), the two theoretical glycosite-determining fragment ions.

We also detected the glycopeptide RPAPGST(Tn)APPAHG (Figure 2A) with IPAD and IPMD values lower than 4% and 4 ppm, respectively. Also observed in this peptide included 18 matched

7

fragment ions in the MS/MS spectrum (Figure 2B, *=GalNAc), among which 9 contain GalNAc (as shown in the read box) including c6 (*m/z* 583.33026) and z7* (*m/z* 837.38580).

A MS/MS library of MUC1 antigen epitope glycopeptides with 8 to 11 amino acids.

We further analyzed the pronase digests of chemically synthesized 13-residue glycopeptides. In contrast to trypsin, or Arg-C-specific endopeptidase clostripain, which are site-specific, the pronase cleaves peptide-bonds without sequence-specificity. The cleaved glycopeptides could be detected as listed in Table 1.

For the pronase digests of RPAPGS(Tn)TAPPAHG, the 11-, 10- and 9-residue glycopeptides were successfully identified with unambiguous glycosite localization (Table 1). The 11-residue was observed with 9 matched fragment ions (Figure S1); b6*/c6* and y5/z5 are the four theoretical glycosite-determining fragment ions, and c6-1+* was actually observed. The 10-residue was observed with 11 matched fragment ions (Figure S2); b6*/c6* and y4/z4 are the four theoretical glycosite-determining fragment ions, and c6-1+* was actually observed. The 9-residue was observed with 10 matched fragment ions (Figure S3); b4*/c4* and y5/z5 are the four theoretical glycosite-determining fragment ions, and c4-1+* was actually observed.

For the pronase digests of RPAPGST(Tn)APPAHG, the 11-, 10-, 9- and 8-residue glycopeptides were successfully identified with unambiguous glycosite localization (Table 1) as well. The 11-residue was observed with 6 matched fragment ions (Figure S4); b6/c6 and y5*/z5* are the four theoretical glycosite-determining fragment ions, and c6-1+ was actually observed. The 10-residue was observed with 5 matched fragment ions (Figure S5); b6/c6 and y4*/z4* are the four theoretical

8

glycosite-determining fragment ions, and c6-1+ was actually observed. The 9-residue was observed with 14 matched fragment ions (Figure S6); b4/c4 and y5*/z5* are the four theoretical glycosite-determining fragment ions, and b4-1+, c4-1+ and y5-1+* were actually observed. The 8-residue was observed with 8 matched fragment ions (Figure S7); b4/c4 and y4*/z4* are the four theoretical glycosite-determining fragment ions, and b4-1+ and c4-1+ were actually observed.

Discussion

Feasibility to distinguish MUC1 glycopeptide isomers shorter than 10 amino acids by LC-MS methods. Our data showed that it is feasible to distinguish glycosite isomers of glycopeptides by LC-MS/MS method. The signature glycosite-determining fragment ions together with differential retention time can clearly separate the paired glycosite isomers of RPAPGS(Tn)TAPPA vs. RPAPGST(Tn)APPA, RPAPGS(Tn)TAPP vs. RPAPGST(Tn)APP, and APGS(Tn)TAPPA vs. APGST(Tn)APPA (Table 1).

Feasibility to generate short MUC1 glycopeptide libraries by pronase. The TR region of MUC1 is resistant to trypsin digestion. Previous attempts to digest MUC1 by Arg-C-specific endopeptidase clostripain could generate heterogenous icosapeptides, but isomeric 20-residue glycopetides could not be separated by liquid chromatography (32-33). To generate minimal epitope length glycopeptides for LC-MS identification, we digested the 13-residue glycopeptides by pronase from *Streptomyces griseus*. We could generate 9-, 10- and 11-residue short glycopeptide isomers which could be clearly identified by LC-MS/MS.

9

Among the pronase digests of RPAPGST(Tn)APPAHG, the 7-residue APGS(Tn)TAP was well observed in the MS spectrum (Figure S8A), and 8 matched fragment ions were observed in its EThcD MS/MS spectrum (Figure S8B); b4*/c4* and y3/z3 are the four theoretical glycosite-determining fragment ions, but none of them was observed. Among the pronase digests of RPAPGST(Tn)APPAHG, the 7-residue APGST(Tn)AP was well observed in the MS spectrum (Figure S9A), and 7 matched fragment ions were observed in its EThcD MS/MS spectrum (Figure S9B); , b4/c4 and y3*/z3* are the four theoretical glycosite-determining fragment ions, but none of them was observed.

The 12-residue digest of both RPAPGS(Tn)TAPPAHG and RPAPGST(Tn)APPAHG, and the 8-residue digest of the former were not identified in the current study; most likely, they are not abundantly produced during pronase digestion. Alternatively, it might be due to their inefficient ionization.

In summary, our data established the LC-MS/MS identities of a clinically-relevant MUC1 glycopeptide neoantigen epitope centering GSTA motif. A library of short MUC1 glycopeptides centered on GSTA motif was created, which is a critical step for analysis of such antigen epitopes in real biological samples.

Acknowledgements

Author contribution statement

DZ designed this study. DZ, KX, and ZT contributed to the collection, analysis and interpretation of data. DZ and ZT wrote the manuscript. All authors read and approved the final manuscript.

References

1. Cheever MA1, Allison JP, Ferris AS, Finn OJ, Hastings BM, Hecht TT, Mellman I, Prindiville SA, Viner JL, Weiner LM, Matrisian LM. The prioritization of cancer antigens: a national cancer institute pilot project for the acceleration of translational research. Clin Cancer Res. 2009 Sep 1;15(17):5323-37.

2. Zhou D, Xu L, Huang W, Tonn T. Epitopes of MUC1 Tandem Repeats in Cancer as Revealed by Antibody Crystallography: Toward Glycopeptide Signature-Guided Therapy. Molecules. 2018 May 31;23(6).

3. Song W, Delyria ES, Chen J, Huang W, Lee JS, Mittendorf EA, Ibrahim N, Radvanyi LG, Li Y, Lu H, Xu H, Shi Y, Wang LX, Ross JA, Rodrigues SP, Almeida IC, Yang X, Qu J,

Schocker NS, Michael K, Zhou D. MUC1 glycopeptide epitopes predicted by computational glycomics. Int J Oncol. 2012 Dec;41(6):1977-84.

4. Qu J, Yu H, Li F, Zhang C, Trad A, Brooks C, Zhang B, Gong T, Guo Z, Li Y, Ragupathi G, Lou Y, Hwu P, Huang W, Zhou D. Molecular basis of antibody binding to mucin glycopeptides in lung cancer. Int J Oncol. 2016 Feb;48(2):587-94.

5. Barnea E, Beer I, Patoka R, Ziv T, Kessler O, Tzehoval E, Eisenbach L, Zavazava N, Admon A. Analysis of endogenous peptides bound by soluble MHC class I molecules: a novel approach for identifying tumor-specific antigens. Eur J Immunol. 2002 Jan;32(1):213-22.

6. Carmon, L., El-Shami, K. M., Paz, A., Pascolo, S., Tzehoval, E., Tirosh, B., Koren, R., Feldman, M., Fridkin, M., Lemonnier, F. A. and Eisenbach, L., Novel breast-tumor-associated MUC1-derived peptides: characterization in Db-/- × beta2 microglobulin null mice transgenic for a chimeric HLA-A2.1/Db-beta2 microglobulin single chain. Int. J. Cancer 2000. 85: 391–397.

7. Hanisch, F.-G., Schwientek, T., Von Bergwelt-Baildon, M.S., Schultze, J.L., Finn, O.J., O-linked glycans control glycoprotein processing by antigen-presenting cells: a biochemical approach to the molecular aspects of MUC1 processing by dendritic cells. 2003. Eur. J. Immunol. 33, 3242–3254.

8. Ninkovic T, Kinarsky L, Engelmann K, Pisarev V, Sherman S, Finn OJ, Hanisch FG. Identification of O-glycosylated decapeptides within the MUC1 repeat domain as potential MHC class I (A2) binding epitopes. Mol Immunol. 2009 Nov;47(1):131-40

9. Tsai PL, Chen SF. A Brief Review of Bioinformatics Tools for Glycosylation Analysis by Mass Spectrometry. Mass Spectrom (Tokyo). 2017;6(Spec Iss):S0064.

10. Woodin CL, Maxon M, Desaire H. Software for automated interpretation of mass spectrometry data from glycans and glycopeptides. Analyst. 2013 May 21;138(10):2793-803.

11. 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 Jun;21(6):660-6.

12. Zhang H, Aebersold R. Isolation of glycoproteins and identification of their N-linked glycosylation sites. Methods Mol Biol. 2006;328:177-85.

13. Nita-Lazar A, Haltiwanger RS. Methods for analysis of O-linked modifications on epidermal growth factor-like and thrombospondin type 1 repeats. Methods Enzymol. 2006;417:93-111.

14. Tian Y, Zhou Y, Elliott S, Aebersold R, Zhang H. Solid-phase extraction of N-linked glycopeptides. Nat Protoc. 2007;2(2):334-9.

15. Zou Z, Ibisate M, Zhou Y, Aebersold R, Xia Y, Zhang H. Synthesis and evaluation of superparamagnetic silica particles for extraction of glycopeptides in the microtiter plate format. Anal Chem. 2008 Feb 15;80(4):1228-34.

16. Carapito C, Klemm C, Aebersold R, Domon B. Systematic LC-MS analysis of labile post-translational modifications in complex mixtures. J Proteome Res. 2009 May;8(5):2608-14.

17. North SJ, Hitchen PG, Haslam SM, Dell A. Mass spectrometry in the analysis of N-linked and O-linked glycans. Curr Opin Struct Biol. 2009 Oct;19(5):498-506.

18. Jensen PH, Kolarich D, Packer NH. Mucin-type O-glycosylation--putting the pieces together. FEBS J. 2010 Jan;277(1):81-94

19. Zauner G, Kozak RP, Gardner RA, Fernandes DL, Deelder AM, Wuhrer M. Protein O-glycosylation analysis. Biol Chem. 2012 Aug;393(8):687-708.

20. Zhang Y, Yin H, Lu H. Recent progress in quantitative glycoproteomics. Glycoconj J. 2012 Aug;29(5-6):249-58.

21. Steentoft C, Vakhrushev SY, Joshi HJ, Kong Y, Vester-Christensen MB, Schjoldager KT, Lavrsen K, Dabelsteen S, Pedersen NB, Marcos-Silva L, Gupta R, Bennett EP, Mandel U, Brunak S, Wandall HH, Levery SB, Clausen H. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. EMBO J. 2013 May 15;32(10):1478-88.

22. Dallas DC, Martin WF, Hua S, German JB. Automated glycopeptide analysis--review of current state and future directions. Brief Bioinform. 2013 May;14(3):361-74.

23. Liu Y, Chen J, Sethi A, Li QK, Chen L, Collins B, Gillet LC, Wollscheid B, Zhang H, Aebersold R. Glycoproteomic analysis of prostate cancer tissues by SWATH mass spectrometry discovers N-acylethanolamine acid amidase and protein tyrosine kinase 7 as signatures for tumor aggressiveness. Mol Cell Proteomics. 2014 Jul;13(7):1753-68.

24. Thaysen-Andersen M, Packer NH. Advances in LC-MS/MS-based glycoproteomics: getting closer to system-wide site-specific mapping of the N- and O-glycoproteome. Biochim Biophys Acta. 2014 Sep;1844(9):1437-52.

25. Sjöström M, Ossola R, Breslin T, Rinner O, Malmström L, Schmidt A, Aebersold R, Malmström J, Niméus E. A Combined Shotgun and Targeted Mass Spectrometry Strategy for Breast Cancer Biomarker Discovery. J Proteome Res. 2015 Jul 2;14(7):2807-18.

26. Chandler KB, Costello CE. Glycomics and glycoproteomics of membrane proteins and cell-surface receptors: Present trends and future opportunities. Electrophoresis. 2016 Jun;37(11):1407-19.

27. Narimatsu H, Kaji H, Vakhrushev SY, Clausen H, Zhang H, Noro E, Togayachi A, Nagai-Okatani C, Kuno A, Zou X, Cheng L, Tao SC, Sun Y. Current Technologies for Complex

Glycoproteomics and Their Applications to Biology/Disease-Driven Glycoproteomics. J Proteome Res. 2018 Oct 25. doi: 10.1021/acs.jproteome.8b00515.

28. Yu A, Zhao J, Peng W, Banazadeh A, Williamson SD, Goli M, Huang Y, Mechref Y. Advances in mass spectrometry-based glycoproteomics. Electrophoresis. 2018 Dec;39(24):3104-3122.

29. Mulagapati S, Koppolu V, Raju TS. Decoding of O-Linked Glycosylation by Mass Spectrometry. Biochemistry. 2017 Mar 7;56(9):1218-1226.

30. Zhang Q, Li Z, Wang Y, Zheng Q, Li J. Mass spectrometry for protein sialoglycosylation. Mass Spectrom Rev. 2018 Sep;37(5):652-680.

31. Yang W, Ao M, Hu Y, Li QK, Zhang H. Mapping the O-glycoproteome using site-specific extraction of O-linked glycopeptides (EXoO). Mol Syst Biol. 2018 Nov 20;14(11):e8486. doi: 10.15252/msb.20188486.

32. Müller S, Alving K, Peter-Katalinic J, Zachara N, Gooley AA, Hanisch FG. High density O-glycosylation on tandem repeat peptide from secretory MUC1 of T47D breast cancer cells. J Biol Chem. 1999 Jun 25;274(26):18165-72.

33. Sihlbom C, van Dijk Härd I, Lidell ME, Noll T, Hansson GC, Bäckström M. Localization of O-glycans in MUC1 glycoproteins using electron-capture dissociation fragmentation mass spectrometry. Glycobiology. 2009 Apr;19(4):375-81.

34. Xiao KJ, Yu F, Tian ZX. Top-down protein identification using isotopic envelope fingerprinting. J Proteomics. 2017;152:41-47.

35. Li SS, Zhou Y, Xiao KJ, Li J, Tian ZX. Selective fragmentation of the N-glycan moiety and protein backbone of ribonuclease B on an Orbitrap Fusion Lumos Tribrid mass spectrometer. Rapid Commun Mass Sp. 2018;32(23):2031-2039.

36. Wu H, Xiao KJ, Tian ZX. Top-down characterization of chicken core histones. J Proteomics. 2018;184:34-38.

37. Kaijie-Xiao, Yu F, Fang HQ, et al. Are neutral loss and internal product ions useful for top-down protein identification? J Proteomics. 2017;160:21-27.

38. Xiao KJ, Tian ZX. Top-down characterization of histone H4 proteoforms with ProteinGoggle 2.0. Chin J Chromatogr. 2016;34(12):1255-1263.

39. Fang HQ, Xiao KJ, Li YH, et al. Intact Protein Quantitation Using Pseudoisobaric Dimethyl Labeling. Anal Chem. 2016;88(14):7198-7205.

Figure legends

Figure 1. MS and MS/MS spectra for RPAPGS(Tn)TAPPAHG. (A) The isotopic envelope fingerprinting map of the precursor ion. Red squares indicate theoretical mass spectra, and black bars indicate experimental mass spectra. (B) The annotated MS/MS spectrum with the graphical fragmentation map; the fragment ions containing GalNAc are marked in red box. The glyco-site determining fragment ions are $c6-1^{+}$* (*=GalNAc, m/z 786.40936) and z7-1+ (m/z 634.30573). Red bars indicate that the experimental mass spectra match with the theoretical mass spectra.
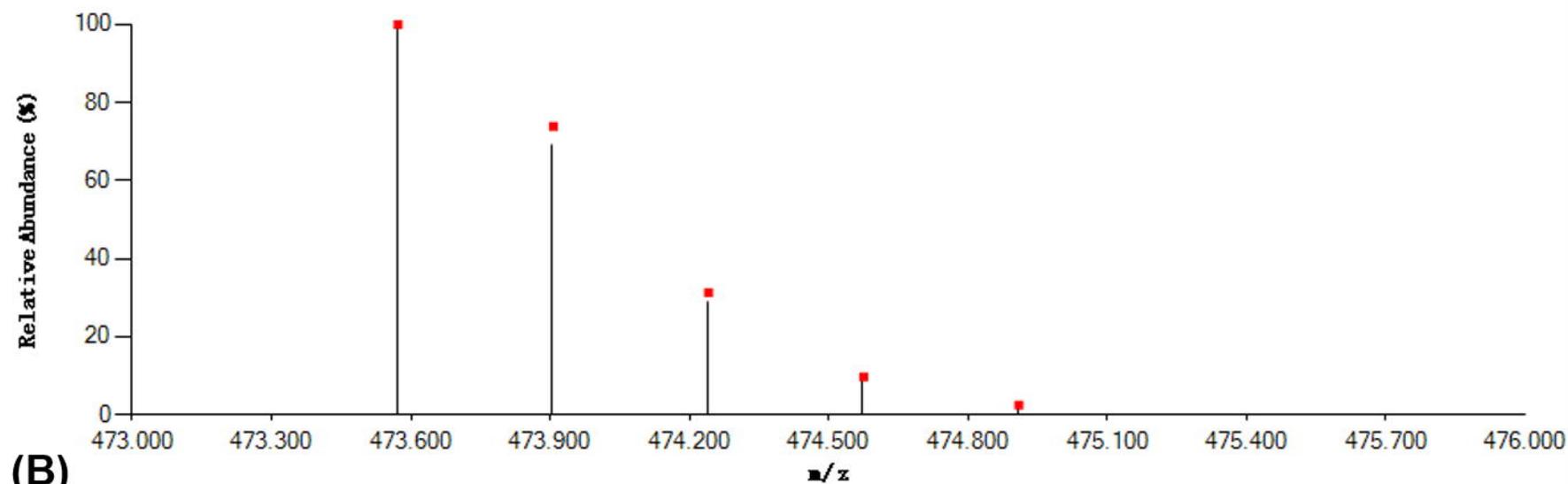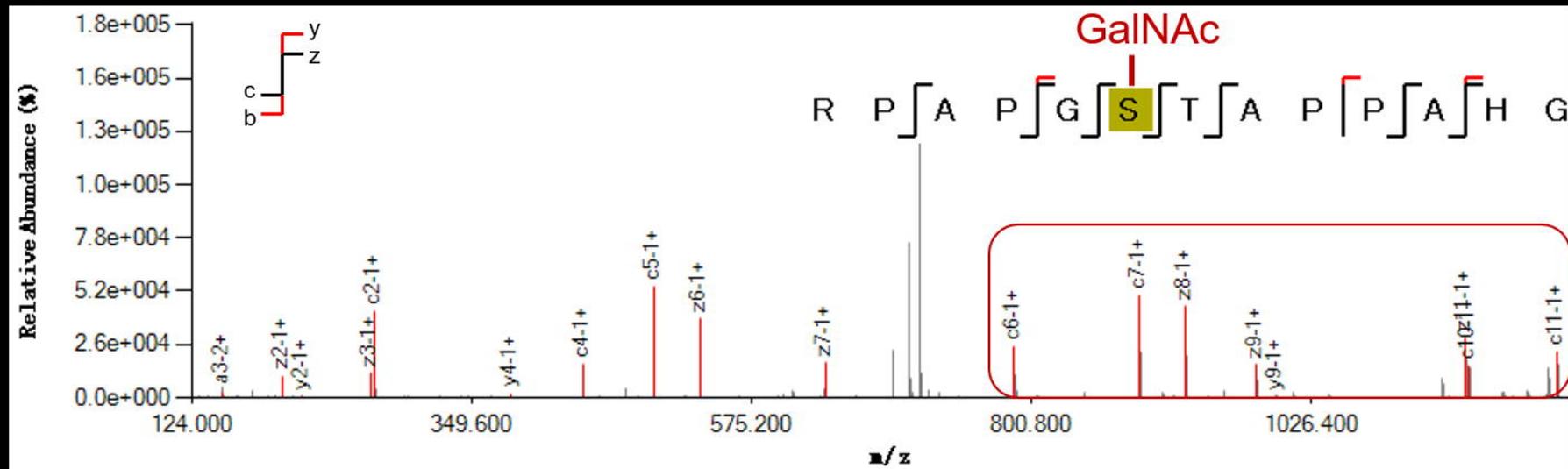
Figure 2. MS and MS/MS profiles for RPAPGST(Tn)APPAHG. (A) The isotopic envelope fingerprinting map of the precursor ion. Red squares indicate theoretical mass spectra, and black bars indicate experimental mass spectra. (B) The annotated MS/MS spectrum with the graphical fragmentation map; the fragment ions containing GalNAc are marked in red box. The glyco-site determining ions are c6-1+ (m/z 583.33026) and z7-1+* (*=GalNAc, m/z 837.38580). Red bars indicate that the experimental mass spectra match with the theoretical mass spectra.

17

Table 1. Short glycopeptides from pronase digests of RPAPGS(Tn)TAPPAHG and RPAPGST(Tn)APPAHG

	Amino acid sequence	Sequence length	Site-determining product ions	MS-MS profile	Retention time (min)
-GS(Tn)TA-	RPAPGS(Tn)TAPPA	11	c6-1+* (m/z 786.41089)	Figure S1	15.24
	RPAPGS(Tn)TAPP	10	c6-1+* (m/z 786.40985)	Figure S2	11.11
	APGS(Tn)TAPPA	9	c4-1+*(m/z 533.25671)	Figure S3	12.17
-GST(Tn)A-	RPAPGST(Tn)APPA	11	c6-1$^+$ (m/z 583.33124)	Figure S4	11.15
	RPAPGST(Tn)APP	10	c6-1$^+$ (m/z 583.33008)	Figure S5	11.10
	APGST(Tn)APPA	9	b4-1+ (m/z 313.15085) c4-1+ (m/z 330.17761) y5-1$^+$* (m/z 659.32452)	Figure S6	12.00
	APGST(Tn)APP	8	b4-1+ (m/z 313.15097) c4-1+ (m/z 330.17856)	Figure S7	12.86

(A)

IPAD   0   -4   -2   0   0
IPMD   -1   -3   -4   -4   -3

(B)

GalNAc

R P⌐A P⌐G⌐S⌐T⌐A P⌐P⌐A⌐H G

(A)

IPAD   0   4   1   0   0
IPMD   -1   -3   -4   -4   -4

(B)

GalNAc

R P⌐A⌐P G⌐S⌐T⌐A P⌐P⌐A⌐H G