Glycoinformatics approach for identifying target positions to inhibit initial binding of SARS-CoV-2 S1 protein to the host cell

Muhammet Uslupehlivan, Ecem Şener*

Ege University

Faculty of Science, Department of Biology

Molecular Biology Section

Bornova/Izmir

Turkey

*Corresponding Author: Ecem Şener

Email: ecem.sener@ege.edu.tr

Abstract

COVID-19 outbreak is still threatening the public health. Therefore, in the middle of the pandemic, all kind of knowledge on SARS-CoV-2 may help us to find the solution. Determining the 3D structures of the proteins involved in host-pathogen interactions are of great importance in the fight against infection. Besides, post-translational modifications of the protein on 3D structure should be revealed in order to understand the protein function since these modifications are responsible for the host-pathogen interaction. Based on these, we predicted O-glycosylation and phosphorylation positions using full amino acid sequence of S1 protein. Candidate positions were further analyzed with enzyme binding activity, solvent accessibility, surface area parameters and the positions determined with high accuracy rate were used to design full 3D glycoprotein structure of the S1 protein using carbohydrate force field. In addition, the interaction between the C-type lectin CD209L and α-mannose residues was examined and carbohydrate recognition positions were predicted. We suggest these positions as a potential target for the inhibition of the initial binding of SARS-CoV-2 S1 protein to the host cell.

Keywords: Coronavirus, CD209L □ α-mannose interaction, 3D glycoprotein structure, O-glycosylation, phosphorylation, carbohydrate force field, molecular docking

Introduction

SARS-CoV-2 emerged in Wuhan, China in December 2019 and spread quickly across the world. It resulted in 575,444 confirmed cases and 26,654 deaths as to World Health Organization data as of 29 March 2020. Therefore, in the middle of the pandemic, all kind of knowledge on SARS-CoV-2 may help us to find the solution.

The coronavirus spike protein (S) plays a key role in the early steps of viral infection. S protein comprises of two sub-units S1 and S2 which are responsible for the binding of the host cell receptor and fusion of the cellular membrane, respectively [1,2,3]. S protein also contains furin cleavage site at the boundary between the S1-S2 sub-units which mediates the membrane fusion and virus infectivity [1,2]. It has been suggested that different domains within a single S protein could bind multiple alternative receptors. Although ACE2 (angiotensin-converting enzyme 2) is known to be the SARS-CoV receptor [2,4], CD209L, a C-type lectin that binds to high-mannose glycans on glycoproteins, has also been found to be as an alternative receptor for SARS-CoV [5]. However, recent studies have mainly focused on the S protein-ACE2 interaction, there is a paucity of information on the SARS-CoV-2 S1 protein-CD209L lectin interaction. It has been found that CD209L lectin binds to SARS-CoV, human coronavirus 229E, Ebola, Hepatitis C, HIV, Influenza virus glycoproteins and may mediate the endocytosis of pathogens [6-11]. Thus, identification of the carbohydrate-lectin interaction sites on the 3D structure for understanding of the initial binding of virus to a host cell is crucial. Besides, as post-translational modifications regulate the host-pathogen interaction [12], identifying SARS-CoV-2 S1 protein modifications may help us to inhibit initial binding of the virus. 3D crystal structure of SARS-CoV-2 S protein with N-glycosylation positions was identified but, sites such as furin cleavage and O-glycosylation are missing. In addition, the detailed structure of N-glycan units was not included in this model [13]. As glycan structures typically exist in solution or on proteins, it is a big challenge to characterize the 3D structure of glycoproteins experimentally. However, computational structural biology allows us to generate 3D protein and glycoprotein modelling with high accuracy rate using all amino acid sequence [14-16]. Herein, we predicted O-glycosylation and phosphorylation positions using full amino acid sequence of S1 protein. Candidate positions were further analyzed with enzyme binding activity, solvent accessibility, surface area parameters and the positions determined with high accuracy rate were used to design full 3D glycoprotein structure of the S1 protein using carbohydrate force field. In addition, the

interaction between the C-type lectin and α-mannose residues was examined and carbohydrate recognition positions were predicted.

Results and Discussion

Determining the 3D structures of the proteins involved in host-pathogen interactions are of great importance in the fight against infection. Besides, post-translational modifications of the protein on 3D structure should be revealed in order to understand the protein function since these modifications are responsible for the host-pathogen interaction. Therefore, researchers primarily have focused on the S protein structure of SARS-CoV-2 virus and revealed the 3D structure with N-glycosylation positions. But, the detailed structure of the N-glycan units in these sites and the O-glycosylation positions are not included in the current model [13]. Also, it is known that the initial binding of other coronaviruses to host cell occurs via the alternative receptor C-type lectin CD209L besides ACE2 protein [5]. Taken together, herein, we first demonstrated SARS-CoV-2 S1 protein high mannose N-glycan structure and its interaction with CD209L. When 3D glycoprotein structure of S1 protein analysed, N-glycosylation positions were found to be located mainly on the N-Terminal Domain (NTD) (Figure 1). Since the Receptor Binding Domain (RBD) is known to be involved in ACE2 binding, we suggest that N-glycan positions on NTD may associated with the binding of C-type lectin. C-type lectins interact with α-linked mannose residues on high mannose N-glycan structures [17]. There has been a structural study on C-type lectin-glycan interaction but terminal sugar is not α-mannose here [18]. Thus, in this study, the interaction between the C-type lectin and α-D-mannose was analysed and interface residues were identified. ZDOCK docking and PDBePISA interface interaction analysis showed in two models with high accuracy rate that mannose sugar interacts with C-type lectin at Met282, Lys307, and Ser345 positions with hydrogen bonding (Fig. 2). We suggest these positions as the α-D-mannose recognizing sites having function on the lectin-sugar binding.

Secondly, we identified O-GalNAc and O-GlcNAc modifications of SARS-CoV-2 S1 protein. O-glycans are known to be involved in protein stability and function [19]. The presence of O-glycans in some viral proteins has been demonstrated and suggested that glycans may play a role on biological activity of viral proteins. In the comparative study on human SARS-CoV-2 and other coronaviruses S proteins have showed that Ser673, Thr678, and Ser686 are conserved O-glycosylated positions and have suggested that SARS-CoV-2 S1 protein may show O-glycosylation at these positions [20]. In this study, S1 protein was found to be O-GalNAcylated at Thr632, Thr678 and O-β-GlcNAcylated at Thr323, Thr638, Ser686

positions (Table 1). When compared our results with the previous study, SARS-CoV-2 S1 protein found not to be O-glycosylated at Thr673, but O-GalNAcylated at Thr678 and O-β-GlcNAcylated at Ser686. Besides, we found additional O-glycosylation positions at Thr323, Thr632, and Thr638 on the S1 protein. On the 3D glycoprotein structure, Thr323 was found to be located at RBD and Thr678 and Ser686 located near Furin Recognizing Site (FRS) (Fig. 1 and 3). Since O-glycans are involved in protein-protein interactions [19], we suggest that O-glycans at Thr323 may play a role on binding to ACE2 and O-glycans at Thr678, Thr686 may responsible for furin protease enzyme binding. It has been known that O-glycans are responsible for the protein stability and creating mucin like domain as glycan shields involved in immunoevasion [21]. Based on this, we suggest that the O-glycosylation positions found may be involved in the stability of the S1 protein and immunoevasion of virus to the host cell.

Phosphorylation of viral proteins are catalyzed by host cell enzymes like glycosylation modifications [12,22]. In this study, phosphorylation modification of SARS-CoV-2 S1 protein was also examined and 36 positions of Ser/Thr residues were found to be phosphorylated on almost all domains of the protein (Table 2). Davidson et al. have shown that S1 protein is phosphorylated at 7 positions [23]. Amongst them, Ser459, Ser637, Ser640 positions were found to be correlated with our results. Phosphorylation modification regulates the protein activity and function in cooperation with glycosylation [24,25]. When we examined the phosphorylation and glycosylation locations of S1 protein on the 3D structure; the phosphorylation positions at Ser161 and Ser162 were found to be located close to glycosylation position Asn165. Likewise, Thr109 with Asn234; Ser680 with Thr678 and FRS; Thr604 with Asn603; Thr630 with Thr632; Thr637 and Thr640 with Thr638; Ser659 with Asn657 were found to be located close to each other on the 3D structure (Fig. 3). Thus, we suggest these sites as critical sites for S1 protein activity.

Conclusion

COVID-19 outbreak is still threatening the public health. At this point, inhibiting the initial binding of virus to the host cell is crucial in order to find treatment. Since post-translational modifications regulate the host-pathogen interaction, identifying SARS-CoV-2 S1 protein modifications may help us to inhibit initial binding of the virus. Therefore, we focused on the CD209L lectin-α-D-mannose interaction of S1 protein and suggested Met282, Lys307, and Ser345 positions as targets for the initial binding. Also, we found that phosphorylation and glycosylation positions where located at close sites may be critical for S1 protein activity.

Therefore, we suggest that, these positions can be targeted for drug development against COVID-19.

Materials and Methods

Prediction of O-Glycosylation and Phosphorylation Positions

Amino acid sequence of SARS-CoV-2 S1 protein was taken from NCBI with QHD43416 ID. The potential O-glycosylation (O-GalNAc) and O-β-GlcNAc sites of S1 protein were analysed via NetOGlyc 4.0 [26] and YinOYang 1.2 Server [27], respectively. The threshold was chosen as 0.5 for both to predict high potential sites. The potential glycosylation positions which are found on NetOGlyc and YinOYang were analysed with the ISOGlyP server for the enzyme binding activity [28]. ISOGlyP was used to calculate all potential positions with ppGalNAc transferase isoforms which showed high enzyme binding activity and calculates the enhancement value product (EVP) values as an indication of glycosylation rates [29]. The potential phosphorylation positions of S1 protein was analysed with NetPhos 3.1 Server [30]. NetSurfP v1.1 was used to assess the surface and solvent accessibility of predicted Ser and Thr positions with glycosylation and phosphorylation positions [31]. The relative solvent accessibility of potential glycosylation positions was analysed using Poly-View 2D-SABLE protein structure prediction server [32].

Glycoprotein Building on 3D Structure using Carbohydrate Force Field

3D structure models involving full amino acid sequence of S1 protein was taken with QHD43416 ID from I-TASSER (Iterative Threading ASSEmbly Refinement) [33]. The structure model has been generated by the C-I-TASSER pipeline, which utilizes deep convolutional neural-network based contact-map predictions to guide the I-TASSER fragment assembly simulations. GLYCAM-Web Server (AMBER carbohydrate force field) was used to screen highly reliable glycosylation sites on the protein 3D structure. GLYCAM-Web is dedicated to simplifying the prediction of three-dimensional structures of carbohydrates and macromolecular structures involving carbohydrates. Glycoprotein Builder runs SASA (Solvent Accessible Surface Area) prediction and finds the most appropriate glycosylation sites for adding glycan units on the 3D structure using molecular dynamics simulation [34,35]. Glycan units skeleton were chosen as Core 1 type that is the most common O-glycan motifs on proteins for O-glycan units. The common shape of high mannose type that is a conserved motif that plays an important role on the N-glycan formation were chosen for N-glycan units [17]. Additionally, all potential phosphorylation sites were analysed for SASA

parameters on the 3D protein structure using GLYCAM and the positions with high accuracy rates were chosen as phosphorylation positions.

Lectin-Carbohydrate Docking

In order to find out the CD209L lectin–mannose interaction, ZDOCK docking server was used [36]. CD209L lectin 3D structure (1.4 Å resolution) was taken from PDB (ID: 1XPH) [37] and α-D-mannose structure was taken from Glyco3D database [38]. PDBePISA server was used for the exploration of macromolecular interfaces between the receptor-ligand [39]. All structures were visualized with PyMOL.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflicts of interest

The authors have no conflict of interest to declare.

References

[1] Belouzard, S., Chu, V. C., & Whittaker, G. R. (2009). Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. *Proceedings of the National Academy of Sciences*, *106*(14), 5871-5876.

[2] Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T., & Veesler, D. (2020). Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell*.

[3] Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., ... & Müller, M. A. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell*.

[4] Wong, S. K., Li, W., Moore, M. J., Choe, H., & Farzan, M. (2004). A 193-amino acid fragment of the SARS coronavirus S protein efficiently binds angiotensin-converting enzyme 2. *Journal of Biological Chemistry*, *279*(5), 3197-3201.

[5] Jeffers, S. A., Tusell, S. M., Gillim-Ross, L., Hemmila, E. M., Achenbach, J. E., Babcock, G. J., ... & Ambrosino, D. M. (2004). CD209L (L-SIGN) is a receptor for severe acute respiratory syndrome coronavirus. *Proceedings of the National Academy of Sciences*, *101*(44), 15748-15753.

[6] Alvarez, C. P., Lasala, F., Carrillo, J., Muñiz, O., Corbí, A. L., & Delgado, R. (2002). C-type lectins DC-SIGN and L-SIGN mediate cellular entry by Ebola virus in cis and in trans. *Journal of virology*, *76*(13), 6841-6844.

[7] Lin, G., Simmons, G., Pöhlmann, S., Baribaud, F., Ni, H., Leslie, G. J., ... & Doms, R. W. (2003). Differential N-linked glycosylation of human immunodeficiency virus and Ebola virus envelope glycoproteins modulates interactions with DC-SIGN and DC-SIGNR. *Journal of virology*, *77*(2), 1337-1346.

[8] Cormier, E. G., Durso, R. J., Tsamis, F., Boussemart, L., Manix, C., Olson, W. C., ... & Dragic, T. (2004). L-SIGN (CD209L) and DC-SIGN (CD209) mediate transinfection of liver cells by hepatitis C virus. *Proceedings of the National Academy of Sciences*, *101*(39), 14067-14072.

[9] Jeffers, S. A., Hemmila, E. M., & Holmes, K. V. (2006). Human coronavirus 229E can use CD209L (L-SIGN) to enter cells. In *The Nidoviruses* (pp. 265-269). Springer, Boston, MA.

[10] Londrigan, S. L., Turville, S. G., Tate, M. D., Deng, Y. M., Brooks, A. G., & Reading, P. C. (2011). N-linked glycosylation facilitates sialic acid-independent attachment and entry of influenza A viruses into cells expressing DC-SIGN or L-SIGN. *Journal of virology*, *85*(6), 2990-3000.

[11] Marzi, A., Gramberg, T., Simmons, G., Möller, P., Rennekamp, A. J., Krumbiegel, M., ... & Steinkasserer, A. (2004). DC-SIGN and DC-SIGNR interact with the glycoprotein of Marburg virus and the S protein of severe acute respiratory syndrome coronavirus. *Journal of virology*, *78*(21), 12090-12095.

[12] Sugrue, R. J. (Ed.). (2007). *Glycovirology protocols* (Vol. 379). Springer Science & Business Media.

[13] Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C. L., Abiona, O., ... & McLellan, J. S. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*, *367*(6483), 1260-1263.

[14] Schultz, M. J., Swindall, A. F., & Bellis, S. L. (2012). Regulation of the metastatic cell phenotype by sialylated glycans. Cancer and Metastasis Reviews, 31(3-4), 501-518.

[15] Uemura, S., Kurose, T., Suzuki, T., Yoshida, S., Ito, M., Saito, M., ... & Inokuchi, J. I. (2006). Substitution of the N-glycan function in glycosyltransferases by specific amino acids: ST3Gal-V as a model enzyme. Glycobiology, 16(3), 258-270.

[16] Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., ... & Schnaar, R. L. Glycoinformatics--Essentials of Glycobiology.

[17] Taylor, M. E., & Drickamer, K. (2006). *Introduction to glycobiology*. Oxford university press, p17.

[18] Feinberg, H., Mitchell, D. A., Drickamer, K., & Weis, W. I. (2001). Structural basis for selective recognition of oligosaccharides by DC-SIGN and DC-SIGNR. *Science*, *294*(5549), 2163-2166.

[19] Cumming, D. A. (2003). Pathways and functions of mammalian protein glycosylation. *New Comprehensive Biochemistry*, *38*, 433-455.

[20] Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C., & Garry, R. F. (2020). The proximal origin of SARS-CoV-2. *Nature Medicine*, 1-3.

[21] Bagdonaite, I., & Wandall, H. H. (2018). Global aspects of viral glycosylation. *Glycobiology*, *28*(7), 443-467.

[22] Bretana, N. A., Lu, C. T., Chiang, C. Y., Su, M. G., Huang, K. Y., Lee, T. Y., & Weng, S. L. (2012). Identifying protein phosphorylation sites with kinase substrate specificity on human viruses. *PloS one*, *7*(7).

[23] Davidson, A. D., Kavangh Williamson, M., Lewis, S., Shoemark, D., Carroll, M. W., Heesom, K., Zambon, M., Ellis, J., Lewis, P. A., Hiscox, J. A., & Matthews, D. A. (2020). Characterisation of the transcriptome and proteome of SARS-CoV-2 using direct RNA sequencing and tandem mass spectrometry reveals evidence for a cell passage induced in-frame deletion in the spike glycoprotein that removes the furin-like cleavage site. BioRxiv, 2020.03.22.002204. https://doi.org/10.1101/2020.03.22.002204

[24] Blom, N., Sicheritz□Pontén, T., Gupta, R., Gammeltoft, S., & Brunak, S. (2004). Prediction of post□translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics*, *4*(6), 1633-1649.

[25] Hart, G. W., Kreppel, L. K., Comer, F. I., Arnold, C. S., Snow, D. M., Ye, Z., ... & Akimoto, Y. (1996). O-GlcNAcylation of key nuclear and cytoskeletal proteins: reciprocity with O-phosphorylation and putative roles in protein multimerization. *Glycobiology*, *6*(7), 711-716.

[26] Steentoft, C., Vakhrushev, S. Y., Joshi, H. J., Kong, Y., Vester□Christensen, M. B., Katrine, T., ... & Marcos□Silva, L. (2013). Precision mapping of the human O□GalNAc glycoproteome through SimpleCell technology. *The EMBO journal*, *32*(10), 1478-1488.

[27] Gupta, R., & Brunak, S. (2001, December). Prediction of glycosylation across the human proteome and the correlation to protein function. In *Pac Symp Biocomput* (Vol. 7, pp. 310-22).

[28] Gerken, T. A., Raman, J., Fritz, T. A., & Jamison, O. (2006). Identification of common and unique peptide substrate preferences for the UDP-GalNAc: polypeptide α-N-acetylgalactosaminyltransferases T1 and T2 derived from oriented random peptide substrates. *Journal of Biological Chemistry*, *281*(43), 32403-32416.

[29] Gerken, T. A., Jamison, O., Perrine, C. L., Collette, J. C., Moinova, H., Ravi, L., ... & Tabak, L. A. (2011). Emerging paradigms for the initiation of mucin-type protein O-glycosylation by the polypeptide GalNAc transferase family of glycosyltransferases. *Journal of Biological Chemistry*, *286*(16), 14493-14507.

[30] Blom, N., Gammeltoft, S., & Brunak, S. (1999). Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *Journal of molecular biology*, *294*(5), 1351-1362.

[31] Petersen, B., Petersen, T. N., Andersen, P., Nielsen, M., & Lundegaard, C. (2009). A generic method for assignment of reliability scores applied to solvent accessibility predictions. *BMC structural biology*, *9*(1), 51.

[32] Adamczak, R., Porollo, A., & Meller, J. (2004). Accurate prediction of solvent accessibility using neural networks–based regression. *Proteins: Structure, Function, and Bioinformatics*, *56*(4), 753-767.

[33] Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC bioinformatics*, *9*(1), 40.

[34] Kirschner, K. N., Yongye, A. B., Tschampel, S. M., González□Outeiriño, J., Daniels, C. R., Foley, B. L., & Woods, R. J. (2008). GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *Journal of computational chemistry*, *29*(4), 622-655.

[35] Nivedha, A. K., Makeneni, S., Foley, B. L., Tessier, M. B., & Woods, R. J. (2014). Importance of ligand conformational energies in carbohydrate docking: Sorting the wheat from the chaff. *Journal of computational chemistry*, *35*(7), 526-539.

[36] Pierce, B. G., Wiehe, K., Hwang, H., Kim, B. H., Vreven, T., & Weng, Z. (2014). ZDOCK server: interactive docking prediction of protein–protein complexes and symmetric multimers. *Bioinformatics*, *30*(12), 1771-1773.

[37] Snyder, G. A., Colonna, M., & Sun, P. D. (2005). The structure of DC-SIGNR with a portion of its repeat domain lends insights to modeling of the receptor tetramer. Journal of molecular biology, 347(5), 979-989.

[38] Pérez, S., Sarkar, A., Rivet, A., Breton, C., & Imberty, A. (2015). Glyco3D: a portal for structural glycosciences. In *Glycoinformatics* (pp. 241-258). Humana Press, New York, NY.

[39] Krissinel, E., & Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *Journal of molecular biology*, *372*(3), 774-797.

Figures and Tables

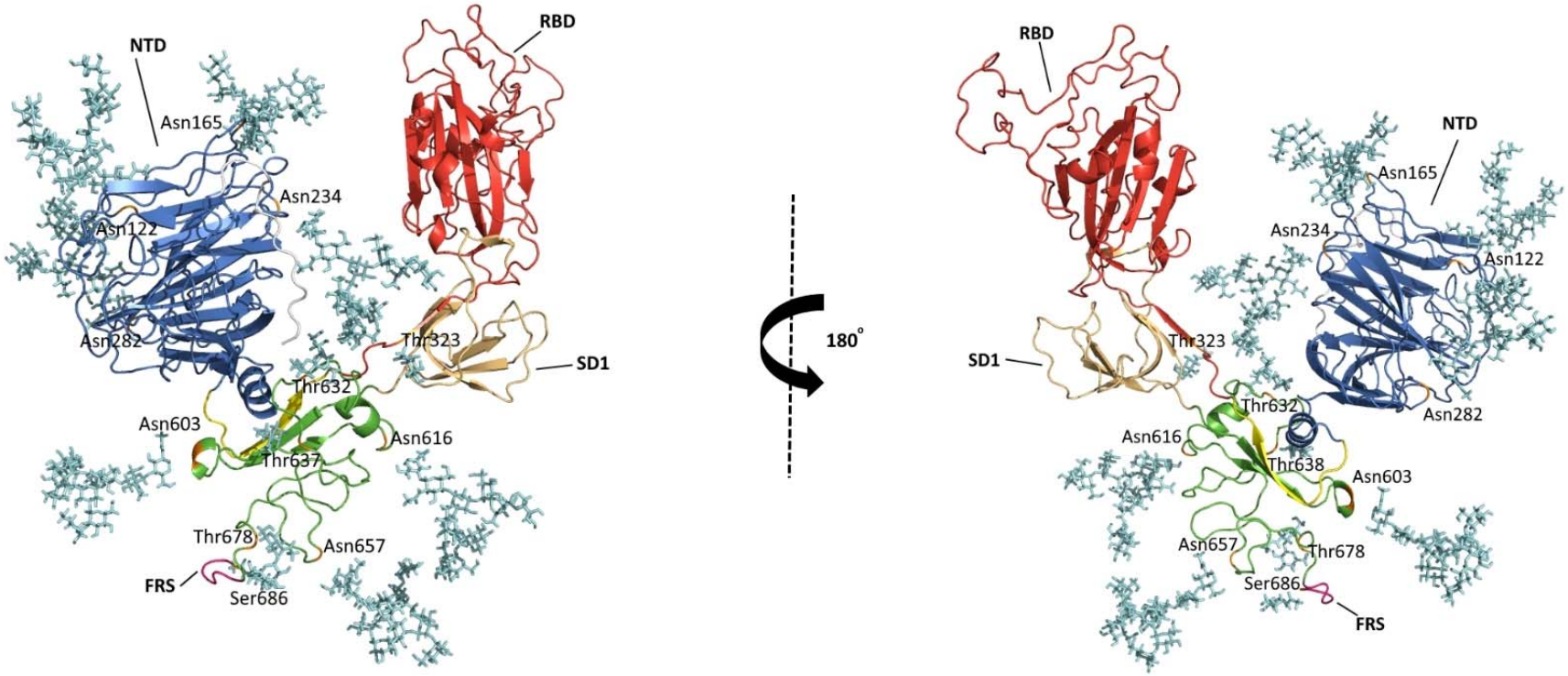

Figure 1. N- and O-glycosylated 3D glycoprotein structure of SARS-CoV-2 S1 protein. High mannose glycans of the N-glycosylation sites and core 1 type glycans of the O-GalNAcylation positions are shown in cyan color. N-glycosylation positions: Asn122, Asn165, Asn234, Asn282, Asn603, Asn616, Asn657; O-GalNAcylation positions: Thr632, Thr678; O-β-GlcNAcylation positions: Thr323, Thr638, Ser686. NTD: N-Terminal Domain, RBD: Receptor Binding Domain, FRS: Furin Recognizing Site. (The domain information was cited from Wong et al., 2004 and Wrapp et al., 2020.)

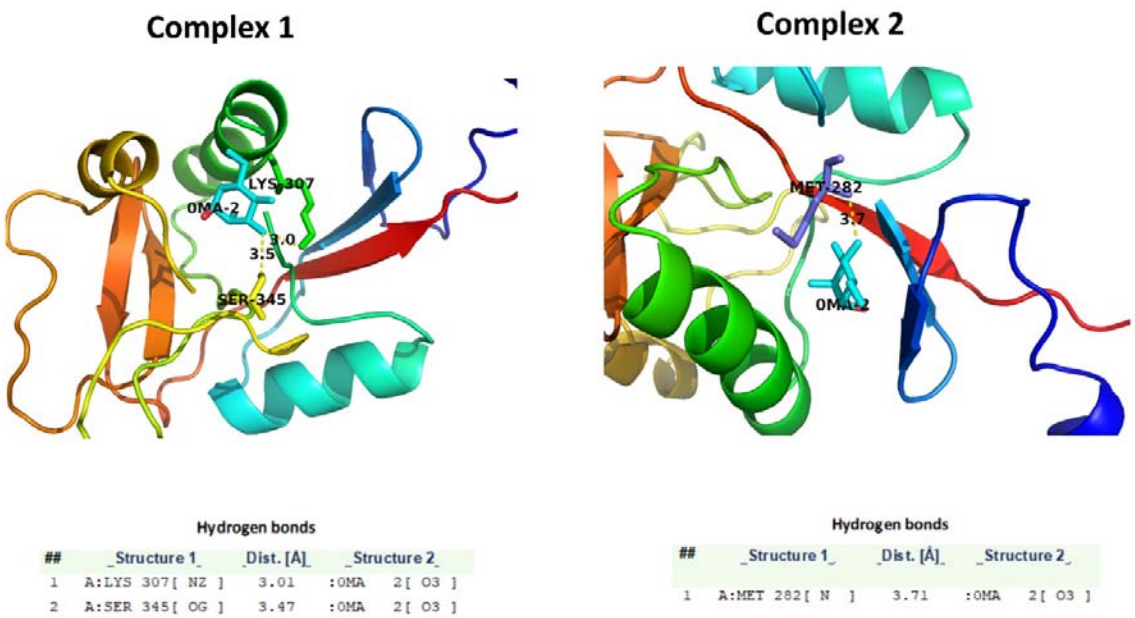

Figure 2. Receptor-ligand interactions of CD209L–α-D-mannose. ZDOCK and PDBePISA results showed that the 2 complexes have receptor-ligand interaction by hydrogen bonds at Lys307, Ser345, and Met282 positions. The mannose ligand is shown in cyan, receptor positions Lys307 in green, Ser345 in yellow and Met282 in blue color. All residues are shown in the stick representation.

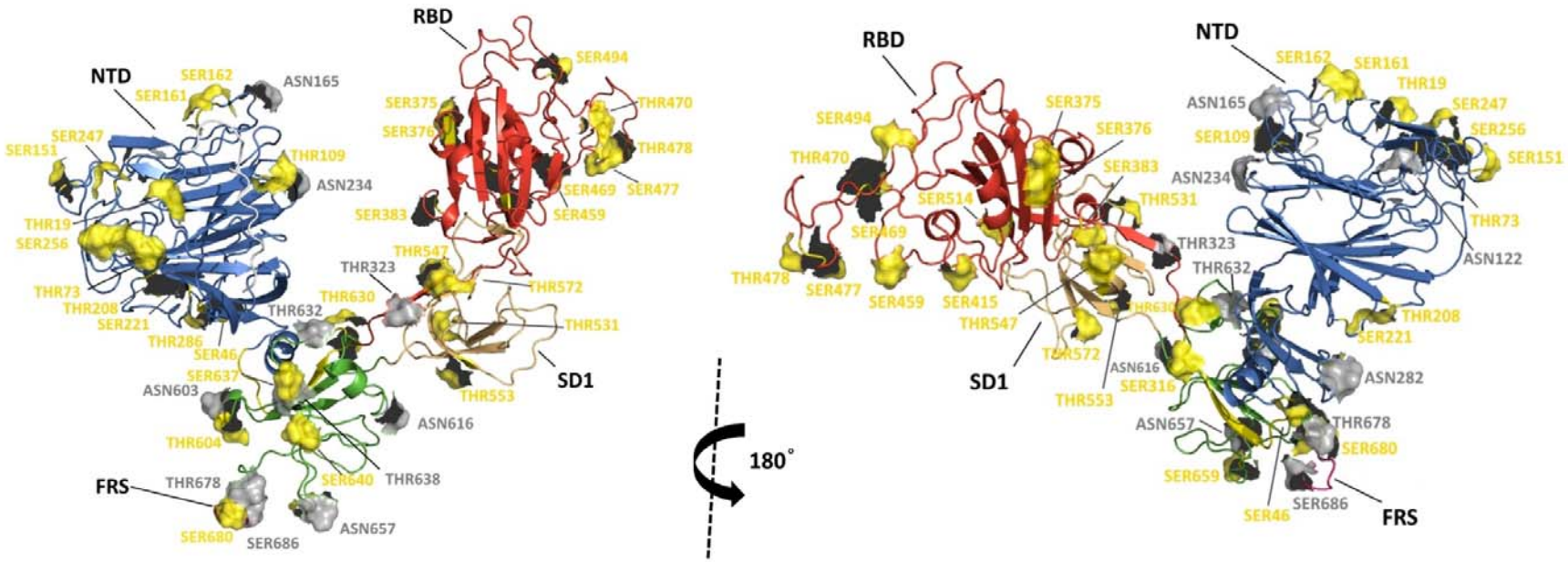

Figure 3. N-, O-glycosylation and Ser/Thr phosphorylation positions of SARS-CoV-2 S1 protein on the 3D structure. The glycosylation and phosphorylation positions are shown in grey and yellow, respectively. NTD: N-Terminal Domain, RBD: Receptor Binding Domain, FRS: Furin Recognizing Site

Table 1. The predicted O-glycosylation positions of SARS-CoV-2 S1 protein.

Glycosylation Position	NetOGlyc/YinOYang Score	ISOGlyP ppGalNAc Tranferase Isoforms Binding Value											Glycosylation Type
		T1	*T2*	*T3*	*T4*	*T5*	*T10*	*T11*	*T12*	*T13*	*T14*	*T16*	
		T/S ratio	*15:1*	*6:61*	*15:6*	*2:8*	*8:01*	*3:0*	*8:15*	*2:5*	*14:5*	*6:48*	*6:35*	
T323[]	0.5301	2.17	**6.96***	0.42	0.73	0.35	0.47	1.38	0.19	1.66	1.41	4.28	O-β-GlcNAc
T632[ζ]	0.5927	1.24	**2.87***	0.93	1.11	2.43	1.80	0.72	0.41	0.90	1.37	1.84	O-GalNAc
T638	0.5329	1.39	1.17	0.69	0.23	0.51	0.44	0.99	0.51	**1.48***	0.50	1.01	O-β-GlcNAc
T678	0.5458	1.34	0.44	1.03	1.94	2.20	0.23	1.51	**6.11***	1.28	0.34	0.60	O-GalNAc
S686[ς]	*0.6770*	*0.09*	*0.15*	*0.05*	*0.05*	*0.42*	*0.08*	*0.71*	*0.05*	***1.28****	*0.11*	*0.22*	*O-β-GlcNAc*

*** indicates the highest enzyme binding score between the transferase isoforms. [] indicates that position localized on Receptor Binding Domain and near the SD1 Variant Residue Glu324. ζ indicates that position is also phosphorylated. ς indicates that position is the first amino acid of S2 protein and localized near the FRS: Furin Recognizing Site (Protease Cleavage Site).**

Table 2. The predicted phosphorylation positions of SARS-CoV-2 S1 protein. All positions were analysed with the kinase binding, surface, and solvent accessibility parameters. The kinase specificity and protein localizations were also given in the table.

	NetPhos		NetSurfP			SABLE-POLYVIEW	Protein Localization
Amino acid Position	Score	Kinase	Buried or Exposed Surface	RSA	ASA	RSA*	
T19	0.503	cdc2	E	0.320	44.426	2	NTD
S46	0.957	unsp	B	0.098	11.486	2	NTD
	0.783	PKC					
	0.721	PKA					
T73	0.642	PKC	E	0.359	49.738	2	NTD
T109	0.799	PKC	B	0.196	27.227	1	NTD
	0.613	unsp					
	0.561	cdc2					
S151	0.985	unsp	E	0.390	45.731	3	NTD
	0.579	CKII					
S161	0.753	PKA	B	0.226	26.475	3	NTD
S162	0.557	PKA	B	0.298	34.984	4	NTD
T208	0.543	p38MAPK	E	0.329	45.618	3	NTD
S221	0.601	cdc2	E	0.312	36.555	3	NTD
S247	0.536	cdc2	E	0.351	41.196	2	NTD
S256	0.509	CKI	E	0.517	60.628	4	NTD
T286	0.549	CKII	B	0.170	23.565	2	NTD
S316	0.531	cdc2	E	0.443	51.931	2	Near RBD
S359	0.945	unsp	B	0.295	34.527	1	RBD
	0.802	PKA					
	0.514	cdc2					
S375	0.707	PKC	E	0.330	38.676	3	RBD
T376	0.846	PKC	E	0.379	52.526	3	RBD
S383	0.501	p38MAPK	E	0.341	39.942	2	RBD
T415	0.873	unsp	B	0.272	37.726	2	RBD
S459	0.895	unsp	E	0.445	52.119	3	RBD (Also VR)
	0.504	RSK					
S469	0.968	unsp	E	0.479	56.139	2	RBD

T470	0.515 CKII	B	0.230	31.929	2	RBD
S477	0.805 PKC	B	0.244	28.620	2	RBD
	0.510 cdc2					
T478	0.733 unsp	B	0.195	26.977	2	RBD (Also VR)
S494	0.538 PKC	E	0.280	32.757	4	RBD (Also VR)
S514	0.587 unsp	B	0.027	3.164	1	RBD
T531	0.625 PKC	B	0.314	43.566	2	SD1
T547	0.502 CKI	E	0.395	54.731	2	SD1
T553	0.521 CKII	E	0.299	41.444	4	SD1
T572	0.874 unsp	E	0.276	38.240	2	SD1
	0.621 cdk5					
	0.502 GSK3					
T604	0.723 unsp	E	0.549	76.146	4	SD2 VR
T630	0.564 p38MAPK	E	0.368	51.000	2	Near GP T632
	0.529 cdk5					
T632	0.859 PKC	B	0.172	23.926	1	Also O-Glycosylated
S637	0.967 unsp	E	0.473	55.459	4	Near GP T638
	0.579 PKA					
S640	0.501 cdc2	E	0.484	56.690	3	Near GP T638
S659	0.957 unsp	E	0.318	37.281	2	
	0.511 CKI					
S680	0.635 unsp	B	0.287	33.613	3	Near FRS and GP T678

B: Buried, E: Exposed, RSA: Relative Surface Accessibility, ASA: Absolute Surface Accessibility, RSA*: Score reliability increases from 1 to 9. NTD: N-terminal Domain, RBD: Receptor Binding Domain (ACE2), VR: Variant Residue, GP: O-Glycosylation Position, FRS: Furin Recognizing Site (Protease Cleavage Site).