Simultaneous enrichment and separation of neutral and sialyl glycopeptides of SARS-CoV-2 spike protein enabled by dual-functionalized Ti-IMAC material

Junfeng Huang¹ · Danqing Wang² · Richard David Shipman¹ · Zexin Zhu¹ · Yuan Liu¹ · Lingjun Li^{1,2}

Received: 1 April 2021 / Revised: 19 May 2021 / Accepted: 25 May 2021 / Published online: 21 June 2021
© Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

The coronavirus disease 2019 (COVID-19) pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) presents a serious threat to human health all over the world. The development of effective vaccines has been focusing on the spike (S) glycoprotein, which mediates viral invasion to human cells through its interaction with the angiotensin-converting enzyme 2 (ACE2) receptor. In this work, we perform analytical characterization of N- and O-linked glycosylation of the SARS-CoV-2 S glycoprotein. We explore the novel use of dual-functionalized titanium (IV)-immobilized metal affinity chromatography (Ti-IMAC) material for simultaneous enrichment and separation of neutral and sialyl glycopeptides of a recombinant SARS-CoV-2 S glycoprotein from HEK293 cells. This strategy helps eliminate signal suppression from neutral glycopeptides for the detection of sialyl glycopeptides and improves the glycoform coverage of the S protein. We profiled 19 of its 22 potential N-glycosylated sites with 398 unique glycoforms using the dual-functional Ti-IMAC approach, which exhibited improvement of coverage by 1.6-fold compared to the conventional hydrophilic interaction chromatography (HILIC) glycopeptide enrichment method. We also identified O-linked glycosylation site that was not found using the conventional HILIC approach. In addition, we reported on the identification of mannose-6-phosphate (M6P) glycosylation, which substantially expands the current knowledge of the spike protein's glycosylation landscape and enables future investigation into the influence of M6P glycosylation of the spike protein on its cell entry.

Keywords SARS-CoV-2 spike protein · Neutral and sialyl glycopeptide separation · O-Linked glycosylation · Mannose-6-phosphate (M6P) glycosylation · Posttranslational modifications · IMAC

Introduction

The global COVID-19 pandemic caused by SARS-CoV-2 has spread rapidly since December 2019 [1]. RNA sequencing of SARS-CoV-2 shows that it belongs to the beta-coronavirus genus and is most closely related to SARS-CoV [2]. The structural proteins of mature SARS-CoV-2 include spike (S) protein, envelope (E) protein, membrane (M) protein, and

nucleocapsid (N) protein. All structural proteins can serve as antigens for vaccine development or targets for anti-viral treatment [3]. Among these proteins, the transmembrane S protein is a homotrimer and protrudes from the virus surface, which enables its binding to the angiotensin-converting enzyme 2 (ACE2) on host cells to promote fusion of viral and host cell membranes [4]. Since both the viral spike protein and human ACE2 are glycoproteins, their glycosylation patterns can affect their interactions and consequent vaccine design. Given its key role in virus entry and infectivity, S protein is an ideal target for developing vaccines and medical treatment to block or interfere the interaction between the virus and the host cell. Therefore, comprehensive profiling of the glycosylation patterns of the S protein is urgently needed to address the challenges posed by the pandemic.

Each spike protein consists of S1 and S2 subunits; the S1 subunit mediates binding of the virus to the ACE2 receptor, while the S2 subunit enables fusion of the virion with the cell

Published in the topical collection *Analytical Characterization of Viruses* with guest editor Joseph Zaia.

✉ Lingjun Li
lingjun.li@wisc.edu

¹ School of Pharmacy, University of Wisconsin-Madison, 5125 Rennebohm Hall, 777 Highland Avenue, Madison, WI 53705, USA

² Department of Chemistry, University of Wisconsin-Madison, Madison, WI 53706, USA

membrane and initiates viral entry. Each monomer of the S protein consists of 22 predicted N-linked glycan sequons and all of them are occupied by complex, hybrid, and high-mannose glycan structures, which have been confirmed by both LC-MS/MS and Cryo-EM [5–7]. Furthermore, three O-glycosylation sites T323, S325, and T678 were identified in published data [8, 9]. Glycans on the S protein are important factors that can affect protein folding, stability, and potential target for immunization. Since glycans can shield the amino acid residues and other epitopes from cells and antibody recognition, glycosylation is able to make coronavirus evade both the innate and adaptive immune responses [10]. Blocking N- and O-glycan modification of the S protein will inhibit the SARS-CoV-2 viral entry [11]. Therefore, comprehensive characterization of the glycoforms and glycosylation sites of the S protein will facilitate the study of SARS-CoV-2 viral entry mechanism and the screening for inhibitors of glycosylation that may help develop the treatment of COVID-19 [12, 13].

In this study, to overcome the ion suppression effects from neutral glycopeptides for the detection of sialylated glycopeptides, we perform dual-functional Ti-IMAC enrichment for a recombinant SARS-CoV-2 S protein expressed in human embryonic kidney (HEK 293) cell line. Ti-IMAC was developed as a technique for enrichment of negatively charged phosphopeptides through electrostatic interaction [14, 15]. We later found that Ti-IMAC material contains numerous hydroxyl, amine, and phosphate groups, which is highly hydrophilic and enables enrichment of glycopeptides through hydrophilic interactions [16]. Therefore, the negatively charged sialylated glycopeptides can be captured by Ti-IMAC through synergetic electrostatic and hydrophilic interactions, while neutral glycopeptides were only captured through hydrophilic interaction. By performing a two-step elution, this strategy enables the simultaneous enrichment and separation of neutral and sialyl glycopeptides, which improves the glycoform coverage of the S protein compared with the conventional HILIC method.

Materials and methods

Protein source and reagents

The recombinant full-length SARS-CoV-2 spike protein (40589-V08H4, Met1-Pro1213-His-tag; R683A, R685A, F817P, A892P, A899P, A942P, K986P, V987P) expressed in the HEK 293 cell line was obtained from Sino Biological US Inc. (Wayne, PA, USA). The purity of the purchased protein was >90%. Trypsin and chymotrypsin, sequencing grade, were obtained from Promega (Madison, WI, USA). Dithiothreitol (DTT), iodoacetamide (IAA), urea, and TEAB

buffer were purchased from Sigma-Aldrich (St. Louis, MO, USA).

Protein digestion

Aliquots of SARS-CoV-2 S protein were dissolved in 8 M urea/50 mM TEAB buffer (pH 8) to a final concentration of 8 mg/mL. The S protein was reduced with 10 mM dithiothreitol (DTT) for 120 min at 37 °C, and then alkylated with 30 mM IAA for 30 min in the dark at room temperature. The protein solution was diluted to 1 M urea with 50 mM TEAB and then trypsin was added to the samples with protein:trypsin ratio at 50:1 and incubated at 37 °C water bath for 8 h. After trypsin digestion, the protein was continually digested by chymotrypsin with protein:chymotrypsin ratio at 50:1 and incubated at 37 °C water bath for 16 h.

Glycopeptide enrichment

Aliquots of 2 µg digested S proteins were dried down in SpeedVac and prepared for either HILIC enrichment or dual-functional Ti-IMAC enrichment:

HILIC enrichment was performed with homemade HILIC stage-tip which used a neutral, polar material (PolyHYDROXYETHYL A: 12 µm and 300 Å from PolyLC Inc. (Columbia, MD, USA)). Briefly, 2 mg of cotton wool was weighed and inserted into a 200-µL pipette tip. The pipette tip was placed on a 2-mL microcentrifuge tube with the help of an adapter unit. HILIC beads were dissolved into 1% trifluoroacetic acid (TFA) solvent and the HILIC beads slurry was vortexed for 15 min to activate the HILIC beads. The HILIC slurry was transferred onto the top of the cotton wool, which was then centrifuged to remove the solvent at 500×g for 2 min. A total of 200 µL 1% TFA was added to flush the beads and removed at 500×g for 2 min. HILIC beads were conditioned with 200 µL 80% ACN/1% TFA three times. Then, 2 µg tryptic digested S protein sample was dissolved in 100 µL 80% ACN/1% TFA and loaded onto the HILIC stage-tip. The stage-tip was centrifuged at 500×g for 2 min and flow through of the sample was re-loaded to the HILIC stage-tip twice. The non-glycopeptides were washed away with 100 µL 80% ACN/1% TFA three times, and the glycopeptides were eluted with 100 µL 0.1% FA solution twice. The sample was immediately dried down in SpeedVac. Samples were stored at −20 °C and resuspended in 0.1% formic acid (FA) before LC-MS/MS analysis and 20% of each sample was loaded for analysis.

Dual-functional Ti-IMAC enrichment was performed in HILIC mode. A total of 2 µg S protein tryptic digest sample was loaded onto 100 µg Ti(IV)-IMAC materials with 100 µL 80% ACN/3%TFA buffer, which was similar to the conventional HILIC enrichment. An aqueous layer was generated across the substantial hydroxyl, amine groups, and phosphate

chelated Ti (IV) ions on the material surface. The beads were washed twice with 100 μ L 80% ACN/0.1% TFA buffer. After enrichment, the standard protein samples were eluted with the following two elution steps: (1) 100 μ L 0.1% FA (v/v) twice, (2) 100 μ L 6% TFA (v/v) twice. Samples were immediately dried down in SpeedVac. Samples were stored at -20°C and resuspended in 0.1% FA. For LC-MS/MS analysis, 20% of each sample was loaded.

LC-MS/MS analysis

Enriched S protein glycopeptides were analyzed by Ultimate 3000 nanoLC coupled to an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific, San Jose, CA). Resuspended samples were injected onto a 75 μm i.d. $\times$ 15 cm length homemade column with an integrated HF etched emitter tip and packed with 1.7 μm , 130 Å, BEH C₁₈ material from a Waters UPLC column (Waters, Milford, MA). Peptides were separated with a gradient that ramped from 97% solvent A (0.1% FA in H₂O) and 3% solvent B (0.1% FA in ACN) to 30% solvent B over 80 min and then kept each of 75% solvent B, 95% solvent B, and 100% solvent A for 10 min. The flow rate was set at 300 nL/min. MS method was set according to the following parameters: MS scan range (m/z) = 400–2000; resolution = 120,000; AGC target = 1.0e^6 ; maximum injection time = 250 ms; included charge state = 2–6; dynamic exclusion duration = 30 s. MS/MS method was a top 20 data-dependent acquisition (DDA) mode in which all MS/MS dissociations were performed with stepped higher energy collisional dissociation (step-HCD): resolution = 60,000; AGC target = 1.0e^5 ; maximum injection time = 150 ms; stepped collision energy = 22%, 30%, 38%.

Data processing Byonic V3.10 (Protein Metrics, San Carlos, CA) was used to analyze the acquired MS and MS/MS spectra of intact glycopeptides. An advantage of this software was that it allowed manually adding M6P glycans in the glycan database for searching. Raw files (Data are available via ProteomeXchange with identifier PXD026450.) were searched against SARS-CoV-2 spike protein sequences downloaded from Uniprot with proper mutation added. Precursor ion mass tolerance of 10 ppm and fragment ion mass tolerance of 20 ppm were selected. Lysine (K), arginine (R), tyrosine (Y), tryptophan (W), phenylalanine (F), leucine (L), and methionine (M) were set as cleavage sites and up to 4 mis-cleavage sites were allowed. The phosphorylation of serine (S), threonine (T), and tyrosine (Y) and the oxidation of methionine (M) were set as variable modifications. Meanwhile, the carbamidomethylation of cysteine (C) was set as fixed modification. Common N-linked glycopeptide searching used a human N-glycome database containing 182 glycans and a human O-glycome database containing 70 glycans. In addition, M6P glycopeptides were searched by

expanding the N-glycome database to include N-linked M6P glycans consisting of HexNAc (2–4)-Hex (3–9)-Phospho (1–2) modification. Peptide identifications were filtered at two-dimensional false discovery rate (2D FDR) <1%, PEP 2D <0.05, |Log Prob| > 1, and Byonic Score > 150. Manual inspection of MS/MS spectra of M6P glycopeptides was performed to examine if Byonic identification results contained phosphorylated hexose diagnostic ions. All searches were allowed for a maximum total of two common modifications. Glycoforms were categorized according to the following rules: (1) high mannose: glycan composition is “HexNAc(2)Hex(5–9)”; (2) paucimannose: glycan composition contains variations in “HexNAc(1–2)Hex(0–4)”; (3) mannose-6-phosphate glycosylation: glycan composition contains “phospho”; (4) sialylation: glycan composition contains “NeuAc”; (5) fucosylation: glycan composition contains “Fuc” but without “NeuAc”; (6) hybrid and complex: any glycan composition not categorized into the above five types.

Results and discussion

N-Glycopeptide profiling

We performed the site-specific glycoform analysis of SARS-CoV-2 S full-length protein with strategies including (1) double enzyme digestion with trypsin and chymotrypsin. Trypsin cleaves carboxy-terminal to lysine and arginine residues, while chymotrypsin prefers to cleave carboxy-terminal to large hydrophobic amino acid (tyrosine (Y), tryptophan (W), and phenylalanine (F)) and leucine (L) and methionine (M) at slower rates, which is complementary to trypsin [17]. The combination of trypsin and chymotrypsin digestion will make the site-specific mapping of N-linked glycopeptide bearing two sites more accurate [18–20]. For example, R.SSVLHSTQDLFLPFFSN₆₁VTWFHAIHVSGTN₇₄GTK.R and R.FPN₃₃₁ITNLC[+57.02146]PFGEVFN₃₄₃ATR.F are generated by trypsin-only digestion and they both have two potential N-glycosylation sites, which are more difficult to be precisely localized. After additional chymotrypsin digestion, both peptides can be cleaved between the two N-glycosylation sites, producing peptides that bear only one N-glycosylation site, which are F.FSN₆₁VTWF.H and F.HAIHVSGTN₇₄GTK.R, and R.FPN₃₃₁ITNLC[+57.02146]PF.G and F.GEVFN₃₄₃ATR.F, respectively. (2) Sample preparation through a novel dual-functional Ti-IMAC enrichment strategy. Samples were loaded on dual-functional Ti-IMAC materials in a HILIC-mode, where neutral glycopeptides were captured by Ti-IMAC material through hydrophilic interaction, and the negatively charged sialylated glycopeptides were captured through synergetic hydrophilic and electrostatic interactions. A two-step elution was performed in this case to release and separate neutral and negatively charged

glycopeptides. Aqueous buffer in mild acidic condition will mainly elute neutral glycopeptides, while aqueous buffer in strong acidic condition with the ability to protonate sialic acid will elute the sialylated glycopeptides. This approach enables simultaneous enrichment and separation of neutral and sialyl glycopeptides (Fig. 1). As a result, the ion suppression effects from neutral glycopeptides in the detection of sialylated glycopeptides would be eliminated and the glycoform coverage of the S protein would be improved compared with conventional HILIC method [16, 21, 22]. (3) Stepped higher energy collisional dissociation (HCD) fragmentation for better fragmentation efficiency. We used high-resolution LC-MS/MS at both MS1 and MS2 levels with stepped HCD fragmentation to enhance the fragmentation efficiency for both peptide backbone and its attaching glycans, which further improve the confidence in glycopeptide identification [23, 24].

As a comparison, the conventional HILIC method enabled profiling of 18 N-linked glycosylation sites among the 22 potential N-linked glycosites, while our dual-functional Ti-IMAC approach enabled profiling of 19 N-linked glycosites. These two methods appear to have similar performance in glycosylation site profiling (Table 1). However, as each glycosylation site bears multiple glycans, the heterogeneity of glycopeptides is much more complex than other PTMs. As shown in Fig. 2a and b, we identified a total of 624 unique glycopeptides that were assigned to 398 unique glycoforms (a unique glycoform means a glycosite with a unique glycan composition) distributed on 19 glycosylation sites of the S

Table 1 Overview of profiled N- and O- glycosylation sites of the SARS-CoV-2 S protein by two different methods

Potential glycosylation site		HILIC	Ti-IMAC
N-Linked	N17	Y	Y
	N61	Y	Y
	N74	Y	Y
	N122	Y	Y
	N149	Y	Y
	N165	Y	Y
	N234	Y	Y
	N282	Y	Y
	N331	Y	Y
	N344	Y	Y
	N616	Y	Y
	N657	Y	Y
	N801	Y	Y
	N1074	Y	Y
	N1098	Y	Y
	N1134	N/A	N/A
	N1158	N/A	N/A
	N1173	N/A	N/A
	N1194	N/A	Y
O-Linked	T323	N/A	Y
	S325	N/A	Y or N/A
	T678	N/A	N/A

Fig. 1 Scheme of Ti-IMAC and HILIC dual-mode affinity enrichment approach for the simultaneous enrichment and separation of neutral and sialyl glycopeptides

protein by performing the dual-functional Ti-IMAC enrichment approach, suggesting that each N-glycosylation site on the S protein possesses an average of 21 unique glycoforms. The conventional HILIC method enabled profiling of 307 unique glycopeptides and 247 unique glycoforms, which was much fewer than the number obtained by the dual-functional Ti-IMAC enrichment approach (see Supplementary Information (ESM) Table S1).

As for the sialylated glycopeptides, we found that only 37 unique sialylated glycopeptides attributed to 32 unique sialylated glycoforms were identified in the weak acidic elution buffer, while 148 unique sialylated glycopeptides assigned to 99 unique sialylated glycoforms were profiled in strong acidic elution buffer (Fig. 2a, b). These data demonstrated that the sialylated glycopeptides were separated well from the neutral glycopeptides in this two-step elution procedures. The overlap between profiled N-glycopeptides and sialylated N-glycopeptides, N-glycoforms, and sialylated N-glycoforms by conventional HILIC method and dual-functional Ti-IMAC enrichment approach was also calculated. As shown in Fig. 3a, 87.0% N-glycopeptides identified by conventional HILIC method were found in the results obtained using the dual-functional Ti-IMAC enrichment approach, while 57.2% N-glycopeptides were uniquely found in the

Fig. 2 Comparison of profiled N-glycopeptides (a) and N-glycoforms (b) by HILIC and dual-functional Ti-IMAC enrichment (Ti-IMAC FA and Ti-IMAC TFA represent the weak)

Fig. 3 Overlap of profiled N-glycopeptides and sialylated N-glycopeptides (a, b); and N-glycoforms and sialylated N-glycoforms (c, d) by HILIC and dual-functional Ti-IMAC enrichment

latter approach. The increased coverage of S protein glycosylation mapping demonstrated the advantages of this new method. The trend of overlaps between N-glycopeptides and sialylated N-glycopeptides, N-glycoforms, and sialylated N-glycoforms by HILIC and dual-functional Ti-IMAC enrichment approach appeared to be quite similar (Fig. 3b, c, d).

In Fig. 4, a representative MS2 spectrum of an annotated N-linked glycopeptide generated by our double-enzyme digestion strategy with trypsin and chymotrypsin was shown. Nonetheless, there is still a limitation of this strategy for SARS-CoV-2 S protein glycosylation profiling. The large tryptic peptide R.AAASVASQSIIAYTMSLGAENSVAYS₇₀₉NSIAIPTN₇₁₇FTISVTTEILPVSM₇₁₇TK.T can be trimmed to a shorter peptide Y.SN₇₀₉NSIAIPTN₇₁₇F.T by chymotrypsin digestion; chymotrypsin still lacks the ability to cleave between the two glycosylation sites on such a short peptide. Therefore, the site-specific localization of glycoform containing these two sites might not be accurate and the site-specific annotation provided by the software can only serve as a reference. In the future, accurate localization of glycosylation sites on this peptide may be achieved in two strategies: (1) Add this peptide in the MS2 inclusion list and use electron-transfer/higher-energy collision dissociation (ET_hCD) fragmentation to fragment the peptide backbone. Fragments with intact glycan on the glycosylation site will be generated, so that the glycoforms can be accurately annotated [25, 26]. (2) Digest protein samples with a third protease, alpha-lytic protease, which cleaves after threonine (T), serine (S), alanine (A), and valine (V) residues. With the complementary ability to cleave between N709 and N717, it may serve as a good supplementary protease to both trypsin and chymotrypsin [6].

Mannose-6-phosphate glycopeptide profiling

It is worth mentioning that we also identified a mannose-6-phosphate (M6P) glycopeptide on S protein expressed in HEK293 cells (Fig. 5). The MS/MS spectrum clearly showed that it contains a phosphorylated hexose oxonium ion, which could serve as a unique diagnostic ion of M6P glycopeptides. M6P is initially added to high mannose glycans in the form of GlcNAc-M6P in *cis*-Golgi apparatus, and then GlcNAc-cap is removed in the *trans*-Golgi. It is mainly added to lysosomal enzymes and can be recognized by M6P receptor, which is used to transfer the enzymes to lysosomal. Recently, it has been shown that β -coronaviruses can use lysosomes for egress instead of the biosynthetic secretory pathway [27]. Given that SARS-CoV-2 is a β -coronavirus and we have found that the S protein of SARS-CoV-2 possesses M6P glycosylation, we hypothesize that M6P glycosylation may be involved in the pathway of transferring whole viruses from the *trans*-Golgi to late endosomes/lysosomes and then egressing through lysosomal pathway. Further investigation is needed to evaluate this hypothesis.

Fig. 4 Overview of profiled N- and O-linked glycosylation sites and the representative MS2 spectrum of an annotated N-linked glycopeptide located at the RBD region. The diagnostic ions of sialic acid (NeuAc) were labeled with a blue arrow

O-Glycopeptide profiling

In addition to N-linked glycosylation, we also found one O-linked glycosylation site at T323 with three glycoforms (Table 2). They were identified using the novel dual-functional Ti-IMAC enrichment approach, but failed to be detected using the conventional HILIC method. However, since the delta modification score of identified O-linked glycopeptides was only 2.8, it was possible that

the other site, S325, on this O-linked glycopeptide could also be a potential O-linked glycosylation site (Fig. 6). It is worth noting that the identified O-linked glycosylation site is located at the receptor-binding domain (RBD, 319–541 residues) together with two other N-linked glycosylation sites, N331 and N343 (Fig. 4). Though the mechanism of N- or O-glycosylation at the RBD region has not been thoroughly investigated, they might be involved in stabilization of the interaction between the RBD and viral entry

Fig. 5 A representative MS2 spectrum of an annotated M6P glycopeptide. The diagnostic ion of M6P is indicated with a blue arrow

Table 2 Summary of identified O-glycopeptide of the SARS-CoV-2 S protein by dual-functional Ti-IMAC approach

Position	Sequence	Glycans	PEP 2D	Log Prob	Score	Delta Mod. Score	z	Calc. m/z	ppm err.	Scan time
T323	R.VQPT[+1312.45523]ESIVR.F	HexNAc(2)Hex(2)NeuAc(2)	0.00028	3.55	351.4	2.8	2	1171.018	-2.4	61.1944
	R.VQPT[+656.22761]ESIVR.F	HexNAc(1)Hex(1)NeuAc(1)	2.90E-05	4.54	258.4	2.8	2	842.9042	-1.7	53.9046
	R.VQPT[+947.32303]ESIVR.F	HexNAc(1)Hex(1)NeuAc(2)	2.10E-05	4.67	495.2	2.8	2	988.4519	-0.4	62.7981

target ACE2 [4]. Another notable finding is that the three O-linked glycoforms were all sialylated, while only part of the N331 and N343 glycoforms were sialylated [8]. As a negatively charged monosaccharide, one or more sialic acids can be modified at the terminal of a glycan structure and serve as shields for epitopes or key residues on the RBD region of S protein and protect it from immune system attack [2, 28]. Experimental investigations demonstrated that blocking N- and O-glycan elaboration in whole S protein region will inhibit the SARS-CoV-2 viral entry [11]. Further experimental studies to investigate the effect of glycosylation on some specific regions like the RBD domain are needed to confirm if any of the glycosylation sites are used in SARS-CoV-2 infection [29].

Conclusions

In this work, we performed a comprehensive analysis of the N- and O-linked glycosylation of SARS-CoV-2 S

glycoprotein. We demonstrated that dual-functionalized Ti-IMAC material enabled simultaneous enrichment and separation of neutral and sialyl glycopeptides. This feature helped eliminate signal suppression from neutral glycopeptides for the detection of sialylated glycopeptides and increase the overall glycoproteome coverage of S protein. We profiled 19 of its 22 potential N-glycosylated sites with 398 unique glycoforms using the dual-functional Ti-IMAC approach, which showed 1.6-fold improvement in glycoform coverage compared to the conventional HILIC method. We also identified O-linked glycosylation site that was only found using the dual-functional Ti-IMAC approach. In addition, we profiled mannose-6-phosphate (M6P) glycosylation, which further expands our current knowledge of the spike protein's glycosylation and enables future investigation into the influence of mannose-6-phosphate on its cell entry. These results will prompt further research to understand the role of N- and O-linked glycosylation in viral infection throughout the SARS-CoV-2 process protein, especially in the region of RBD. Furthermore, the developed N- and O-glycosylation

Fig. 6 A representative MS2 spectrum of an annotated O-linked glycopeptide located at the RBD region. The diagnostic ion of sialic acid (NeuAc) is indicated with a blue arrow

profiling approach will provide useful knowledge for elucidating the pathology of viral infection and offer possible strategies for future therapeutics as well as the development of immunogens in the design of suitable vaccines.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00216-021-03433-1>.

Acknowledgements Lingjun Li acknowledges a Vilas Distinguished Achievement Professorship and Charles Melbourne Johnson Distinguished Chair Professorship with funding provided by the Wisconsin Alumni Research Foundation and University of Wisconsin-Madison School of Pharmacy. The authors wish to pay tribute to the late Prof. Hanfa Zou in memory of his great contribution that led the development of Ti(IV)-IMAC materials.

Funding This work was supported, in part, by the National Institutes of Health Grants U01CA231081, R01AG052324, and R01 DK071801. The Orbitrap instruments were purchased through the support of an NIH Shared Instrument Grant (NIH-NCRR S10RR029531) and the University of Wisconsin-Madison, Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation.

Declarations

Conflict of interest The authors declare no competing interests.

References

- Wu F, Zhao S, Yu B, Chen Y-M, Wang W, Song Z-G, et al. A new coronavirus associated with human respiratory disease in China. *Nature*. 2020;579:265–9.
- Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. *Nat Med*. 2020;26:450–2.
- Asghari A, Naseri M, Safari H, Saboori E, Parsamanesh N. The novel insight of SARS-CoV-2 molecular biology and pathogenesis and therapeutic options. *DNA Cell Biol*. 2020;39:1741–53.
- Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell*. 2020;181:271–80.e8.
- Bangaru S, Ozorowski G, Turner HL, Antanasijevic A, Huang D, Wang X, et al. Structural analysis of full-length SARS-CoV-2 spike protein from an advanced vaccine candidate. *Science*. 2020;370:1089–94.
- Watanabe Y, Allen JD, Wrapp D, McLellan JS, Crispin M. Site-specific glycan analysis of the SARS-CoV-2 spike. *Science*. 2020;369:330–3.
- Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh C-L, Abiona O, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*. 2020;367:1260–3.
- Shajahan A, Supekar NT, Gleinich AS, Azadi P. Deducing the N- and O-glycosylation profile of the spike protein of novel coronavirus SARS-CoV-2. *Glycobiology*. 2020;30:981–8.
- Sanda M, Morrison L, Goldman R. N- and O-glycosylation of the SARS-CoV-2 spike protein. *Anal Chem*. 2021;93:2003–9.
- Grant OC, Montgomery D, Ito K, Woods RJ. Analysis of the SARS-CoV-2 spike protein glycan shield reveals implications for immune recognition. *Sci Rep*. 2020;10:14991.
- Yang Q, Hughes TA, Kelkar A, Yu X, Cheng K, Park S, et al. Inhibition of SARS-CoV-2 viral entry upon blocking N- and O-glycan elaboration. *eLife*. 2020;9:e61552.
- Arbeitman CR, Auge G, Blaustein M, Bredeson L, Corapi ES, Craig PO, et al. Structural and functional comparison of SARS-CoV-2-spike receptor binding domain produced in *Pichia pastoris* and mammalian cells. *Sci Rep*. 2020;10:21779.
- Huang Y, Yang C, Xu X-f XW, S-w L. Structural and functional properties of SARS-CoV-2 spike protein: potential antiviral drug development for COVID-19. *Acta Pharmacol Sin*. 2020;41:1141–9.
- Yu Z, Han G, Sun S, Jiang X, Chen R, Wang F, et al. Preparation of monodisperse immobilized Ti(4+) affinity chromatography microspheres for specific enrichment of phosphopeptides. *Anal Chim Acta*. 2009;636:34–41.
- Huang J, Qin H, Dong J, Song C, Bian Y, Dong M, et al. In situ sample processing approach (iSPA) for comprehensive quantitative phosphoproteome analysis. *J Proteome Res*. 2014;13:3896–904.
- Huang J, Dong J, Shi X, Chen Z, Cui Y, Liu X, et al. Dual-functional titanium(IV) immobilized metal affinity chromatography approach for enabling large-scale profiling of protein mannose-6-phosphate glycosylation and revealing its predominant substrates. *Anal Chem*. 2019;91:11589–97.
- Mohammed S, Lorenzen K, Kerkhoven R, Bv B, Vannini A, Cramer P, et al. Multiplexed proteomics mapping of yeast RNA polymerase II and III allows near-complete sequence coverage and reveals several novel phosphorylation sites. *Anal Chem*. 2008;80:3584–92.
- Chen R, Jiang XN, Sun DG, Han GH, Wang FJ, Ye ML, et al. Glycoproteomics analysis of human liver tissue by combination of multiple enzyme digestion and hydrazide chemistry. *J Proteome Res*. 2009;8:651–61.
- Bian Y, Song C, Cheng K, Dong M, Wang F, Huang J, et al. An enzyme assisted RP-RPLC approach for in-depth analysis of human liver phosphoproteome. *J Proteome*. 2014;96:253–62.
- Zhu J, Sun Z, Cheng K, Chen R, Ye M, Xu B, et al. Comprehensive mapping of protein N-glycosylation in human liver by combining hydrophilic interaction chromatography and hydrazide chemistry. *J Proteome Res*. 2014;13:1713–21.
- Glover MS, Yu Q, Chen ZW, Shi XD, Kent KC, Li LJ. Characterization of intact sialylated glycopeptides and phosphorylated glycopeptides from IMAC enriched samples by EThcD fragmentation: toward combining phosphoproteomics and glycoproteomics. *Int J Mass Spectrom*. 2018;427:35–42.
- Cui Y, Yang K, Tabang DN, Huang J, Tang W, Li L. Finding the sweet spot in ERLIC mobile phase for simultaneous enrichment of N-glyco and phosphopeptides. *J Am Soc Mass Spectrom*. 2019;30:2491–501.
- Cao Q, Zhao X, Zhao Q, Lv X, Ma C, Li X, et al. Strategy integrating stepped fragmentation and glycan diagnostic ion-based spectrum refinement for the identification of core fucosylated glycoproteome using mass spectrometry. *Anal Chem*. 2014;86:6804–11.
- Riley NM, Malaker SA, Driessen MD, Bertozzi CR. Optimal dissociation methods differ for N- and O-glycopeptides. *J Proteome Res*. 2020;19:3286–301.
- Yu Q, Wang B, Chen Z, Urabe G, Glover MS, Shi X, et al. Electron-transfer/higher-energy collision dissociation (EThcD)-enabled intact glycopeptide/glycoproteome characterization. *J Am Soc Mass Spectrom*. 2017;28:1751–64.

26. Wang D, Baudys J, Bundy JL, Solano M, Keppel T, Barr JR. Comprehensive analysis of the glycan complement of SARS-CoV-2 spike proteins using signature ions-triggered electron-transfer/higher-energy collisional dissociation (ET_hcD) mass spectrometry. *Anal Chem*. 2020;92:14730–9.
27. Ghosh S, Dellibovi-Ragheb TA, Kerviel A, Pak E, Qiu Q, Fisher M, et al. β -Coronaviruses use lysosomes for egress instead of the biosynthetic secretory pathway. *Cell*. 2020;183:1520–35.e14.
28. Bagdonaite I, Wandall HH. Global aspects of viral glycosylation. *Glycobiology*. 2018;28:443–67.
29. Casalino L, Gaieb Z, Goldsmith JA, Hjorth CK, Dommer AC, Harbison AM, et al. Beyond shielding: the roles of glycans in the SARS-CoV-2 spike protein. *ACS Central Science*. 2020;6:1722–34.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.