Impact of glycan cloud on the B-cell epitope prediction of SARS-CoV-2 Spike protein

René Wintjens1, Amanda Makha Bifani2 and Pablo Bifani1,3,4,5

The SARS-CoV-2 outbreak originated in China in late 2019 and has since spread to pandemic proportions. Diagnostics, therapeutics and vaccines are urgently needed. We model the trimeric Spike protein, including flexible loops and all N-glycosylation sites, in order to elucidate accessible epitopes for antibody-based diagnostics, therapeutics and vaccine development. Based on published experimental data, six homogeneous glycosylation patterns and two heterogeneous ones were used for the analysis. The glycan chains alter the accessible surface areas on the S-protein, impeding antibody-antigen recognition. In presence of glycan, epitopes on the S1 subunit, that notably contains the receptor binding domain, remain mostly accessible to antibodies while those present on the S2 subunit are predominantly inaccessible. We identify 28 B-cell epitopes in the Spike structure and group them as non-affected by the glycan cloud versus those which are strongly masked by the glycan cloud, resulting in a list of favourable epitopes as targets for vaccine development, antibody-based therapy and diagnostics.

npj Vaccines (2020) 5:81; https://doi.org/10.1038/s41541-020-00237-9

INTRODUCTION

A cluster of viral pneumonia emerged in Wuhan, Hubei Province, China in December 2019. Shortly after, the aetiologic agent was determined to be a novel coronavirus now referred to as Severe Acute Respiratory Syndrome 2 (SARS-CoV-2) which causes Corona Virus Disease 2019 (COVID-19). Since its emergence, SARS-CoV-2 has spread across the globe, establishing local outbreaks in over 200 countries. Close to 17.5 million reported infections and over 675,000 death have been reported as of August 1st, and these numbers are likely greater due to insufficient testing in parts of the world and asymptomatic carriers.

SARS-CoV-2 is a member of the betacoronavirus genus with a single stranded RNA genome. Encoded in the genome is the immunodominant Spike protein (S). The S-protein is translated as a single protein and subsequently cleaved into two subunits. Three S-proteins assemble on the surface of the virion to form a trimeric spike. The S1 subunit is involved in receptor recognition and binding, while the S2 subunit contains a fusion peptide and necessary machinery for fusion, typically observed in class 1 fusion proteins. The S-protein binds host receptor angiotensin converting enzyme 2 (ACE2) and mediates viral entry into the host cell, similar to 2002 SARS-CoV. The receptor binding domain (RBD) of SARS-CoV and SARS-CoV-2 shares 72% identity at the amino acid level. Thus, it is no surprise that some antibodies are species specific, while other cross-reactive antibodies between SARS-CoVs have also been reported. However, SARS-CoV-2 has a flexible loop which projects into the hydrophobic pocket of ACE2, conferring a stronger interaction with the receptor than observed for SARS-CoV. While this S-protein protrusion out of the viral envelop facilitates receptor recognition and binding, it also leaves the protein exposed to recognition by the host immune system.

Consequently, the S1-subunit epitopes of SARS-CoV and SARS-CoV-2 are target candidates for induction of a B-cell response. A candidate monoclonal antibody isolated from SARS-CoV patients during the 2002 outbreak can cross-neutralise SARS-CoV2. Accordingly, potential SARS-CoV-2 epitopes have been suggested based on homologous regions of the SARS-CoVs S-protein that are known to be immunogenic. Strictly relying on known SARS-CoVs epitopes shared with SARS-CoV-2 limits the selection of antigenic sequences to non-SARS-CoV-2 specific epitopes. A limitation of these published studies is the failure to consider the impact of glycosylation, which could shield some of the selected epitopes.

As the number of COVID-19 cases continues to increase around the world, there is an urgent need for diagnostics, therapeutics and vaccines. This may be hampered by heavy glycosylation of the SARS-CoV-2 S-protein trimer, hindering exposure to the host adaptive immune system. With the critical need for a vaccine, it is crucial to identify epitopes that are conserved, accessible and unrestricted by the glycan chains to elicit a robust neutralising antibody response from the host. Thus, we aimed to identify epitopes, which are readily available to the host immune system. We identify various epitopes in the S-protein, considering the protein structure, the flexible loops and stearic hindrance resulting from protein glycosylation in the hope that this may advance vaccine development and antibody-based therapies and diagnostics.

RESULTS

N-glycosylation sites in the SARS-CoV-2 S-protein
Twenty-two experimentally determined N-glycosylation sites were evaluated to differentiate SARS-CoV-2 S-protein antigenic regions exposed to B cells from epitopes that are shielded by a glycan chain. The 22 N-glycosylation positions are unevenly distributed across the two S-protein subunits, accounting for 66 sites in the trimeric structure (Fig. 1a and Supplementary Fig. 1). A 3D
A glycosylation pattern of the trimeric spike with the flexible loop regions

We sought to build a glycosylation pattern model of the SARS-CoV-2 S-protein to identify accessible antigenic epitopes that were not masked with glycosylation by (i) retrieving the trimeric S-protein from the protein data bank (PDB) (PDB id 6VXX; 2.80 Å resolution; segment 27-1147) \(^{14}\) (ii) reconstructing the flexible loops not present in the starting structure and (iii) model glycan chains attached on the N-linked glycosylation sites in order to be able to (iv) compute the solvent-accessible surface area (SASA) and antibody-accessible surface area (AASA) and (v) predict the B-cell epitopes and glycan masking (Fig. 2a). Six different N-glycan types were considered in our study to model homogenous glycosylation patterns, including three high-mannose N-glycan types (Man3, Man5, and Man9), one hybrid type (Hyb8), and two complex types (NAc(4)Man(3)Fuc(1)Gal(2)Neu(2) (Complex-12) and NAc(6)Man(3)Fuc(1)Gal(4)(Neu(3) (Complex-15)) were also included, the former of which was frequently identified in experimental data and the latter glycan served to illustrate the effects of a large glycan chains on epitope recognition restriction\(^{11-13}\) (Fig. 2b). Two additional heterogeneous glycosylation patterns were included, based on Watanabe et al.\(^{12}\) (pat1) and Shajahan et al. (pat2)\(^{11}\). However, it is important to note that glycosylation patterns and glycan species can vary in different cell types\(^{15}\) and are altered in malignant cells compared to non-malignant cells\(^{16}\). As these previous S-protein glycosylation studies were performed in HEK293 cells\(^{11,12}\), which were immortalised by adenovirus, it is possible that the glycosylation pattern may not reflect a natural infected cell. Consequently, we chose to additionally include common glycosylation types found in proteins of healthy human cells. Noteworthy, these sugars are highly flexible dynamic structures which adopt various conformations and hence, rendering a single structural arrangement misleading when representing an in vivo structure. Here, we chose to use two different conformations to model each of the glycan chains.

Profiling accessibility to the surface area

Eighteen glycosylation sites almost fully decorate the SARS-CoV-2 S-protein in glycan chains, creating a glycan cloud (Supplementary Fig. 3). The accessible surface area (ASA) was calculated in order to determine which protein regions remain accessible to antibody
binding. Succinctly, FreeSASA\textsuperscript{17} was used to calculate the ASA for a de-glycosylated model, the cryo-EM structure, and the eight glycosylated models elucidated in this study (six homogenous N-glycan types and two composite patterns comprising each of ten models, i.e. five loop models times two glycan conformations). Firstly, the solvent-ASA (SASA) was calculated by using the probe radius of a water molecule to be used as reference. As expected, the SASA profile in the de-glycosylated and glycosylated models do not differ greatly (Supplementary Fig. 4). This is expected as water molecules can pass in between the glycan chains on the S-protein surface.

The antibody-ASA (AASA) was subsequently computed using a large probe radius comparable in size to the recognition domain of an antibody, allowing surface points available for protein-protein contacts, such as antibody-antigen\textsuperscript{18} to be identified. Unlike the SASA profiles, several domains of the S-protein were not accessible following glycosylation (Fig. 3a). Notably, the antibody accessibility of the S1 subunit (the NTD, subdomain 1 (SD1), subdomain 2 (SD2) and cleavage loop (CL)), were less affected by the glycosylation as was the case for the RBD and CL which were only partially obstructed by glycosylation (Fig. 3a). In contrast, except for the fusion peptide (FP), the AASA of all structural domains in the S2 subunit decreased rapidly in relation with the presence of glycans of increasing size, with the upstream helix (UH) and connecting region (CR) being the most affected domains and the central helix (CH) found to be completely inaccessible to antibodies even in the absence of glycans (Fig. 3a and Supplementary Table 1). The AASA of the total protein was estimated at 44% with a glycosylation pattern from Watanabe et al.\textsuperscript{12}, of which three domains, namely NTD, RBD and CL retained about 30% of the antibody accessibility, even in the presence of the large complex chain Complex-15 (Fig. 3a). While the conserved AASA for RBD and CL could result from their function as receptor binding and cleavage site respectively; a plausible function for the
protein-accessibility of NTD is less clear. Importantly, the AASA was not more impacted in the presence of longer glycan chains, likely due to the outwards orientation of the chains. These longer glycan chains project further outwards rather than folding back into the protein obstructing additional sites.

**B-cell epitope prediction and glycan shield impact**

Based on the AASA profile of the de-glycosylated model (Fig. 3b), we identified 28 protruding epitopes and evaluated their accessibility based on position and glycan shielding (Fig. 4). As expected most of epitopes were localised in the loop regions and
42% (12/28 epitopes) were found within NTD and RBD; the two critical domains for antibody recognition. In contrast, the epitopes in the S2 domain or the C-terminal close to the membrane were shielded by a glycan cloud. Accordingly, the AASA values of these latter epitopes were very low, dropping close to zero with the largest glycans (<10%) (Fig. 4). The 28 predicted B-cell epitopes were categorised into three groups based on the degree of accessibility lost upon glycosylation with sugar of increasing size: slightly shielded (epi2, epi3, epi5, and epi6, in the NTD; epi10, epi11 and epi12 in the RBD; and epi19 in the CL); moderately shielded (epi1, epi4 in the NTD; epi9 in the RBD, epi13 in the SD1; epi17 in the SD2; epi21 and epi23) or strongly masked (epi7, epi8, epi14, epi15, epi16, epi18, epi20, epi22, epi24, epi25, epi26, epi27 and epi28) (Fig. 3c). All the slightly and moderately shielded epitopes were localised within the S1 domain or N-terminal part of the protein, except epi21 and epi23. This underlines the importance of S1 domain for antibody recognition. It is critical to highlight that the loop spanning positions 828-853, was not modelled in this study and may likely harbour additional epitopes. Furthermore, conformational epitope (discontinuous epitopes), that form when distal parts of the amino acid sequence assemble in the tertiary structure, are found here comprising of the following clusters, epi1-epi2-epi-5-epi6, epi3-epi6, epi4-epi5, and epi10-epi12, among the epitopes less affected by glycosylation (Fig. 3c and Supplementary Table 2).

DISCUSSION

The COVID-19 pandemic has spread rapidly worldwide. It has exhausted health systems, burdened the global economy and brought many countries to a standstill. The international community is now looking to researchers to alleviate the impacts of SARS-CoV-2 through potential drugs, antibody therapy and most importantly, novel vaccine candidates. Here, we accounted for the N-glycosylation sites in different conformations on the SARS-CoV-2 trimeric S-protein in order to identify promising antibody accessible epitopes.

Recent published experimental data shows that the SARS-CoV-2 trimeric spike is highly glycosylated12,13. The degree of glycosylation is a key feature to consider when selecting potential B-cell epitopes and designing antibody-based diagnostics, therapeutics or vaccines as the number of accessible epitopes is limited. Noteworthy, is the differential distribution of these glycan along the protein. The S2 subunit glycosylation is more concentrated, especially towards the membrane while the S1 subunit contains more exposed domains possibly due to accessibility required for RBD and CL (Fig. 1a).

We determined that the epitopes in the S2 subunit become completely shielded by the glycan cloud, whereas this phenomenon is not observed with epitopes localised to the S1 subunit. This may be explained by the fact that the S2 subunit is a more compact structure, consisting mostly of a-helices (Fig. 1b), thereby generating a denser glycan cloud (Fig. 3c). The biological reasons why the S2 subunit is more protected by glycans remains to be determined. In a marked contrast, the subunit S1 has a more extended structure, with several B-sheets, creating a diffuse glycan cloud mitigating shielding effect of the epitopes (Fig. 3c).

We modelled the N-glycosylation pattern garnishing the S-protein in its trimeric conformation considering the flexible loop regions which were absent in the published experimental cryo-EM structure14. Our modelling approach, however has a few limitations. First, as glycan chains are dynamic structures, a single confirmation is insufficient to represent the complex structure. Thus, our model benefits from using two glycan chain conformations to account for the dynamic nature of N-glycosylation. Despite this approach, one has to bear in mind the flexibility of the glycan chains, as well as of that of the loops. In order to account the variability associated with the flexibility of the loops, here five different loop conformations have been considered in the analysis. A more realistic model based on molecular dynamics simulations would necessitate huge computational resources. Second, we based our model strictly on the closed structure (triple down protomers of the RBD) of the S-protein, rather than the open structure observed during receptor binding19. Nevertheless, this model still allowed us to capture a previously identified epitope S230-CoV-15,19 from the open conformation, that corresponds in part to the epi11, supporting the validity of our model.

Our model is in-part corroborated by recent reports of experimentally determined epitopes10,20-22 (Supplementary Table S5). For example, three neutralising monoclonal antibodies COV2-2130, COV2-2165 and COV2-2196 contain respectively the following key residues K444 and G447 (epi10), N487 and F486 (epi11), and N487 found in epi11 as well17. Likewise, 15/17 epitopes reported by Ravichandran and colleagues were identified in our work21. Another study using sera from COVID-19 convalescent patients led to the identification of two immunogenic peptides derived from the SARS-CoV-2 S-protein24. These two immunogenic peptides, S14P15 (region 553-570, SD1, partially impacted by glycan shield), and S21P2 (region 809-826, FP, moderately impacted by glycan shielding), correspond to epi14 and epi21, respectively, described in this study (Fig. 4). Other previously reported structures of the SARS-CoV S-protein RBD in complex with the three antibodies, namely 80R (PDB id 2GHW)25, m396 (PDB id 2DD8)26 and F26G19 (PDB id 3BGF)27 show the involvement of epitopes corresponding to SARS-CoV ep10 to epi12 antibody recognition sites found here. The partial match between our RBD epitopes and the antibody-binding regions in SARS-CoV RBD emphasises the dissimilarities of the two SARS-CoVs RBDs as confirmed by the poor cross-reactivity of these three antibodies28. Nevertheless, cross-reactivity has been reported with the monoclonal antibody S309 derived from memory B-cells dark from a 2003 SARS-CoV convalescent patient that recognises discontinuous epitopes10. Our model correctly identified fragments 333-346 and 440-441 (corresponding to epi9 and epi10) which are approximately 25 Å apart, but did not identify fragment 354-361 in the discontinuous epitopes recognised by antigen S309 (Supplementary Table 3). The majority of discontinuous epitopes are composed of 1-5 linear segments comprising of 1-6 amino acids. As such we have provided a distance matrix highlighting the space between the 28 epitopes identified here.
The model presented here is in agreement with others linear-based and discontinuous epitope predictions. Noteworthy, the interaction between RBD and ACE2 involves residues 445-456, 473-477, and 484-505 corresponding to our epi10, epi11 and epi11 + epi12, respectively. The glycan masking of epitopes may address why some patients’ sera is limited in its ability to prevent pseudovirus entry into host cells. A similar epitope masking phenomenon has been observed in related alphacoronavirus, NL63 by cryo-EM as well as human immune deficiency virus (HIV), where the receptor binding proteins are masked by heavy glycosylation shielding antigen exposure referred to as a glycan shield. De-glycosylation of sites on HIV’s glycoprotein gp120 results in enhanced immunogenicity and neutralising antibody production. Interestingly, a fraction of HIV positive patients are capable of developing antibodies to the glycan shield itself. Although, the SARS-CoV-2 S-protein is not as glycosylated as the HIV glycan shield, the glycan chains mask critical areas of the S-protein; hence the use of the term “glycan cloud”. It will be interesting to ascertain whether convalescent sera of SARS-CoV-2 individuals recognise the glycan cloud.

In conclusion, we constructed a 3D model of the SARS-CoV-2 trimeric S-protein, completed with the flexible loops and N-linked glycan chains. This structure successfully enabled protruding epitopes, unmasked by the glycan cloud to be elucidated. Such epitopes can serve as targets for antibody therapeutics or to be incorporated in the urgently needed vaccines.

METHODS

Study design

This study used a structure-based approach of the SARS-CoV-2 Spike protein, taking into account the glycan chains, with the objective of identify B-cell epitopes unshielded by glycosylation. A complete trimeric 3D model of the S-protein was first built and used a based on which several glycosylated models were generated. As glycosylation is a dynamic process, six different carbohydrate chains were used, assuming complete and homogeneous glycosylation site occupancy. In addition, two heterogeneous glycosylation patterns were included (see Supplementary Table 4). Epitopes were identified based on accessible surface area calculations in absence or in presence of the different glycan chains.

Fig. 4. Several predicted B-cell epitopes are partially or fully shielded by glycan chains. Figure depicting the 28 B-cell epitopes predicted based on the S-protein trimeric spike structure. Epitope region, length, position in the spike protein and 3D structure are shown. The percentage of antibody-accessible surface area of said epitopes under different glycosylation patterns is also presented. "CryoEM" is for the glycosylation pattern found in the initial S trimer structure and the labelling of other glycan chains are defined in Methods. The glycan chains are arranged in ascending order of their global SASA in the models (see Supplementary Table 5).
Building fully glycosylated trimer model of SARS-CoV-2 S-protein

The GlyProt tool (Glycosciences.DB portal40 (www.glycosciences.de)) was employed to attach carbohydrate chains at each N-linked glycosylation sites onto the full trimeric S-protein structure39. Six different glycans were considered for full homogeneously glycosylation patterns that are schematically depicted in Fig. 2b: three high-mannose type, one hybrid and two complex types (NAC(4)Man(3)Fuc(1)Gal(2)Neu(21) (Complex-12) and NAC(6)Man(3)Fuc(1)Gal(4)Neu(3) (Complex-15)). In this study, it was assumed that all possible N-glycosylation sites were occupied. A further two patterns of N-glycosylation were considered based on the most frequent glycan found at each N-site of experimentally determined glycosylation data. The numbers of glycan residues besides the common structure of two N-acetylgalactosamine moieties is also noted (see Fig. 2b). As glycans form a dynamic landscape on the surface of the protein, two different glycan conformations were modelled at each given glycosylation site. The first conformation extended the existing glycan in the orientation elucidated in the cryo-EM structure16, and the second conformation was built from a statistical evaluation of the occurrence of N-glycan orientations in the PDB39. Thus, ten models were proposed for each of the SARS-CoV-2 glycosylation patterns. All the 3D models are available from the authors upon request.

Accessible surface area (ASA) calculations

In order to predict potential epitope exposed to the immune system unshielded by glycosylation, the ASA was computed for each model as well as de-glycosylated model using FreeSASA according to the Lee & Richards algorithm17. Two different probe radiiues were considered for each model during ASA calculations. Firstly, the solvent-ASA (SASA) was determined using a 1.4 Å radius to model a water molecule. Subsequently, a 10 Å radius was use as this size is comparable to an antibody recognizes loop models and over the final models exhibit a suitable quality determined by MolProbity scores between 2.06-2.1630. Images of 3D model were produced with PyMOL (The PyMOL Molecular Graphics System, Version 2.3.0 Schrödinger, LLC).

DATA AVAILABILITY

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Received: 10 May 2020; Accepted: 20 August 2020;
Published online: 04 September 2020

REFERENCES

1. WHO. Coronavirus disease 2019 (COVID-19) Situation Report - 66 (WHO, Switzerland, 2020).
2. Mizumoto, K., Kagaya, K., Zarebski, A. & Chowell, G. Estimating the asymptomatic proportion of 2019 novel coronavirus onboard the Princess Cruises Ship. 2020. medRxiv https://doi.org/10.1101/2020.02.20.20025866 (2020).
3. Yan, R. et al. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. Science 367, 1444–1448 (2020).
4. Wan, Y., Shang, J., Graham, R., Baric, R. S. & Liu, F. Receptor recognition by novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS. J. Virol. 2020.05.22.111005 (2020).
5. Wrapp, D. et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science https://doi.org/10.1126/science.abb2507 (2020).
6. Ou, X. et al. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. Nat. Commun. 11, 1620 (2020).
7. Chen, Y., Guo, Y., Pan, Y. & Zhao, Z. J. Structure analysis of the receptor binding of 2019-nCoV. Biochim. Biophys. Res. Commun. https://doi.org/10.1016/j.bbr.2020.02.071 (2020).
8. Ahmed, S. F., Quadeer, A. A. & McKay, M. R. Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. Viruses 12, https://doi.org/10.3390/v12030254 (2020).
9. Grifoni, A. et al. A sequence homology and bioinformatic approach can predict candidate targets for immune responses to SARS-CoV-2. Cell Host Microbe https://doi.org/10.1016/j.chom.2020.03.002 (2020).
10. Pinto, D. et al. Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. Nature https://doi.org/10.1038/s41586-020-2349-y (2020).
11. Shajahan, A., Supekar, N. T., Gleinich, A. S. & Azadi, P. Deducing the N- and O-glycosylation profile of the spike protein of novel coronavirus SARS-CoV-2. Glycobiology https://doi.org/10.1093/glycob/ccaa042 (2020).
12. Watanabe, Y., Allen, J. D., Wrapp, D., McLellan, J. S. & Crispin, M. Site-specific glycan analysis of the SARS-CoV-2 spike. Science https://doi.org/10.1126/science.abh9983 (2020).
13. Zhang, Y. et al. Site-specific N-glycosylation characterization of recombinant SARS-CoV-2 spike proteins. bioRxiv https://doi.org/10.1101/2020.03.28.013276 (2020).
14. Walls, A. C. et al. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. Cell https://doi.org/10.1016/j.cell.2020.02.058 (2020).
15. Gugliotta, A. et al. Glycosylation and antigenic activity of hyperglycosylated IFN-alpha2a potentiate HEK293 cells as biofactories. Eur. J. Pharm. Biopharm. 112, 119–131 (2017).
16. Ruaahk, L. R. et al. Differential N-glycosylation patterns in lung adenocarcinoma tissue. J. Proteome Res. 14, 4538–4549 (2015).
17. Mitternacht, S. FreeSASA: an open source C library for solvent accessible surface area calculations. F1000Res 5, 189 (2016).
18. Novotny, J. et al. Antigenic determinants in proteins coincide with surface regions accessible to large probes (antibody domains). Proc. Natl Acad. Sci. USA 83, 226–230 (1986).
19. Walls, A. C. et al. Unexpected receptor functional mimicry elucidates activation of coronavirus fusion. Cell 176, 1026-1039 e1015 (2019).
20. Grifoni, A. et al. Targets of T cell responses to SARS-CoV-2 coronavirus in humans with COVID-19 disease and unexposed individuals. Cell https://doi.org/10.1016/j.cell.2020.05.015 (2020).
21. Ravichandran, S. et al. Antibody signature induced by SARS-CoV-2 spike protein bioRxiv https://doi.org/10.1101/2020.05.22.111005 (2020).
22. Zost, S. J. et al. Potent neutralizing antibodies against SARS-CoV-2 identified by high-throughput single-cell sequencing of convalescent patients’ B cells. Cell https://doi.org/10.1016/j.cell.2020.05.025 (2020).
23. Poh, C. M. et al. Two linear epitopes on the SARS-CoV-2 spike protein that elicit neutralizing antibodies in COVID-19 patients. Nat. Commun. 11, 2806 (2020).
24. Hwang, W. C. et al. Structural basis of neutralization by a human anti-severe acute respiratory syndrome spike protein antibody, 80R. J. Biol. Chem. 281, 34610–34616 (2006).
26. Prabakaran, P. et al. Structure of severe acute respiratory syndrome coronavirus receptor-binding domain complexed with neutralizing antibody. J. Biol. Chem. 281, 15829–15836 (2006).
27. Pak, J. E. et al. Structural insights into immune recognition of the severe acute respiratory syndrome coronavirus S protein receptor binding domain. J. Mol. Biol. 388, 815–823 (2009).
28. Tian, X. et al. Potent binding of 2019 novel coronavirus spike protein by a SARS coronavirus-specific human monoclonal antibody. Emerg. Microbes Infect. 9, 382–385 (2020).
29. Vashi, Y., Jagrit, V. & Kumar, S. Understanding the B and T cell epitopes of spike protein of severe acute respiratory syndrome coronavirus-2: a computational way to predict the immunogens. Infect. Genet. Evol. 84, 104382 (2020).
30. Baruah, V. & Bose, S. Immunoinformatics-aided identification of T cell and B cell epitopes in the surface glycoprotein of 2019-nCoV. J. Med. Virol. 92, 495–500 (2020).
31. Lan, J. et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature 581, 215–220 (2020).
32. Walls, A. C. et al. Glycan shield and epitope masking of a coronavirus spike protein observed by cryo-electron microscopy. Nat. Struct. Mol. Biol. 23, 899–905 (2016).
33. Wei, X. et al. Antibody neutralization and escape by HIV-1. Nature 422, 307–312 (2003).
34. Bolmstedt, A. et al. Enhanced immunogenicity of a human immunodeficiency virus type 1 env DNA vaccine by manipulating N-glycosylation signals. Effects of elimination of the V3 N306 glycan. Vaccine 20, 397–405 (2001).
35. Doores, K. J. The HIV glycan shield as a target for broadly neutralizing antibodies. FEBS J. 282, 4679–4691 (2015).
36. Dearlove, B. L. et al. A SARS-CoV-2 vaccine candidate would likely match all currently circulating strains. bioRxiv https://doi.org/10.1101/2020.04.27.064774 (2020).
37. Sali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234, 779–815 (1993).
38. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. Acta Crystallogr. Sect. D 60, 2126–2132 (2004).
39. Williams, C. J. et al. MolProbity: more and better reference data for improved all-atom structure validation. Proteins Sci. 27, 293–315 (2018).
40. Böhm, M. et al. Glycosciences.DB: an annotated data collection linking glycomics and proteomics data (2018 update). Nucleic Acids Res. 47, D1195–D1201 (2018).
41. Bohne-Lang, A. & von der Lieth, C.-W. GlyProt: in silico glycosylation of proteins. Nucleic Acids Res. 33, W214–W219 (2005).
42. Novotny, J. & Auffray, C. A program for prediction of protein secondary structure from nucleotide sequence data: application to histocompatibility antigens. Nucleic Acids Res. 12, 243–255 (1984).

ACKNOWLEDGEMENTS
R.W. is a Research Associate with the Belgian National Funds for Scientific Research (FRS–FNRS). A.M.B. is supported by graduate scholarships from Duke-NUS and Singapore Ministry of Education Tier 3 grant, Singapore. P.B. is supported by the National University of Singapore Start-up Funds, and the Singapore Immunology Network (SIgN); Agency for Science Technology and Research, Singapore. Schematics were generated using BioRender.

AUTHOR CONTRIBUTIONS
R.W. developed and performed the calculations. A.M.B., P.B. and R.W. analysed the data and contributed to the interpretation of the results. A.M.B. and R.W. designed and produced the figures. All authors drafted the original paper and all authors reviewed and edited the final manuscript.

COMPETING INTERESTS
The authors declare no competing interests.

ADDITIONAL INFORMATION
Supplementary information is available for this paper at https://doi.org/10.1038/s41541-020-00237-9.

Correspondence and requests for materials should be addressed to R.W. or P.B.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.