Systematic analysis and comparison of *O*-glycosylation of five recombinant spike proteins in β-coronaviruses

Xuefang Dong [a], Xiuling Li [a, b, **], Cheng Chen [a], Xiaofei Zhang [a], Xinmiao Liang [a, b, *]

[a] *Key Laboratory of Separation Science for Analytical Chemistry, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian, 116023, PR China*
[b] *Ganjiang Chinese Medicine Innovation Center, Nanchang, 330000, PR China*

HIGHLIGHTS

- *O*-glycosylation of β-CoVs S1 proteins was characterized with the revealed macro- and micro-heterogeneity.
- Differently exposed sialylated *O*-glycosylation on β-CoV S1 proteins on visual model was displayed.
- The *O*-glycosylation differences between SARS-CoV-2 and its mutant D614G S1 proteins on functional domains were uncovered.

GRAPHICAL ABSTRACT
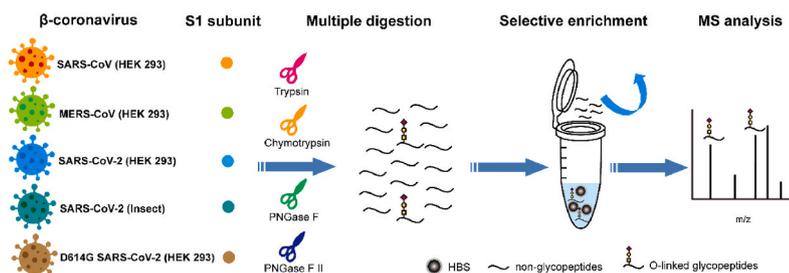
ABSTRACT

β-coronaviruses (β-CoVs), representative with severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), depend on their highly glycosylated spike proteins to mediate cell entry and membrane fusion. Compared with the extensively identified *N*-glycosylation, less is known about *O*-glycosylation of β-CoVs S proteins, let alone its biological functions. Herein we comprehensively characterized *O*-glycosylation of five recombinant β-CoVs S1 subunits and revealed the macro- and micro-heterogeneity nature of site-specific *O*-glycosylation. We also uncovered the *O*-glycosylation differences between SARS-CoV-2 and its natural D614G mutant on functional domains. This work describes the systematic *O*-glycosylation analysis of β-CoVs S1 proteins and will help to guide the related vaccines and antiviral drugs development.

1. Introduction

COVID-19 pandemic, which was caused by SARS-CoV-2 belonging to β-coronaviruses (β-CoVs), brings a severe threat to human health and economy all over the world [1]. Other β-CoVs including severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV) emerged in past two decades are also infectious in the human population and cause acute respiratory system diseases [2].

β-CoVs utilize the extensively glycosylated spike (S) proteins to bind

---

the host receptors, accomplish attachment and invasion of host cells [2]. SARS-CoV-2 and SARS-CoV S proteins recognize angiotensin-converting enzyme 2 (ACE2) of host cells whereas MERS-CoV S proteins use dipeptidyl peptidase 4 (DPP4) as receptors to accomplish related invasion machinery [3,4]. The β-CoVs S have a similar 3D structure and contains two functional subunits, S1 and S2 [2]. S1 subunit is responsible for receptor binding and S2 subunit is involved in mediating membrane fusion [5,6]. S1 subunit utilizes receptor binding domain (RBD) to recognize and complex with the specific protein receptors on host cells [7–9], and N-terminal domain (NTD) to recognize the specific sugar receptor in some β-CoVs [10,11]. Another crucial region on S1 subunit is the cleavage site at the S1/S2 boundary, responsible for activation of β-CoV S proteins [12,13]. Consequently, S1 subunit is the main target of neutralizing antibodies and developing vaccines in clinical trials, and the determinant of tissue tropism and host ranges in CoVs [14,15].

Viral glycosylation determines the adaptive immunity and infectivity [16,17]. Generally β-CoVs need to hijack host cellular and molecular machinery to glycosylate their proteins in the endoplasmic reticulum (ER) apparatus and further process in the ER-Golgi compartment [18]. The well-processed glycans can mask viral immune epitopes expediently through imitating host innate glycans and facilitate the viruses to evade cunningly. The level of glycosylation processing indicates the viral evolutionary adaptation capacity [19].

N- and *O*-glycosylation are two principal glycosylation types in viruses [20]. The *N*-glycosylation of S proteins in SARS-CoV (23 N-glycosylation sites), MERS-CoV (23 *N*-glycosylation sites) and SARS-CoV-2 (22 N-glycosylation sites) has been comprehensively investigated [5,19,21]. Compared with SARS-CoV and MERS-CoV S proteins [16], SARS-CoV-2 has less *N*-glycosylation sites [19]. Quite strikingly, even with sparser glycan shielding and more vulnerable exposed area, SARS-CoV-2 is still more pathogenic and infectious than other β-CoVs [22,23]. Encoding this phenomenon is critical to develop antivirals and vaccines. Akin to viral *N*-glycosylation, viral *O*-glycosylation has the fundamental roles in viral entry, propagation and glycan shielding as well, which is associated with evolutionary adaptation and the viral infectivity [24–26]. Unlike the *N*-glycosylation sites with canonical sequons, a common core pentasaccharide and high occupancy (95%), *O*-glycosylation sites without conserved sequon are modified with diverse *O*-linked glycans in a very low occupancy (0.01%~4%) manner on SARS-CoV-2 S proteins [19,27,28], which severely impedes the research on the viral *O*-glycosylation in depth. Therefore, more efforts are urgently needed to explore the *O*-glycosylation of β-CoVs S proteins.

In this study, we systematically compared the *O*-glycosylation of five recombinant β-CoVs S1 subunit proteins through LC-MS/MS analysis. Four recombinant S1 proteins of SARS-CoV, MERS-CoV, SARS-CoV-2, D614G SARS-CoV-2 are expressed by HEK 293 cells and one S1 proteins of SARS-CoV-2 are expressed by insect baculovirus cells, which are utilized to compare the difference of *O*-glycosylation from different expression systems. The *O*-Glycosylation landscape of these β-CoVs S1 proteins exhibited that the *O*-glycosylation are highly heterogeneous. We identified the *O*-glycosylation sites, compared the distribution of *O*-glycosylation sites in different functional domains, and exhibited different *O*-glycosylation abundances on visual models. Collectively, these results presented the detailed *O*-glycosylated structure information of β-CoVs S proteins and pave the way for antivirals development and vaccines optimization.

2. Materials and methods

2.1. Reagents and materials

Bovine fetuin, bovine serum albumin, trypsin and chymotrypsin were purchased from Sigma Aldrich (USA). PNGase F was purchased from New England Biolabs (Ipswich, MA) and PNGase F II was kindly donated by Prof. Chen Li (Fudan University). Recombinant SARS-CoV-2 S1 proteins (expressed from HEK293 cells, the purity > 90%), SARS-CoV-2 S1 proteins (expressed from baculovirus-insect cells, the purity > 92%), D614G SARS-CoV-2 S1 proteins (expressed from HEK293 cells, the purity > 94.6%), SARS-CoV S1 proteins (expressed from HEK293 cells, the purity > 90%) and MERS-CoV S1 proteins (expressed from HEK293 cells, the purity > 96.6%) were all purchased from Sino Biological and general information was attached in supplementary information. Chemical reagents of iodacetamide (IAA), 1, 4-dithiothreitol (DTT), acetonitrile (ACN, HPLC grade), ammonium bicarbonate, urea ammonium hydroxide, formic acid and trifluoroacetic acid (TFA) were obtained from Sigma. Deionized water was prepared with Milli-Q system (18.2 MΩ cm, Bedford, MA, USA).

2.2. Digestion of protein samples

Each standard protein (BSA or bovine fetuin) of 1 mg was dissolved with 100 μL of urea (6 M) in 50 mM ammonium bicarbonate (NH$_4$HCO$_3$). The protein solution was treated with 20 μL of 200 mM dithiothreitol (DTT) for 45 min at 56 °C. 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$_4$HCO$_3$. Trypsin and elastase were used for *N*-glycopeptides and *O*-glycopeptides digestion respectively at the enzyme to proteins mass ratio of 1: 40 at 37 °C overnight. After that, PNGase F and PNGase F II were used to remove *N*-glycans at 37 °C overnight for fetuin.

S1 subunit was dissolved with 100 μL of urea (6 M) in 50 mM NH$_4$HCO$_3$. The protein solution was treated with 20 μL of 200 mM DTT for 45 min at 56 °C. 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 at 37 °C overnight. After that, PNGase F and PNGase F II were used to remove *N*-glycans at 37 °C overnight for S1.

2.3. Enrichment of O-linked glycopeptides with the histidine-bonded silica (HBS) materials

In this work, we utilized the histidine-bonded silica (HBS) materials, which was previously developed to enrich *N*- and *O*-glycopeptides [29, 30]. Five milligrams of HBS material were suspended in acetonitrile (CH$_3$CN) and packed into GELoader tip. The GELoader tip was conditioned and equilibrated with 30 μL of 50% CH$_3$CN/0.1%TFA and 80% CH$_3$CN/0.1%TFA, respectively. The 50 μg peptides sample dissolved in 100 μL of 80% CH$_3$CN/0.1% TFA, incubated with HBS and loaded onto the GELoader tip. The bound peptides were rinsed three times with 30 μL 80% CH$_3$CN/0.1%TFA, and subsequently eluted with 40 μL of 10% ammonium hydroxide. The eluent was collected and dried for further LC-MS/MS analysis.

2.4. Mass spectrometry analysis

The enriched *O*-glycopeptides from each β-CoVs S1 proteins were analyzed by using Q Exactive coupled with Accela 600 HPLC System (Thermo, CA, USA). For the separation of peptides with reverse-phase liquid chromatography, 0.1% formic acid (FA, pH 2.59) aqueous solution and 0.1% FA in 80% CH$_3$CN were 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 600 nL/min. Gradient elution was performed with 2–8% B in 0.2 min, 8–50% B in 45 min, 50–90% B in 0.5 min, and 90% B in 5 min. Full mass scans were set the range from *m/z* 500 to 1500 (R = 70,000 at *m/z* 400). The AGC was set to 3E6 and ion injection time was set to 36 ms. The data-dependent acquisition (DDA) mode was employed and the 15 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

17.5%, 25%, 32.5%) at a resolution of 17,500.

2.5. MS data analysis

All the MS RAW data files obtained were searched against the β-CoVs sequences 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. 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. The *O*-linked glycan database was a human *O*-linked glycan database containing human 70 *O*-linked glycans in Byonic software.

3. Results and discussion

3.1. The strategy for comprehensive analysis of O-glycosylation on recombinant β-CoVs S1 proteins
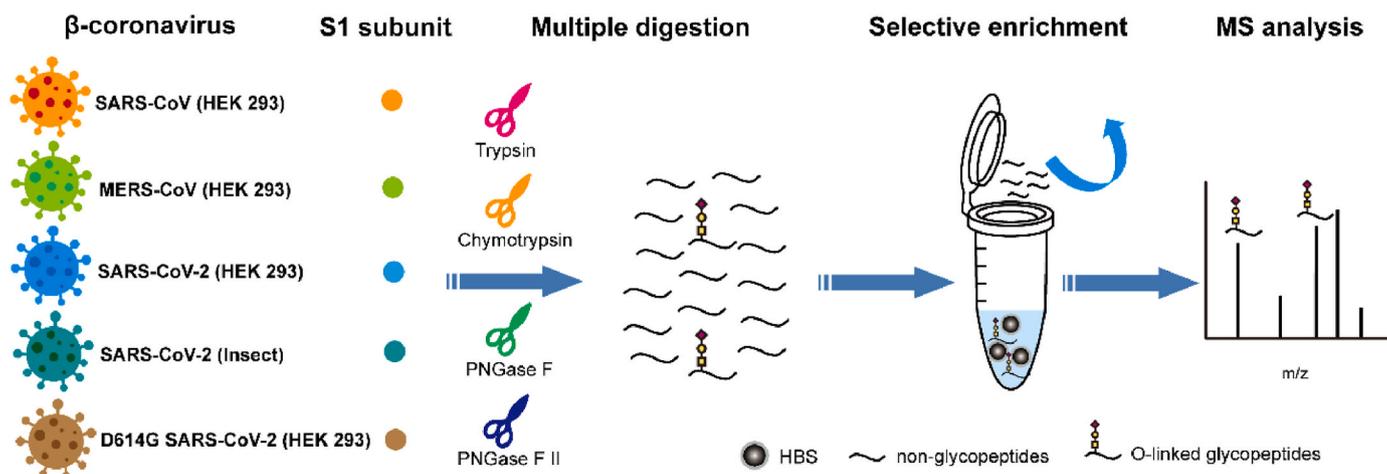
To systematically explore the *O*-glycosylation of different β-CoVs S1 proteins, we established a novel strategy for *O*-glycosylation identification (Scheme 1). We chose recombinant MERS-CoV, SARS-CoV, SARS-CoV-2 and D614G SARS-CoV-2 S1 proteins expressed by HEK 293 cells to compare the *O*-glycosylation differences amongst them. Furthermore, we used recombinant SARS-CoV-2 S1 proteins expressed by insect baculovirus cells (insect SARS-CoV-2 S1) to investigate the *O*-glycosylation differences across different expression systems. Given that *O*-glycosylation is often of low stoichiometry and low abundance, highly selective enrichment of *O*-glycopeptides is essential before comprehensive *O*-glycosylation profiling. In our previous work, we synthesized and used HBS materials to enrich *N*-glycopeptides [29]. In this work, we developed a novel strategy for *O*-glycopeptide enrichment based on HBS materials (Scheme 1). This HBS-based strategy exhibited high selectivity towards the *O*-glycopeptides from bovine fetuin even mixed with 100-M fold bovine serum albumin (Fig. S1 and Table S1). In addition, digestion with trypsin and chymotrypsin followed by removing the *N*-linked glycans and core-fucosylated *N*-linked glycans with PNGase F and PNGase II was adopted to generate peptide samples [30,31]. The enriched *O*-glycopeptides were further analyzed by high-resolution liquid chromatography tandem mass spectrometry (LC-MS/MS). The raw data were searched against the individual S proteins sequence using Byonic software. Each annotated spectrum of *O*-glycopeptides was manually validated.

3.2. The comprehensive characterization of site-specific O-glycosylation

Based on the abovementioned strategy, we demonstrated the site-specific *O*-glycosylation macroheterogeneity of β-CoVs S1 proteins with 36, 36, 22, 19 and 11 identified *O*-glycosylation sites on SARS-CoV-2, D614G SARS-CoV-2, insect SARS-CoV-2, SARS-CoV and MERS-CoV S1 proteins, respectively (Fig. 1A). Among these identified *O*-glycosylation sites, 16, 14, 6, 5, and 5 *O*-glycosylation sites were unambiguously identified for SARS-CoV-2, D614G SARS-CoV-2, insect SARS-CoV-2, SARS-CoV and MERS-CoV S1 proteins, respectively (Fig. 1B). The distributions of *O*-glycosylation sites on SARS-like β-CoVs (SARS-CoV, insect SARS-CoV-2, SARS-CoV-2 and D614G SARS-CoV-2) S1 proteins were much denser than MERS-CoV S1 proteins. Among the SARS-like β-CoVs S1 proteins, SARS-CoV-2 and D614G SARS-CoV-2 possessed notably more *O*-glycosylation sites than insect SARS-CoV-2 and SARS-CoV (Fig. 1A and B). Curiously, the number of identified *O*-glycosylation sites on insect SARS-CoV-2 S1 proteins was much lower than that expressed with HEK 293 cells, which could be attributed to the different expression systems. Except for the identified *O*-glycosylation sites from HEK 293 cells or insect cells in the pioneering studies [27,32,33], our results also provided additional *O*-glycosylation sites not reported. For example, S161/S162/T167 (1), S383/T385 (1), S530/T531, S596/T599/T602/T604/T605 (2) from HEK293-derived recombinant SARS-CoV-2 S1; T167, S371, S375/T376 (1), S383/T385 (1), S438/S443 (1), and T604/S605 (1) from insect-derived recombinant SARS-CoV-2 S1. Taken together, these β-CoVs S1 proteins, except MERS S1 proteins, almost exhibit extensive O-glycosylation sites distribution and lead to the macro-heterogeneity. In general, β-CoVs S proteins is often large and highly N-glycosylated. The widely distributed O-glycosylation of these β-CoVs S1 proteins might provide the possible compensation to the less than 25 N-glycosylation sites distributed on individual β-CoVs S proteins.

According to the database searching results, 28, 61, 16, 62 and 61 *O*-linked glycans were identified from MERS-CoV, SARS-CoV, insect SARS-CoV-2, D614G SARS-CoV-2 and SARS-CoV-2 S1 proteins, respectively. In comparison, these five β-CoVs S1 proteins were equipped with certain unique *O*-linked glycans, such as HexNAc(2)Hex(1)Fuc(2) for MERS-CoV S1 proteins and HexNAc(3)Hex(2)Fuc(3) for SARS-CoV S1 proteins (Table S9). *O*-glycosylation locates mainly on serine (Ser) or threonine (Thr) residue, especially in the areas riched with Ser/Thr residues [34]. We calculated the proportion of total Ser and Thr (the

Scheme 1. The workflow for comprehensive analysis of *O*-glycosylation from recombinant β-CoVs S1 proteins. β-CoVs S1 proteins were digested using trypsin, chymotrypsin, subsequently PNGase F and PNGase F II. The *O*-glycopeptides were then enriched from the peptide mixtures by HBS materials. The enriched glycopeptides were identified by LC-MS/MS.
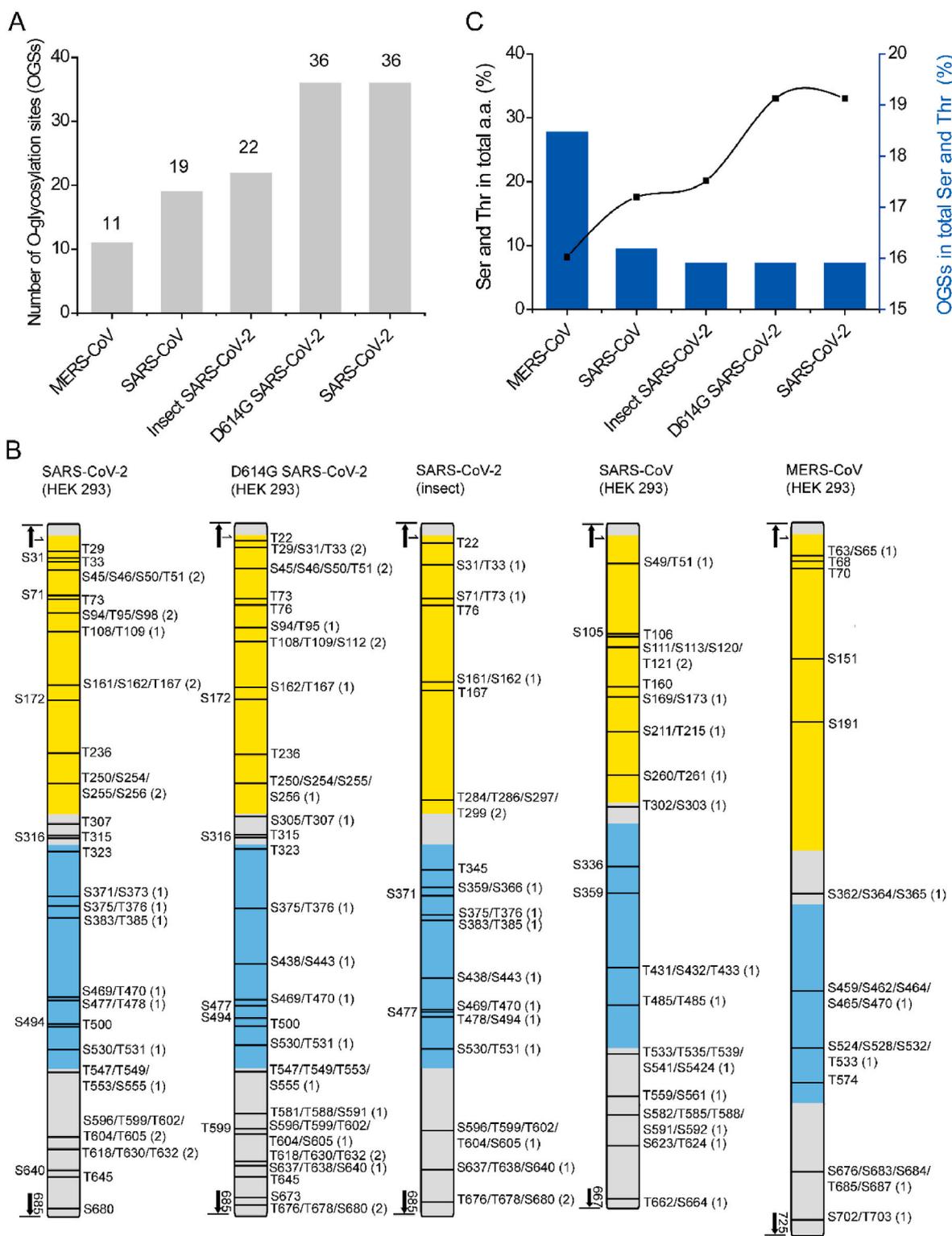
Fig. 1. The comprehensive characterization of *O*-glycosylation on five recombinant β-CoVs S1 proteins corresponding to SARS-CoV-2, D614G SARS-CoV-2, insect SARS-CoV-2, SARS-CoV and MERS-CoV. (A) The numbers of the totally identified *O*-glycosylation sites (OGSs) of five recombinant β-CoVs S1 proteins. (B) The comprehensive mapping of OGSs on five recombinant β-CoVs S1 proteins. NTD, N-terminal domain (labeled in yellow); RBD, receptor binding domain (labeled in blue). The unambiguously identified OGSs are labeled with serine (S) or threonine (T) and its possible amino acid position in S1 sequence. The ambiguous identified OGSs are labeled with potential S/T with the possible number of OGSs in bracket. (C) The occurrence ratio of OGSs and the ratio of S/T on five recombinant β-CoVs S1 proteins. The detailed information was in Tables S4–S8. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

number of total Ser and Thr divided by the number of amino acids in S1 proteins sequence) and the proportion of total *O*-glycosyated Ser and Thr (the number of total *O*-glycosyated Ser and Thr divided by the number of total Ser and Thr in S1 proteins sequence) in each S1 protein sequence. Intriguingly, MERS-CoV S1 proteins has the highest proportion of total Ser and Thr (18.5%), but the lowest proportion of *O*-glycosyated Ser and Thr (8.2%, Fig. 1C). In sharp contrast to this, SARS-CoV-2 and D614G SARS-CoV-2 has the highest proportion of *O*-glycosyated Ser and Thr (33.0% and 33.0%) in spite of their lowest proportion of total Ser and Thr (15.9%). The results implied that SARS-CoV-2 and D614G SARS-CoV-2 S1 proteins have higher occurrence ratios of *O*-glycosylation than that of other β-CoVs.

3.3. The distribution of O-glycosylation sites on different domains

We further investigated the distribution of *O*-glycosylation sites on three functional domains and one region. These domains and rigion encompass NTD, RBD, the receptor binding motif (RBM) and the cleavage site region (designated as the region within 100 residues from the cleavage site). Among the selected β-CoVs S1 proteins, NTD and RBD of β-CoVs were commonly responsible for neutralizing antibody and receptor binding, respectively [5,9]. RBM was the crucial functional motif contained in the RBD and mediated the contacts with receptors [22]. The cleavage site between S1 and S2 subunit played a role in membrane fusion activation [5]. Herein we calculated the *O*-glycosylation sites distribution through dividing the number of *O*-glycosylation sites by the number of total amino acid residues on each domain or

region. The results demonstrate that the *O*-glycosylation sites are not evenly distributed on each domain or region (Fig. 2). Four SARS-like CoVs S1 proteins have generally higher *O*-glycosylation sites distribution while MERS-CoV S1 proteins exhibits the sparser *O*-glycosylation sites distribution for these domains. More *O*-glycosylation sites are emerged on NTD and RBD of SARS-CoV-2 S1 proteins than D614G SARS-CoV-2. On the contrary, D614G SARS-CoV-2 S1 proteins possessed more *O*-glycosylation sites on RBM and cleavage site region than SARS-CoV-2.

In addition, we compared the distribution of the *O*-glycosylation sites with *N*- glycosylation sites on these domains and region. To our knowledge, all potential *N*-glycosylation sites on SARS-CoV, MERS-CoV, insect SARS-CoV-2, SARS-CoV-2 and D614G SARS-CoV-2 have been thoroughly identified [16,19,35]. Interestingly, the distribution of *O*-glycosylation sites (green bars) exceeds that of *N*-glycosylation (orange bars) on each domain or region of the five β-CoVs S1 proteins except on MERS-CoV NTD (Fig. 2A). The ratio of *O*-glycosylation on RBD of these five β-CoVs S1 proteins is 1.3–5 folds higher than that of N-glycosylation (Fig. 2B). Strikingly, the *N*-glycosylation rarely occurs on RBM on SARS-like β-CoVs S1 proteins while only MERS S1 proteins possess *N*-glycosylation on this domain [2]. Contrarily, the *O*-glycosylation is relatively abundant in this motif (Fig. 2C). The cleavage site of β-CoVs S plays a critical role in mediating the hydrolysis of S proteins and it was reported that the furin cleavage of SARS-CoV-2 S proteins was modulated by *O*-glycosylation recently [13,36]. Herein, we also compared the distribution of the *O*-glycosylation sites and *N*-glycosylation sites approximal to the cleavage site (within 100 amino acids).
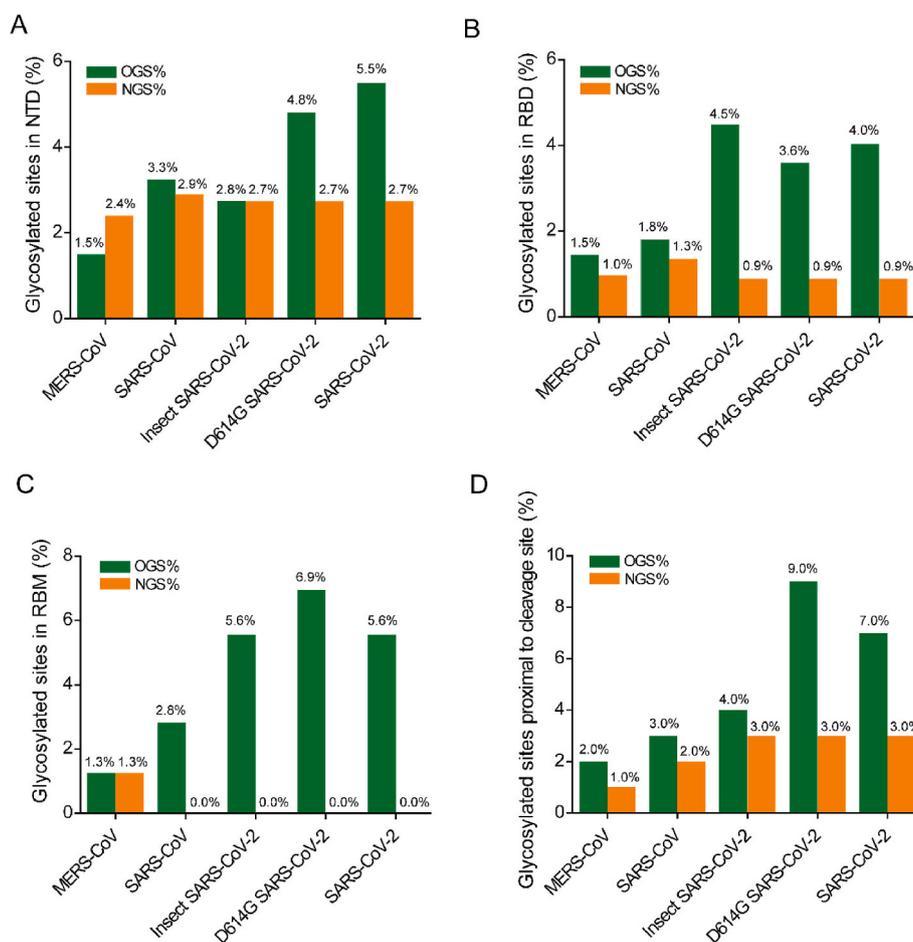

Fig. 2. The distribution of *O*-glycosylation sites (green bars) and *N*-glycosylation (orange bars) sites on (A) NTD, (B) RBD, (C) RBM, and (D) cleavage site region. The ratio is calculated by the number of *O*-/*N*-glycosylation sites (OGSs/NGSs) divided by the total amino acid number of each domain or region. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Overall, the number of *O*-glycosylation sites is more than that of *N*-glycosylation sites approximal to the cleavage site. Each domain or region on SARS-CoV-2 and D614G SARS-CoV-2 S1 proteins all present more *O*-glycosylation sites than the other three β-CoVs S1 proteins. These results demonstrate that *O*-glycosylation sites is mostly widely distributed in comparison with *N*-glycosylation sites on these β-CoVs S1 proteins key functional domains.

3.4. The sialylated O-glycosylation differences on five recombinant β-CoVs S1 proteins

The sialic acid plays a crucial role in virus pathogenesis and inter-species transmission [4,37]. The sialylation, as the symbol of mature glycosylation, is involved in the CoV infection, intercellular expansion and cell-cell spread [38]. Thus in-depth characterization of viral sialy-lation is of great needs and the sialylation abundances of each β-CoV S1 proteins are deserved to be investigated. We first compared the normalized abundances of these five S1 proteins and visualized their exposed normalized sialylated *O*-glycosylation with the PDB models. According to the identified *O*-linked glycans for each β-CoV S1 proteins in our results (Table S9), the proportion of the sialylated *O*-linked gly-cans in total *O*-linked glycans is from 18.7% to 44%.

To further explore the site-specific sialylated *O*-glycosylation on different β-CoVs S1 proteins, we calculated the total normalized *O*-glycosylation abundances of these five S1 proteins and the site-specific sialylated *O*-glycosylation on each *O*-glycosylation site, respectively. And their exposed normalized sialylated *O*-glycosylation of each site was indicated on the visual models. The normalized sialylated *O*-glycosyla-tion abundance was calculated by dividing the total sialylated *O*-glycosylation abundance of each *O*-glycosylation site by the total *O*-glycosylation abundance of each β-CoV S1 proteins. Considering these S1 proteins are sialylated to varying degrees, we defined the six orders of magnitudes to represent different normalized sialylated *O*-glycosylation abundances as denoted in the keys (Fig. 3). The red color represented the highest normalized sialylated *O*-glycosylation abundance, the limon color represented the least ones.

We also investigated the site-specific normalized *O*-glycosylation abundance on each *O*-glycosylation site of each β-CoV S1 proteins and the results were shown in Figure S3~S7. As presented in Fig. 3, SARS-CoV-2 and D614G SARS-CoV-2 S1 proteins (Fig. 3A and B) possess the most exposed sialylated *O*-glycosylation while the insect SARS-CoV-2 S1 proteins possesses the least one (Fig. 3C), which can be ascribed to the different cellular expression system utilized [27]. Compared with SARS-CoV-2, D614G SARS-CoV-2 S1 proteins displays variable sialy-lated *O*-glycosylation. Remarkably, the sialylated *O*-glycosylation of S640 and T307 on D614G SARS-CoV-2 S1 proteins disappeared but the sialylated *O*-glycosylation of S316 on D614G SARS-CoV-2 S1 proteins enhanced. Even with high sequence identity with SARS-CoV-2, SAR-S-CoV S1 displays relatively less exposed sialylated *O*-glycosylation than SARS-CoV-2 and D614G SARS-CoV-2 (Fig. 3D). SARS-CoV S1 proteins have four exposed sialylated *O*-glycosylation sites on RBD but little on NTD. The site-specific sialylated *O*-glycosylation of MERS-CoV S1 pro-teins is still weaker and sparser in spite of expression from HEK 293 (Fig. 3E). Together, the exposed site-specific sialylated *O*-glycosylation on each β-CoV S1 proteins is distinct from each other. We then studied the correlation between the total sialylated *O*-glycosylation abundance of each β-CoV S1 protein and the *O*-glyocsylation sites ratios (Fig. 3F). A positive correlation was observed, and SARS-CoV-2 and D614G SARS-CoV-2 both display high sialylated *O*-glycosylation and *O*-glyco-sylation sites distribution. The identified *O*-linked glycans in our study were also summarized in Table S9 and the top five sialylated *O*-glyco-sylation abundances ratios on β-CoVs S1 proteins were listed in Fig. 3G. There are only three types of sialylated *O*-linked glycans on insect SARS-CoV-2 S1 proteins that can be identified in our study. Although sialylation plays a vital role in viral pathogenesis, the specific function of sialylated *O*-glycosylation on β-CoV S1 proteins needs further biological

verification.

3.5. The comparison of sialylated O- and N-glycosylation abundance onrecombinant β-CoVs S1 proteins

As *O*- and *N*-glycosylation are both primary glycosylation types, it is worthy to make the comparison between sialylated *O*- and *N*-glycosyl-ation. Though the sialylated *N*-glycosylation of some β-CoVs S proteins, like SARS-CoV-2, SARS-CoV and MERS-CoV S proteins, have been re-ported [16,19], the sialylated *O*-glycosylation of β-CoVs S proteins haven't been widely paid attention.

As we have accomplished *O*-glycosylation abundance identification of the above β-CoVs S1 proteins, we further used the visualized model to make a contrast with sialylated *O*- and *N*-glycosylation abundance be-tween MERS-CoV, SARS-CoV, and SARS-CoV-2 S1 proteins simulta-neously. Herein, the normalized sialylated *O*-glycosylation abundance was calculated through dividing the sialylation abundance by total glycosylation abundance on each site. The data of sialylated *N*-glyco-sylation abundance was derived from the report from Crispin's labora-tory [16,19]. And the sialylated glycosylation abundances on *O*- and *N*-glycosylation sites were highlighted according to the normalized sialylation proportions as denoted in the keys (Fig. 4).

According to the visualized comparison, we found that SARS-CoV-2 and SARS-CoV S1 proteins displayed more remarkable increased sialy-lated *O*-glycosylation abundance than *N*-glycosylation. MERS-CoV S1 proteins had more widely distributed sialylated glycosylation on *N*-glycosylation sites than *O*-glycosylation sites. The result demonstrates the three β-CoVs S1 proteins provided unique sialylated N- and O-glycosylation. Specific biological effects of different sialylated N- and O-glycosylation require more deeper biological validation in the future.
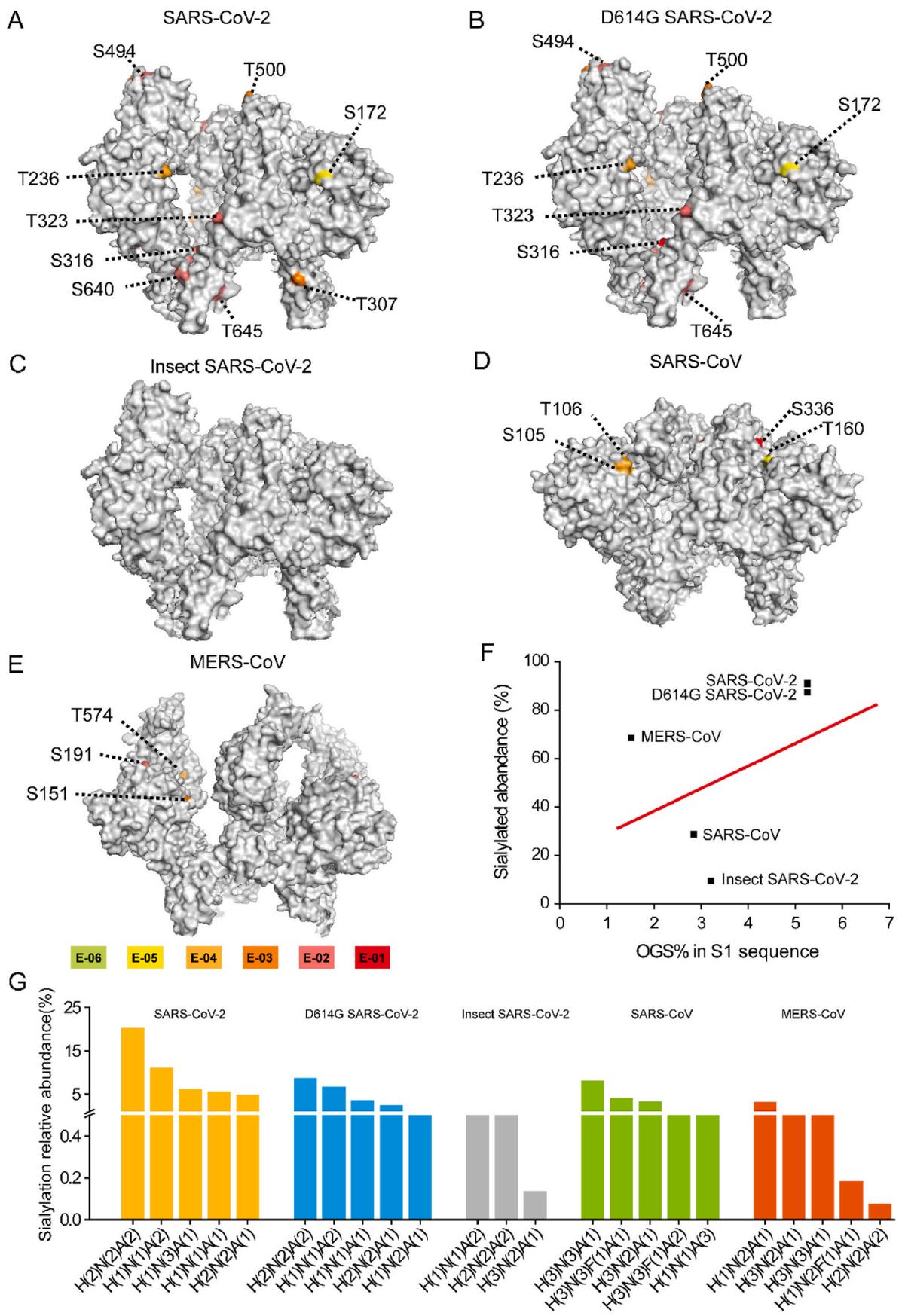
3.6. The differentiation of O-glycosylation between SARS-CoV-2 and D614G SARS-CoV-2 S1 proteins

SARS-CoV-2 and its variant D614G SARS-CoV-2 is the most infec-tious pathogen of these β-CoVs [38,39]. Only one amino acid mutation on S triggers increased infection as well as the changed in S proteins conformation for D614G SARS-CoV-2 [39]. Here we explored the dif-ferentiation of the sialylated *O*-glycosylation between SARS-CoV-2 and D614G SARS-CoV-2 S1 proteins, especially on NTD, RBD and the cleavage site region in an attempt to present some valuable structural information.

Nine unambiguous *O*-glycosylation sites are common between SARS-CoV-2 and D614G SARS-CoV-2 S1 proteins with 6 and 3 unique ones, respectively (Fig. 5A). In order to compare more deeply, three individual parts with remarkable sialylated *O*-glycosylation distinction from NTD, RBD and cleavage site region were shown in Figure 5B, 5C and 5D. The variations sialylated *O*-glycosylation appeared on these *O*-glycosylation sites. T22, T323, S371/S373, T500, T581/T588/S591, S596/T599/T602/T604/S605, T618/T630/T632, T618/T630/T632, S637/T638/S640 and S673 on D614G SARS-CoV-2 S1 proteins exhibited increasing *O*-glycosylation abundances while decreasing on other sites. Despite only one amino acid mutation, obvious differentiation on sialylated *O*-glycosylation took place. The practical biological effect on virology needs further validation and exploration.

3.7. The concomitance of O-glycosylation and phosphorylation on β-CoVs S1 proteins

Viral phosphorylation modulates viral replication, infectivity and pat hogenesis [40,41] and occurs on Ser and Thr in the same manner of protein *O*-glycosylation [42]. The proteins of SARS-CoV-2 have been reported phosphorylated extensively in host cell [43]. However, the concomitance of *O*-glycosylation and phosphorylation has not yet been reported for β-CoVs. Here we listed the identified concomitant sites with *O*-glycosylation and phosphorylation on β-CoVs S1 proteins according to

(caption on next page)

Fig. 3. The sialylated *O*-glycosylation differentiations of five recombinant β-CoVs S1 proteins. and the correlation between normalized sialylated *O*-glycosylation abundance and *O*-glycosylation sites (OGSs) ratio in β-CoVs S1 proteins sequences. (**A-E**) The normalized sialylated *O*-glycosylation abundance on unambigous OGSs mapped on (**A**) SARS-CoV-2 S1 proteins, (**B**) D614G SARS-CoV-2 S1 proteins, (**C**) insect SARS-CoV-2 S1 proteins, (**D**) SARS-CoV S1 proteins, (**E**) MERS-CoV S1 proteins with model structures. PDB code: 6vsb (SARS-CoV-2, insect SARS-CoV-2 and D614G SARS-CoV-2), 5x59 (MERS), 5x58 (SARS). Six orders of magnitudes were selected to label the normalized sialoglycan abundance with limon (E−06), yellow (E−05), bright orange (E−04), orange (E−03), deep salmon (E−02) and red (E−01), respectively. (**F**) The correlation between normalized sialoglycan abundance and *O*-glycosylation sites proportion on S1 proteins. (**G**) The ratios of top five sialylated *O*-glycosylation abundances on β-CoVs S1 proteins. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)
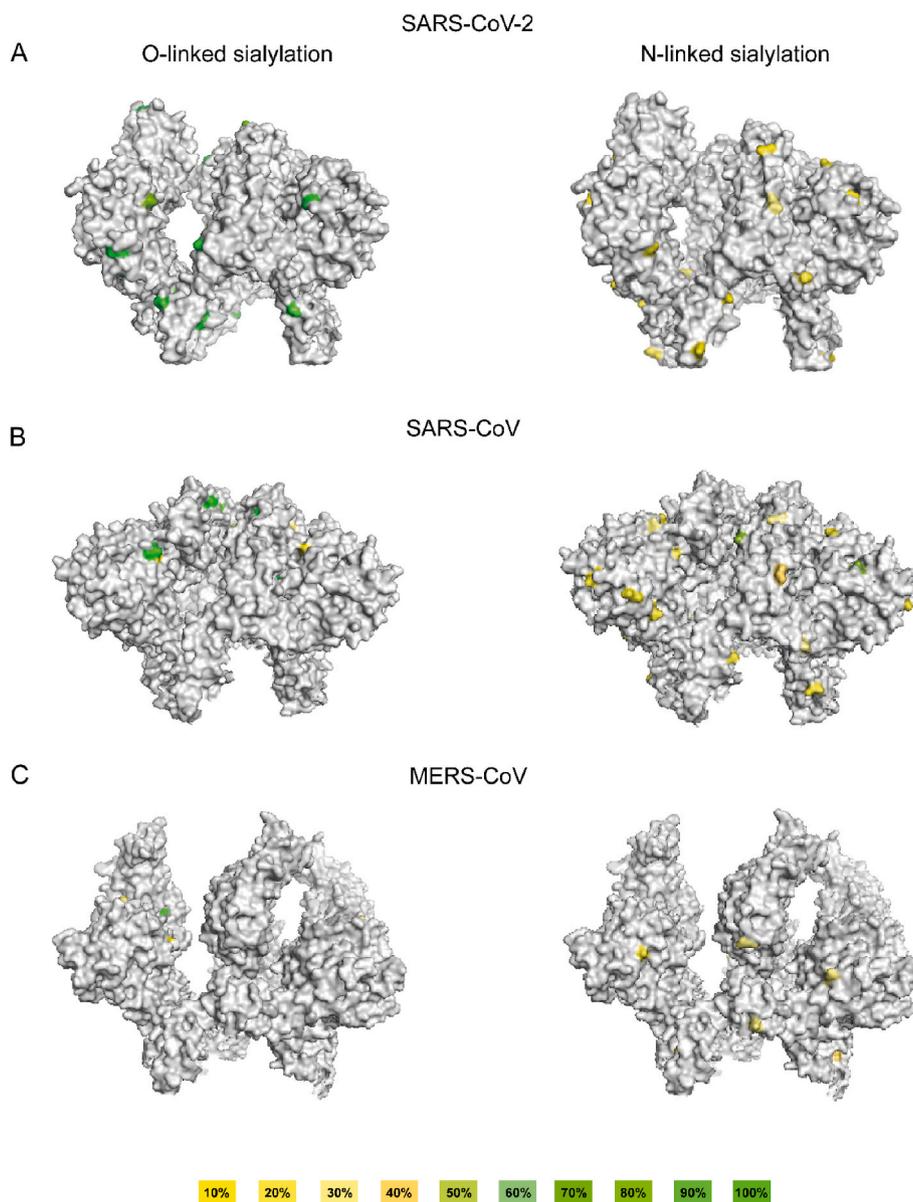

Fig. 4. The comparison of sialylated *O*- and *N*-glycosylation abundance on recombinant β-CoVs S1 proteins. Ten color gradations were set to label the normalized sialylated *O*- and *N*-glycosylation abundance. (A). SARS-CoV-2 (B). SARS-CoV (C). MERS-CoV. PDB code: 6vsb (SARS-CoV-2), 5x59 (MERS-CoV), 5x58 (SARS-CoV). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

our results (Table 1). SARS-CoV-2 and D614G SARS-CoV-2 S1 proteins has 4 and 2 concomitant sites respectively. SARS-CoV and insect SARS-CoV-2 has only one concomitant site while none concomitant site exists on MERS-CoV S1 proteins. These finding may help to investigate the *O*-glycosylation and phosphorylation crosstalk in the subsequent biological research.

4. Conclusion

In summary, β-CoVs S1 proteins are highly *O*-glycosylated in site-specific macro- and micro-heterogeneity manner based on the results in this study. In spite of emphasizing the significance of *N*-glycosylation of β-CoVs S proteins in recent studies, some vulnerabilities of β-CoVs S proteins still exist. Whether the *O*-glycosylation can complement the sparse *N*-glycosylation of β-CoVs deserves to be studied in the future. However, considerable efforts are still in urgent needs to fully
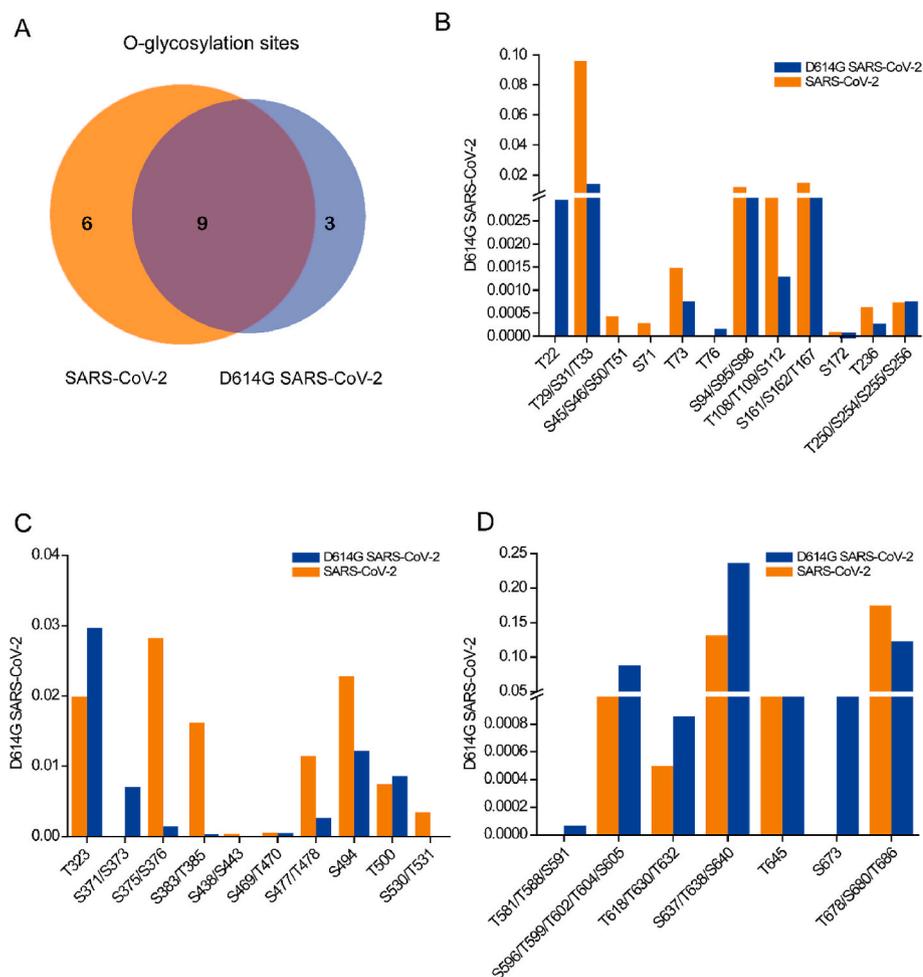
Fig. 5. The site-specific normalized sialylated *O*-glycosylation abundances on SARS-CoV-2 and D614G SARS-CoV-2 S1 proteins. (A) The overlap of unambiguous *O*-glycosylation sites on SARS-CoV-2 S1 proteins and D614G SARS-CoV-2 S1 proteins. The specific *O*-glycosylation sites on SARS-CoV-2 and D614G SARS-CoV-2 S1 proteins are listed in Table S2. The differences of normalized sialylated *O*-glycosylation abundances on individual parts of NDT (B), RBD (C) and approximal to cleavage site region (D).

Table 1

The list of *O*-glycosylation and phosphorylation concomitant sites on β-CoVs S1 proteins.

	SARS-CoV-2	D614G SARS-CoV-2	insect SARS-CoV-2	SARS-CoV	MERS-CoV
1	S31	S316	T76	T359	–
2	S316	T323			
3	S640				
4	S680				

understand the biological role of *O*-glycosylation on β-CoVs S proteins. Finally, these detailed *O*-glycosylation analysis of β-CoVs S1 proteins will guide developing the related subunit vaccines and help to combat β-CoVs.

CRediT authorship contribution statement

Xuefang Dong: finished the enrichment work, and, Writing – original draft. **Xiuling Li:** Funding acquisition, and, contributed to the experimental design, and, Writing – review & editing. **Cheng Chen:** pretreated the proteins samples, and, optimized the enrichment strategy, and. **Xiaofei Zhang:** pretreated the proteins samples, and, optimized the enrichment strategy, All authors contributed in data collection and interpretation of the results. **Xinmiao Liang:** Funding acquisition, and, contributed to the experimental design, and, Writing – review & editing, and.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.aca.2022.340394.

References

[1] F. Zhou, G. Fan, Z. Liu, B. Cao, SARS-CoV-2 shedding and infectivity Reply, Lancet 395 (2020), 1340-1340.

[2] F. Li, Structure, function, and evolution of coronavirus spike proteins, Annu. Rev. Virol. 3 (2016) 237–261.

[3] A.J. Thompson, R.P. de Vries, J.C. Paulson, Virus recognition of glycan receptors, Curr. Opin. Virol. 34 (2019) 117–129.

[4] E. Qing, M. Hantak, S. Perlman, T. Gallagher, Distinct roles for sialoside and protein receptors in coronavirus infection, mBio 11 (2020) e02764-19.
[5] Y. Yuan, D. Cao, Y. Zhang, J. Ma, J. Qi, Q. Wang, G. Lu, Y. Wu, J. Yan, Y. Shi, X. Zhang, G.F. Gao, Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains, Nat. Commun. 8 (2017), 15092.
[6] A. Shajahan, L.E. Pepi, D.S. Rouhani, C. Heiss, P. Azadi, Glycosylation of SARS-CoV-2: structural and functional insights, Anal. Bioanal. Chem. 413 (2021) 7179–7193.
[7] R.N. Kirchdoerfer, C.A. Cottrell, N. Wang, J. Pallesen, H.M. Yassine, H.L. Turner, K. S. Corbett, B.S. Graham, J.S. McLellan, A.B. Ward, Pre-fusion structure of a human coronavirus spike protein, Nature 531 (2016) 118–121.
[8] T.-J. Yang, Y.-C. Chang, T.-P. Ko, P. Draczkowski, Y.-C. Chien, Y.-C. Chang, K.-P. Wu, K.-H. Khoo, H.-W. Chang, S.-T.D. Hsu, Cryo-EM analysis of a feline coronavirus spike protein reveals a unique structure and camouflaging glycans, Proc. Natl. Acad. Sci. U.S.A. 117 (2020) 1438–1446.
[9] M. Yuan, N.C. Wu, X. Zhu, C.-C.D. Lee, R.T.Y. So, H. Lv, C.K.P. Mok, I.A. Wilson, A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV, Science 368 (2020) 630–633.
[10] F. Kunkel, G. Herrler, Structural and functional-analysis of the surface protein of human coronavirus OC43, Virology 195 (1993) 195–202.
[11] G. Peng, L. Xu, Y.-L. Lin, L. Chen, J.R. Pasquarella, K.V. Holmes, F. Li, Crystal structure of bovine coronavirus spike protein lectin domain, J. Biol. Chem. 287 (2012) 41931–41938.
[12] C.A. Reis, C. Tauber, V. Blanchard, Glycosylation is a key in SARS-CoV-2 infection, J. Mol. Med. (Berl) 99 (2021) 1023–1031.
[13] X. Liu, Q. Wu, Z. Zhang, Global diversification and distribution of coronaviruses with furin cleavage sites, Front. Microbiol. 12 (2021), 649314.
[14] B. Ju, Q. Zhang, J. Ge, R. Wang, J. Sun, X. Ge, J. Yu, S. Shan, B. Zhou, S. Song, X. Tang, J. Yu, J. Lan, J. Yuan, H. Wang, J. Zhao, S. Zhang, Y. Wang, X. Shi, L. Liu, J. Zhao, X. Wang, Z. Zhang, L. Zhang, Human neutralizing antibodies elicited by SARS-CoV-2 infection, Nature (2020) 115–119.
[15] G.O. Schaefer, C.C. Tam, J. Savulescu, T.C. Voo, COVID-19 vaccine development: time to consider SARS-CoV-2 challenge studies? Comment, Vaccine 38 (2020) 5085–5088.
[16] Y. Watanabe, Z.T. Berndsen, J. Raghwani, G.E. Seabright, J.D. Allen, O.G. Pybus, J. S. McLellan, I.A. Wilson, T.A. Bowden, A.B. Ward, M. Crispin, Vulnerabilities in coronavirus glycan shields despite extensive glycosylation, Nat. Commun. 11 (2020) 2688.
[17] Y. Watanabe, J. Raghwani, J.D. Allen, G.E. Seabright, S. Li, F. Moser, J. T. Huiskonen, T. Strecker, T.A. Bowden, M. Crispin, Structure of the Lassa virus glycan shield provides a model for immunological resistance, Proc. Natl. Acad. Sci. U.S.A. 115 (2018) 7320–7325.
[18] Y. Watanabe, T.A. Bowden, I.A. Wilson, M. Crispin, Exploitation of glycosylation in enveloped virus pathobiology, Biochim. Biophys. Acta Gen. Subj. 1863 (2019) 1480–1497.
[19] Y. Watanabe, J.D. Allen, D. Wrapp, J.S. McLellan, M. Crispin, Site-specific glycan analysis of the SARS-CoV-2 spike, Science 369 (2020) 330–333.
[20] A.A. Hargett, M.B. Renfrow, Glycosylation of viral surface proteins probed by mass spectrometry, Curr. Opin. Virol. 36 (2019) 56–66.
[21] A.C. Walls, M.A. Tortorici, B. Frenz, J. Snijder, W. Li, F.A. Rey, F. DiMaio, B.-J. Bosch, D. Veesler, Glycan shield and epitope masking of a coronavirus spike protein observed by cryo-electron microscopy, Nat. Struct. Mol. Biol. 23 (2016) 899–905.
[22] J. Shang, G. Ye, K. Shi, Y. Wan, C. Luo, H. Aihara, Q. Geng, A. Auerbach, F. Li, Structural basis of receptor recognition by SARS-CoV-2, Nature 581 (2020) 221–224.
[23] D. Wrapp, N. Wang, K.S. Corbett, J.A. Goldsmith, C.-L. Hsieh, O. Abiona, B. S. Graham, J.S. McLellan, Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation, Science 367 (2020) 1260–1263.
[24] Z.A. Silver, A. Antonopoulos, S.M. Haslam, A. Dell, G.M. Dickinson, M.S. Seaman, R.C. Desrosiers, Discovery of O-linked carbohydrate on HIV-1 envelope and its role in shielding against one category of broadly neutralizing antibodies, Cell Rep. 30 (2020) 1862–1869, e1864.
[25] I. Bagdonaite, R. Norden, H.J. Joshi, S. Dabelsteen, K. Nystrom, S.Y. Vakhrushev, S. Olofsson, H.H. Wandall, A strategy for O-glycoproteomics of enveloped viruses-the O-glycoproteome of herpes simplex virus type 1, PLoS Pathog. 11 (2015), e1004784.
[26] S. Olofsson, O. Blixt, T. Bergstrom, M. Frank, H.H. Wandall, Viral O-GalNAc peptide epitopes: a novel potential target in viral envelope glycoproteins, Rev. Med. Virol. 26 (2016) 34–48.
[27] I. Bagdonaite, A.J. Thompson, X. Wang, M. Søgaard, C. Fougeroux, M. Frank, J. K. Diedrich, J.R. Yates, A. Salanti, S.Y. Vakhrushev, J.C. Paulson, H.H. Wandall, Site-specific O-glycosylation analysis of SARS-CoV-2 spike protein produced in insect and human cells, Viruses 13 (2021) 551.
[28] P. Zhao, J.L. Praissman, O.C. Grant, Y. Cai, T. Xiao, K.E. Rosenbalm, K. Aoki, B. P. Kellman, R. Bridger, D.H. Barouch, M.A. Brindley, N.E. Lewis, M. Tiemeyer, B. Chen, R.J. Woods, L. Wells, Virus-receptor interactions of glycosylated SARS-CoV-2 spike and human ACE2 receptor, Cell Host Microbe 28 (2020) 586–601.
[29] X. Dong, H. Qin, J. Mao, D. Yu, X. Li, A. Shen, J. Yan, L. Yu, Z. Guo, M. Ye, H. Zou, X. Liang, In-depth analysis of glycoprotein sialylation in serum using a dual-functional material with superior hydrophilicity and switchable surface charge, Anal. Chem. 89 (2017) 3966–3972.
[30] Y. Cui, X. Dong, X. Zhang, C. Chen, D. Fu, X. Li, X. Liang, Deciphering the O-glycosylation of HKU1 spike protein with the dual-functional hydrophilic interaction chromatography materials, Front. Chem. 9 (2021), 707235.
[31] X. Dong, C. Chen, J. Yan, X. Zhang, X. Li, X. Liang, Comprehensive O-glycosylation analysis of the SARS-CoV-2 spike protein with biomimetic trp-arg materials, Anal. Chem. 93 (2021) 10444–10452.
[32] Y. Wang, Z. Wu, W. Hu, P. Hao, S. Yang, Impact of expressing cells on glycosylation and glycan of the SARS-CoV-2 spike glycoprotein, ACS Omega 6 (2021) 15988–15999.
[33] Y. Zhang, W. Zhao, Y. Mao, Y. Chen, S. Zheng, W. Cao, J. Zhu, L. Hu, M. Gong, J. Cheng, H. Yang, O-glycosylation landscapes of SARS-CoV-2 spike proteins, Front. Chem. 9 (2021), 689521.
[34] W. Yang, A. Song, M. Ao, Y. Xu, H. Zhang, Large-scale site-specific mapping of the O-GalNAc glycoproteome, Nat. Protoc. 15 (2020) 2589–2610.
[35] D. Wang, B. Zhou, T.R. Keppel, M. Solano, J. Baudys, J. Goldstein, M.G. Finn, X. Fan, A.P. Chapman, J.L. Bundy, A.R. Woolfitt, S.H. Osman, J.L. Pirkle, D. E. Wentworth, J.R. Barr, N-glycosylation profiles of the SARS-CoV-2 spike D614G mutant and its ancestral protein characterized by advanced mass spectrometry, Sci. Rep. 11 (2021), 23561.
[36] L. Zhang, M. Mann, Z.A. Syed, H.M. Reynolds, E. Tian, N.L. Samara, D.C. Zeldin, L. A. Tabak, K.G. Ten Hagen, Furin cleavage of the SARS-CoV-2 spike is modulated by O-glycosylation, Proc. Natl. Acad. Sci. U.S.A. 118 (2021), e2109905118.
[37] J.E. Stencel-Baerenwald, K. Reiss, D.M. Reiter, T. Stehle, T.S. Dermody, The sweet spot: defining virus-sialic acid interactions, Nat. Rev. Microbiol. 12 (2014) 739–749.
[38] B. Korber, W.M. Fischer, S. Gnanakaran, H. Yoon, J. Theiler, W. Abfalterer, N. Hengartner, E.E. Giorgi, T. Bhattacharya, B. Foley, K.M. Hastie, M.D. Parker, D. G. Partridge, C.M. Evans, T.M. Freeman, T.I. de Silva, C. McDanal, L.G. Perez, H. Tang, A. Moon-Walker, S.P. Whelan, C.C. LaBranche, E.O. Saphire, D. C. Montefiori, C.-G.G. Sheffield, Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus, Cell 182 (2020) 812–827.
[39] Q. Li, J. Wu, J. Nie, L. Zhang, H. Hao, S. Liu, C. Zhao, Q. Zhang, H. Liu, L. Nie, H. Qin, M. Wang, Q. Lu, X. Li, Q. Sun, J. Liu, L. Zhang, X. Li, W. Huang, Y. Wang, The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity, Cell 182 (2020) 1284–1294.
[40] H. Yan, J. Shobahah, M. Wei, E. Obeng, S. Xue, D. Hu, Y. Quan, W. Yu, Phosphorylation of nucleopolyhedrovirus 39K is essential for the regulation of viral gene transcription in silkworm cells, Acta Virol. 63 (2019) 469–474.
[41] H.S. Hoover, J.C.-Y. Wang, S. Middleton, P. Ni, A. Zlotnick, R.C. Vaughan, C. C. Kao, Phosphorylation of the brome mosaic virus capsid regulates the timing of viral infection, J. Virol. 90 (2016) 7748–7760.
[42] A.C. Leney, D. El Atmioui, W. Wu, H. Ovaa, A.J.R. Heck, Elucidating crosstalk mechanisms between phosphorylation and O-GlcNAcylation, Proc. Natl. Acad. Sci. U.S.A. 114 (2017) E7255–E7261.
[43] K. Klann, D. Bojkova, G. Tascher, S. Ciesek, C. Munch, J. Cinatl, Growth factor receptor signaling inhibition prevents SARS-CoV-2 replication, Mol. cell 80 (2020) 164–174.