bioRxiv preprint doi: https://doi.org/10.1101/2020.04.01.020966; this version posted April 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Deducing the N- and O- glycosylation profile of the spike protein of novel coronavirus SARS-CoV-2

Asif Shajahan, Nitin T. Supekar, Anne S. Gleinich, and Parastoo Azadi*.

Complex Carbohydrate Research Center, The University of Georgia, Athens, GA 30602

*Corresponding Author:

Parastoo Azadi, Ph.D.

Analytical Service and Training Laboratory

Complex Carbohydrate Research Center,

University of Georgia,

315 Riverbend Road, Athens, GA 30602

Keywords: COVID-19 glycosylation, Spike protein, Coronavirus vaccine, O-glycosylation. Subunit S1, S2.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.01.020966; this version posted April 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

The current emergence of the novel coronavirus pandemic caused by SARS-CoV-2 demands the development of new therapeutic strategies to prevent rapid progress of mortalities. The coronavirus spike (S) protein, which facilitates viral attachment, entry and membrane fusion is heavily glycosylated and plays a critical role in the elicitation of host immune response. The spike protein is comprised of two protein subunits (S1 and S2), which together possess 22 potential N-glycosylation sites. Herein, we report the mapping of glycosylation on spike protein subunits S1 and S2 expressed on human cells through high resolution mass spectrometry. We have characterized the quantitative N-glycosylation profile on spike protein and interestingly, observed two unexpected O-glycosylation modifications on the receptor binding domain (RBD) of spike protein subunit S1. Even though O-glycosylation has been predicted on the spike protein of SARS-Cov-2, this is the first report of experimental data for both the site of O-glycosylation and identity of the O-glycans attached on the subunit S1. Our data on the N- and O- glycosylation is strengthened by extensive manual interpretation of each glycopeptide spectra in addition to using bioinformatics tools to confirm the complexity of glycosylation in the spike protein. The elucidation of the glycan repertoire on the spike protein provides insights into the viral binding studies and more importantly, propels research towards the development of a suitable vaccine candidate.

Introduction

The current major health crisis is caused by the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) that rapidly spread globally within weeks in early 2020. This highly transmissible infectious disease causes a respiratory illness named COVID-19.[1, 2] As of the 31st of March 2020, 750890 cases of COVID-19 and 36405 COVID-19-related deaths have been confirmed globally by the World Health Organization (WHO).[3]

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.01.020966; this version posted April 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

To date, no specific medical treatments or vaccines for COVID-19 have been approved.[4, 5] Therefore, the scientific community is expending great effort in compiling data regarding the virus, as well as the respiratory illness caused by it, to find effective ways of dealing with this health crisis.

The pathogenic SARS-CoV-2 enters human target cells via its viral transmembrane spike (S) glycoprotein. The spike protein is a trimeric class I fusion protein, and consists of two subunits, namely S1 and S2. The S1 subunit facilitates the attachment of the virus, and subsequently the S2 subunit allows for the fusion of the viral and human cellular membranes.[6-8] The entry receptor for SARS-CoV-2 has been identified as the human angiotensin-converting enzyme 2 (hACE2), and recent studies determined a high binding affinity to hACE2.[6, 7, 9] Given its literal key role, the S protein is one of the major targets for the development of specific medical treatments or vaccines: Neutralizing antibodies targeting the spike proteins of SARS-CoV-2 could prevent the virus from binding to the hACE2 entry receptor and therefore from entering the host cell.[9]

Each monomer of the S protein is highly glycosylated with 22 predicted N-linked glycosylation sites. Furthermore, four O-glycosylation sites were also predicted.[10] Cryo-electron microscopy (Cryo-EM) provides evidence for the existence of 14–16 N-glycans on 22 potential sites in the SARS-CoV-2 S protein.[7, 11] The glycosylation pattern of the spike protein is a crucial characteristic to be considered regarding steric hindrance, chemical properties and even as a potential target for mutation in the future. The N-glycans on S protein play important roles in proper protein folding and priming by host proteases. Since glycans can shield the amino acid residues and other epitopes from cells and antibody recognition, glycosylation can enable the coronavirus to evade both the innate and adaptive immune responses.[7, 12] Elucidating the glycosylation of the viral S protein can aid in understanding viral binding with receptors, fusion, entry, replication and also in designing suitable antigens for vaccine development.[13-15] Strategies for vaccine development aim to elicit such adaptive immunity through an antibody response at the sites of viral entry.[16]

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.01.020966; this version posted April 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Here, we report the site-specific quantitative N-linked and O-linked glycan profiling on SARS-CoV-2 subunit S1 and S2 protein through glycoproteomics using high resolution LC-MS/MS. We used recombinant SARS-CoV-2 subunit S1 and S2 expressed in human cells, HEK 293 and observed partial N-glycan occupancy on 17 out of 22 N-glycosylation sites. We found that the remaining five N-glycosylation sites were unoccupied. Remarkably, we have unambiguously identified 2 unexpected O-glycosylation sites at the receptor binding domain (RBD) of subunit S1. O-glycosylation on the spike protein of SARS-CoV-2 is predicted in several recent reports and most of these predictions are for sites in proximity to furin cleavage site (S1/S2) as similar sites are O-glycosylated in SARS-CoV-1.[10] However, we observed O-glycosylation at two sites on the RBD of spike protein subunit S1 and this is the first report on the evidence for such glycan modification at a crucial viral attachment location. Site-specific analysis of N- and O- glycosylation information of SARS-CoV-2 spike protein provides basic understanding of the viral structure, crucial for the identification of immunogens for vaccine design. This in turn has the potential of leading to future therapeutic intervention or prevention of COVID-19.

Materials and methods

Dithiothreitol (DTT) and iodoacetamide (IAA) were purchased from Sigma Aldrich (St. Louis, MO). Sequencing-grade modified trypsin and chymotrypsin were purchased from Promega (Madison, WI). All other reagents were purchased from Sigma Aldrich unless indicated otherwise. Data analysis was performed using Byonic 3.5 software and manually using Xcalibur 4.2. The SARS-CoV-2 spike protein subunit 1 (Cat No. 230-20407) and subunit 2 (Cat No. 230-20408) were purchased from RayBiotech (Atlanta, GA).

Protease digestion and extraction of peptides from SDS-PAGE

The protein subunits S1 and S2 as HEK 293 culture supernatants were fractionated on separate lanes using SDS-PAGE. The gel was stained by Coomassie dye and the bands

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.01.020966; this version posted April 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

corresponding to subunit 1 (200 to 100 kDa) and subunit 2 (150 to 80 kDa) were cut into smaller pieces (1 mm squares approx.) and transferred to clean tubes. The gel pieces were de-stained by adding 100 µL acetonitrile (ACN): 50mM NH_4HCO_3 (1:1) and incubated at room temperature (RT) for about 30 min. Tubes were centrifuged, the supernatant was discarded, and 100 µL ACN was added before incubation for 30 min. The proteins on gel pieces were reduced by adding 350 µL 25 mM DTT and incubating at 60 °C for 30 min. The tubes were cooled to RT and the supernatant removed. The gels were washed with 500 µL of ACN, 350 µL 90 mM IAA was added, and the mixture was incubated at RT for 20 min in the dark. Proteins were digested by adding sequence-grade trypsin and/or chymotrypsin (we performed digests with both enzymes individually and as a cocktail) in digestion buffer (50mM NH_4HCO_3) for 18 h at 37 °C separately. The peptides were extracted out from the gel by addition of 1:2 H_2O:ACN containing 5% Formic acid (500 µL), and the released peptides were speed-dried. The samples were reconstituted in aqueous 0.1% formic acid for LC-MS/MS experiments.

Data acquisition of protein digest samples using nano-LC-MS/MS

The glycoprotein digests were analyzed on an Orbitrap Fusion Tribrid mass spectrometer equipped with a nanospray ion source and connected to a Dionex binary solvent system (Waltham, MA). Pre-packed nano-LC columns of 15 cm length with 75 µm internal diameter (id), filled with 3 µm C18 material (reverse phase) were used for chromatographic separation of samples. The precursor ion scan was acquired at 120,000 resolution in the Orbitrap analyzer and precursors at a time frame of 3 s were selected for subsequent MS/MS fragmentation in the Orbitrap analyzer at 15,000 resolution. The LC-MS/MS runs of each digest were conducted for both 72 min and 180 min in order to separate the glycopeptides. The threshold for triggering an MS/MS event was set to 1000 counts, and monoisotopic precursor selection was enabled. MS/MS fragmentation was conducted with HCD (Higher-energy Collisional Dissociation) product triggered CID (Collision-Induced Dissociation) (HCDpdCID) program. Charge state screening was

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.01.020966; this version posted April 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

enabled, and precursors with unknown charge state or a charge state of +1 were excluded (positive ion mode). Dynamic exclusion was enabled (exclusion duration of 30 secs).

Data analysis of glycoproteins

The LC-MS/MS spectra of tryptic, chymotryptic and combined tryptic /chymotryptic digests of glycoproteins were searched against the .fasta sequence of spike protein S1 and S2 subunit using the Byonic software by choosing appropriate peptide cleavage sites (semi-specific cleavage option enabled). Oxidation of methionine, deamidation of asparagine and glutamine, possible common human N-glycans and O-glycan masses were used as variable modifications. The LC-MS/MS spectra were also analyzed manually for the glycopeptides with the support of the Xcalibur software. The HCDpdCID MS^2 spectra of glycopeptides were evaluated for the glycan neutral loss pattern, oxonium ions and glycopeptide fragmentations to assign the sequence and the presence of glycans in the glycopeptides.

Results and Discussion

Studies over the past two decades have shown that glycosylation on the protein antigens can play crucial roles in the adaptive immune response. Thus, it is obvious that the glycosylation on the protein antigen is relevant for the development of vaccines, and it is widely accepted that the lack of information about the glycosylation sites hampers the design of such vaccines.[17]

Mapping N-glycosylation on SARS-CoV-2 spike protein

We have procured culture supernatants of HEK 293 cells expressing SARS-CoV-2 subunit 1 and subunit 2 separately. The proteins were expressed with a His tag with Val16 to Gln690 for subunit 1 and Met697 to Pro1213 for subunit 2. According to manufacturers, SDS-PAGE of the proteins showed a higher molecular weight than the predicted 75 and 60 kDa, respectively, due to glycosylation. Since the proteins were unpurified, we fractionated them through SDS-PAGE on separate lanes and cut the bands corresponding to subunit 1 and subunit 2. The gels were stained

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.01.020966; this version posted April 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

with Coomassie dye, and gel bands were cut into small pieces, de-stained, reduced, alkylated and subjected to in-gel protease digestion. We employed trypsin, chymotrypsin, and both trypsin-chymotrypsin in combination to generate glycopeptides that contain a single N-linked glycan site. The glycopeptides were further analyzed by high resolution LC-MS/MS, using a glycan oxonium ion product dependent HCD triggered CID program. The LC-MS/MS data were analyzed using Byonic software, each detected spectrum was manually validated, and false detections eliminated.

We identified the glycan compositions at 17 out of the 22 predicted N-glycosylation sites of the SARS-CoV-2 S1 and S2 proteins and found the remaining five sites unoccupied (Figure 2, 3). We observed both high mannose and complex-type glycans across the N-glycosylation sites but found no hybrid type N-glycans. We quantified the relative intensities of glycans at each site by comparing the area under the curve of each glycopeptide peak on the LC-MS chromatogram. A recent preprint investigated the N-glycosylation on SARS-CoV-2 and reported prevalence of hybrid-type glycans.[18] In contrast, we observed a combination of high mannose and complex-type, but no hybrid-type glycans on most of the sites. We discovered predominantly highly processed sialylated complex-type glycans on sites N165, N282, N801, and N1098 (Figure 3, 4). The highly sialylated glycans at N234 and N282, which are at the RBD can act as determinant in viral binding with hACE2 receptors.[6, 7, 19] Similar to one recent report, we observed $Man_5GlcNAc_2$ as a predominant structure across all sites.[18] However, we observed significant unoccupied sites on both S1 and S2 subunits. Sites N17, N603, N1134, N1158 and N1173 were completely unoccupied, although further studies with higher concentration and purity of proteins are required to validate this finding (Figure 1, 2). On subunit S2, the detection of N-glycosylation at sites N709, N717, and N1134 was ambiguous as the quality of the MS/MS spectra was not satisfactory and we are currently evaluating the possibilities of other post translational modifications adjacent to these sites.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.01.020966; this version posted April 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Identification of O-Glycosylation on SARS-CoV-2 spike protein

O-glycans, which are involved in protein stability and function, have been observed on some viral proteins and have been suggested to play roles in the biological activity of viral proteins.[10, 20] A comparative study of human SARS-CoV-2 S protein with other coronavirus S proteins has shown that Ser673, Thr678, and Ser686 are conserved O-glycosylation locations, and SARS-CoV-2 S1 protein was suggested to be O-glycosylated at these locations.[10] Although it is unclear what function these predicted O-linked glycans perform, they have been suggested to create a 'mucin-like domain' which could shield SARS-CoV-2 spike protein epitopes or key residues.[20] Since some viruses can utilize mucin-like domains as glycan shields for immunoevasion, researchers have highlighted the importance of experimental studies for the determination of predicted O-linked glycosylation sites.[10, 20] We evaluated the O-glycosylation site prediction using widely accepted tool Net-O-Gly server 4.0.[21] However, did not find any strong prediction for the O-glycosylation except for the sites Ser673, Thr678 and Ser686.

Nevertheless, we have identified O-glycosylation at sites Thr323 and Ser325 on the S1 subunit of SARS-CoV-2 spike protein through high resolution mass spectrometry glycoproteomic profiling (Figure 5, 6). Since O-glycosylation at Thr323 and Ser325 on the spike protein has not been reported before and is not indicated based on cryo-electron microscopy data of SARS-CoV-2 S protein, we evaluated the detected O-glycopeptide manually.[7] We observed very strong evidence for the presence of O-glycosylation at sites Thr323 and Ser325 as b and y ions of the peptide ^{320}VQPTESIVR^{328} with high mass accuracy. Upon manual validation of the fragment ions of O-glycopeptide ^{320}VQPTESIVR^{328} we observed that Thr323 is the predominantly occupied site. This conclusion was based on the b (b_1- m/z 228.13) and y (y_1- m/z 175.12, y_2- m/z 274.19, y_4- 474.30, y_5- 603.34 m/z, and y_7+glycan- 1748.77 m/z) ions we detected upon fragmentation of the glycopeptide (Figure 6). In addition, neutral losses and the detection of oxonium ions also confirmed the presence of glycosylation on these peptides. Core-1 mucin type O-glycans such as

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.01.020966; this version posted April 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

HexNAc, HexNAcHex and HexNAcHexNeuAc$_2$ were observed on site Thr323. Interestingly, Ser325 is found occupied with HexNAcHexNeuAc only when Thr323 is glycosylated with the same HexNAxHexNeuAc glycan (Figure 5). Intriguingly, the possible O-glycosylation at Thr323 of SARS-CoV-2 subunit 1 glycoprotein has been predicted by computational analysis in a very recent preprint report.[22] The accuracy of our observation of O-glycosylation at Thr323 is further confirmed by the presence of proline at location 322, considering the well-established fact that the frequency of occurrence of proline residues is higher adjacent to O-glycosylation sites.[23]

Cryo-EM studies on SARS-CoV-2 indicate that the binding of S protein to the hACE2 receptor primarily involves extensive polar residue interactions between RBD and the peptidase domain of hACE2.[6, 7] The S protein RBD located in the S1 subunit of SARS-CoV-2 undergoes a hinge-like dynamic movement to enhance the capture of the receptor RBD with hACE2, displaying 10–20-fold higher affinity for the human ACE2 receptor than SARS-CoV-1, which partially explains the higher transmissibility of the new virus.[11, 24] The residues Thr323 and Ser325 are located at the RBD of the S1 subunit of SARS-Cov-2, and thus the O-glycosylation at this location could play a critical role in viral binding with hACE2 receptors.[6, 10] Our observation will pave the way for future studies to understand the implication of O-glycosylation at the RBD of S1 protein in viral attachment with hACE2 receptors.

Two very recent preprints reporting N-glycosylation on spike protein of SARS-CoV-2 showed different glycosylation profiles, and both reports were different from our results although all three studies utilized recombinant S protein from an HEK 293-based expression system.[18, 25] This indicates the caution to be exercised in determining the glycosylation on viral antigens generated from various sources as the changes in glycosylation pattern can influence the efficacy of potential vaccine candidates.[26]

Our comprehensive N- and O-glycosylation characterization of SARS-CoV-2 expressed in a human cell system provides insights into site-specific N- and O- glycan decoration on the trimeric

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.01.020966; this version posted April 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

spike protein. We have employed extensive manual interpretation strategy for the assignment of each glycopeptide structure in order to eliminate possibilities of ambiguous software based annotation. We are currently working on elucidating other potential post translational modifications on SARS-CoV-2 spike protein as understanding the protein modifications in detail is important to guide future researches on disease interventions involving spike protein. Detailed glycan analysis is important for the development of glycoprotein-based vaccine candidates as a means to correlate the structural variation with immunogenicity. Glycosylation can serve as a measure to evaluate antigen quality as various expression systems and production processes are employed in vaccine manufacture. The understanding of complex sialylated N-glycans and sialylated mucin type O-glycans, particularly in the RBD domain of the spike protein of SARS-CoV-2, provides basic knowledge useful for elucidating the viral infection pathology in future therapeutic possibilities, as well as in the design of suitable immunogens for vaccine development.

Acknowledgements

Financial support from the US National Institutes of Health (S10OD018530) is gratefully acknowledged. This work was also supported in part by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, under Award DE-SC0015662 to DOE - Center for Plant and Microbial Complex Carbohydrates at the Complex Carbohydrate Research Center, USA.

Contributions

A.S. and P.A. conceived of the paper; A.S., N.S. and A.G. contributed equally and performed experiments; everyone contributed toward writing the paper; P.A. monitored the project.

Competing interests

The authors certify that they have no competing interests.

Supplementary material: Annotated MS/MS spectra are incorporated as a separate supplementary material file.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.01.020966; this version posted April 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

1. Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., Cheng, Z., Yu, T., Xia, J., Wei, Y., Wu, W., Xie, X., Yin, W., Li, H., Liu, M., Xiao, Y., Gao, H., Guo, L., Xie, J., Wang, G., Jiang, R., Gao, Z., Jin, Q., Wang, J., and Cao, B. (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, *Lancet 395*, 497.
2. Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., Hu, Y., Tao, Z. W., Tian, J. H., Pei, Y. Y., Yuan, M. L., Zhang, Y. L., Dai, F. H., Liu, Y., Wang, Q. M., Zheng, J. J., Xu, L., Holmes, E. C., and Zhang, Y. Z. (2020) A new coronavirus associated with human respiratory disease in China, *Nature 579*, 265-269.
3. World Health Organization. (2020) WHO COVID-19 Situation Report - 71, Licence: CC BY-NC-SA 3.0 IGO.
4. Li, G. D., and De Clercq, E. (2020) Therapeutic options for the 2019 novel coronavirus (2019-nCoV), *Nat Rev Drug Discov 19*, 149-150.
5. World Health Organization. (2020) Coronavirus, Licence: CC BY-NC-SA 3.0 IGO.
6. Hoffmann, M., Kleine-Weber, H., Schroeder, S., Kruger, N., Herrler, T., Erichsen, S., Schiergens, T. S., Herrler, G., Wu, N. H., Nitsche, A., Muller, M. A., Drosten, C., and Pohlmann, S. (2020) SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor, *Cell*.
7. Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T., and Veesler, D. (2020) Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein, *Cell*.
8. Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., Si, H. R., Zhu, Y., Li, B., Huang, C. L., Chen, H. D., Chen, J., Luo, Y., Guo, H., Jiang, R. D., Liu, M. Q., Chen, Y., Shen, X. R., Wang, X., Zheng, X. S., Zhao, K., Chen, Q. J., Deng, F., Liu, L. L., Yan, B., Zhan, F. X., Wang, Y. Y., Xiao, G. F., and Shi, Z. L. (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin, *Nature 579*, 270-273.
9. Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., Geng, Q., Auerbach, A., and Li, F. (2020) Structural basis of receptor recognition by SARS-CoV-2, *Nature*.
10. Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C., and Garry, R. F. (2020) The proximal origin of SARS-CoV-2, *Nature Medicine*.
11. Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C. L., Abiona, O., Graham, B. S., and McLellan, J. S. (2020) Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation, *Science 367*, 1260-1263.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.01.020966; this version posted April 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

12. Walls, A. C., Xiong, X., Park, Y. J., Tortorici, M. A., Snijder, J., Quispe, J., Cameroni, E., Gopal, R., Dai, M., Lanzavecchia, A., Zambon, M., Rey, F. A., Corti, D., and Veesler, D. (2019) Unexpected Receptor Functional Mimicry Elucidates Activation of Coronavirus Fusion, *Cell 176*, 1026-1039 e1015.
13. Watanabe, Y., Bowden, T. A., Wilson, I. A., and Crispin, M. (2019) Exploitation of glycosylation in enveloped virus pathobiology, *Biochim Biophys Acta Gen Subj 1863*, 1480-1497.
14. Chakraborti, S., Prabakaran, P., Xiao, X., and Dimitrov, D. S. (2005) The SARS coronavirus S glycoprotein receptor binding domain: fine mapping and functional characterization, *Virol J 2*, 73.
15. Zheng, J., Yamada, Y., Fung, T. S., Huang, M., Chia, R., and Liu, D. X. (2018) Identification of N-linked glycosylation sites in the spike protein and their functional impact on the replication and infectivity of coronavirus infectious bronchitis virus in cell culture, *Virology 513*, 65-74.
16. Afrough, B., Dowall, S., and Hewson, R. (2019) Emerging viruses and current strategies for vaccine intervention, *Clin Exp Immunol 196*, 157-166.
17. Wolfert, M. A., and Boons, G. J. (2013) Adaptive immune activation: glycosylation does matter, *Nat Chem Biol 9*, 776-784.
18. Watanabe, Y., Allen, J. D., Wrapp, D., McLellan, J. S., and Crispin, M. (2020) Site-specific analysis of the SARS-CoV-2 glycan shield, *bioRxiv*, 2020.2003.2026.010322.
19. Tortorici, M. A., Walls, A. C., Lang, Y., Wang, C., Li, Z., Koerhuis, D., Boons, G. J., Bosch, B. J., Rey, F. A., de Groot, R. J., and Veesler, D. (2019) Structural basis for human coronavirus attachment to sialic acid receptors, *Nat Struct Mol Biol 26*, 481-489.
20. Bagdonaite, I., and Wandall, H. H. (2018) Global aspects of viral glycosylation, *Glycobiology 28*, 443-467.
21. Steentoft, C., Vakhrushev, S. Y., Joshi, H. J., Kong, Y., Vester-Christensen, M. B., Schjoldager, K. T., Lavrsen, K., Dabelsteen, S., Pedersen, N. B., Marcos-Silva, L., Gupta, R., Bennett, E. P., Mandel, U., Brunak, S., Wandall, H. H., Levery, S. B., and Clausen, H. (2013) Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology, *EMBO J 32*, 1478-1488.
22. Uslupehlivan, M., and Sener, E. (2020) Glycoinformatics approach for identifying target positions to inhibit initial binding of SARS-CoV-2 S1 protein to the host cell, *bioRxiv*, 2020.2003.2025.007898.
23. Thanka Christlet, T. H., and Veluraja, K. (2001) Database analysis of O-glycosylation sites in proteins, *Biophys J 80*, 952-960.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.01.020966; this version posted April 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

24. Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., and Zhou, Q. (2020) Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2, *Science 367*, 1444-1448.
25. Zhang, Y., Zhao, W., Mao, Y., Wang, S., Zhong, Y., Su, T., Gong, M., Lu, X., Cheng, J., and Yang, H. (2020) Site-specific N-glycosylation Characterization of Recombinant SARS-CoV-2 Spike Proteins using High-Resolution Mass Spectrometry, *bioRxiv*, 2020.2003.2028.013276.
26. Chang, D., and Zaia, J. (2019) Why Glycosylation Matters in Building a Better Flu Vaccine, *Mol Cell Proteomics 18*, 2348-2358.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.01.020966; this version posted April 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figures and legends

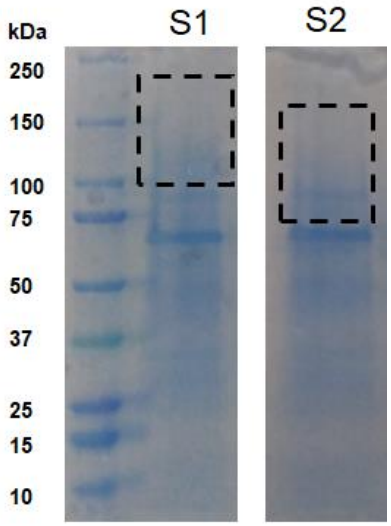

Figure 1: The expression of SARS-Cov-2 spike protein subunits S1 and S2 on HEK 293 culture supernatant showed higher molecular weight upon SDS-PAGE than expected, because of glycosylation. Thus, the gel bands corresponding to the molecular weight of 200 kDa to 100 kDa for S1 and 150 kDa to 80 kDa for S2 were cut, proteins were lysed after reduction-alkylation and analyzed by LC-MS/MS.

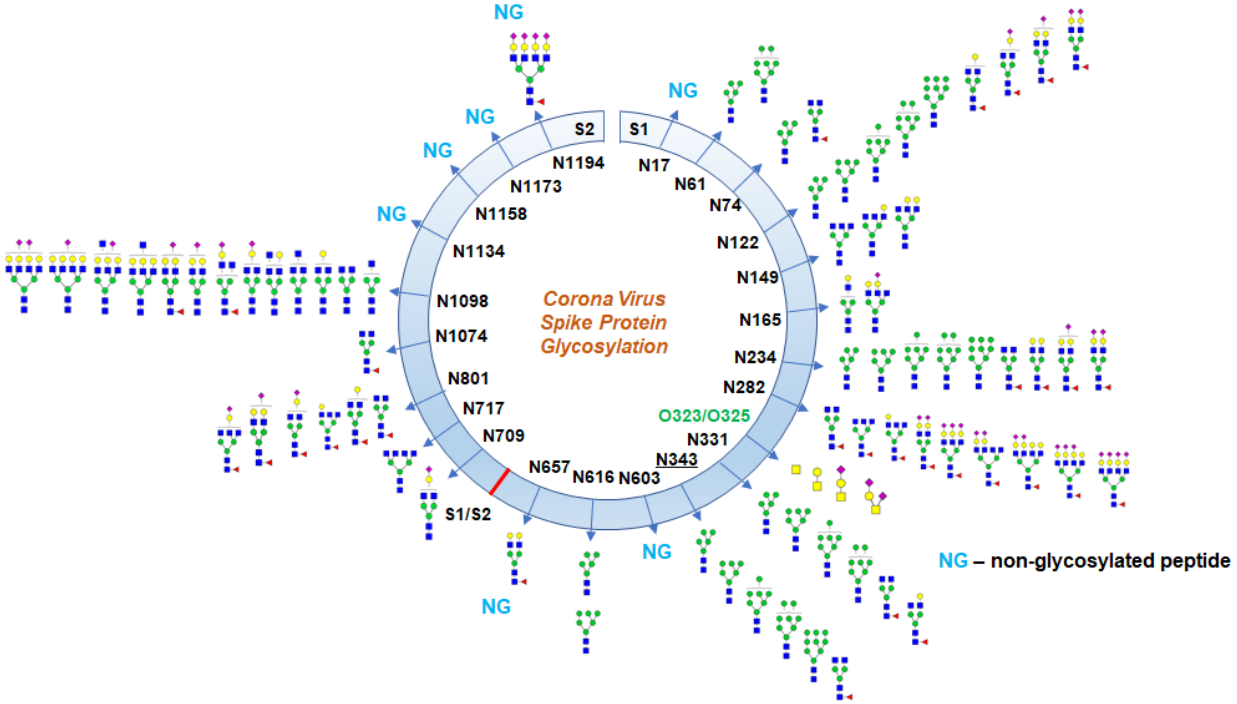

Figure 2: Glycosylation profile on coronavirus SARS-CoV-2 characterized by high-resolution LC-MS/MS. About 17 N-glycosylation sites were found occupied out of 22 potential sites along with two O-glycosylation sites bearing core-1 type O-glycans. Some N-glycosylation sites were partially glycosylated.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.01.020966; this version posted April 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 3: Quantitative glycosylation profile of N- glycans on coronavirus SARS-CoV-2 spike protein characterized by high-resolution LC-MS/MS. **a.** 13 sites on subunit S1; **b.** 9 sites on subunit S2.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.01.020966; this version posted April 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

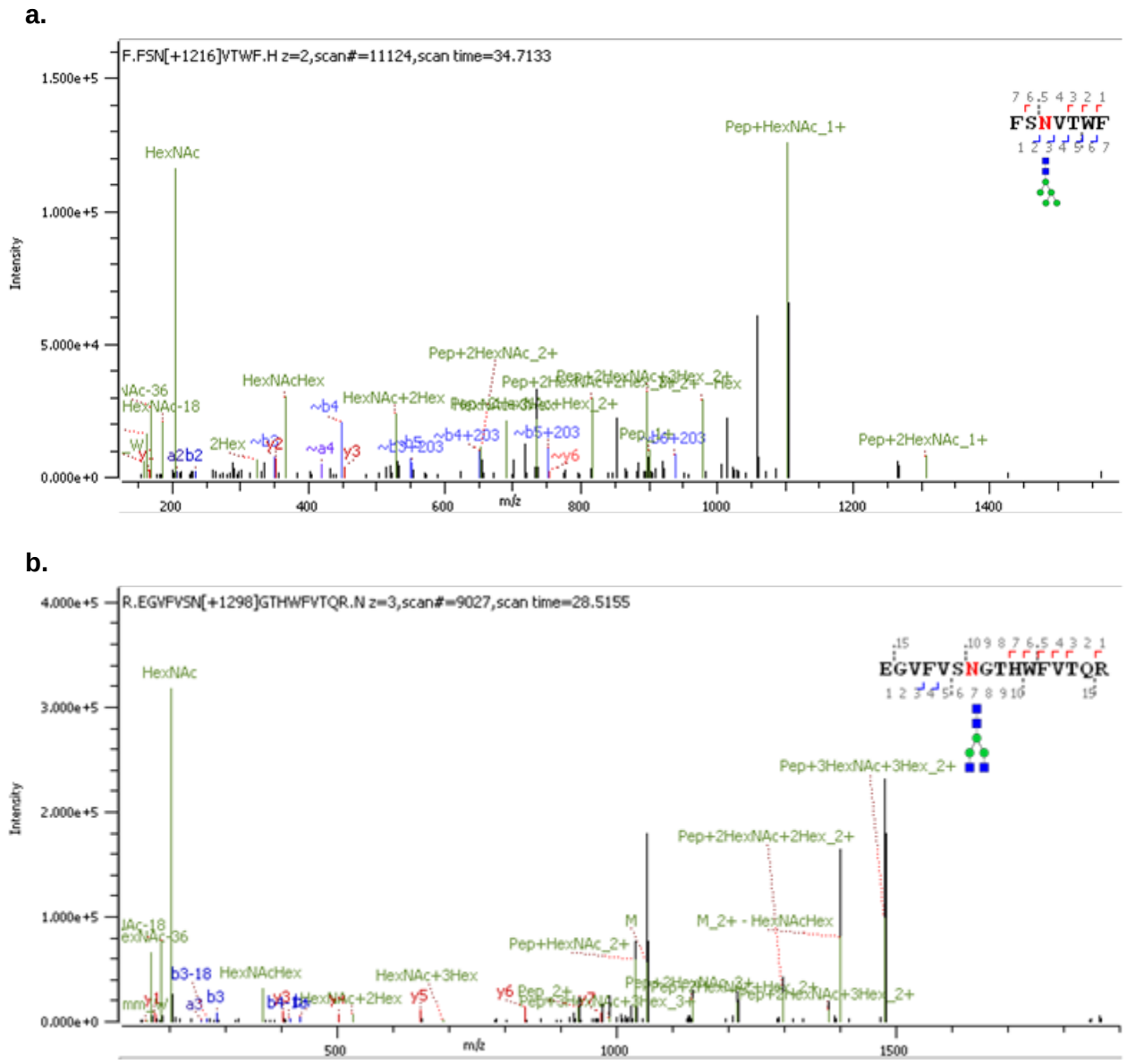

Figure 4: HCD MS/MS spectrum showing glycan neutral losses, oxonium ions and peptide fragments of **a.** representative N-glycopeptide FSNVTWF (site N61) of spike protein subunit 1; **b.** representative N-glycopeptide EGVFVSNGTHWFVTQR (site N1098) of spike protein subunit 2.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.01.020966; this version posted April 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

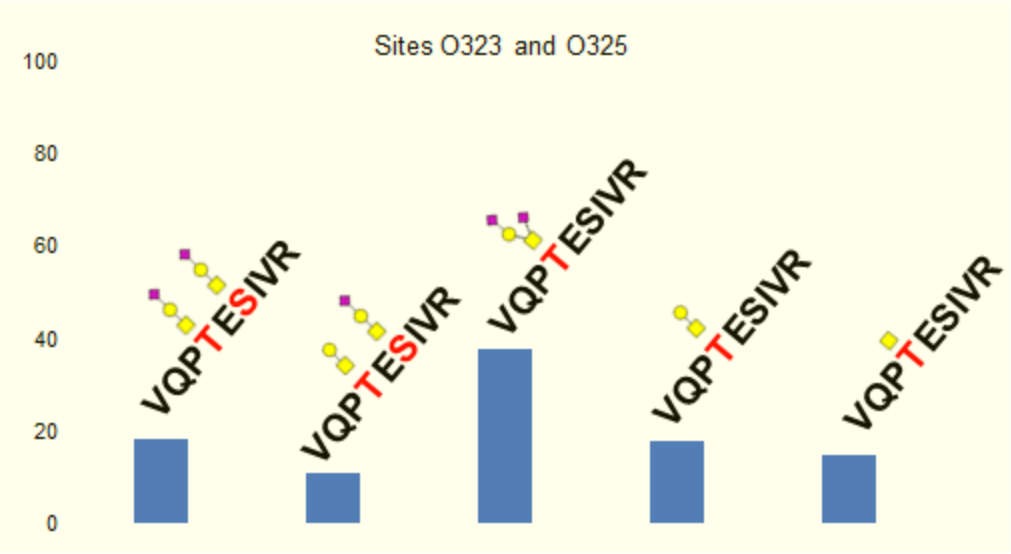

Figure 5: Quantitative glycosylation profile of O-glycans on sites T323 and S325 of coronavirus SARS-Cov-2 spike protein characterized by high-resolution LC-MS/MS.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.01.020966; this version posted April 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

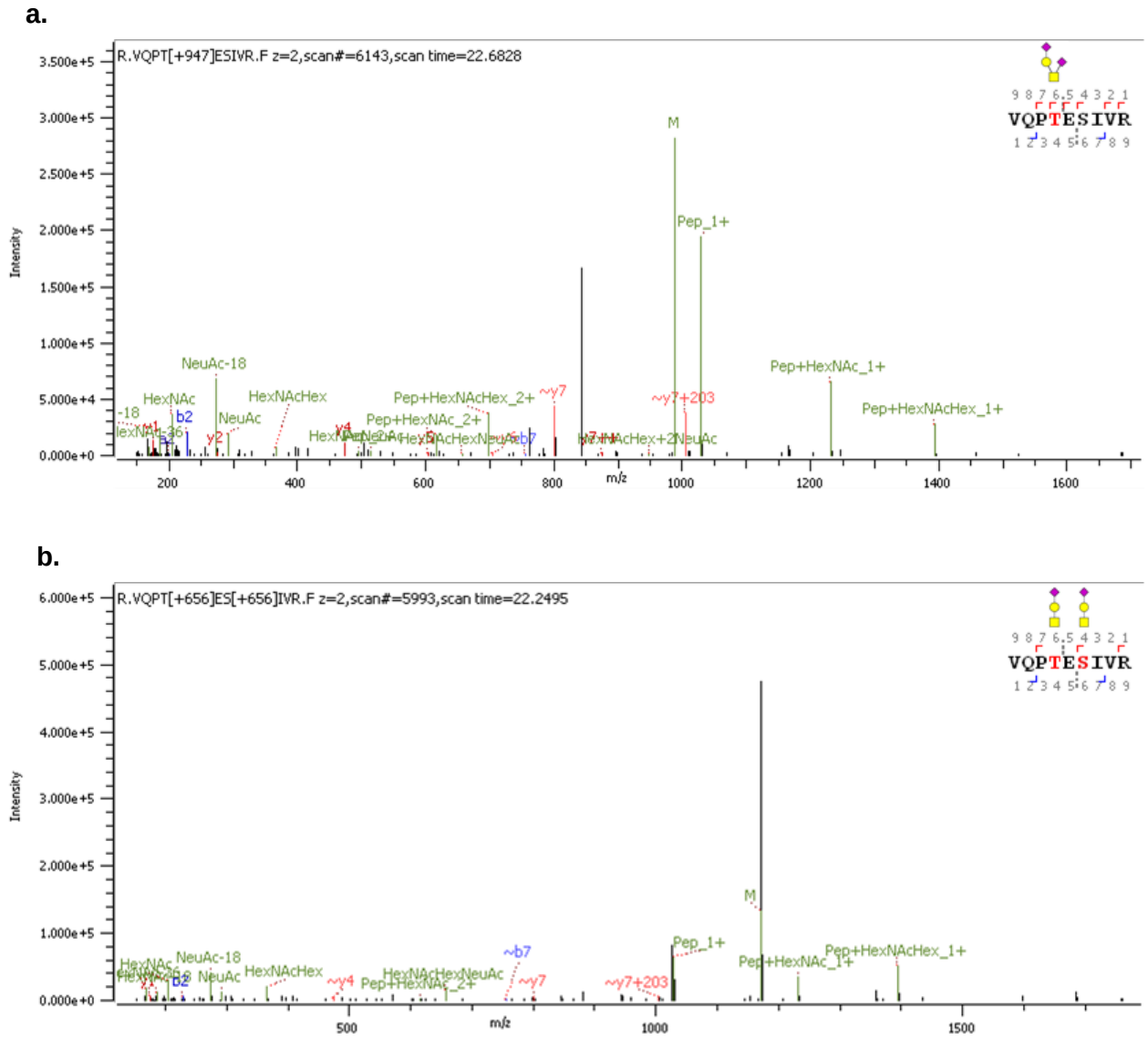

Figure 6: HCD MS/MS spectrum showing glycan neutral losses, oxonium ions and peptide fragments of **a.** representative O-Glycopeptide ^{320}VQPTESIVR^{328} with core 1 type HexNAcHexNeuAc$_2$ glycan detected on site Thr323 of spike protein subunit 1; **b.** representative O-Glycopeptide ^{320}VQPTESIVR^{328} with core 1 type HexNAcHexNeuAc$_2$ glycan detected on sites Thr323 and Ser325 of spike protein subunit 1.