Comprehensive O-glycosylation analysis of the SARS-CoV-2 spike protein

Xuefang Dong
Dalian Institute of Chemical Physics
Cheng Chen
Dalian Institute of Chemical Physics, Chinese Academy of Sciences
Jingyu Yan
Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Key Laboratory of Separation Science for Analytical Chemistry
Xiuling Li
Dalian Institute of Chemical Physics, Chinese Academy of Sciences
Xinmiao Liang (liangxm@dicp.ac.cn)
Dalian Institute of Chemical Physics

Article

Keywords: COVID-19, Vaccine, O-glycosylation Analysis

Posted Date: August 28th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-62499/v1

Comprehensive O-glycosylation analysis of the SARS-CoV-2 spike protein

Xuefang Dong[1], Cheng Chen[1], Jingyu Yan[1], Xiuling Li[1]* & Xinmiao Liang[1]*

[1]Key Laboratory of Separation Science for Analytical Chemistry, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, P. R. China

*Corresponding author. Email: liangxm@dicp.ac.cn (X. L.); lixiuling@dicp.ac.cn (X. L.)

Abstract:

The COVID-19 pandemic caused by severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) is a serious public health threat. Most vaccines being developed against SARS-CoV-2 target the highly glycosylated spike protein (S). A good knowledge of the glycosylation profile of this protein is key for successful vaccine development. Unlike the 22 confirmed N-glycosylation sites (NGSs) on the SARS-CoV-2 S, only a few O-glycosylation sites (OGSs) on this protein have been reported. This difference is mainly ascribed to an extremely low O-glycosylation stoichiometry. Herein, we comprehensively analyzed the O-glycosylation profile of recombinant SARS-CoV-2 S employing biomimetic polymer consisting of Trp-Arg monomer. Twenty-six OGSs and 33 O-linked glycans (OLGs) in the SARS-CoV-2 S were unambiguously identified, among which 24 OGSs and 25 OLGs are novel to the SARS-CoV-2 S. Our study reveals the comprehensive O-glycosylation profile of the SARS-CoV-2 S, which might shed light on viral pathobiology and assist in vaccine development.

Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), a pathogen and the causative agent of the COVID-19 pandemic, has triggered a global public health crisis[1,2]. As of 18th August 2020, 21,756,357 confirmed cases and 771,635 deaths from COVID-19 worldwide had been reported to the World Health Organization (https://www.who.int/). However, no specific medical treatments or efficient vaccines for SARS-CoV-2 have been developed. The development of a vaccine for the effective prevention of SARS-CoV-2 infection and control the spread of the COVID-19 pandemic is urgently needed.

Several vaccines for COVID-19 are currently being developed. Most of these vaccines target the SARS-CoV-2 spike (S) protein[3-5], which is the main target of the humoral immune response. The S protein protrudes from the SARS-CoV-2 surface and consists of two subunits S1 and S2[2,6]. S1 binds angiotensin-converting enzyme 2 on host cells through its receptor-binding domain (RBD) and S2 mediates membrane fusion during viral entry into host cells[7,8]. In addition, the high level of glycosylation of the SARS-CoV-2 S protein plays a role in epitope shielding to promote immune evasion and facilitates receptor binding[9,10]. Mutations of both N-and O-linked glycosylation sites in SARS-CoV-2 S protein markedly reduced the viral infectivity and are resistant to neutralizing antibodies[11]. The SARS-CoV-2 S protein is less densely glycosylated than HIV-1 and MERS-CoV and exhibits numerous vulnerabilities[12]. Hence, a comprehensive analysis of S protein glycosylation is pivotal to investigate the viral pathogenesis of SARS-CoV-2 in depth and rationally design a vaccine.

Glycosylation of SARS-CoV-2 S protein consists of N- and O-linked glycosylation, in which glycans are linked to a peptide through the side chains of asparagine and serine/threonine residues, respectively[13-15]. The N-linked glycans (NLGs) in SARS-CoV-2 S protein contain a trimannosyl chitobiose pentasaccharide core and is often larger, while the reported O-linked glycans (OLGs) in SARS-CoV-2 S have relatively diverse structures and are much smaller[9,10]. Twenty-two N-glycosylation sites (NGSs) as well as a few O-linked glycosylation sites (OGSs) have been

predicted on the SARS-CoV-2 S protein[9,10,16,17]. N-glycosylation of SARS-CoV-2 S has been investigated in-depth, and 22 NGSs on the S protein have been confirmed[9,10]. Unlike the confirmed NGSs on the SARS-CoV-2 S protein, the OGSs on SARS-CoV-2 S protein have been less studied due to methodological constrains. High-resolution cryo-electron microscopy and mass spectrometry can map the bulkier NLGs[8-10], but these techniques are unsuitable for mapping smaller, diversely-structured OLGs. Only two OGSs have been directly discovered from the digested SARS-CoV-2 S protein with high-resolution liquid chromatography-tandem mass spectrometry (LC-MS/MS)[9,10]. The low number of OGSs identified in the SARS-CoV-2 S protein has been ascribed to the extremely low O-glycosylation stoichiometry, with only 0.01% of the specific sites being actually O-glycosylated determined by LC-MS/MS[9]. However, the lack of knowledge on OGSs and OLGs in SARS-CoV-2 has translated in a lack of knowledge of the functions of O-glycosylation in this virus. To aid in vaccine design through providing an O-linked glycosylation profile, and determine the functions of SARS-CoV-2 S protein O-glycosylation, the specific enrichment of O-linked glycopeptides is required for their identification by mass spectrometry.

O-linked glycopeptides can be enriched with hydrazide chemistry[18], hydrophilic interaction chromatography[19], and lectin affinity chromatography methods[20]. However, the oxidation step in hydrazide chemistry method destroys the intact glycan structure; the hydrophilic interaction chromatography method cannot enrich efficiently the O-linked glycopeptides with small glycans[21]. Lectins are widely used for saccharide recognition[22] and can selectively enrich O-linked glycopeptides[23,24]. The recognition of specific saccharides by lectins is accomplished via a series of cooperative noncovalent interactions, such as hydrogen bond interaction and CH-pi interaction between specific amino acids on lectins and saccharides[22,25,26]. Comparison of the tendencies of amino acids in lectins to be involved in lectin-saccharide complexes in the UniProt database[27] revealed that Trp accounts for the highest distribution of CH-pi interactions between lectins and saccharide molecules, as Arg does for hydrogen

bonding interactions. Inspired by the specificity of lectin Trp and Arg amino acids for saccharides, we prepared a polymer consisting of Trp-Arg (WR) monomer-grafted silica microspheres (designated WR-SiO_2) and utilized WR-SiO_2 for O-glycosylation profiling of the SARS-CoV-2 S protein.

Results and discussions

The interaction between WR monomer and N-Acetylneuraminic acid. To investigate the interaction between WR monomer and N-acetylneuraminic acid (Neu5Ac, one of the most common monosaccharides), we use isothermal titration calorimetry (ITC) and hydrogen nuclear magnetic resonance (^{1}HNMR). Fig. 1a shows an isothermal calorimetric profile of 10 mM WR titrated with 50 mM Neu5Ac in ammonium formate solution (pH 3) at 25 °C. The strong interaction between WR monomer and Neu5Ac is revealed and corresponds to stepwise binding constants of K_{a1}, $1.17\times10^5\pm7.6\times10^3$ and K_{a2}, $9.30\times10^4\pm7.4\times10^3$ $L\cdot mol^{-1}$.

To further investigate the banding details between WR and Neu5Ac, we perform different molar ratios of Neu5Ac and WR for ^{1}H-NMR titration. As illustrated (Fig. 1b), the methylene signal of Neu5Ac undergoes remarkable changes towards high field after binding different molar ratio of WR monomer respectively. Meanwhile, chemical shift changes of the methylene signal of arginine is also observed in 1.7~1.8 ppm. CH-pi interactions supplied by indole group on tryptophan in 7.0~7.3 ppm towards low field could be also observed. Evidential changes in chemical shifts for the C-H protons of both WR host and Neu5Ac guest on the basis of hydrogen bonding and CH-pi interactions are observed, which indicated binding behavior between them.

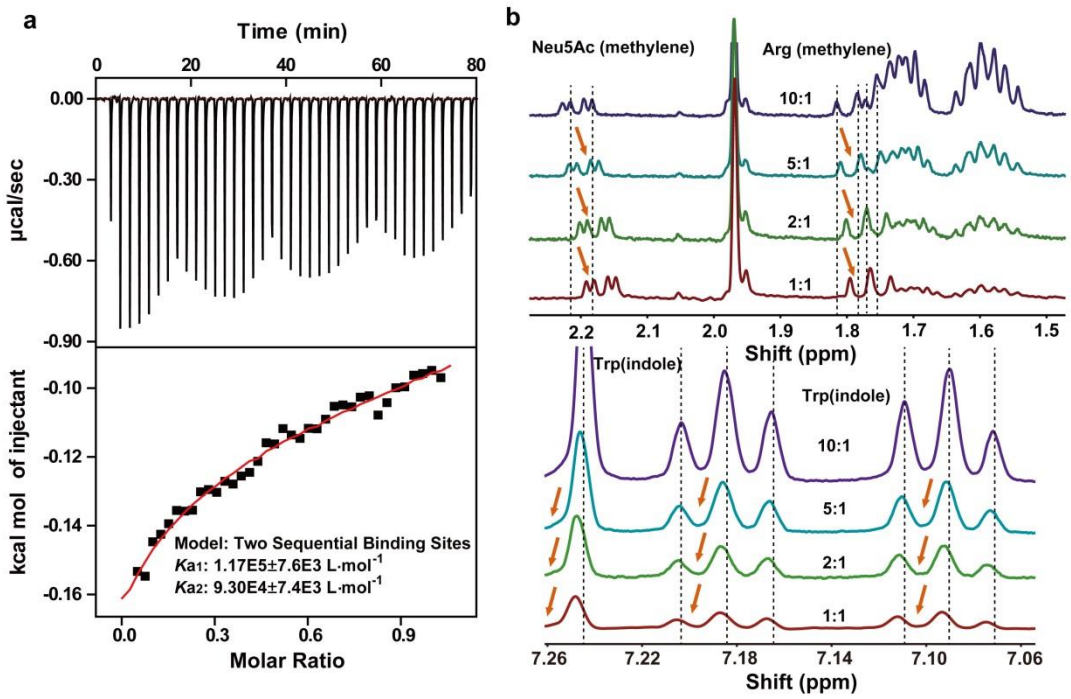


Fig. 1 The interactions between WR monomer and Neu5Ac. a Isothermal titration calorimetric data (in the upper panel) of 50 mM Neu5Ac (pH 3.0) added to a 10 mM WR aqueous solution (pH 3.0) at 20 °C. The red line (in the lower panel) represents a nonlinear fitting curve based on a sequential binding site model (N=2). **b** Partial hydrogen nuclear magnetic resonance (^{1}H NMR) spectra of the mixture of Neu5Ac and WR at molar ratios of 10:1, 5:1, 2:1 and 1:1 in D_2O at 20 °C; chemical shift changes of different groups of WR and Neu5Ac are indicated by orange arrows.

The synthesis and characterizations of WR-SiO_2. Based on the interactions between WR and Neu5Ac, WR-SiO_2 are successfully prepared using reversible addition-fragmentation chain-transfer polymerization method[28] (the synthesis route see Supplementary Fig. 1) and characterized with the thermogravimetric analysis (TGA), the infrared spectrometry, the scanning electron microscope (SEM) and the N_2 adsorption and desorption experiment (Supplementary Fig. 2). TGA curve of WR-SiO_2 shows a sharp decline above 150◦C and a platform with mass loss of 1.72% above 900 ◦C (Supplementary Fig. 2b). The infrared spectrum of WR-SiO_2 exhibits two obvious bands at about 3000 cm^{-1} and 1600 cm^{-1}, which correspond to the characteristic groups of WR (i.e., methylene and amine in Supplementary Fig. 2c).

Additionally, SEM image (Supplementary Fig. 2a) and nitrogen adsorption results show that WR-SiO_2 presents the surface morphology similar to bare SiO_2, and pores are preserved with average pore diameter of 300 Å (Supplementary Fig. 2d). All the results demonstrated that the polymers were successfully modified on SiO_2.

Highly selective enrichment of glycopeptides from model proteins. Based on the prepared WR-SiO_2 microspheres, we optimize a method for glycopeptide enrichment with standard protein digests (Fig. 2a). WR-SiO_2 demonstrates high selectivity for N-linked glycopeptides and enriches 35 N-linked glycopeptides (detailed information in Supplementary Table 1) from the tryptic digests of bovine fetuin (a standard glycoprotein) and bovine serum albumin (non-glycosylated protein) at a molar ratio of 1:100 (Supplementary Fig. 3a), compared to 18 N-linked glycopeptides enriched with the commercial resin ZIC-HILIC (Supplementary Fig. 3b). In particular, WR-SiO_2 is highly selective for the enrichment of O-linked glycopeptides and 33 O-linked glycopeptides (detailed information in Supplementary Table 2) are enriched from fetuin elastase digests with WR-SiO_2 (Fig. 2d), in sharp contrast to 0 O-glycopeptide before enrichment (Fig. 2b) and 18 O-linked glycopeptides with ZIC-HILIC (Fig. 2c). These data demonstrated the higher selectivity of WR-SiO_2 toward both N- and O-linked glycopeptides than commercial resin ZIC-HILIC.

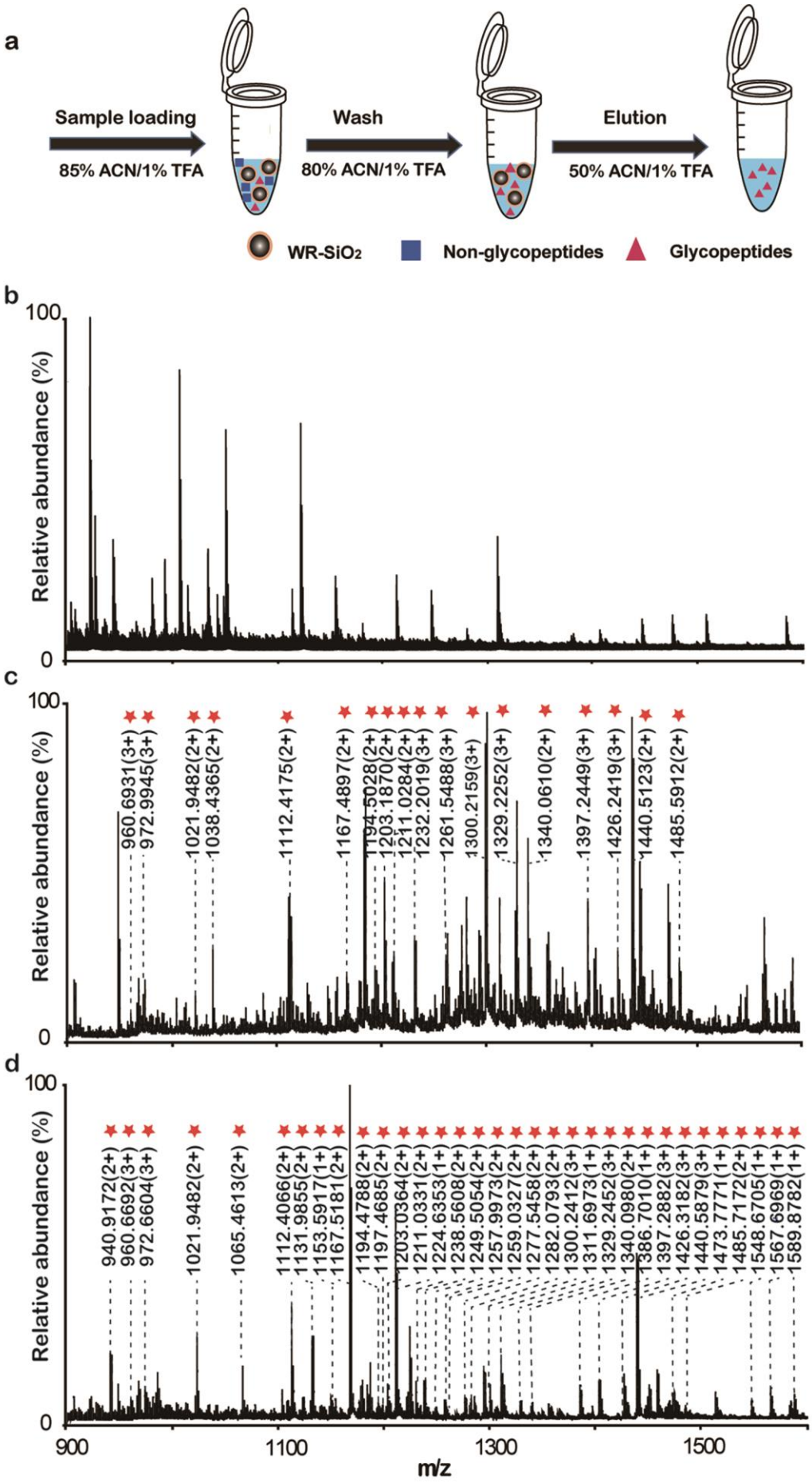
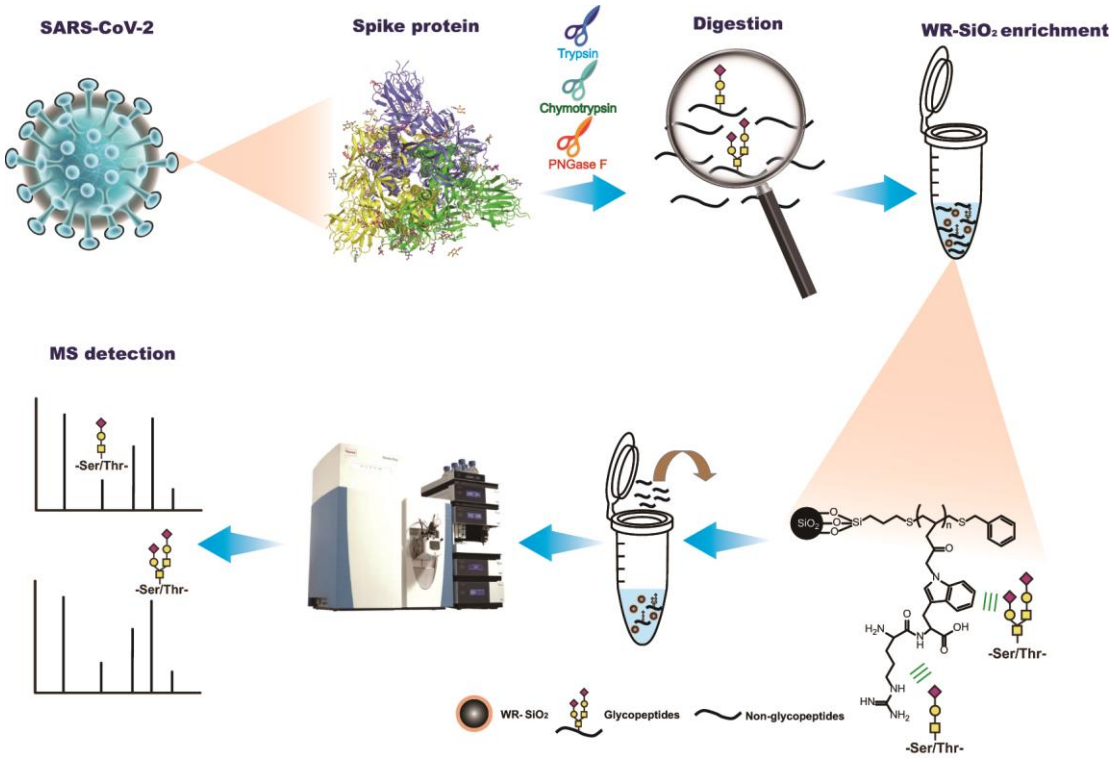

a
Sample loading
85% ACN/1% TFA
Wash
80% ACN/1% TFA
Elution
50% ACN/1% TFA
WR-SiO2
Non-glycopeptides
Glycopeptides
b
c
d
Relative abundance (%)
100
0
960.6931(3+)
972.9945(3+)
1021.9482(2+)
1038.4365(2+)
1112.4175(2+)
1167.4897(2+)
1194.5028(2+)
1203.1870(2+)
1211.0284(2+)
1232.2019(3+)
1261.5488(3+)
1300.2159(3+)
1329.2252(3+)
1340.0610(2+)
1397.2449(3+)
1426.2419(3+)
1440.5123(2+)
1485.5912(2+)
940.9172(2+)
960.6692(3+)
972.6604(3+)
1021.9482(2+)
1065.4613(2+)
1112.4066(2+)
1131.9855(2+)
1153.5917(1+)
1167.5181(2+)
1194.4788(2+)
1197.4685(2+)
1203.0364(2+)
1211.0331(2+)
1224.6353(1+)
1238.5608(2+)
1249.5054(2+)
1257.9973(2+)
1259.0327(2+)
1277.5458(2+)
1282.0793(2+)
1300.2412(3+)
1311.6973(1+)
1329.2452(3+)
1340.0980(2+)
1386.7010(1+)
1397.2882(3+)
1426.3182(3+)
1440.5879(3+)
1473.7771(1+)
1485.7172(2+)
1548.6705(1+)
1567.6969(1+)
1589.8782(1+)
900
1100
1300
1500
m/z

Fig 2. Enrichment of O-linked glycopeptides from elastase digested fetuin with WR-SiO_2 and commercial resin ZIC-HILIC. a The enrichment procedure of O-linked glycopeptides based on WR-SiO_2. **b** Mass spectrum of elastase digested fetuin before enrichment. **c** mass spectrum of O-glycopeptides enriched from elastase digested fetuin with ZIC-HILIC. **d** mass spectrum of O-glycopeptides enriched from elastase digested fetuin with WR-SiO_2. Detailed information about peptide sequences and glycosylation is shown in Supplementary Table 2. The enriched O-linked glycopeptides are labeled with red asterisks.

The strategy for glycosylation profiling of SARS-CoV-2 spike protein. The excellent performance of WR-SiO_2 for O-linked glycopeptides enrichment encourages us to apply it to analysis of SARS-CoV-2 S protein glycosylation. To this end, recombinant SARS-CoV-2 S protein expressed in HEK 293 is digested with trypsin and chymotrypsin and subsequently deglycosylated with PNGase F to remove N-linked glycans. The resultant peptides are incubated with WR-SiO_2 and the enriched O-linked glycopeptides are analyzed by LC-MS/MS. Each detected spectrum is manually validated (Scheme 1).



Scheme 1. Schematic representation of a strategy based on a biomimetic WR-SiO_2 polymer to enrich O-linked glycopeptides from the SARS-CoV-2 spike protein (PDB ID: 6VSB). Recombinant SARS-CoV-2 S protein expressed in HEK 293 was digested with trypsin and chymotrypsin and subsequently deglycosylated with PNGase F to remove N-linked glycans. The resultant peptides samples were incubated with WR-SiO_2. Attributing to the interactions between WR monomer on silica and O-linked glycans on O-linked glycopeptides, O-linked glycopeptides can be enriched effectively from the peptides samples. The enriched O-linked glycopeptides were further analyzed by LC-MS/MS.

The OGSs on SARS-CoV-2 spike protein. With this above strategy, 26 OGSs and 33 OLGs on SARS-CoV-2 S are unambiguously identified and from SARS-CoV-2 S, and among these OGSs and OLGs, 24 and 25, respectively, are novel (Fig. 3a, Supplementary Table 3). A total of 15 OGSs and 11 OGSs are located on the S1 and S2 subunits, respectively. The percentage of OGSs distributed on subunit S1 (61.5%) is consistent with that of NGSs distributed on subunit S1 (59.1%). Some

coronaviruses exhibit strong similarity and conservation in amino acid sequences of SARS-CoV-2 S, SARS-CoV S and RaTG13 S[16]. In this study we find that 18 OGSs are conserved among the SARS-CoV-2 S, SARS-CoV S and RaTG13 S proteins (Fig. 3b), with a higher percentage of conserved OGSs located on subunit S2 than on subunit S1. It is possible that the increased number of conserved OGSs as well as their attached glycans on the S2 subunit shield viral particles from the host immune responses, similar to what has been observed for NGSs and NLGs in other coronaviruses[12]. Intriguingly, 16 identified OGSs are located 1-4 amino acids away from the adjacent NGSs (Fig. 3b). For example, T333 is near to N331 and T345 is adjacent to N346. The NLGs on SARS-CoV-2 S are less dense than the viral glycoproteins on HIV-1 and MERS-CoV[9], and exhibit numerous vulnerabilities[12]. The close proximity of OGSs to NGSs in the SARS-CoV-2 S protein might compensate for the lower density of NLGs to form a solid glycan shield.

Among the novel OGSs, two (T333, T345) and two previously reported OGSs (T323, S325) are located in the receptor-binding domain (RBD) (Fig. 3a), which is the main target of the SARS-CoV-2 candidate vaccines currently in clinical trials[29-31]. In addition, three novel OGSs (T73, S161, and S221) are located in the NTD (Fig. 3a), which bind the monoclonal antibody isolated from COVID-19 patients[5]. Furthermore, two novel OGSs (S1170, S1175) are located in the HR2 region and 4 OGSs (S673, T676, S810 and S813) are proximal to S1/S2 cleavage sites (R685/R686)[16]. Notably, our method confirms the previously predicted SARS-CoV-2 S protein OGSs T618[17] and S673[16] are experimentally confirmed in this study.

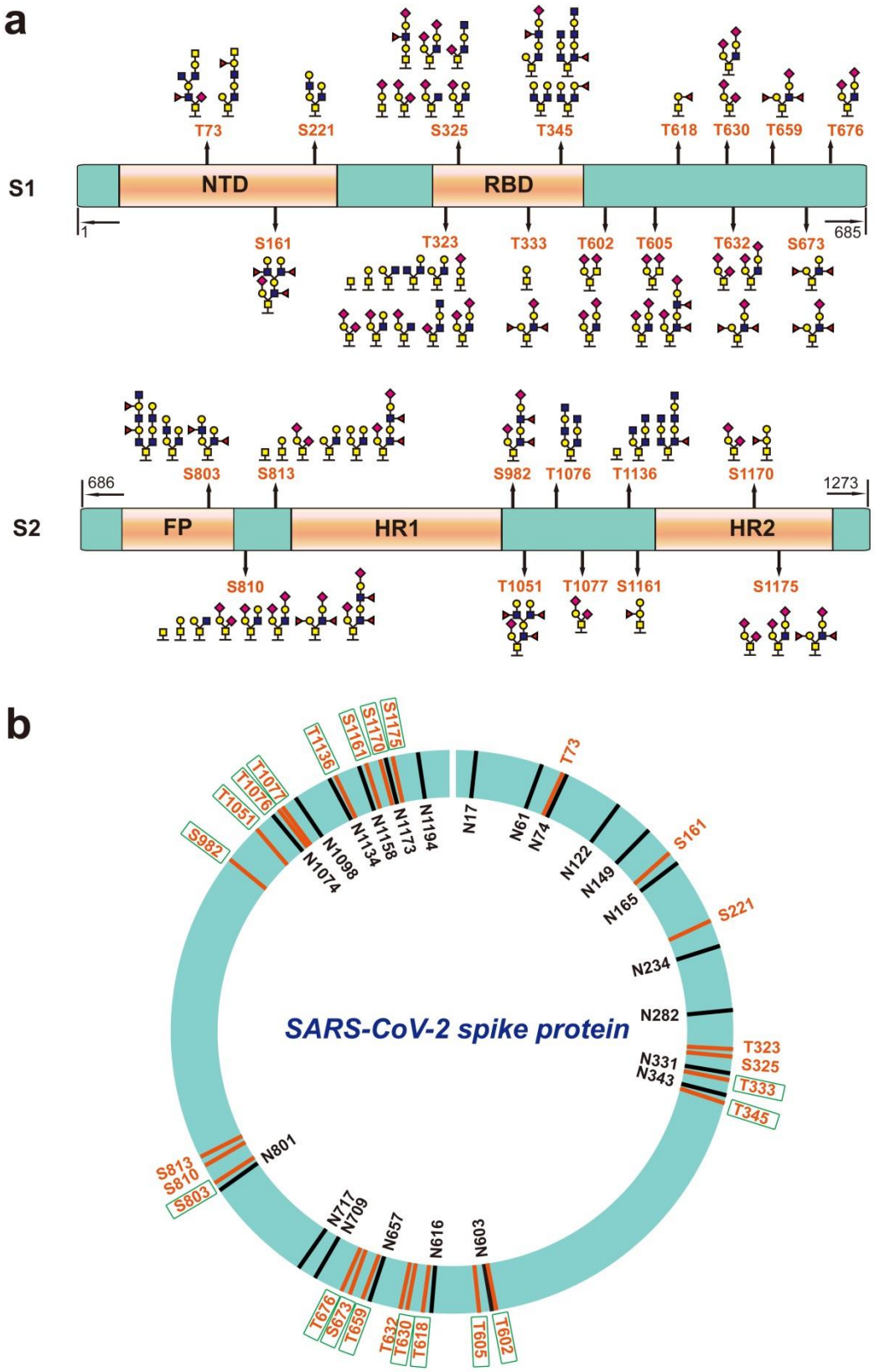


Fig. 3 O-Glycosylation profiling of SARS-CoV-2 S with WR-SiO_2-based enrichment and LC-MS/MS. a Schematic representation of O-glycosylation sites

(OGSs, orange) and O-linked glycans identified from the SARS-CoV-2 S protein. Protein domains in the S1 and S2 subunit are colored orange: N-terminal domain (NTD), receptor-binding domain (RBD), fusion peptide (FP), heptad repeat 1 (HR1) and heptad repeat 2 (HR2); □, GalNAc; ■, GlcNAc; ○, galactose; ◆, N-acetylneuraminic acid (Neu5Ac); and ▲, fucose. Detailed information about the peptide sequences and glycosylation is shown in Supplementary Tables 3. **b** The proximity of OGSs (in orange) and N-glycosylation sites (in black) and the conserved OGSs (outlined in green) on SARS-CoV-2 S protein, compared with SARS-CoV S and RaTG13 S proteins.

The OLGs and the abundances on SARS-CoV-2 spike protein. In addition to the 24 novel OGSs identified on the S protein, 33 OLGs were identified, and 25 of them are novel. The distribution and normalized abundance of site-specific OLGs are shown in Fig. 4. Among the identified OLGs, 15 are sialylated and 14 are fucosylated, accounting for 45.4% and 42.4% of the total number of identified OLGs; this is in contrast to the percentage of sialylated and fucosylated NLGs on SARS-CoV-2 S (15% and 52% , respectively)[9]. To our knowledge, it is the first time that fucosylated OLGs are reported on SARS-CoV-2 S.

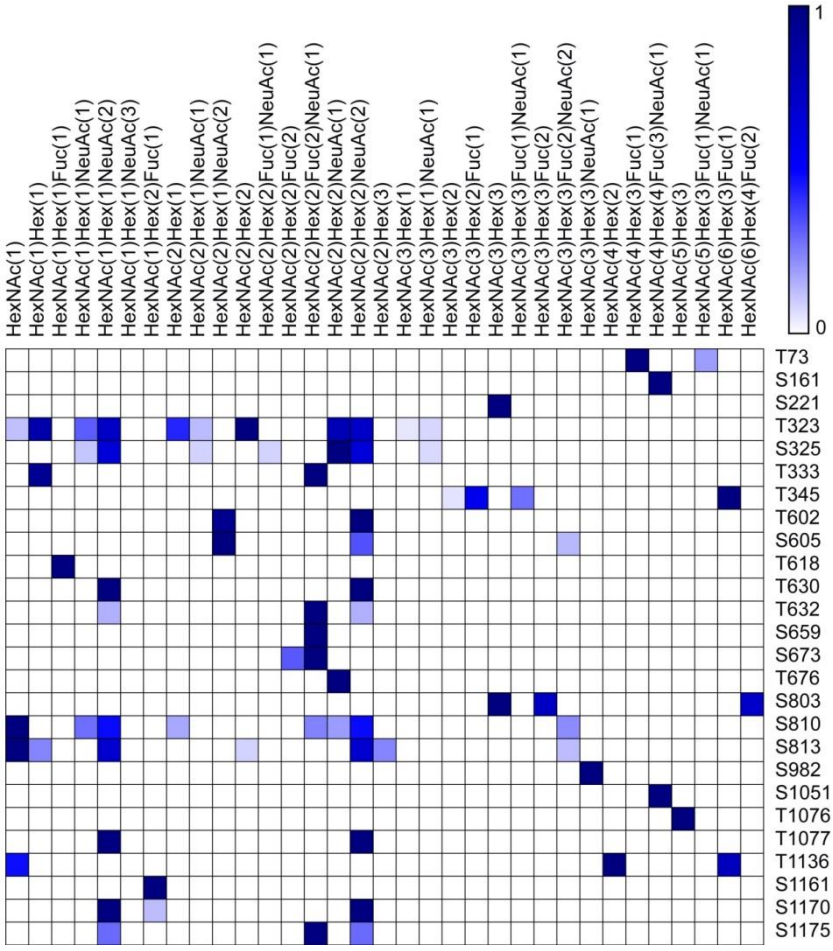


Fig. 4 The distribution of site-specific glycans and the normalized abundance of each O-linked glycopeptide.

We find that 5 OLGs are widely distributed on different OGSs, such as HexNAc(1)Hex(1)NeuAc(2) on 9 OGSs (Fig. 5a). The abundance of the 5 most abundant OLGs distributed on different OGSs is shown in Fig. 5b-5f. Three sialylated OLGs on residues T323 and S325 in the RBD display the highest abundance (Fig. 5b, c and e), explaining why these 2 OGSs were the first identified ones. Notably, 8 OLGs are uniquely distributed on 8 different OGSs, emphasizing the site-specific nature of SARS-CoV-2 S protein glycosylation (Supplementary Fig. 4).

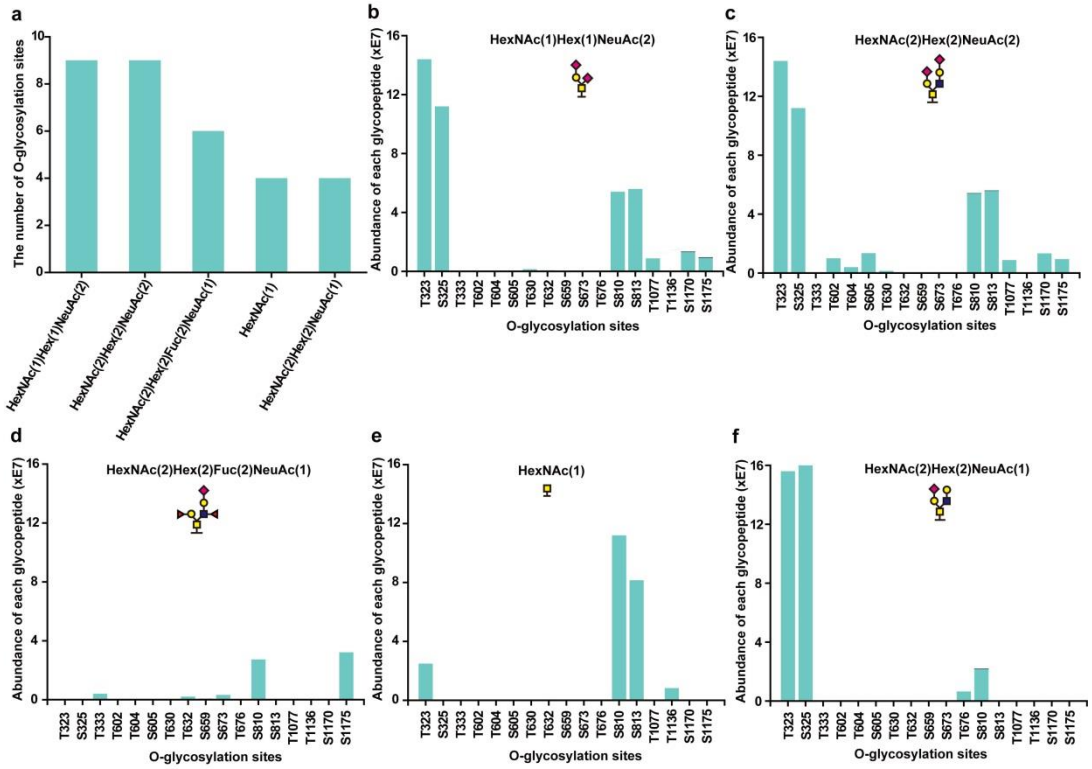


Fig. 5 The five most highly abundant O-linked glycans (OLGs) on the SARS-CoV-2 S protein are distributed on different O-glycosylation sites (OGSs). **a** The number of OGSs occupied by the five most highly abundant OLGs. (**b-f**) The abundance of each individual OLG (with the corresponding glycan structure inserted) on different OGSs.

Enrichment of N-linked glycopeptides from SARS-CoV-2 S. Besides OLGs, we also used WR-SiO_2 to enrich N-linked glycopeptides from SARS-CoV-2 S digested with trypsin and chymotrypsin. Using this method we identified 80 NLGs, including 15 novel NLGs and 20 NGSs (Supplementary Table 4). Compared with the previously reported 22 NGSs and 107 NLGs[9,10], the slight lower number of NLGs and NGSs ascribes to the different enzymes used for digestion. These results provide more detailed structural information on the O-glycosylation of SARS-CoV-2 S, which may facilitate the vaccine development.

In summary, the comprehensive analysis of SARS-CoV-2 S protein O-glycosylation conducted in this study reveals the novel glycosylation of SARS-CoV-2 S protein. An enrichment strategy based on WR-SiO_2 displayed the

capacity to enrich O-glycosylated peptides for mass spectrometry-based identification. This lectin inspired design idea could be expanded to other oligopeptides-based materials, which may have specific adsorption toward O-linked glycopeptides, solving knotty enrichment challenges in glycoprotoemics. Furthermore, detailed information on SARS-CoV-2 S protein glycosylation will facilitate precise vaccine design and allow vaccine quality control, as well as pave the way for exploring the viral pathobiology.

Methods

Synthesis and characterization of WR-SiO_2. WR-SiO_2 was prepared using reversible addition-fragmentation chain-transfer polymerization method[28]. Scanning electron microscope (SEM) images were obtained on JEOL JSM-7800F scanning electron microscope (JEOL, Tokyo, Japan). FT-IR was performed on a Therno Scientific FT-IR spectrometer. Thermogravimetric analysis (TGA) was performed under air atmosphere at a heating rate of 10 °C/min from room temperature to 900 °C (Netzsch, Selb, Germany). The samples were dried at 80 °C under vacuum overnight prior to measurement. Surface areas and pore size distribution were calculated using the Brunauer-Emmett-Teller (BET).

ITC and NMR titration. ITC measurements were made with a MicroCal VP-ITC200 calorimeter (MicroCal, Northampton, MA). To avoid air bubbles, all solutions were degassed under vacuum prior to use. Neu5Ac titration were performed by injecting 10 μL (50 mM) into the calorimeter cell containing WR monomer (10 mM), at constant time intervals of 2 min. A control titration in which Neu5Ac was titrated into 50 mM solution (pH 3.0) was used to correct for the heat of dilution. Raw data were processed using Origin graphing software provided with the instrument. To test the detailed binding behavior between WR and Neu5Ac by NMR, different molar ratios of WR monomer were added to the host solution of Neu5Ac (10 mM). After equilibration overnight at ambient temperature, the chemical shifts of C-H protons were recorded and analyzed on Bruker AVANCE III 400 MHz spectrometer.

Digestion of protein samples. Each standard protein (bovine serum albumin, and bovine fetuin) of 1 mg was dissolved with 100 μL of urea (6 M) in 50 mM ammonium bicarbonate (NH_4HCO_3). The protein solution was treated with 20 μL of 200 mM dithiothreitol (DTT) for 45 min at 56 ℃. Then 80 μL of 200 mM iodoacetamide (IAA) was added and the mixture was incubated in the dark for 30 min at room temperature. The mixture was diluted to 1 mL with 50 mM NH_4HCO_3 and digested with different enzymes (trypsin for N-glycopeptide enrichment, PNGase F and elastase for O-glycopeptide enrichment) at the enzyme to proteins mass ratio of 1:25 and incubated at 37 ℃ overnight.

Spike protein was dissolved with 100 μL of urea (6 M) in 50 mM NH_4HCO_3. The protein solution was treated with 20 μL of 200 mM DTT for 45 min at 56 ℃. Then 80 μL of 200 mM IAA was added and the mixture was incubated in the dark for 30 min at room temperature. Trypsin and chymotrypsin were used for digestion at the enzyme to proteins mass ratio of 1:20. After that, PNGase F was used to remove N-glycans for the following enrichment of O-linked glycopeptides.

Enrichment of O-linked glycopeptides. GELoader tips were packed with 1 mg of material suspended in 30 μL of ACN. The tip was equilibrated with 30 μL of 85% ACN/1%TFA. 30 μL of peptides sample (equivalent to 5 μg of fetuin or 30 μg of SARS-CoV-2 spike protein) in 85% ACN/1% TFA was loaded on the sorbent, washed with 80% ACN/1% TFA (30 μL) twice, and eluted with 50 % ACN/1% TFA (30 μL) successively. The eluent was collected, concentrated and redissolved for ESI-Q/TOF-MS analysis.

Enrichment of N-linked glycopeptides. The peptides mixtures of fetuin and BSA at a molar ratio of 1:100 were first mixed with 2 mg of material (in 85% ACN/1% TFA) and shaken for 1 h followed by centrifugation at 10,000 g for 2 min. After that, the supernatant was removed and the precipitation was washed with 85% ACN/1% TFA (100 μL*4) and centrifuged at 10,000 g for 2 min. The precipitate was transferred into the GELoader tip and washed with 80% ACN/1% TFA (30 μL) for three times. Then, the bound peptides on sorbent was eluted with 50% ACN/1% TFA (30 μL). The eluent was collected and desalted for analysis by ESI Q/TOF-MS or nanoLC-MS/MS.

Mass spectrometry analysis. Glycopeptides enriched from protein samples were analyzed with a nano ESI Q-TOF MS (Waters, Milford, MA, USA). The samples were infused into the ESI source with Nano Acquity UPLC (Waters). The MS analysis was performed under the positive ion mode. Capillary voltage was 2.3 kV and source temperature was 100 °C. Full scan MS data and tandem MS/MS data were acquired at m/z 600-2000 and 100-2000, respectively. The glycopeptides and deglycopeptides obtained from SARS-CoV-2 spike protein were separated and characterized using Orbitrap coupled with Accela 600 HPLC System (Thermo, CA, USA). For the reverse-phase liquid chromatography separation, 0.1% FA (pH: 2.59) in water and in CH_3CN was used as mobile phases A and B, respectively. The analytical column with an inner diameter of 75 μm was packed in-house with Daisogel C18 AQ particles (3 μm, 120 Å) to 12 cm length. The flow rate was set at 300 nL/min. The 53 min gradient elution was performed with a gradient of 0-2% B in 10 min, 2-6% B in 1 min, 6-50% B in 35 min, 50-90% B in 2 min and 90% B in 5 min. Full mass scans were carried out on the Orbitrap with acquisition range from m/z 400 to 2000 (R = 60,000 at m/z 400). Full mass scans were carried out on the Orbitrap with acquisition range from m/z 400 to 2000 (R= 60,000 at m/z 400). The 20 most intense ions from the full scan were selected for fragmentation via high–energy collisional dissociation (HCD) in the ion trap (relative collision energy for HCD was set to 27%). The dynamic exclusion function was set as follows: repeat count 1, repeat duration 30 s and exclusion duration of 60 s.

MS data analysis. All the RAW data files obtained from Orbitrap were searched against the SARS-CoV-2 S protein sequence using Byonic software (version 3.6.0, Protein Metrics, Inc.) with the mass tolerance for precursors and fragment ions set at 10 ppm and 20 ppm, respectively. Two missed cleavage sites were allowed for trypsin or/and chymotrypsin digestion. The fixed modification was carbamidomethyl (C), and variable modifications included oxidation (M), acetyl (protein N-term), and deamidation (N). Trypsin and chymotrypsin were set as the specific proteolytic enzyme with up to two missed cleavages allowed. Peptides with charge states of 2, 3

and 4 were chosen for further fragmentation. The FDR were all set as <1%. Moreover, the data were searched against reverse and contaminant sequences.

Data availability

All data is available in the main text or the Supplementary informations.

References

1. Kissler, S.M., Tedijanto, C., Goldstein, E., Grad, Y.H. & Lipsitch, M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science* **368**, 860-868 (2020).
2. Yan, R., et al. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* **367**, 1444-1448 (2020).
3. Mullard, A. COVID-19 vaccine development pipeline gears up. *Lancet* **395**, 1751-1752 (2020).
4. Dagotto, G., Yu, J. & Barouch, D.H. Approaches and Challenges in SARS-CoV-2 Vaccine Development. *Cell Host Microbe.*, (2020). doi: 10.1016/j.chom.2020.08.002.
5. Chi, X., et al. A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. *Science (New York, N.Y.)* **369**, 650-655 (2020).
6. Wrapp, D., et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science (New York, N.Y.)* **367**, 1260-1263 (2020).
7. Hoffmann, M., et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **181**, 271-280 (2020).
8. Walls, A.C., et al. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **181**, 281-292 (2020).
9. Watanabe, Y., Allen, J.D., Wrapp, D., McLellan, J.S. & Crispin, M. Site-specific glycan analysis of the SARS-CoV-2 spike. *Science (New York, N.Y.)* **369**, 330-333 (2020).

10. Shajahan, A., Supekar, N.T., Gleinich, A.S. & Azadi, P. Deducing the N- and O-glycosylation profile of the spike protein of novel coronavirus SARS-CoV-2. *Glycobiology*, (2020). doi: 10.1093/glycob/cwaa042.
11. Li, Q., et al. The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity. *Cell*, (2020). doi.org/10.1016/j.cell.2020.07.012
12. Watanabe, Y., et al. Vulnerabilities in coronavirus glycan shields despite extensive glycosylation. *Nat. Commun.* **11**, 2688-2688 (2020).
13. Joshi, H.J., et al. SnapShot: O-Glycosylation Pathways across Kingdoms. *Cell* **172**, 632-632 (2018).
14. de Haan, N., Falck, D. & Wuhrer, M. Monitoring of immunoglobulin N- and O-glycosylation in health and disease. *Glycobiology* **30**, 226-240 (2020).
15. Ruhaak, L.R., Xu, G., Li, Q., Goonatilleke, E. & Lebrilla, C.B. Mass Spectrometry Approaches to Glycomic and Glycoproteomic Analyses. *Chem. Rev.* **118**, 7886-7930 (2018).
16. Andersen, K.G., Rambaut, A., Lipkin, W.I., Holmes, E.C. & Garry, R.F. The proximal origin of SARS-CoV-2. *Nat.Med.* **26**, 450-452 (2020).
17. Vankadari, N. & Wilce, J.A. Emerging WuHan (COVID-19) coronavirus: glycan shield and structure prediction of spike glycoprotein and its interaction with human CD26. *Emerg. Microbes. Infect.* **9**, 601-604 (2020).
18. Yang, S., Hu, Y., Sokoll, L. & Zhang, H. Simultaneous quantification of N- and O-glycans using a solid-phase method. *Nature Protocols* **12**, 1229-1244 (2017).
19. Zhang, L., *et al.* Facile Fabrication of Biomimetic Chitosan Membrane with Honeycomb-Like Structure for Enrichment of Glycosylated Peptides. *Anal. Chem.* **91**, 2985-2993 (2019).
20. Anan, G., *et al.* The Impact of Glycosylation of Osteopontin on Urinary Stone Formation. *Int. J. Mol. Sci.* **21**, (2020). doi: 10.3390/ijms21010093.
21. You, X., Qin, H. & Ye, M. Recent advances in methods for the analysis of protein o-glycosylation at proteome level. *J. Sep. Sci.* **41**, 248-261 (2018).

22. Lis, H. & Sharon, N. Lectins: Carbohydrate-specific proteins that mediate cellular recognition. *Chem. Rev.* **98**, 637-674 (1998).
23. Singh, Y., et al. Positional Scanning MUC1 Glycopeptide Library Reveals the Importance of PDTR Epitope Glycosylation for Lectin Binding. *J. Org. Chem.* **85**, 1434-1445 (2020).
24. Narimatsu, Y.*, et al.* Exploring Regulation of Protein O-Glycosylation in Isogenic Human HEK293 Cells by Differential O-Glycoproteomics. *Mol. Cell. Proteomics* **18**, 1396-1409 (2019).
25. Kitov, P.I., et al. Shiga-like toxins are neutralized by tailored multivalent carbohydrate ligands. *Nature* **403**, 669-672 (2000).
26. Barwell, N.P. & Davis, A.P. Substituent Effects in Synthetic Lectins-Exploring the Role of CH-pi Interactions in Carbohydrate Recognition. *J. Org. Chem.* **76**, 6548-6557 (2011).
27. Hudson, K.L., et al. Carbohydrate-Aromatic Interactions in Proteins. *J. Am. Chem. Soc.* **137**, 15152-15160 (2015).
28. Yu, D., et al. A controlled thiol-initiated surface polymerization strategy for the preparation of hydrophilic polymer stationary phases. *Chem. Commun.* **51**, 14778-14780 (2015).
29. Yang, J., et al. A vaccine targeting the RBD of the S protein of SARS-CoV-2 induces protective immunity. *Nature*, (2020). doi: 10.1038/s41586-020-2599-8
30. Iqbal Yatoo, M., et al. COVID-19-Recent advancements in identifying novel vaccine candidates and current status of upcoming SARS-CoV-2 vaccines. *Hum. Vaccines Immunother.*, (2020). doi: 10.1080/21645515.2020.1788310
31. Pang, J., et al. Potential Rapid Diagnostics, Vaccine and Therapeutics for 2019 Novel Coronavirus (2019-nCoV): A Systematic Review. *J. Clin. Med.* **9**, 623-652 (2020).

Acknowledgments

This work was supported by the National Natural Science Foundation of China (21934005 and 21775148).

Author contributions: X. M. L. and X. L. L. secured fundings and contributed to the experimental design and revised the manuscript. X. F. D. synthesized, characterized the materials and wrote the initial draft of the manuscript. C. C. pretreated the proteins samples and finished the control experiments. J. Y. Y. finished the N-glycopeptides enrichment experiments. All authors contributed in data collection and interpretation of the results.

Competing interests

Authors declare no competing interests.

Figures

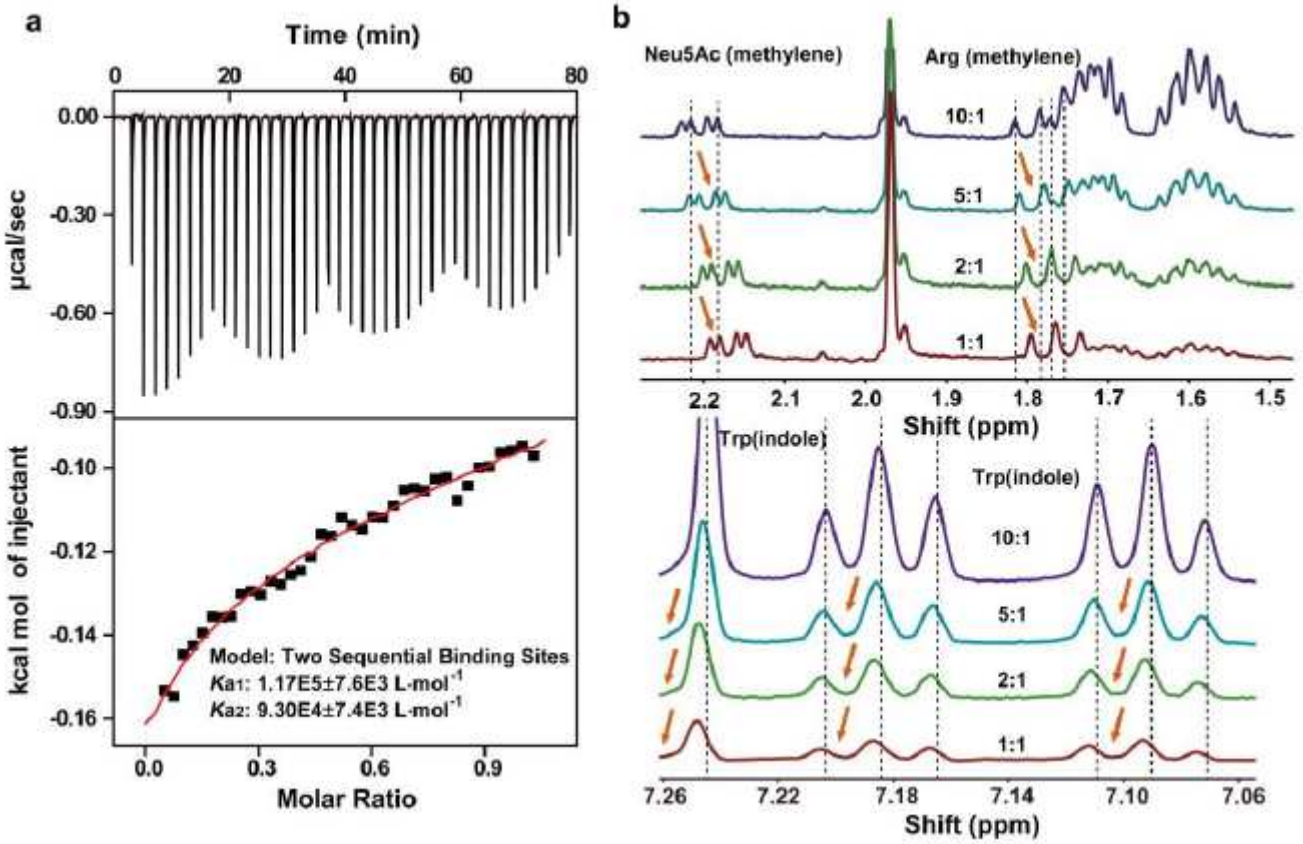


Figure 1

The interactions between WR monomer and Neu5Ac. a Isothermal titration calorimetric data (in the upper panel) of 50 mM Neu5Ac (pH 3.0) added to a 10 mM WR aqueous solution (pH 3.0) at 20 °C. The red line (in the lower panel) represents a nonlinear fitting curve based on a sequential binding site model (N=2). b Partial hydrogen nuclear magnetic resonance (1H NMR) spectra of the mixture of Neu5Ac and WR at molar ratios of 10:1, 5:1, 2:1 and 1:1 in D2O at 20 °C; chemical shift changes of different groups of WR and Neu5Ac are indicated by orange arrows.

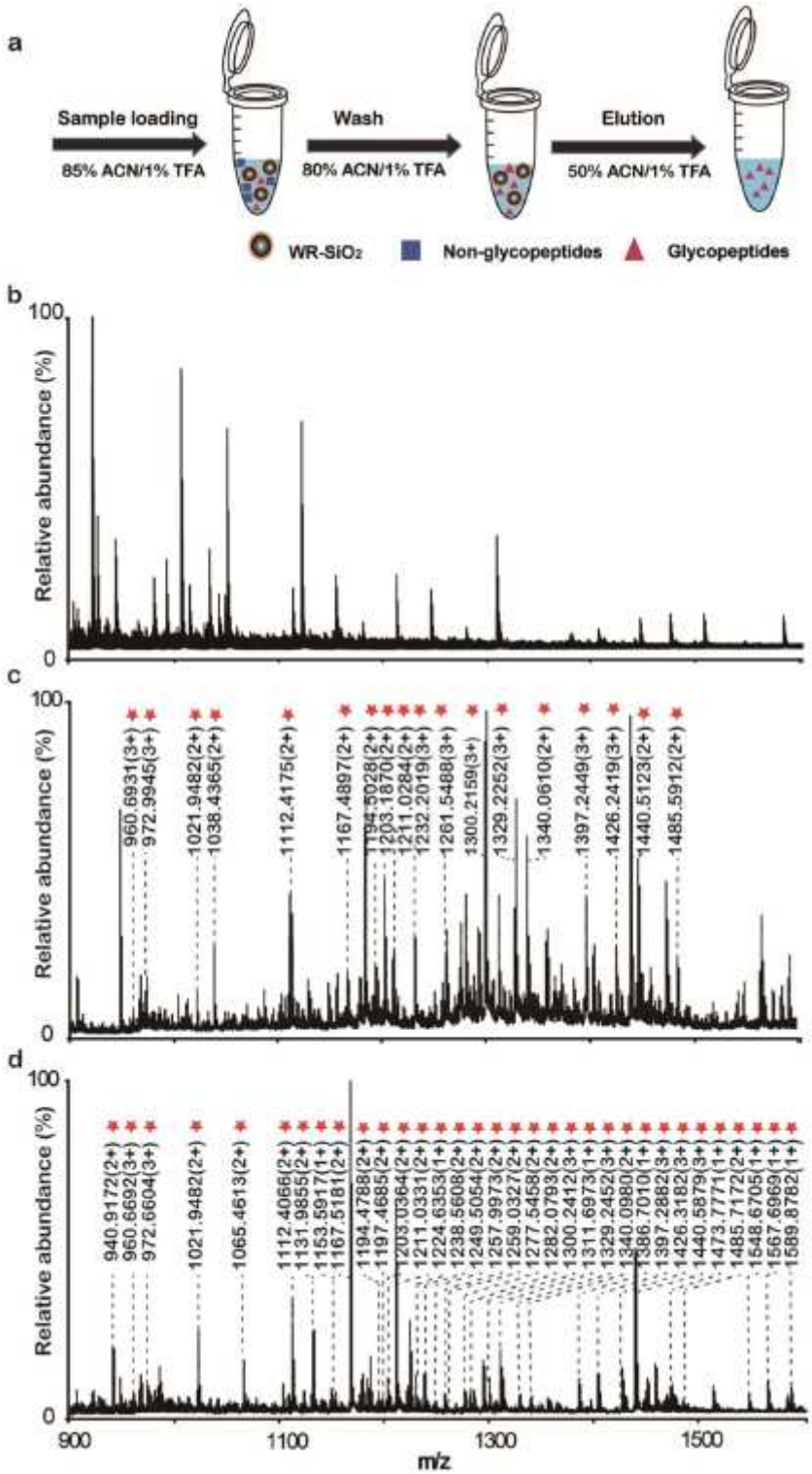


Figure 2

Enrichment of O-linked glycopeptides from elastase digested fetuin with WR-SiO2 and commercial resin ZIC-HILIC. a The enrichment procedure of O-linked glycopeptides based on WR-SiO2. b Mass spectrum of elastase digested fetuin before enrichment. c mass spectrum of O-glycopeptides enriched from elastase digested fetuin with ZIC-HILIC. d mass spectrum of O-glycopeptides enriched from elastase digested fetuin with WR-SiO2. Detailed information about peptide sequences and glycosylation is shown in Supplementary Table 2. The enriched O-linked glycopeptides are labeled with red asterisks.

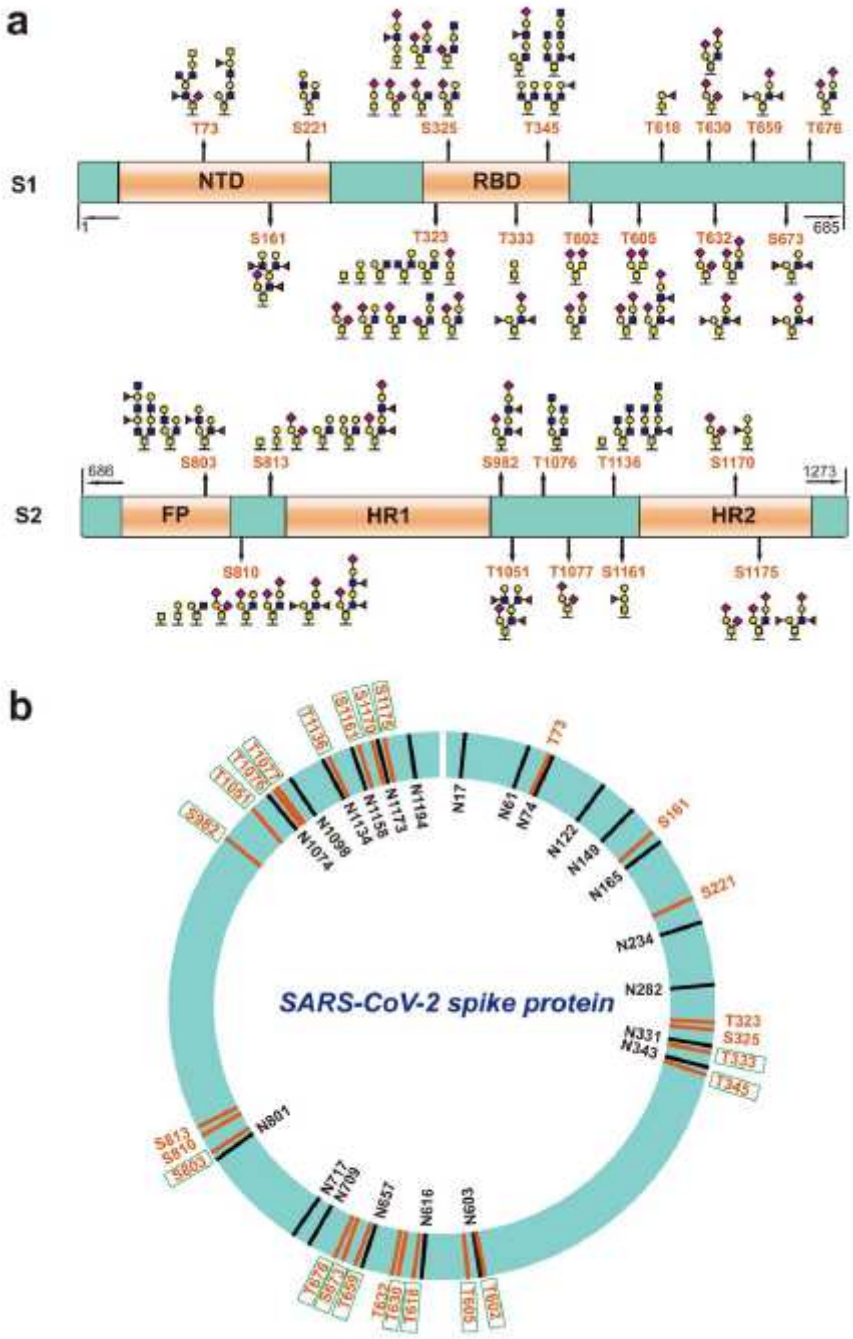


Figure 3

O-Glycosylation profiling of SARS-CoV-2 S with WR-SiO2-based enrichment and LC-MS/MS. a Schematic representation of O-glycosylation sites (OGSs, orange) and O-linked glycans identified from the SARS-CoV-2 S protein. Protein domains in the S1 and S2 subunit are colored orange: N-terminal domain (NTD), receptor-binding domain (RBD), fusion peptide (FP), heptad repeat 1 (HR1) and heptad repeat 2 (HR2); , GalNAc; , GlcNAc; , galactose; , N-acetylneuraminic acid (Neu5Ac); and , fucose. Detailed information about the peptide sequences and glycosylation is shown in Supplementary Tables 3. b The proximity of OGSs (in orange) and N-glycosylation sites (in black) and the conserved OGSs (outlined in green) on SARS-CoV-2 S protein, compared with SARS-CoV S and RaTG13 S proteins.

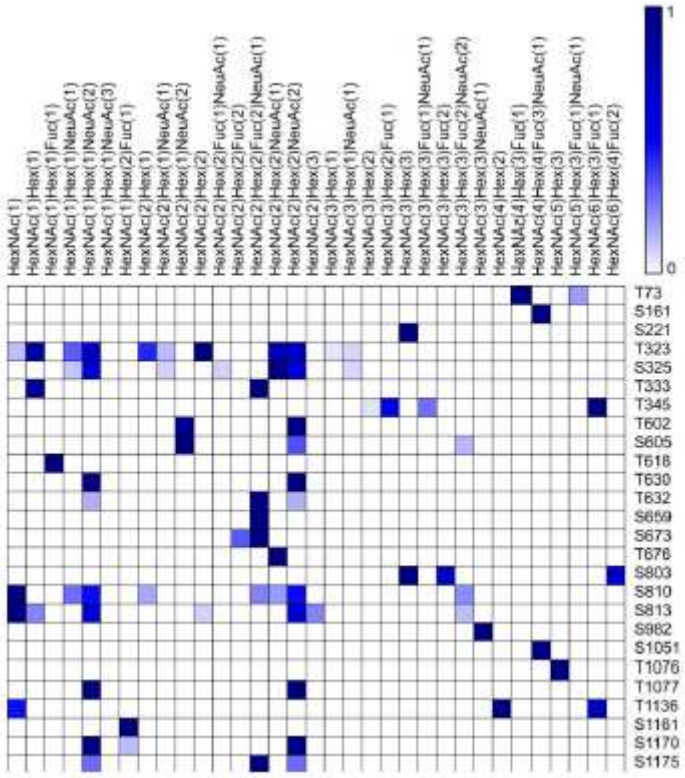


Figure 4

The distribution of site-specific glycans and the normalized abundance of each O-linked glycopeptide.

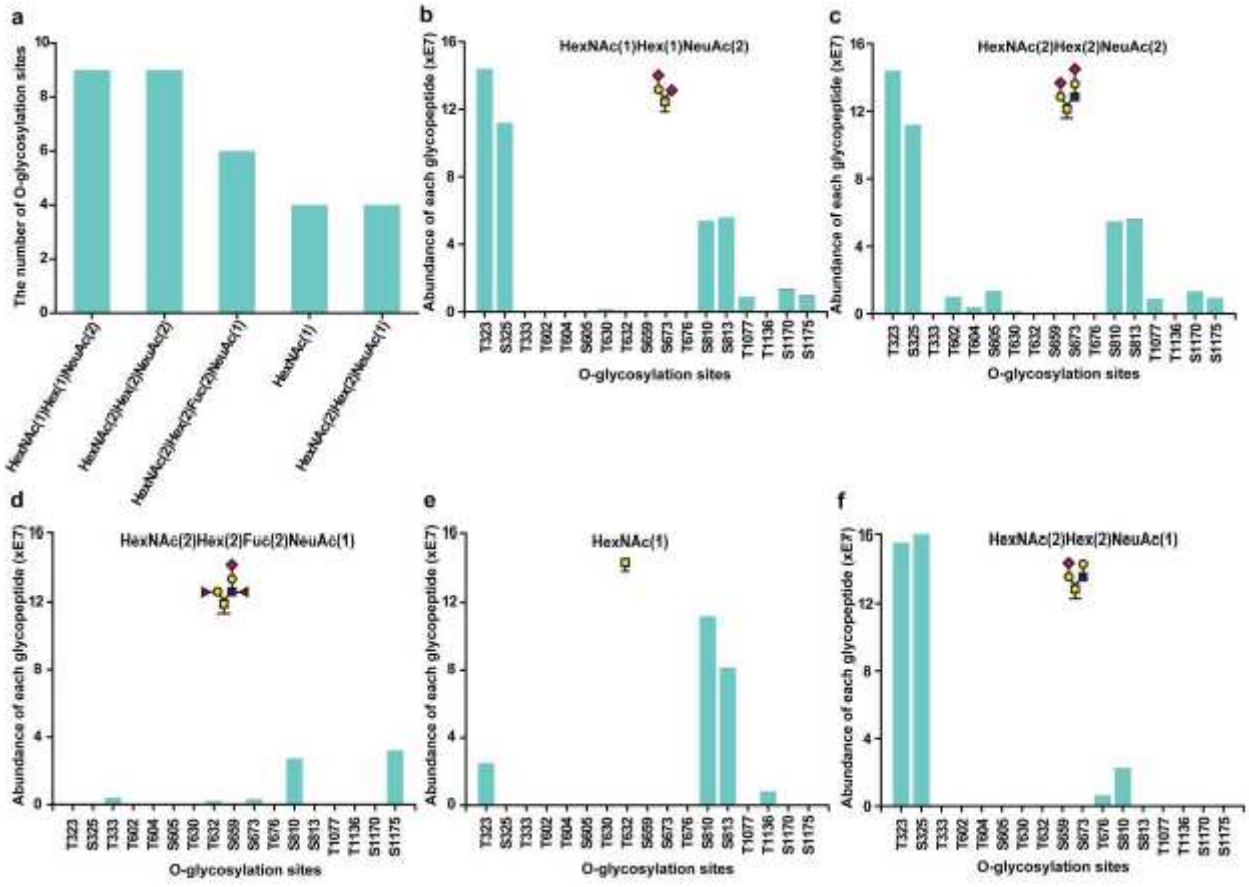


Figure 5

The five most highly abundant O-linked glycans (OLGs) on the SARS-CoV-2 S protein are distributed on different O-glycosylation sites (OGSs). a The number of OGSs occupied by the five most highly abundant OLGs. (b-f) The abundance of each individual OLG (with the corresponding glycan structure inserted) on different OGSs.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- NatureCommunicationSupplementaryInformations.pdf
- TableS1.NlinkedglycopeptidesinfetuindigestwithfoldBSA.xlsx
- TableS2.OlinkedglycopeptidesinfetuindigesttreatedbyelastaseTableS3.OlinkedglycosylationinSARSCoV2spikeprotein.xls
- TableS4.NlinkedglycosylationinSARSCoV2spikeprotein.xls