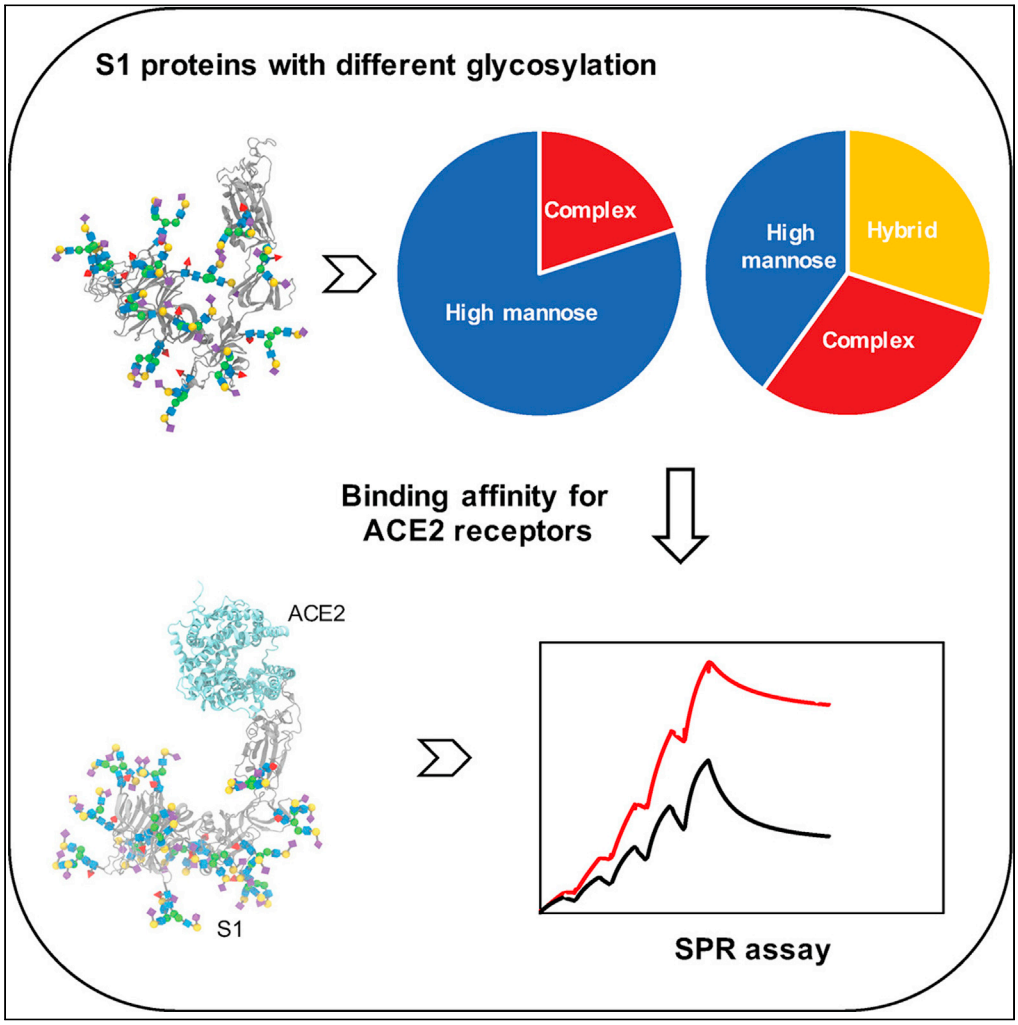
iScience

CellPress
OPEN ACCESS

Article

The effect of *N*-glycosylation of SARS-CoV-2 spike protein on the virus interaction with the host cell ACE2 receptor



Chuncui Huang, Zeshun Tan, Keli Zhao, ..., Dongbo Bu, Wengang Chai, Yan Li

w.chai@imperial.ac.uk (W.C.)
yanli@ibp.ac.cn (Y.L.)

Highlights

SARS-CoV-2 interaction with ACE2 receptor is affected by glycosylation of S protein

De-glycosylated S1 receptor binding subunit showed stronger binding activity to ACE2

N-glycosylation profiles are different for S1 produced in different expression systems

N-glycans have steric hindrance and Coulombic repulsion effects to S protein binding

Huang et al., iScience *24*, 103272
November 19, 2021
https://doi.org/10.1016/j.isci.2021.103272

iScience

CellPress
OPEN ACCESS

Article

The effect of *N*-glycosylation of SARS-CoV-2 spike protein on the virus interaction with the host cell ACE2 receptor

Chuncui Huang,[1,5] Zeshun Tan,[1,3] Keli Zhao,[1,3] Wenjun Zou,[1,3] Hui Wang,[2,3] Huanyu Gao,[1] Shiwei Sun,[2,3] Dongbo Bu,[2,3] Wengang Chai,[4,6,*] and Yan Li[1,3,5,*]

SUMMARY

The densely glycosylated spike (S) protein highly exposed on severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) surface mediates host cell entry by binding to the receptor angiotensin-converting enzyme 2 (ACE2). However, the role of glycosylation has not been fully understood. In this study, we investigated the effect of different *N*-glycosylation of S1 protein on its binding to ACE2. Using real-time surface plasmon resonance assay the negative effects were demonstrated by the considerable increase of binding affinities of de-*N*-glycosylated S1 proteins produced from three different expression systems including baculovirus-insect, Chinese hamster ovarian and two variants of human embryonic kidney 293 cells. Molecular dynamic simulations of the S1 protein-ACE2 receptor complex revealed the steric hindrance and Coulombic repulsion effects of different types of *N*-glycans on the S1 protein interaction with ACE2. The results should contribute to future pathological studies of SARS-CoV-2 and therapeutic development of Covid-19, particularly using recombinant S1 proteins as models.

INTRODUCTION

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) uses the heavily glycosylated transmembrane spike (S) protein to gain entry into the host cell by its interaction with the angiotensin-converting enzyme 2 (ACE2) receptor on the host cell to promote the fusion of viral and host cellular membranes (Wrapp et al., 2020; Letko et al., 2020). S protein is a homo-trimeric class I fusion protein composed of two functional subunits, S1 responsible for receptor binding and S2 for membrane fusion. S protein is highly *N*-glycosylated on the virus surface which possesses a total of 22 potential *N*-glycosylation sites (Asn–X–Ser/Thr, X≠Pro) in each monomer (Wang et al., 2020), and 13 *N*-glycosylation sites (N17-N657) were located on the S1 subunit.

N-glycosylation of SARS-CoV-2 S proteins have been investigated extensively (Zhang et al., 2021; Shajahan et al., 2020; Watanabe et al., 2020a), but these were exclusively carried out with recombinant S proteins produced in different expression systems because of the difficulty in wild type viral proteins of sufficient amount until the recent publication on the molecular architecture of the authentic SARS-CoV-2 virus and mass spectrometry (MS) analysis of compositions of the *N*-glycans on the native S protein (Yao et al., 2020). Site specific *N*-glycosylation characterization manifested different glycosylation features for the proteins obtained from different expression systems, e.g. complex type *N*-glycans were found to be the predominant glycosylation form in S1 protein produced by human embryonic kidney (HEK) 293 cells (Zhang et al., 2021) and high-mannose *N*-glycans were identified as the main type in insect-cell expressed S1 (Zhang et al., 2021). However, using the same expression system, e.g., HEK 293, different *N*-glycosylation profiles have been reported. In one of these reports both high-mannose and complex-type glycans across the *N*-glycosylation sites on S1 were observed as the prevalent *N*-glycans (Shajahan et al., 2020), whereas others showed all three types of *N*-glycans (high-mannose, hybrid and complex type) on S1 proteins expressed in HEK 293 cells (Watanabe et al., 2020b). Detailed knowledge of glycosylation of S protein, particularly its S1 subunit, and its role in viral binding with receptors are important not only in understanding the pathology but also in the design of suitable antigens for vaccine development and therapeutic target.

[1]Institute of Biophysics, Chinese Academy of Sciences, 15 Datun Road, Beijing 100101, China

[2]Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, 6 Kexueyuan South Road, Beijing 100080, China

[3]University of Chinese Academy of Sciences, 19 Yuquan Road, Beijing 100049, China

[4]Glycosciences Laboratory, Faculty of Medicine, Imperial College London, London W12 0NN, UK

[5]Bioland Laboratory (Guangzhou Regenerative Medicine and Health Guangdong Laboratory), Guangzhou 510005, China

[6]Lead contact

*Correspondence: w.chai@imperial.ac.uk (W.C.), yanli@ibp.ac.cn (Y.L.)

https://doi.org/10.1016/j.isci.2021.103272

N-glycans on viral S proteins have been considered to be an underlying mechanism for coronavirus to evade both the innate and adaptive immune response due to shielding the amino acid residues of its polypeptide backbone from cell and antibody recognition (Casalino et al., 2020). *N*-glycosylation is involved in the invasive process of virus into host cells, and virus relies on the biosynthesis pathway of *N*-glycosylation in host cells to complete its replication (Watanabe et al., 2020a; Casalino et al., 2020; Leemans et al., 2019; Raman et al., 2016). It has been considered that *N*-glycosylation can play a pivotal role in the recognition process between viruses and receptors (Han et al., 2007; Jeffers et al., 2004; Parsons et al., 2019). For virus SARS-CoV, which caused the SARS outbreak in early 2000s, *N*-glycans on S protein are important for DC/L-SIGN-mediated virus infection, and seven glycosylation sites on the S protein have been identified to be critical for DC/L-SIGN-mediated virus entry (Han et al., 2007; Jeffers et al., 2004). However, for S protein of SARS-CoV-2 it has been demonstrated experimentally that the glycosylation status does not contribute to the binding activity between S protein and the ACE2 receptor (Sun et al., 2020) whereas by full-length modeling, molecular dynamics (MD) simulation of the glycosylated spike protein has indicated an important role of *N*-glycosylation, particularly at N165 and N234 of the S1 subunit, in its binding to ACE2, and this was corroborated experimentally by biolayer interferometry using S protein with deleted *N*-glycans through N165A and N234A mutations (Casalino et al., 2020).

Combing glycomic analysis, MD simulation of SARS-CoV-2 S protein embedded in the viral membrane revealed that a glycan shield was formed and utilized to thwart the host immune response in the context of vaccine design (Casalino et al., 2020). Beyond shielding effect to immune response, a possible structural role of *N*-glycans at specific glycosylation sites in S1 subunit was also proposed to modulate and stabilize the conformational dynamics of the receptor binding domain of the spike protein (Casalino et al., 2020). MD simulations of S protein interacting with ACE2 revealed that *N*-glycans at the residues N74 and N165 located in the S1 subunit are important for the interaction, and clearly highlighted the roles for the *N*-glycans directly modulating spike-ACE2 interactions (Zhao et al., 2020). However, the importance of *N*-glycans of SARS-CoV-2 S protein, or S1 in particular, in its binding to ACE2 has not been fully demonstrated experimentally, although a critical role of the glycan shield in hiding the S protein surface from molecular recognition has been well recognized.

In the present study, the role of S1 *N*-glycosylation on ACE2 receptor binding and the implication on viral entry and invasion to host cells were investigated using SARS-CoV-2 S1 proteins expressed in three commonly used systems, baculovirus-infected insect (BII), Chinese hamster ovarian (CHO) and HEK 293 cells. For the latter, two variants, HEK 293E and HEK 293F (referred to as HEK-E and HEK-F, respectively, thereafter), were used. Real-time surface plasmon resonance (SPR) assays were carried out to obtain binding affinities of S1 to ACE2 before and after de-*N*-glycosylation in order to assess the overall influence of different *N*-glycosylation on S1 protein interaction with ACE2 receptor. *N*-glycan analysis of the recombinant proteins produced in the three systems was then performed to corroborate their different *N*-glycan profiles. Finally, MD simulation of the complexes of ACE2 with monomeric S1 subunit or S trimeric proteins carrying different *N*-glycans was used to understand the effects of different types of *N*-glycans on their binding to ACE2 reception.

RESULTS AND DISCUSSION

The effect of N-glycans of SARS-CoV-2 S1 protein on its binding affinity with ACE2 receptor

Real-time SPR assays using S1 proteins expressed from the three commonly used expression systems, including BII, CHO and two variants of HEK 293 cells, were carried out to investigate the binding activities of S1 proteins to the virus ACE2 receptor. The equilibrium dissociation constant K_D values obtained were used to define the strengths of bindings of S1 proteins to ACE2. As shown in Figure 1, the binding strengths of the four S1 proteins were all different, K_D 2.05 nM for BII- and 10.61 nM for CHO-expressed S1 proteins (Figures 1A and 1B, respectively). Even with the same HEK 293 cell system, the two variants using two different expression media, the equilibrium dissociation constants were different; K_D 23.9 nM for S1-HEK-F and K_D 6.43 nM for S1-HEK-E (Figures 1C and 1D, respectively).

The different strength of S1 protein binding to the ACE2 receptor is likely caused by the different glycosylation of S1 expressed in different systems. We then looked further at the effect of *N*-glycosylation by measuring the K_D after de-*N*-glycosylated by PNGase F of the four recombinant S proteins. For comparison, the S1 proteins were also pretreated with inactivated PNGase F under identical conditions as controls.

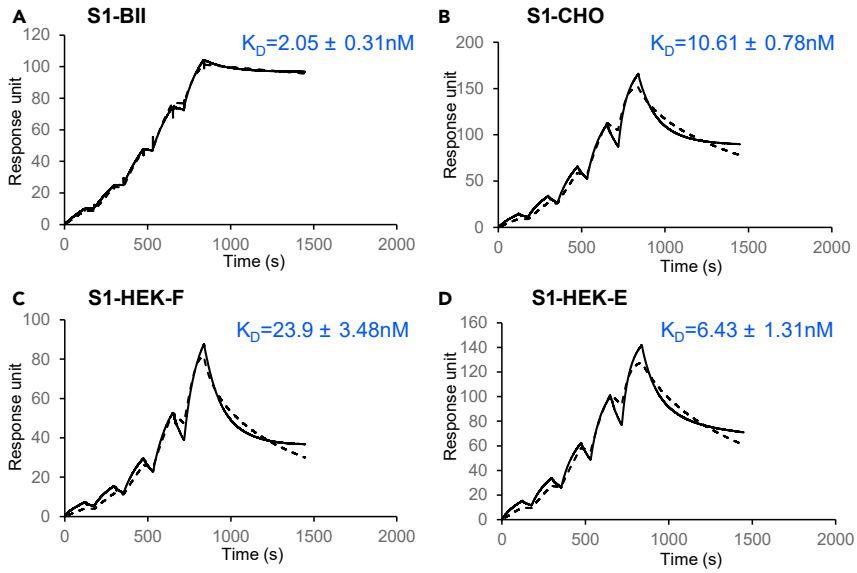


Figure 1. SPR binding curves of immobilized human ACE2 with S1 proteins of SARS-CoV-2 produced in different expression systems

(A–D) (A) Baculovirus-infected insect (BII) cells, (B) CHO cells, (C) HEK 293-F cells, and (D) HEK 293-E cells. The best fit of the data to a 1:1 binding model is shown in the dash line, and the K_D values are the mean ± SD of three independent experiments.

The K_D value obtained for the binding to ACE2 by S1 protein expressed in BII cells after de-*N*-glycosylation was 9.63 nM (dS1-BII, Figure 2, column a) compared with 15.7 nM from the control protein pretreated with deactivated PNGase F (Figure 2, column b), indicating only a slight improvement of its binding affinity to ACE2 after deglycosylation, in broad agreement with a recent observation of binding activity between the intact S protein and the ACE2 receptor (Sun et al., 2020). However, the K_D values for the binding of CHO cell-expressed S1 protein changed significantly, from 12.69 nM to 0.39 nM (dS1-CHO, Figure 2) after the *N*-glycans were removed. Similar major changes were also found for the two HEK 293 produced proteins. After de-*N*-glycosylation, the K_D values for the two HEK 293 expressed proteins were 9.32 nM and 5.68 nM (dS1-HEK-F and dS1-HEK-E, respectively, Figure 2, column a) compared with 171 nM and 43.35 nM for the control proteins (Figure 2, column b). More than an order of magnitude increase in binding affinities of de-glycosylated proteins indicated the important role of *N*-glycosylation on binding activity of S1 protein to the ACE2 receptor.

The SPR assays demonstrated the overall negative effect of *N*-glycosylation on SARS-CoV-2 S1 protein binding to ACE2. The difference of binding affinity detected for the proteins produced from different expression systems and the different degree of improvement of binding activity after removal of *N*-glycans was considered to be the effects of different types of *N*-glycosylation, as it has been well established that different cell expression systems can produce different glycosylation.

Analysis N-glycosylation of S1 proteins produced from different expression systems and under different conditions

As the SPR results indicated the S1 proteins obtained from different expression systems had different binding strengths to ACE2, we then performed *N*-glycan analysis of the four S1 proteins (Figures S1 and S2; Table S1) to demonstrate the different *N*-glycosylation profiles.

By MALDI-MS profiling and MALDI-MS^n branching pattern analysis (Sun et al., 2018), we found that *N*-glycans released from S1 expressed in BII cells were significantly different from *N*-glycans of the proteins expressed in HEK 293 and CHO cells. High-mannose *N*-glycans were the dominant form of the recombinant S1 from BII cells together with some minor hybrid type *N*-glycans, whereas complex type *N*-glycans were found to be as the main form in the proteins produced in HEK 293 and CHO cells (Figures 3A and 3D). A

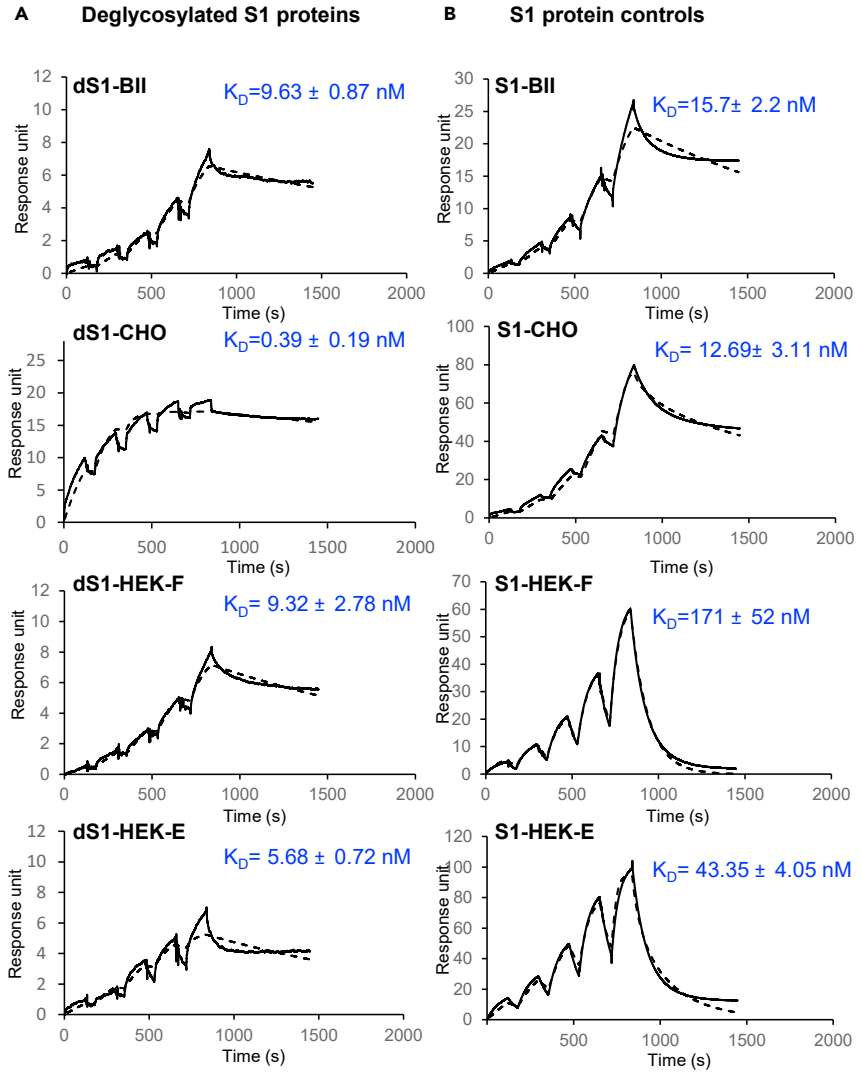


Figure 2. SPR binding curves of immobilized human ACE2 with de-*N*-glycosylated S1 proteins of SARS-CoV-2

(A and B) Column (A): deglycosylated S1 proteins by PNGase F, and column (B): the same S1 proteins pretreated with inactive PNGase F under identical conditions as controls. The best fit of the data to a 1:1 binding model is shown in the dash line, and the K_D values are the mean ± SD of two independent experiments.

small proportion of high-mannose structures were also detected in HEK 293- and CHO-produced S1 proteins. There was an overall high degree (40%–60%) of fucosylation in all four S1 proteins (Figure 3B) but sialylation of the complex type of *N*-glycans in S1 proteins were apparently different.

It is interesting to note that the *N*-glycans of the S1 proteins expressed in HEK 293 and CHO cells but using similar Gibco expression media (S1-HEK-F and S1-CHO, respectively) were very similar not only in the *N*-glycan types (Figure 3A) but also in fucosylation and sialylation levels (Figure 3B) and in the proportions of bi-antennary, tri- antennary, and tetra-antennary structures (Figure 3D). However, using similar cell systems but different expression media, e.g. S1-HEK-F and S1-HEK-E, the *N*-glycans produced were apparently different (Figures 3A and 3D). Appreciable amount of hybrid structures was detected in S1-HEK-E but not in S1-HEK-F. Within the complex type *N*-glycans, the contents of bi-antenary and tri-antennary structures were similar between HEK-E and HEK-F produced proteins but HEK-E derived S1 contained much less tetra-antennary *N*-glycans (Figure 3D). The major difference was in the degree of sialylation. Some 35% of *N*-glycans were sialylated in HEK-E produced protein, whereas only ~5% in HEK-F derived S1 (Figure 3B).

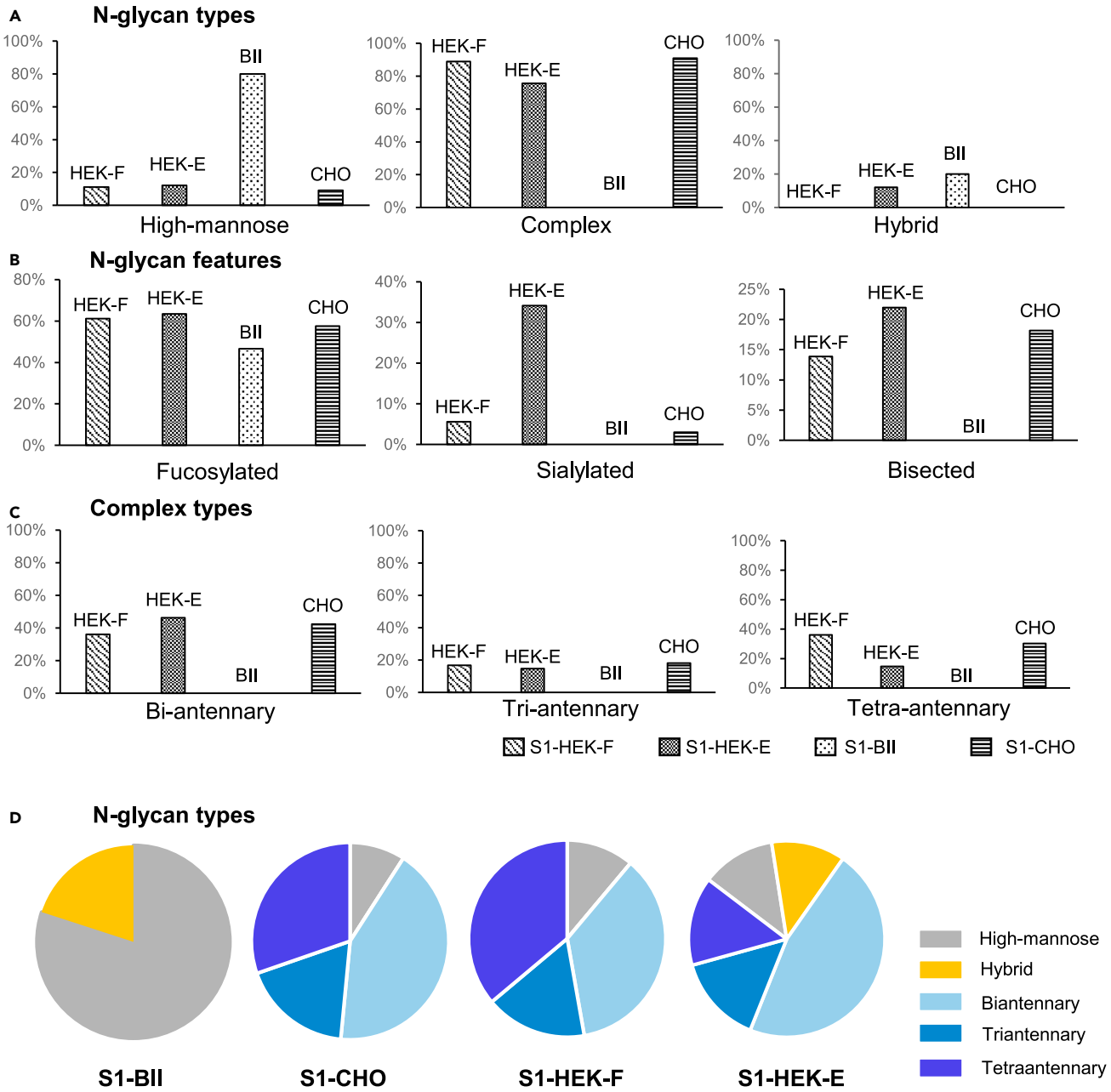


Figure 3. *N*-glycan features of the four recombinant S1 proteins of SARS-CoV-2

(A–D) (A) *N*-glycan types, (B) *N*-glycan features, (C) sub-classes of complex type, and (D) summary of *N*-glycan features of S1 proteins expressed in HEK 293F (HEK-F), HEK 293-E (HEK-E), BII, and CHO cells.

Apart from *N*-glycan types and forms, the relative abundances of individual *N*-glycans detected in all the S1 proteins were also compared after integration of their relative intensities (Figure S3). As S1-BII contains mainly high-mannose structures which are of generally lower molecular masses compared with the complex type *N*-glycans, the average molecular mass of S1-BII protein was lower than any of other S1 proteins. Among the other three proteins, S1-HEK-F was calculated to be the one with the highest average molecular mass (Figure S3B).

The *N*-glycan profiling results further indicated that indeed different *N*-glycans were produced from different expression systems and the difference in binding activity was likely to be caused by the different types of *N*-glycans and their relative abundance.

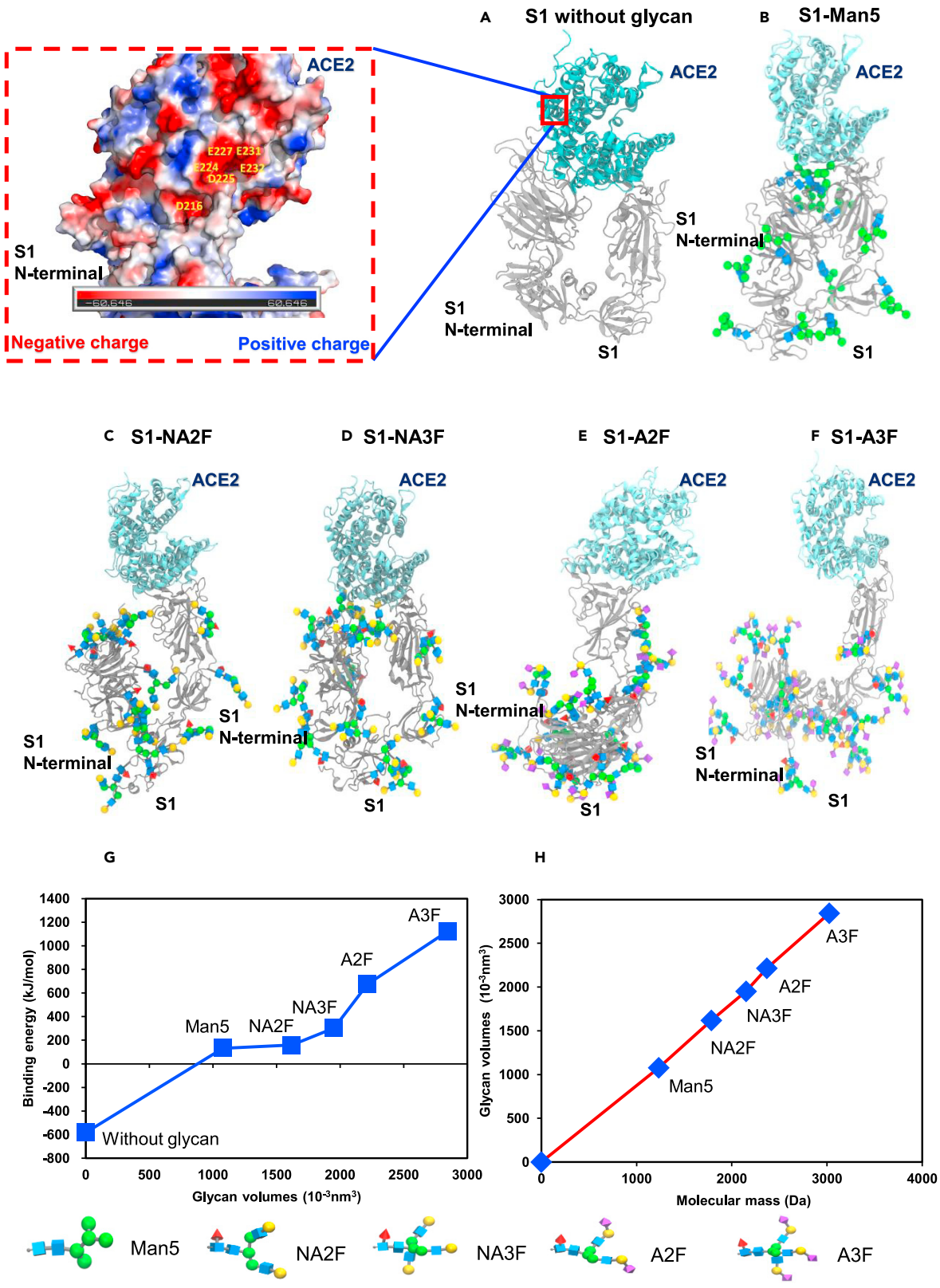

A
S1 without glycan
B
S1-Man5
ACE2
S1
N-terminal
S1
Negative charge
Positive charge
E227 E231
E224 E237
D225
D216
C
S1-NA2F
D
S1-NA3F
E
S1-A2F
F
S1-A3F
G
Binding energy (kJ/mol)
Glycan volumes (10⁻³nm³)
Without glycan
Man5
NA2F
NA3F
A2F
A3F
H
Glycan volumes (10⁻³nm³)
Molecular mass (Da)
Man5
NA2F
NA3F
A2F
A3F

Figure 4. 3D-Structural modeling and MD simulation of interactions of SARS-CoV-2 S1 proteins with ACE2 receptor

(A–H) 3D models of non-glycosylated (A) and differently *N*-glycosylated S1 proteins: (B) Man 5, (C) NA2F, (D) NA3F, (E) A2F, and (F) A3F; (G) the relationship between required binding energy and the glycan molecular volumes; (H) the relationship between glycan molecular volumes and glycan molecular mass.

Steric hindrance and Coulombic repulsion effects of *N*-glycosylation of S protein on its interactions with ACE 2 receptor

To understand the specific effects of different *N*-glycosylation on the binding affinity of S1 protein, we performed MD simulation of ACE2 complex with monomeric S1 proteins containing different *N*-glycans. We constructed the initial S1-ACE2 simulation system with 5 selected *N*-glycans with different molecular masses, ranging from high-mannose type in the low mass end, and fucosylated and sialylated tri-antennary complex-type structure in the high mass end, including Man 5 (1,234 Da), NA2F (1,786 Da), NA3F (2,151 Da), A2F (2,368 Da), and A3F (3,024 Da). In this proof-of-concept study, all of the 13 *N*-glycosites of S1 were occupied by identical glycans, thus yielding five S1-ACE2 complexes as the initial models. Next, using these S1-ACE2 models as the starting points, we performed MD simulation by running GROMACS (Abraham et al., 2015). After 100 ns simulations, the S1-ACE2 system became stable, i.e., the root-mean-square deviation (RMSD) between structures of simulated system and initial system converged (Figure S4). Prolonged sampling time to 150 ns (Figure S5) or 200 ns (Figure S6) did not show any difference. Finally, we sampled the last 20 stable systems (simulated time: 81 to 100 ns, Figure S4), and calculated their binding energy using gmx_mmpbsa (Ou et al., 2020). The average binding energy was used as the measure of binding affinity; a larger binding energy value indicates a lower binding affinity between S1 protein and ACE2 receptor.

Compared with the non-glycosylated, glycosylated S1 proteins showed lower binding affinities to ACE2 (Table S2), no matter which types of glycans were attached, in good agreement with the binding affinities measured by SPR assays (Figure 2).

We further leveraged the MD simulation results to interpret the lower binding affinity of glycosylated S1 proteins. In-depth examination suggested that the lower binding affinity of glycosylated S1 is likely due to steric hinderance effects of the attached glycans. As shown in Figure 4A, the *N*-terminal region of non-glycosylated S1 protein is in close proximity to the ACE2 receptor, with a distance between T20 of S1 and N90 of ACE2 at 6.1 angstrom. In contrast, the presence of Man 5, NA2F, NA3F, A2F, and A3F leads to significant steric effects, thus hindering the interaction between the *N*-terminal sequence of S1 and ACE2 (Figures 4B–4F). *N*-glycans with larger molecular volume (molecular size) contribute to greater steric effects, and therefore cause significant increase of binding energy (Figure 4G). As shown in Figure 4H, there is a good correlation between glycan's molecular volume and mass.

MD simulation of ACE2 complex with trimeric S protein was also conducted but only with limited glycan forms: non-glycosylated and two main types of *N*-glycoforms, high-mannose (*N*-glycan Man-5 as a representative) and complex (A2F *N*-glycan as a representative. Triplicate simulations were performed from different starting configurations of *N*-glycans (Figure S7), and the interface RMSD of several amino acids showed no appreciable difference (Figure S8), indicating that the protein-protein interaction is stable. The result demonstrated the order of the binding affinity of the trimeric S protein: non-glycosylated > *N*-glycosylated by Man-5 > *N*-glycosylated by A2F (Figure 5). The general tendency was identical to that obtained using S1 protein, indicating that the inherent instability of S1 subunit -does not interfere with the *N*-glycosylation effects on S protein interaction with ACE2 receptors.

Apart from steric hindrance effects, Coulombic repulsion between the negative charge of sialic acid residues of complex type *N*-glycans and the negative charge of the acidic amino acids also affects the binding affinity. As shown in Figure 4A, the Y196-K232 region of ACE2 close to the *N*-terminal region of S1 protein is mainly composed of acidic amino acids including aspartic acid (D) and glutamic acid (E). Consequently, when S1 protein is glycosylated with fully sialylated bi-antennary and tri-antennary A2F and A3F, the Coulombic repulsion drives S1 protein away from ACE2, thus reducing the binding affinity.

Conclusions

SARS-CoV-2 interaction with the host ACE2 receptor is affected by glycosylation of both the viral spike protein and ACE2. The former was demonstrated by the binding affinity changes before and after

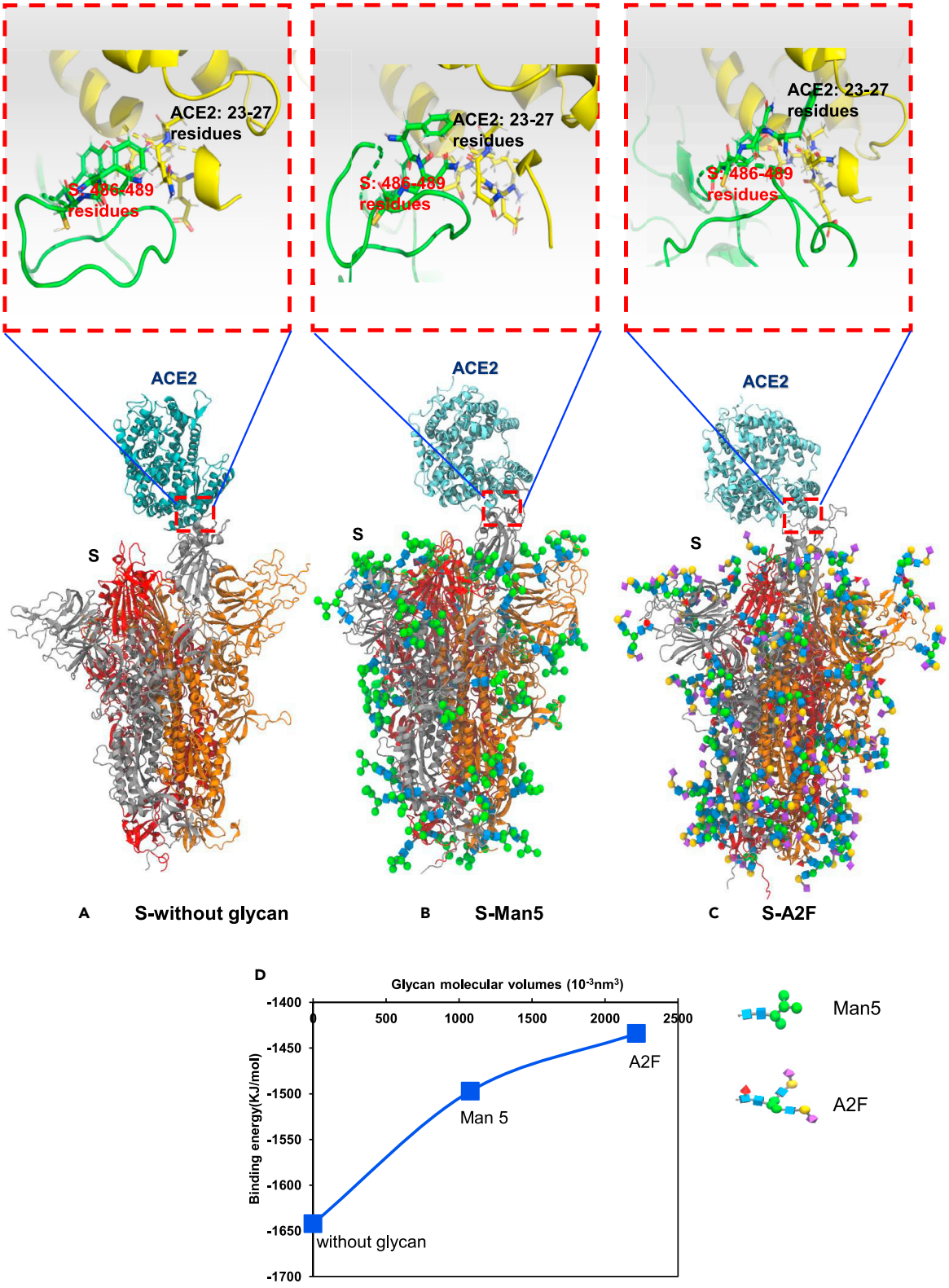


Figure 5. 3D-Structural modeling and MD simulation of interactions of SARS-CoV-2 trimeric S proteins with ACE2 receptor

(A–D) 3D models of non-glycosylated (A) and differently *N*-glycosylated S proteins: (B) Man 5, and (C) A2F; (D) the relationship between required binding energy and the glycan molecular volumes.

removal of *N*-glycans of the recombinant S1 proteins produced in the different expression systems. Deglycosylated S1 proteins showed stronger binding activities in all the cases. *N*-glycan analysis by mass spectrometry demonstrated different *N*-glycosylation profiles of S1 proteins obtained from different expression systems. MD simulation of S1 subunit and S trimeric protein complexed with the host receptor ACE2 using simplified and idealized *N*-glycosylation models indicated that the different *N*-glycan forms of the spike proteins can have steric hindrance and Coulombic repulsion effects in the binding process, resulting in conformation change of the spike protein and reduction of S1 binding affinity to the receptor. It has been considered that *N*-glycans on viral S protein might be an underlying mechanism for coronavirus to evade immune response due to shielding the amino acid residues from antibody recognition [9]. As demonstrated in the present study, *N*-glycans of different types impose different effects on interactions with receptors. A recent study also indicated the role of *N*-glycosylation of ACE2 on its interaction with receptor binding domain of the spike protein by MD simulation (Mehdipour and Hummer, 2021), e.g. the *N*-glycan chains at two glycosylation sites can have opposite effects on spike protein binding with N90 interfering whereas N322 strengthening the binding. For the virus spike protein, beyond the shielding effects attributed by different structures of *N*-glycans, *N*-glycans may cause different degrees of evasion from immune response, which is crucial for vaccine design. In the present work we used recombinant spike proteins to investigate the effect of *N*-glycosylation, the native glycans on spike proteins analyzed recently revealed that the *N*-glycosylation of the native S protein is highly similar to that expressed in HEK 293 cells (Yao et al., 2020; Cai et al., 2020). Therefore, the results obtained from the present study should provide further evidence for the important role of glycosylation of the viral S protein which can be considered in the design of antiviral drugs and vaccines. However, further efforts are required before a complete understanding of the role of glycosylation in SARS-CoV-2 infection and pathogenicity.

Limitations of the study

SARS-CoV-2 interaction with the host ACE2 receptor is affected by glycosylation of the viral spike protein which was demonstrated by the binding affinity changes before and after removal of *N*-glycans of the recombinant S1 subunit. However, because of the unavailability the intact spike trimeric proteins were not used for the SPR studies. Although both S1 subunit and S trimeric protein complexed with the host receptor ACE2 were used for MD simulation, only simplified and idealized *N*-glycosylations were considered, and therefore, the simulation may not reflect the real complex situation.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Material availability
  - Data and code availability
- METHOD DETAILS
  - Materials
  - N-Glycan release and MALDI-MS analysis
  - Deglycosylation of SARS-CoV-2 S1 proteins
  - Surface plasmon resonance experiments
  - Molecular dynamics simulations

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2021.103272.

ACKNOWLEDGMENTS

The work was supported in part by the National Natural Science Foundation of China (31600650, 31671369, and 62072435), Biological Resources Program, Chinese Academy of Sciences (KFJ-BRP-017-76), the March of Dimes Prematurity Research Center grant (22-FY18-82), and Wellcome Trust Biomedical Resource grant (218304/Z/19/Z). We also appreciate the support from the project of "Dengfeng Plan" from Foshan Hospital of Traditional Chinese Medicine.

AUTHOR CONTRIBUTIONS

YL, CH, WC and DB conceived the study. CH, KZ and WZ designed and performed the SPR assays, and measured the binding affinities. CH, SS and WC designed the mass spectrometry methodology, and analyzed the *N*-glycosylation based on mass spectral data. CH, WC, YL, SS and DB wrote the manuscript. All authors discussed the results and commented on the manuscript.

DECLARATION OF INTERESTS

The authors declare that they have no conflict of interest.

Received: May 24, 2021
Revised: August 30, 2021
Accepted: October 12, 2021
Published: November 19, 2021

REFERENCES

Abraham, M.J., Murtola, T., Schulz, R., Pall, S., Smith, J.C., Hess, B., and Lindahl, E. (2015). GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX *1-2*, 19–25.

Cai, Y., Zhang, J., Xiao, T., Peng, H., Sterling, S.M., Walsh, R.M., Jr., Rawson, S., Rits-Volloch, S., and Chen, B. (2020). Distinct conformational states of SARS-CoV-2 spike protein. Science *369*, 1586–1592.

Casalino, L., Gaieb, Z., Goldsmith, J.A., Hjorth, C.K., Dommer, A.C., Harbison, A.M., Fogarty, C.A., Barros, E.P., Taylor, B.C., McLellan, J.S., et al. (2020). Beyond shielding: the roles of glycans in the SARS-CoV-2 spike protein. ACS Cent. Sci. *6*, 1722–1734.

Han, D.P., Lohani, M., and Cho, M.W. (2007). Specific asparagine-linked glycosylation sites are critical for DC-SIGN- and L-SIGN-mediated severe acute respiratory syndrome coronavirus entry. J. Virol. *81*, 12029–12039.

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. J. Mol. Graph. *14*, 33–38.

Jeffers, S.A., Tusell, S.M., Gillim-Ross, L., Hemmila, E.M., Achenbath, J.E., Babcock, G.J., Thomas, W.D., Jr., Thackray, L.B., Young, M.D., Mason, R.J., et al. (2004). CD209L (L-SIGN) is a receptor for severe acute respiratory syndrome coronavirus. Proc. Natl. Acad. Sci. U. S. A. *101*, 15748–15753.

Jo, S., Song, K.C., Desaire, H., MacKerell, A.D., Jr., and Im, W. (2011). Glycan reader: automated sugar identification and simulation preparation for carbohydrates and glycoproteins. J. Comput. Chem. https://doi.org/10.1002/jcc.21886.

Leemans, A., Boerena, M., VanderGucht, W., Martinet, W., Caljon, G., Maes, L., Cos, P., and Delputte, P. (2019). Characterization of the role of N-glycosylation sites in the respiratory syncytial virus fusion protein in virus replication, syncytium formation and antigenicity. Virus Res. *266*, 58–68.

Letko, M., Marzi, A., and Munster, V. (2020). Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. Nat. Microbiol. *5*, 562–569.

Mehdipour, A.R., and Hummer, G. (2021). Dual nature of human ACE2 glycosylation in binding to SARS-CoV-2 spike. Proc. Natl. Acad. Sci. U. S. A. *118*, e2100425118.

Ou, J., Zhou, Z., Dai, R., Zhang, J., Lan, W., Zhao, S., Wu, J., Seto, D., Cui, L., Zhang, G., et al. (2020). Emergence of RBD mutations in circulating SARS-CoV-2 strains enhancing the structural stability and human ACE2 receptor affinity of the spike protein. bioRxiv preprint. https://doi.org/10.1101/2020.03.15.991844.

Parsons, L.M., Bouwman, K.M., Azurmendi, H., de Vries, R.P., Cipollo, J.F., and Verheije, M.H. (2019). Glycosylation of the viral attachment protein of avian coronavirus is essential for host cell and receptor binding. J. Biol. Chem. *294*, 7797–7809.

Raman, R., Tharakaraman, K., Sasisekharan, V., and Sasisekharan, R. (2016). Glycan–protein interactions in viral pathogenesis. Curr. Opin. Struct. Biol. *40*, 153–162.

Shajahan, A., Supekar, N.T., Gleinich, A.S., and Azadi, P. (2020). Deducing the N- and O-glycosylation profile of the spike protein of novel coronavirus SARS-CoV-2. Glycobiology *30*, 981–988.

Sun, S., Huang, C., Wang, Y., Liu, Y., Zhang, J., Zhou, J., Gao, F., Yang, F., Chen, R., Mulloy, B., et al. (2018). Toward automated identification of glycan branching patterns using multistage mass spectrometry with intelligent precursor selection. Anal. Chem. *90*, 14412–14422.

Sun, Z., Ren, K., Zhang, X., Chen, J., Jiang, Z., Jiang, J., Ji, F., Ouyang, X., and Li, L. (2020). Mass spectrometry analysis of newly emerging coronavirus HCoV-19 spike protein and human ACE2 reveals camouflaging glycans and unique post-translational modifications. Engineering. https://doi.org/10.1016/j.eng.2020.07.014.

Wang, D., Baudys, J., Bundy, J.L., Solano, M., Keppel, T., and Barr, J.R. (2020). Comprehensive analysis of the glycan complement of SARS-CoV-2 spike proteins using signature ions-triggered electron-transfer/higher-energy collisional dissociation (EThcD) mass spectrometry. Anal. Chem. *92*, 14730–14739.

Watanabe, Y., Berndsen, Z.T., Raghwani, J., Seabright, G.E., Allen, J.D., Pybus, O.G., McLellan, J.S., Wilson, I.A., Bowden, T.A., Ward, A.B., et al. (2020a). Vulnerabilities in coronavirus glycan shields despite extensive glycosylation. Nat. Commun. *11*, 2688.

Watanabe, Y., Allen, J.D., Wrapp, D., McLellan, J.S., and Crispin, M. (2020b). Site-specific glycan analysis of the SARS-CoV-2 spike. Science *369*, 330–333.

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.

Yao, H., Song, Y., Chen, Y., Wu, N., Xu, J., Sun, C., Zhang, J., Weng, T., Zhang, Z., Wu, Z., et al. (2020). Molecular architecture of the SARS-Cov-2 virus. Cell *183*, 730–738.

Zhang, Y., Zhao, W., Mao, Y., Chen, Y., Wang, S., Zhong, Y., Su, T., Gong, M., Du, D., Lu, X., et al. (2021). Site-specific N-glycosylation characterization of recombinant SARS-CoV-2 spike proteins. Mol. Cell. Proteom. *20*, 100058.

Zhao, P., Praissman, J.L., Grant, O.C., Cai, Y., Xiao, T., Rosenbalm, K.E., Aoki, K., Kellman, B., Bridger, R., Barouch, D.H., et al. (2020). Virus-receptor interactions of glycosylated SARS-CoV-2 spike and human ACE2 receptor. Cell Host Microbe *28*, 1–16.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Cells and S1 proteins expression		
CHO cells	Thermo Fisher	Cat#ExpiCHO-S
HEK 293F cells	Thermo Fisher	Cat#Expi293F
ExpiCHO expression kits	Thermo Fisher	Cat#A29133
Expi293F expression kits	Thermo Fisher	Cat#A14635
Chemicals and recombinant proteins		
ACE2	Sino Biological	Cat#10108-H02H
Mouse anti-human IgG1-Fc	Sino Biological	Cat#10702-MM01T; RRID: AB_2860220
DL-dithiothereitol (DTT)	Sigma-Aldrich	Cat#43815
Iodoacetamide (IAA)	Sigma-Aldrich	Cat#I1149
2,5-dihydroxybenzoic acid (DHB)	Sigma-Aldrich	Cat#39319
Ammonium bicarbonate	Sigma-Aldrich	Cat#09830
Sodium hydroxide	Sigma-Aldrich	Cat#S8045
Tween-20	Sinopharm	Cat#20200326
Dulbecco's phosphate buffer saline (DPBS)	Macgene	Cat# CC010
Trypsin	Promega	Cat#V5113
PNGase F	New England Biolabs	Cat#P0704L
Cartridge and plates		
Sep-PakC18 96-well	Waters	Cat#186003966
Oasis HLB 96-well plate	Waters	Cat#WAT058951
µfocus MALDI plate	Hudson Surface	Cat#PL-PC-000050-P
Software and algorithms		
Gromacs 2019.6	Abraham et al., 2015	https://www.gromacs.org/
VMD 1.9.3	Humphrey et al. (1996)	http://www.ks.uiuc.edu/Research/vmd/
Pymol 2.4.0		https://pymol.org/2/
CHARMM-GUI	Jo et al. (2011)	https://www.charmm-gui.org/
GIPS 1.0	Sun et al., 2018	http://glycan.ict.ac.cn/GIPS/
QtGrace 2.6		http://sourceforge.net/p/qtgrace/wiki/Home/
SciDAVis 2.4.0		http://scidavis.sourceforge.net/index.html
Gmxtool		http://github.com.Jerkwin/gmxtool

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Wengang Chai (w.chai@imperial.ac.uk).

Material availability

This study did not generate new unique reagents.

Data and code availability

The MD trajectories and related files have been deposited in the https://covid.molssi.org/, and are publicly available as the data of publication. The processed data is provided as supplemental information, and the

raw data in this paper will be shared by the lead contact Dr. Wengang Chai (w.chai@imperial.ac.uk). No original code was produced in this paper. Any additional information required to reanalyse the data reported is available from the lead contact upon request.

METHOD DETAILS

Materials

All recombinant SARS-CoV-2 S1 proteins were His tagged. S1-CHO and S1-HEK-F were provided, as gifts, by our collaborator AnyGo Technology (Beijing, China), and produced in CHO and human embryonic kidney (HEK) 293F cells, using the Gibco method and the cell-specific expression kits, ExpiCHO and Expi293F (ThermoFisher, Waltham, MA, USA), respectively. S1-BII and S1-HEK-E were produced in baculovirus-infected insect (BII) and HEK 293-EBNA1 cells, respectively, and were purchased from Sino Biological (Beijing, China). Recombinant Fc-tagged human ACE2 and mouse anti-human IgG1-Fc were from Sino Biological. DL-dithiothreitol (DTT), iodoacetamide (IAA), 2,5-dihydroxybenzoic acid (DHB), ammonium bicarbonate and sodium hydroxide were purchased from Sigma-Aldrich (St. Louis, MO, USA). Tween-20 and Dulbecco's phosphate buffered saline (DPBS) were purchased from Sinopharm. (Shanghai, China). Trypsin was from Promega (Madison, WI, USA). PNGase F was obtained from New England Biolabs (Ipswich, MA, USA). C18 cartridge and 96-well plate were purchased from Waters (Milford, MA, USA). All the chemicals were of analytical grade or better, and used as received without further purification.

N-Glycan release and MALDI-MS analysis

SARS-CoV-2 S1 proteins were denaturized in the presence of DTT and IAA before digestion with trypsin. The resulting glycopeptides were then incubated with PNGase F, and the obtained *N*-glycans were purified and permethylated as described (Sun et al., 2018).

Permethylated *N*-glycans were analyzed on an Axima MALDI Resonance mass spectrometer with a QIT-TOF configuration (Shimadzu, Kyoto, Japan). A nitrogen laser was used to irradiate samples at 337 nm, with an average of 200 shots accumulated. Permethylated *N*-glycans were dissolved in methanol and applied to a μfocus MALDI plate target (900 μm, 384 circles, Hudson Surface Technology, Old Tappan, NJ, USA). A matrix solution (0.5 μL) of 2,5-dihydroxybenzoic acid (20 mg/mL) in a mixture of methanol/water (1:1) containing 0.1% trifluoroacetic acid and 1 mM NaCl was added to the plate and mixed with samples. The mixture was air dried at room temperature before analysis. MALDI-MS profiling and MALDI-CID-MS^n branching pattern analysis were performed as reported previously (Sun et al., 2018).

N-glycan types (high-mannose, complex and hybrid) and features (fucosylation, sialylation and bisecting), and specific antenna subtypes of complex *N*-glycans were described as percentage of total *N*-glycans counted. The relative abundances of each type of *N*-glycans were the percentages of summed intensities.

Deglycosylation of SARS-CoV-2 S1 proteins

SARS-CoV-2 S1 proteins (100 μg) were incubated with 0.5 μL PNGase F (New England Biolab, Ipswich, MA,1:50) or deactivated PNGase F at 37°C in PBS buffer overnight. The digestion was stopped by heating to 75°C for 10 min, and the enzyme was removed using a millipore 50 kD ultrafilter tube (Burlington, MA, USA). For controls, parallel digestions of the four S1 proteins were also carried out under identical conditions using PNGase F pre-deactivated by heating of the enzyme solution at 75°C for 10 min. De-*N*-glycosylation experiments were repeated twice for SPR experiments. Aliquots (1–2 μL) of SARS-CoV-2 S1 protein solutions were incubated with the fluorescent agent of the Qubit protein assay kit (Life Technologies, ThermoFisher) at room temperature for 15 min, and then the concentrations of S1 proteins were measured by fluorescence using Qubit 3 Fluorometer (Invitrogen, ThermoFisher).

Surface plasmon resonance experiments

The Fc-tagged ACE2 proteins were captured by mouse anti-human IgG1-Fc secondary antibodies that were immobilized on the chip and tested for binding with gradient concentrations of the soluble S1 proteins. The binding kinetics and affinity of the ACE2 immobilized at the chip surface to glycosylated and deglycosylated SARS-CoV-2 S1 proteins were measured.

Binding affinities of S1 proteins and ACE2 were measured using Biacore 8K SPR system (GE Healthcare, Pittsburgh, PA, USA) with sensor chip CM5 in a running buffer of DPBS with 0.005% Tween-20. To prepare the capture surface, mouse anti-human IgG Fc was diluted to 50 μg/mL in immobilization buffer (10 mM sodium acetate, pH 5.0), and then amine-coupled at a flow rate of 10 μL/min for 420 s. The chip was activated with a freshly prepared 1:1 (v/v) mixture of aqueous stocks of 0.4 M 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) and 0.1M N-hydroxysuccinimide (NHS) for 420 s, and blocked with 1 M ethanolamine for 420 s. The ACE2 was diluted to 5 μg/mL in running buffer and then injected to sample channel to reach a capture level of about 600 RU. S1 proteins were prepared at concentrations of 0, 12.5, 25, 50, 100, 200 nM, and then injected at a flow rate of 30 μL/min for an association phase of 120 s, followed by 350 s dissociation. Samples were injected in a multi-cycle manner over freshly captured S1 proteins by regenerating the capture surfaces with two injections of glycine-HCl (10 mM, pH 1.7, 20–60 s) at a flow rate of 30 μL/min. Repeated injections were performed to assess the assay's reproducibility. The data was processed and analyzed with Biacore 8K Evaluation Software Version 1.0 (GE Healthcare) to calculate the affinity constant K_D values based on the values of association constant Ka and disassociation constant Kd. The best fit of the data to a 1:1 binding model is shown in dash line, and the K_D values are the mean ± SD of three independent experiments for the proteins produced from different expression systems, and two independent experiments for the deglycosylated S1 protein controls.

Molecular dynamics simulations

The monomeric S1-ACE2 complex model was built based on the fully N-glycosylated SARS-CoV-2 S protein (PDB : 6VSB) and ACE2 receptor (PDB : 6VW1) from CHARMM-GUI COVID-19 protein library. All the heteroatoms including native glycans, ions and water were removed from the original fully-glycosylated SARS-CoV-2 S protein (PDB : 6VSB) and ACE2 receptor (PDB : 6VW1) complex model. A monomer was trimmed from the homotrimeric S protein by Pymol and was defined as Chain A which has the "up" state receptor binding domain and could be capable of binding to ACE2 receptor. Monomeric S1 subunit of chain A was retained, and ACE2 receptor was kept as its original conformation. Glycosylated S1-ACE2 complex system was generated through the functional modules in CHARMM-GUI. Glycan Reader was used to attach glycans to the 13 *N*-glycosites of the S1 protein, and 5 selected *N*-glycans with different molecular masses including Man 5 (1,234 Da), NA2F (1,786 Da), NA3F (2,151 Da), A2F (2,368 Da), and A3F (3,024 Da) were attached. Ions were added to neutralize the system using Monte-Carlo method, and then the system was solvated with explicit water molecules described by the TIP3P model. Finally, we used CHARMM36 all-atom force field to parameterize the generated S1-ACE2 complex system.

For the complex of trimeric S protein and ACE2 receptor (S-ACE2), non-glycosylated S proteins and S proteins attached with Man 5 as high mannose-type *N*-glycan representative and A2F as complex N-glycan representative were utilized in the simulation. Triplicate simulations were performed from different starting configurations of *N*-glycans.

The simulations were performed with GROMACS, a prevalent molecular dynamics software. The simulation of S1-ACE2 complex simulation was performed on Sugon High Performance Computing clusters with 192 CPU cores, and the simulation of S-ACE2 complex simulation was completed on CAS Xiandao-1 computing environment with 100 nodes. Firstly, the binary input files of complex systems were generated with LINCS, a constraint algorithm, which converted all the bonds with hydrogen-atoms to constraints. Next, under periodic boundary conditions, we initiated simulation consisting of three main steps which are energy minimization, equilibration and production respectively. The long-range electrostatic interactions were calculated by the particle-mesh Ewald method and the van der Waals interactions were smoothly switched off over 10–12 Å by a force-based switching function.

During the energy minimization step, the S1-ACE2 and S-ACE2 complex, glycans, solvent and ions were kept unchanged and the system was subjected to an initial minimization using the steepest descent algorithm. To equilibrate the system, we run two steps of equilibration simulations. To maintain the temperature (303.15 K), a Nose-Hoover temperature coupling method with a time constant for coupling of 1 ps was used. For pressure coupling (1 bar), an isotropic Parrinello–Rahman method with a time constant for pressure coupling of 5 ps and a compressibility of 4.5 × 10–5 bar–1 was implemented. During two equilibration steps, NVT (constant particle number, volume, and temperature) dynamics was first applied with a 1 fs time step for 0.5 ns. Subsequently, the NPT (constant particle number, pressure, and temperature) ensemble was applied with a 2 fs time step for 1ns. After the two equilibration steps, the system was well-equilibrated

at the desired temperature and pressure, 303.15 K and 1 bar. The position restrains were released, and production MD was run to obtain trail collection. Specifically, production MD simulation for the S1-ACE2 complex was performed at 2 fs time step for 100 ns, 150 ns and 200 ns, and we found that the system was well-equilibrate around 100 ns. The MD simulation for the S-ACE2 complex was performed for 100 ns, and all the coordinates were saved every 50 ps.