

## Emerging WuHan (COVID-19) coronavirus: glycan shield and structure prediction of spike glycoprotein and its interaction with human CD26

Naveen Vankadari  and Jacqueline A. Wilce 

Monash Biomedicine Discovery Institute and Department of Biochemistry and Molecular Biology, Monash University, Clayton, Australia

### ABSTRACT

The recent outbreak of pneumonia-causing COVID-19 in China is an urgent global public health issue with an increase in mortality and morbidity. Here we report our modelled homo-trimer structure of COVID-19 spike glycoprotein in both closed (ligand-free) and open (ligand-bound) conformation, which is involved in host cell adhesion. We also predict the unique N- and O-linked glycosylation sites of spike glycoprotein that distinguish it from the SARS and underlines shielding and camouflage of COVID-19 from the host the defence system. Furthermore, our study also highlights the key finding that the S1 domain of COVID-19 spike glycoprotein potentially interacts with the human CD26, a key immunoregulatory factor for hijacking and virulence. These findings accentuate the unique features of COVID-19 and assist in the development of new therapeutics.

**ARTICLE HISTORY** Received 4 February 2020; Revised 1 March 2020; Accepted 1 March 2020



**KEYWORDS** Coronavirus; CD26; glycosylation; spike glycoprotein; docking


An outbreak of potentially lethal coronavirus (named COVID-19) in Wuhan, China, is spreading globally and impacting millions of people geographically linked with the Chinese population [1]. Current evidence suggests that the virus originated from wild animals and birds (<https://www.cdc.gov/coronavirus/>) [2]. To date, more than 2,800 deaths and 87,000 confirmed positive cases have been reported around the world, making COVID-19 a major health concern. As a first line of treatment, along with the antiviral drugs, clinicians are using SARS-CoV and MERS-CoV neutralizing antibodies targeting the S1 domain of the COVID-19 spike glycoprotein [1]. Very recently (25 January 2020) the first and complete genome sequence of COVID-19 was deposited in the NCBI (GenBank: MN908947.3) providing the key to the likely structure and glycosylation pattern of the viral proteins and consequent mode of interaction with the host cell. Similar to most other coronaviruses, the outer membrane spike glycoprotein, known for its glycosylation [3], is the prime host interacting protein with host cell targets (such as ACE2, CD26, Ezrin, cyclophilins and other cell adhesion factors) important for cell adhesion and virulence [4,5]. However, the specific host cell factors or proteins that facilitate the novel COVID-19 remain elusive. The current study was thus undertaken to predict the COVID-19 spike glycoprotein structure and glycan shield pattern that has great implications for

understanding the viral camouflage and mode of cell entry, potentially assisting the development of new vaccines, antibodies, small-molecule drugs and screening of the human host targets.

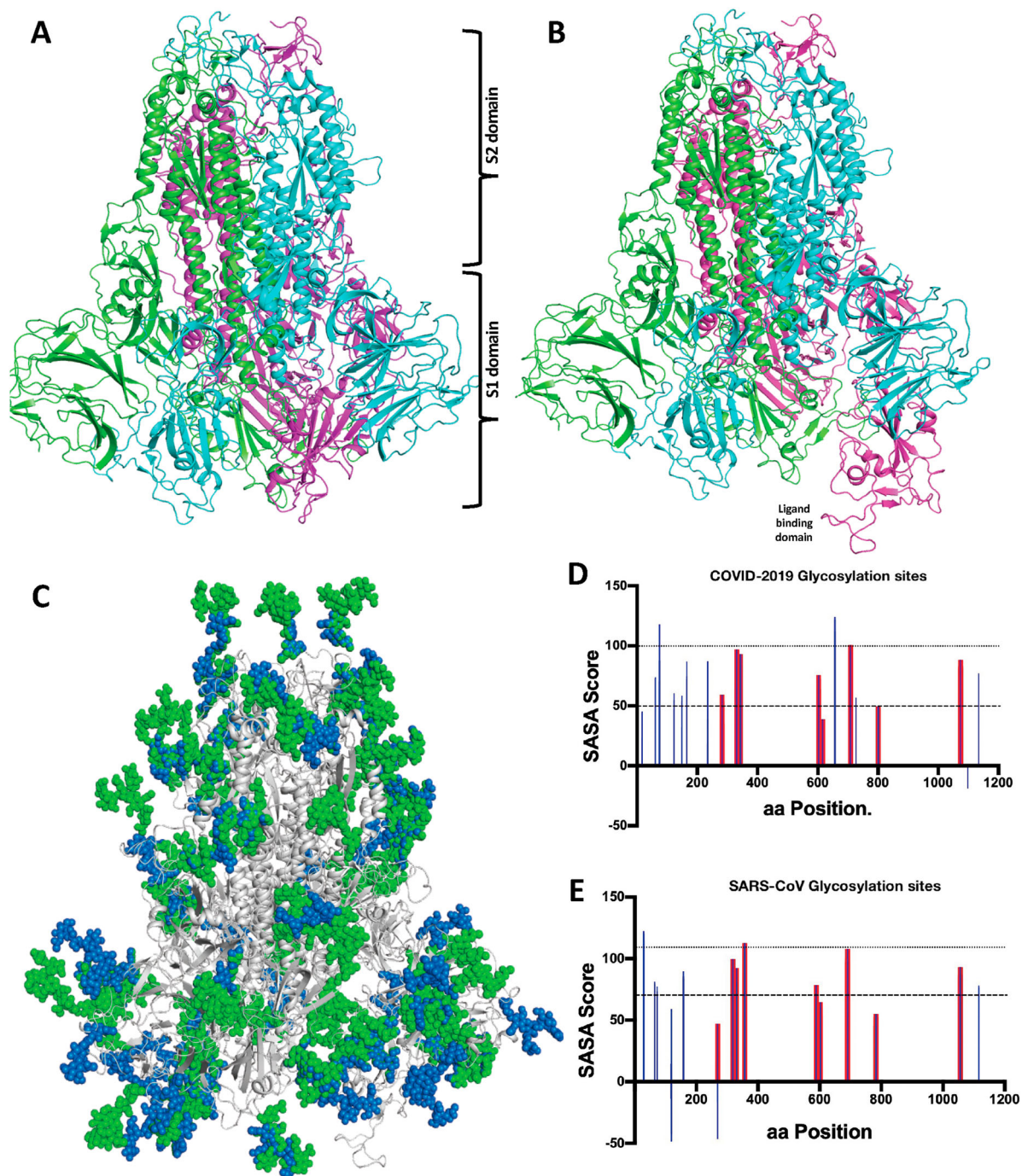
The Clustal-W sequence alignment of COVID-19 and SARS-CoV spike glycoproteins (Figure S1) shows ~91% identity in the S2 domain region (aa570–aa1278), however it lacks similarity in three regions (aa677–690, wing), (aa877–884 and aa930–943, stalk). A larger sequence difference (~55% identity), was found in the S1 domain (aa01–aa550), which is known for its host cell target interaction underlying cell adhesion and virulence [4,5]. Despite sequence dissimilarity in the S1 domains there are conserved residues involved in ternary folding which were conserved. This suggests that the COVID-19 might interact with some of the previously known host targets (ACE2, CD26, Ezrin, cyclophilins), albeit via slightly varied molecular interactions. Recent studies also support the possibility of COVID-19 and ACE-2 interaction [6].

To better understand the structure of COVID-19, including the position and orientation of unique residues involved in target binding, we modelled the homo-trimer structure of COVID-19 S1 and S2 domains (spike glycoprotein) using SWISS-MODEL (<https://swissmodel.expasy.org/>) using the structure of SARS-CoV (PDB: 6ACD) [4]. This model was

**CONTACT** Naveen Vankadari  [Naveen.vankadari@monash.edu](mailto:Naveen.vankadari@monash.edu)  Monash Biomedicine Discovery Institute and Department of Biochemistry and Molecular Biology, Monash University, Clayton, VIC 3800, Australia

 Supplemental data for this article can be accessed <https://doi.org/10.1080/22221751.2020.1739565>

© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group, on behalf of Shanghai Shangyixun Cultural Communication Co., Ltd  
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

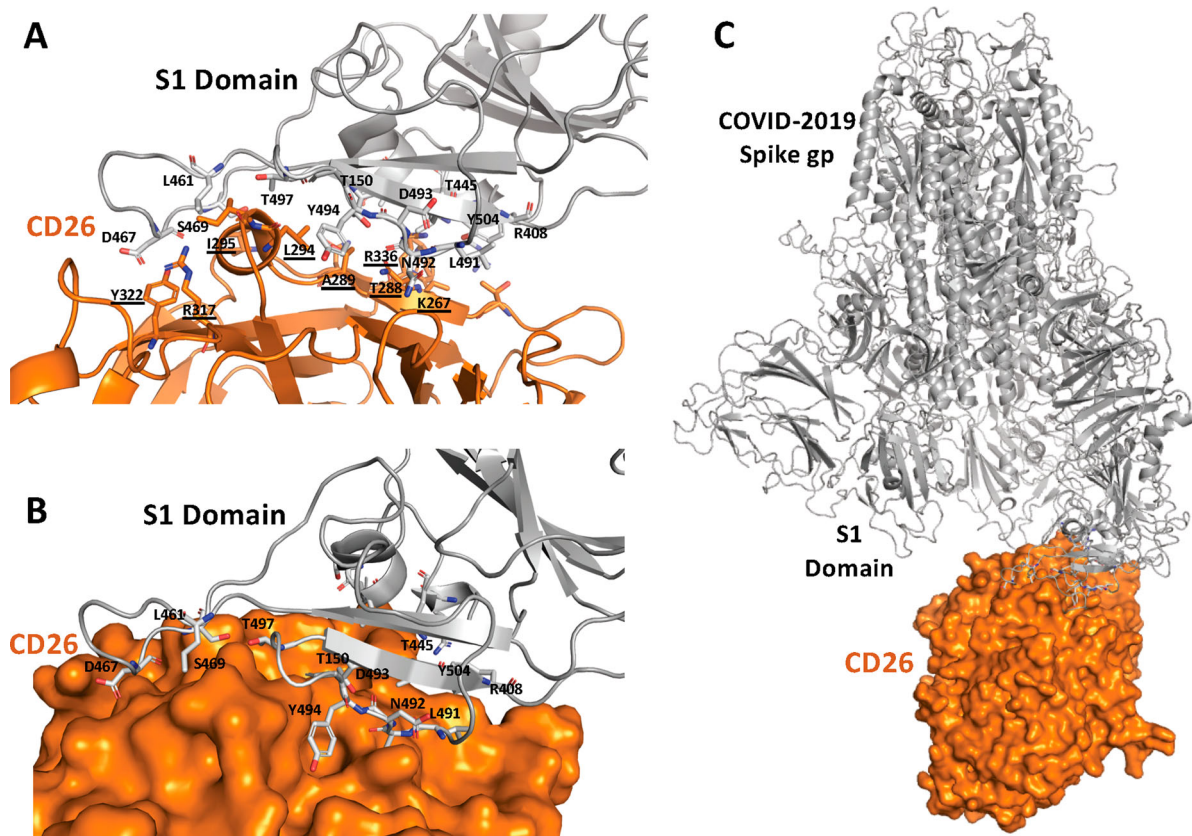


**Figure 1.** Overall homo-trimer model structure of the COVID-19 spike glycoprotein (A) ligand unbound conformation (B) ligand-bound conformation. The three protomers are coloured pink, green and cyan. S1- and S2- domains labelled. Receptor-binding induced hinge motion of S1 is distinguishable. (C) Predicted Glycan shield (spheres) of COVID-19 (green) and SARS-CoV (blue) spike glycoproteins. predicted 3C-like proteinase cleavage site (yellow). Predicted N-linked glycosylation sites for COVID-19 (D) and SARS-CoV (E). Unique glycosylation sites are coloured in *Blue*, and shared sites are shaded in *Red*.

validated using the C-Score (confidence score) and TM score (structural similarity) (Figure S2) demonstrating the most correct fold and confidence of the predicted structure. Further validation and refinement was completed by ensuring that the residues occupied Ramachandran favoured positions using Coot ([www.mrc-imb.cam.ac.uk/](http://www.mrc-imb.cam.ac.uk/)) (Figure S2). All amino acid residues were positioned according to their lowest energy possible orientation in the final model. The final modelled homotrimer structure of COVID-19 in C3 symmetry

(Figure 1(A)) superimposes with SARS-CoV with a  $0.85\text{\AA}$  C $\alpha$  RMSD and with a number of unique residues exposed on the surface COVID-19. A second modelled structure of COVID-19 spike glycoprotein, in ligand-bound conformation (Figure 1(B)) was also predicted based on the SARS-CoV/ACE2 complex structure (PDB:6ACG) [4]. This shows S1 domains in an open conformation, enabling it to interact with target host proteins. As is the case for other coronaviruses [7], we also identified 3C-like proteinase cleavage site





**Figure 2.** (A and B) Ribbon and a surface diagram showing the docking interface of modelled COVID-19 (grey) and human CD26 (orange)(PDB: 4QZV) complex. Predicted key residues involved in the interaction are shown in sticks (CD26 residues are underlined) (C) Overall docking results showing the surface model of CD26 with COVID-19 predicted homo-trimer structure (ligand-bound conformation).

(TGRLQ<sup>^</sup>SLQTY) (aa 997–1007) in COVID-19 spike glycoprotein using a server (<https://services.healthtech.dtu.dk/>). This 3C-like proteinase cleavage site could represent a site for drug discovery as currently being proposed for SARS-CoV [7].

To understand the glycosylation pattern and glycan shield of viral camouflage we used the (<https://services.healthtech.dtu.dk/>) and (<http://glycam.org/>) servers to predict N- and O-linked glycosylation sites on the surface of the modelled homo-trimer structure of COVID-19 spike glycoprotein and verified them according to their Solvent Accessible Surface Area (SASA) (Table S1). The spike glycoprotein trimer was then subjected to a surface glycosylation builder ([http://glycam.org/glycoprotien\\_builder/](http://glycam.org/glycoprotien_builder/)) for the predicted sites and visualized in PyMol. We also performed the same analysis for the SARS-CoV spike protein, to identify significant differences in glycosylation patterns (Figure 1(D,E)). The built glycosylation shield structures of COVID-19 and SARS-CoV spike glycoproteins were superimposed and are shown in Figure 1(C). As shown in Figure 1(C) and Table S1, there are a number of conserved glycosylation sites between these two viral strains, however there are also several unique glycosylation sites in COVID-19 compared to SARS-CoV spike glycoprotein. This suggests a different shielding or glycan camouflage pattern of the spike proteins,

which may underlie differences in host immunity. This leads to the intriguing question of whether COVID-19 could be responsive to a similar therapeutic approach to SARS [8].

Coronavirus trafficking into and hijacking the host cell is primarily driven by the N-terminal S1 domain of spike glycoprotein that interacts with several host cell proteins [4,5]. The host CD26 receptor cleaves amino-terminal dipeptides from polypeptides with either L-proline or L-alanine in the penultimate position, leading to T-cell activation and thus acting as a key immunoregulatory factor in viral infections [9]. Considering the current public health crisis, we considered the potential molecular interactions between COVID-19 spike protein and human CD26, with an interest to explore the structural differences or similarities between SARS-CoV and COVID-19 spike protein interactions. To this end, a computational model based selective docking was performed using the server Cluspro protein–protein docking ([Www.cluspro.bu.edu](http://www.cluspro.bu.edu)) and Frodock (<http://frodock.chaconlab.org/>) for further validation using our modelled 3D homotrimer structure of COVID-19 Spike glycoprotein (Figure 2) and the human CD26 (PDB: 4QZV) [10]. The binding free energies were taken into consideration for selecting the best possible model. The final rigid docked complex structure was

compared with the initial full-length COVID-19 spike glycoprotein and CD26 and their overall RMSD's were found to be 1.34 and 0.28 Å for Ca atoms, respectively.

The docked complex model of COVID-19 spike glycoprotein and CD26 (Figure 2) shows a large interface between the proteins. This suggests a possible tight interaction between the S1 domain loops in the modelled structure and the CD26 surface. Previous studies of CD26 binding have shown that residues K267, T288, A289, A291, L294, I295, R317, Y322 and D542 interact with Bat-CoV (MERS) spike protein [10]. Interestingly our docked model supports this despite the variability between these spike proteins' S1 domains, with the same CD26 residues in close proximity to the active region of S1 domain in COVID-19. We also observed additional residues (Q286, I287, N338, V341, R336) of CD26 predicted to interact with the S1 domain of the spike protein via van der Waals or by hydrogen bonding. However, regarding the COVID-19 spike glycoprotein, we noticed many different and unique residues (R408, Q409, T445, V417, L461, D467, S469, L491, N492, D493, Y 494, T497, T150, Y504) predicted to interact with CD26. Some of these unique residues of S1 domain are also predicted interact with the ACE2 protein [6]. This underlines the novelty and uniqueness of COVID-19 and its interaction with human target proteins. This observation guides us to suggest that COVID-19 may share infection modes with that of SARS-CoV and MERS-CoV and that interactions with other targets also warrant investigation.

### Acknowledgements

We thank the Monash University Software Platform for licence access to the concerned software. I also acknowledge Joseph Polidano of Monash University for editing and proof reading the manuscript.

### Disclosure statement

No potential conflict of interest was reported by the author(s).

### Funding

This work was supported by the National Health and Medical Research Council of Australia with grant APP1161916 awarded to J.A.W.

### ORCID

Naveen Vankadari  <http://orcid.org/0000-0001-9363-080X>  
 Jacqueline A. Wilce  <http://orcid.org/0000-0002-8344-2626>

### References

- [1] Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*. 2020;395(10223):497–506.
- [2] Lu H, Stratton CW, Tang YW. Outbreak of pneumonia of unknown etiology in Wuhan China: the mystery and the miracle. *J Med Virol*. 2020;92(4):401–402.
- [3] Xiong XL, Tortorici MA, Snijder J, et al. Glycan shield and fusion activation of a deltacoronavirus spike glycoprotein fine-tuned for enteric infections. *J Virol*. 2018;92(4):e01628–17.
- [4] Song WF, Gui M, Wang X, et al. Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. *PLoS Pathog*. 2018;14(8):e1007236.
- [5] Millet JK, Kien F, Cheung C-Y, et al. Ezrin interacts with the SARS coronavirus spike protein and restrains infection at the entry stage. *Plos One*. 2012;7(11):e49566.
- [6] Xu X, Chen P, Wang J, et al. Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. *SCIENCE CHINA Life Sciences*; 2020;63(3):457–460.
- [7] Grum-Tokars V, Ratia K, Begaye A, et al. Evaluating the 3C-like protease activity of SARS-coronavirus: recommendations for standardized assays for drug discovery. *Virus Res*. 2008;133(1):63–73.
- [8] Watanabe Y, Bowden TA, Wilson IA, et al. Exploitation of glycosylation in enveloped virus pathobiology. *Biochim Biophys Acta Gen Subj*. 2019;1863(10):1480–1497.
- [9] Morimoto C, Schlossman SF. The structure and function of CD26 in the T-cell immune response. *Immunol Rev*. 1998;161:55–70.
- [10] Wang Q, Qi J, Yuan Y, et al. Bat origins of MERS-CoV supported by bat coronavirus HKU4 usage of human receptor CD26. *Cell Host Microbe*. 2014;16(3):328–337.