

In silico analysis of SARS-CoV-2 spike glycoprotein and insights into antibody binding

Victor Padilla-Sanchez ‡, §, |, ¶

‡ The Catholic University of America, Washington, United States of America

§ Washington Metropolitan University, Washington, United States of America

| Pittsburgh Supercomputing Center, Pittsburgh, United States of America

¶ Texas Advanced Computing Center, Austin, United States of America

Corresponding author: Victor Padilla-Sanchez (70padillasan@cua.edu)

Reviewed

v1

Received: 09 Jun 2020 | Published: 16 Jun 2020

Citation: Padilla-Sanchez V (2020) *In silico* analysis of SARS-CoV-2 spike glycoprotein and insights into antibody binding. Research Ideas and Outcomes 6: e55281. <https://doi.org/10.3897/rio.6.e55281>

Abstract

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) emerged in Wuhan, China in December 2019. Since then, COVID-19, the disease caused by SARS-CoV-2, has become a rapidly spreading pandemic that has reached most countries in the world. So far, there are no vaccines or therapeutics to fight this virus. Here, I present an *in silico* analysis of the virus spike glycoprotein (recently determined at atomic resolution) and provide insights into how antibodies against the 2002 virus SARS-CoV might be modified to neutralize SARS-CoV-2. I ran docking experiments with Rosetta Dock to determine which substitutions in the 80R and m396 antibodies might improve the binding of these to SARS-CoV-2 and used molecular visualization and analysis software, including UCSF Chimera and Rosetta Dock, as well as other bioinformatics tools, including SWISS-MODEL. Supercomputers, including Bridges Large, Stampede and Frontera, were used for macromolecular assemblies and large scale analysis and visualization.

Keywords

COVID-19, SARS-CoV-2, immunology, computational biology, *in silico*, structural biology, coronavirus, antiviral therapy

Introduction

COVID-19, the disease caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), presents with symptoms of fever, severe respiratory illness and pneumonia. As of April 15, 2020, SARS-CoV-2 has infected approximately 2,000,000 people, with more than 120,000 deaths. The virus is spreading exponentially worldwide and has been declared a public health emergency by the World Health Organization (WHO) (World Health Organization 2020). SARS-CoV-2 belongs to the genus Betacoronavirus and is closely related to several bat coronaviruses (Chen et al. 2020). It uses a spike glycoprotein (Chand et al. 2017), the structure of which has recently been determined at atomic resolution (PDB ID: 6VSB) to enter its target cell (Wrapp et al. 2020). The published structure shows that the spike protein has a receptor binding domain (RBD), which binds to the angiotensin-converting enzyme 2 (ACE2) receptor on host cells (Yan et al. 2020). The RBD can exist in two conformations, referred to as “up” (receptor-accessible) and “down” (receptor-inaccessible) (Wrapp et al. 2020). The *in silico* structural analysis showed that ACE2 and potential antibodies bind in the same location on the spike protein (Hwang et al. 2006, Sui et al. 2004). An antibody should thus be very effective in preventing viral spread by blocking the ACE2 binding site in the RBD. Two potent neutralizing human monoclonal antibodies, 80R and m396 (Prabakaran et al. 2006, Hwang et al. 2006), which bind to the RBD of SARS-CoV, the virus responsible for the 2002–2004 outbreak of severe acute respiratory syndrome, did not, however, bind to the RBD of SARS-CoV-2. Here, I report insights into sequence differences that affect the ability of 80R and m396 to bind to the SARS-CoV-2 RBD (Cuesta et al. 2010). Understanding why 80R and m396 do not bind to the SARS-CoV-2 spike protein could pave the way to engineering new antibodies that are effective against SARS-CoV-2. Mutated versions of the 80R and m396 antibodies could then be produced and administered as a therapeutic to fight the disease and prevent infection (Norman et al. 2019, Zhao et al. 2018). The ACE2 dimeric structure in complex with the SARS-CoV-2 RBD has also recently been determined at atomic resolution (PDB ID: 6M17) (Yan et al. 2020). I have now built a structural model of SARS-CoV-2 infection (Fig. 1), which provides further insights into antibody binding and should contribute to solving the complex problem of preventing viral spread (Pettersen et al. 2004).

Material and methods

Structural analysis was performed using UCSF Chimera software (Pettersen et al. 2004) on a desktop computer and on high-performance computers, including Bridges Large (Pittsburgh Supercomputing Center) and Stampede2 and Frontera (Texas Advanced Computing Center). Structural models were built using the SWISS-MODEL (Waterhouse et al. 2018) server at the University of Basel, Switzerland. Rosetta Dock and Snugdock software (Lyskov and Gray 2008, Lyskov et al. 2013, Sircar and Gray 2010, Weitzner et al. 2017) were used for docking experiments between the RBD of SARS-CoV-2 and *in silico*-mutated 80R and m396 antibodies according to their described protocols. PDBEPIA server was used to analyze the structure energies and residue-residue interactions while

the Therapeutic Antibody Profiler server was used for developability studies (Dunbar et al. 2016).

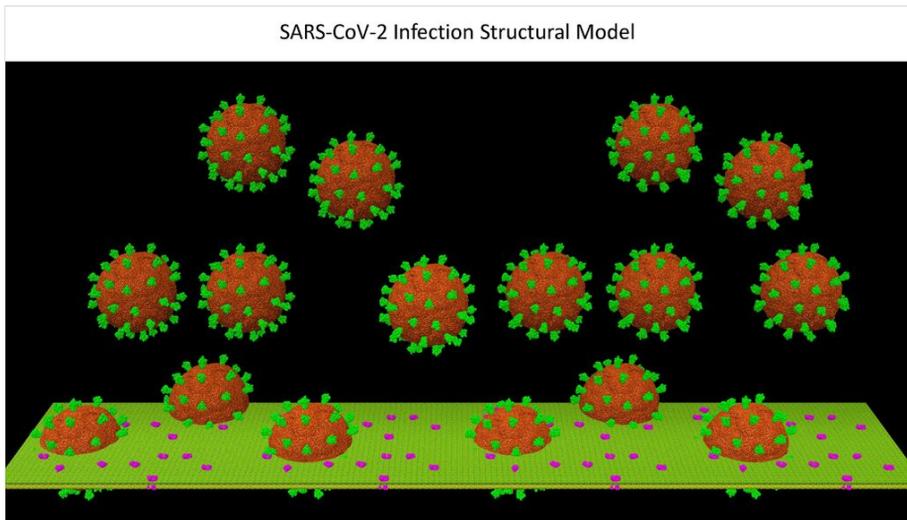


Figure 1. [doi](#)

Structural model of SARS-CoV-2 infection. This structural model was built with UCSF Chimera using high-performance computers (Bridges Large and Frontera). The model shows 16 viruses, with the spike proteins shown in green (PDB ID: 6VSB) and an actual lipid bilayer membrane, with ACE2 dimers shown in magenta. All these structures are at atomic resolution. The length of the membrane is approximately 1 micrometer.

Results

Analysis of SARS-CoV-2 spike glycoprotein

The tertiary structure of the SARS-CoV spike protein bound to ACE2 (Song et al. 2018, Wan et al. 2020), obtained from [RCSB](#) (PDB ID: 6ACG), and the tertiary structure of the SARS-CoV RBD bound to 80R (PDB ID: 2GHW) were superimposed, and the primary sequences were aligned (Fig. 2). A structural model for the SARS-CoV-2 RBD was constructed using SWISS-MODEL and superimposed by sequence alignment and tertiary structures (Fig. 2). In the secondary structures of the spike proteins, helices are shown in red, strands are shown in yellow and loops are shown in green; ACE2 is brown and 80R is cyan. I found that the 80R antibody and ACE2 bind in the same location, which perfectly explains how the antibody prevents infection by competing with ACE2. The structure of the SARS-CoV-2 spike protein (PDB ID: 6VSB) RBD has many loops missing (data not shown). Because these are essential for binding to ACE2, the model was constructed with a valid structure to run docking experiments. Furthermore, glycans in the PDB 6VSB show no interference with 80R binding even in predicted sites but not in the crystal structure (Fig. 2).

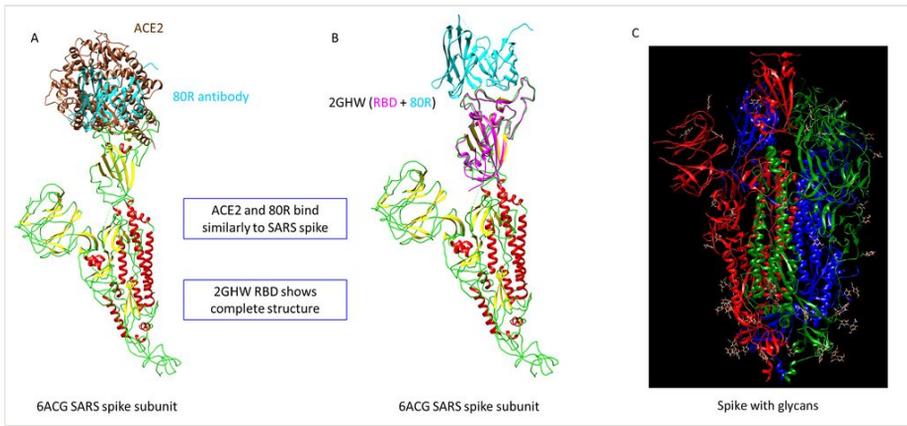


Figure 2. [doi](#)

Structural analysis of SARS-CoV spike glycoprotein. In A the SARS-CoV spike protein (PDB ID: 6ACG) is shown bound to ACE2 (brown) and 80R antibody (cyan), superimposed on the same binding site. In B the spike protein is shown bound only to the 80R antibody (PDB ID: 2GHW), with the structural model of the RBD of the SARS-CoV-2 spike protein (magenta) containing the missing loops. This homology model served as the basis for the docking experiments. In C it is shown a spike colored by subunit and showing the glycans. There are only two possible glycans in RBD region at 331 and 343 and neither of these sites affect the 80R binding.

Introduction of *in silico* mutations into the 80R antibody

Sequence alignments between SARS-CoV RBD and SARS-CoV-2 RBD were built in UCSF Chimera. There are many sequence differences between SARS-CoV and SARS-CoV-2 in the 80R-RBD interface, which explains why the 80R antibody binds with high affinity to the spike protein of SARS-CoV but not to the spike protein of SARS-CoV-2. In the SARS-CoV-2 RBD, polar residues are replaced by neutral residues, which disrupts the binding interactions between 80R and the RBD in the SARS-CoV-2 spike protein. Insertion of a glycine residue at position 482 also twists one of the interacting loops located at E484, which then clashes with the 80R antibody in the superimpositions. To avoid this clash and allow better antibody binding to the RBD in SARS-CoV-2, the 80R partner residue should be replaced by a different residue. I carefully selected six alternative residues, introduced these one at a time, and ran docking experiments using Rosetta Dock (Lyskov and Gray 2008, Lyskov et al. 2013) to explore the new interactions (Fig. 3). Aromatic-aromatic interactions between residues that are within 8 Å of each other are very important for protein structure and protein-protein recognition (Lanzarotti et al. 2011). According to the structural analysis in Chimera, there are many aromatic residues in the RBD of SARS-CoV-2 and I therefore introduced aromatic residues into 80R to match the RBD aromatic residues. My strategy was to thus increase aromatic-aromatic interactions, avoid the clash with E484 and maintain the solubility of the antibody by introducing substitutions, without introducing major changes in the 80R antibody. In support of my findings for the 80R

antibody epitope, an epitope for another SARS-CoV-2 antibody (CR3022) was recently reported to contain many aromatic residues (Yuan et al. 2020).

80R Mutant: R102F, S103F, R152F, S186F, T206G, S210F, N227F

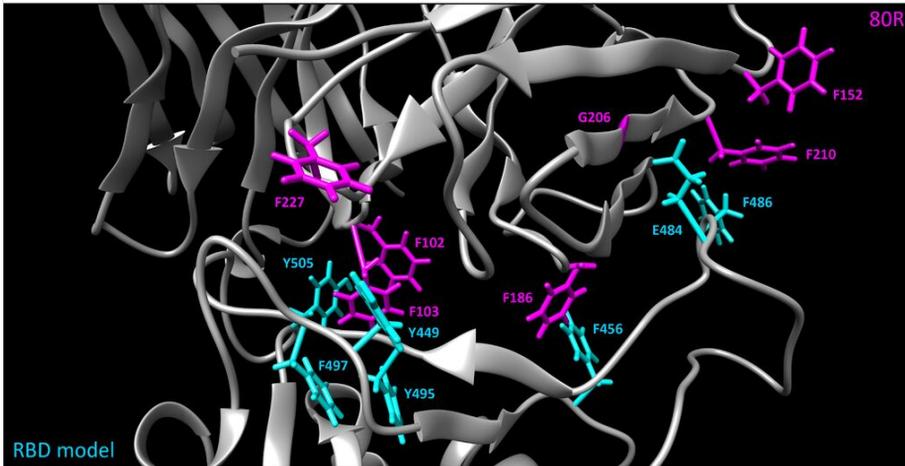


Figure 3. [doi](#)

Docking interface between the modified 80R antibody and the RBD of the SARS-CoV-2 spike protein. The model shows the structural interface with the 80R antibody above and the RBD below. The seven substitutions in 80R are shown in magenta and RBD residues are shown in cyan. Notice how the substitutions in 80R allow new aromatic-aromatic interactions that improve binding to the RBD and are not present in wild type 80R. E484 is shown pointing towards the beta strand of 80R and a glycine substitution was therefore introduced to avoid clashes.

I made a structural model of mutated 80R with SWISS-MODEL and positioned this model close to the RBD region. I used Rosetta Dock in Stampede2 supercomputer to run docking experiments with the following mutations in 80R R102F, S103F, R152F, S186F, T206G, S210F and N227F. These mutations were intended mainly to increase aromatic-aromatic interactions and to avoid contact with E484 (T206G) (Fig. 3). The following aromatic interactions were achieved: F102-F103-Y505, F186-F456, F152-F210-F486, and F227-Y449. These interactions increased the binding affinity between the two partners, as demonstrated by the funnel type charts from Rosetta Dock results (Fig. 4), which show how the structures converge to the lowest possible energy state compare to wild type. Furthermore, PDBEPISA (Krissinel and Henrick 2007) server was used to get the energies of the complexes with satisfactory results for the new mutated 80Rm. Finally, the Therapeutic Antibody Profiler (TAP) (Dunbar et al. 2016) tool was used to analyze the developability of this new mutated antibody and it shows all green flags for five important characteristics to take in consideration such as hydrophobicity and charges.

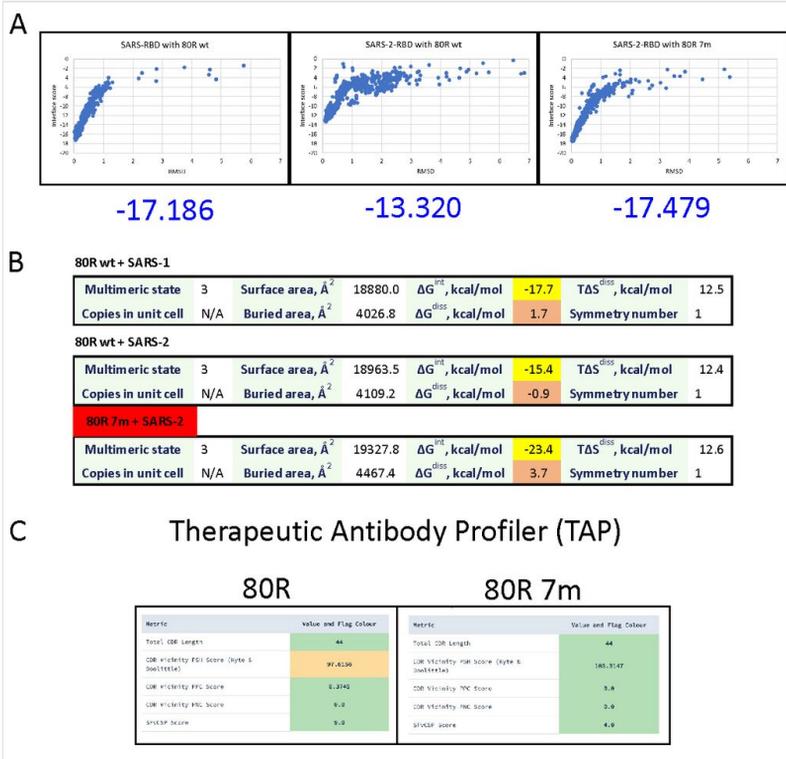


Figure 4. [doi](#)

Docking energies and interface score charts. **A** shows the Rosetta Dock results, binding energies from PDBEPIA server in **B** shows good results for ΔG . Developability of this antibody shows all green flags in **C**.

Introduction of *in silico* mutations into the m396 antibody

Similarly, m396 antibody that neutralizes SARS-CoV binding to its RBD domain was structurally analyzed and five mutations were introduced. In this case, the analysis pointed to electrostatic interactions to maximize and one aromatic. Therefore, five mutations were introduced: in the heavy chain T52F, I56E, N58K, Q61E; and in the light chain S94E. A structure with these mutations was put close to SARS-CoV-2 RBD and docking experiments were run in Stampede2. The results (Fig. 5) show that these mutations improve binding to the level of binding seen with SARS-CoV and wild type antibody (Fig. 6). m396 mutant shows formation of 9 new salt bridges which provide great binding to the target SARS-CoV-2 RBD (data not shown). Energy analysis from PDBEPIA server shows that binding is improved significantly compared to a wild type m396 and SARS-CoV-2 RBD. Furthermore, the developability flags show green even though one is yellow but just for a short margin which may be addressed after testing in wet lab experiments if it is an issue.

Discussion

Docking experiments showed that appropriate amino acid substitutions in 80R and m396 should increase binding interactions between the antibodies and the SARS-CoV-2 RBD, thus providing new antibodies with sufficient affinity for the SARS-CoV-2 spike protein to neutralize the virus. This new antibody should be expressed *in vitro* to study its solubility, stability, specificity and binding kinetic, and these results would be the basis for further mutations to correct some of these parameters or even improve the affinity. This methodology could be the basis for a rapid and effective generation of neutralizing therapeutic antibodies against COVID-19. *In silico* analysis is a very useful tool that structural bioinformaticians can use to guide mutagenesis to achieve a goal, in this case better affinity for the RBD of the SARS-CoV-2 spike protein. The results obtained using this relatively new branch of science must be taken with caution, but the method is becoming increasingly successful with the rapid improvement in bioinformatics tools. This type of analysis was not possible a decade ago, when scientists had to conduct mutagenesis experiments in wet laboratories. It is now possible for many scientists to use a bioinformatics approach to shorten the time needed to find new therapeutics. Further *in silico* experiments, including molecular dynamics simulations, can be performed to analyze interactions in real time, and in the future, I plan to conduct these types of experiment. Humanity is in a race to find therapeutics and/or vaccines against COVID-19 and my experimental analysis and findings should help the scientific community to quickly discover novel therapeutics.

Acknowledgements

Molecular graphics and analyses were performed with UCSF Chimera, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311. The author also wishes to acknowledge the Pittsburgh Supercomputing Center and Texas Advanced Computing Center for providing supercomputers (Bridges, Stampede2 and Frontera) for the *in silico* analyses.

Ethics and security

Only *in silico* tools were used.

Author contributions

Dr Victor Padilla-Sanchez, PhD designed, executed and wrote this research.

Conflicts of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Chand S, Messina E, AlSalmi W, Ananthaswamy N, Gao G, Uritskiy G, Padilla-Sanchez V, Mahalingam M, Peachman K, Robb M, Rao M, Rao V (2017) Glycosylation and oligomeric state of envelope protein might influence HIV-1 virion capture by $\alpha 5 \beta 7$ integrin. *Virology* 508: 199-212. <https://doi.org/10.1016/j.virol.2017.05.016>
- Chen Y, Liu Q, Guo D (2020) Emerging coronaviruses: Genome structure, replication, and pathogenesis. *Journal of Medical Virology* 92 (4): 418-423. <https://doi.org/10.1002/jmv.25681>
- Cuesta Á, Sainz-Pastor N, Bonet J, Oliva B, Álvarez-Vallina L (2010) Multivalent antibodies: when design surpasses evolution. *Trends in Biotechnology* 28 (7): 355-362. <https://doi.org/10.1016/j.tibtech.2010.03.007>
- Dunbar J, Krawczyk K, Leem J, Marks C, Nowak J, Regep C, Georges G, Kelm S, Popovic B, Deane CM (2016) SAbPred: a structure-based antibody prediction server. *Nucleic Acids Research* 44 (W1): W474-W478. <https://doi.org/10.1093/nar/gkw361>
- Hwang W, Lin Y, Santelli E, Sui J, Jaroszewski L, Stec B, Farzan M, Marasco W, Liddington R (2006) Structural basis of neutralization by a human anti-severe acute respiratory syndrome spike protein antibody, 80R. *Journal of Biological Chemistry* 281 (45): 34610-34616. <https://doi.org/10.1074/jbc.m603275200>
- Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology* 372 (3): 774-797. <https://doi.org/10.1016/j.jmb.2007.05.022>
- Lanzarotti E, Biekofsky R, Estrin D, Marti M, Turjanski A (2011) Aromatic–aromatic interactions in proteins: Beyond the dimer. *Journal of Chemical Information and Modeling* 51 (7): 1623-1633. <https://doi.org/10.1021/ci200062e>
- Lyskov S, Gray JJ (2008) The RosettaDock server for local protein-protein docking. *Nucleic Acids Research* 36 <https://doi.org/10.1093/nar/gkn216>
- Lyskov S, Chou F-C, Conchúir SÓ, Der BS, Drew K, Kuroda D, Xu J, Weitzner BD, Renfrew PD, Sripakdeevong P, Borgo B, Havranek J, Kuhlman B, Kortemme T, Bonneau R, Gray JJ, Das R (2013) Serverification of molecular modeling applications: The Rosetta Online Server That Includes Everyone (ROSIE). *PLOS One* 8 (5): e63906. <https://doi.org/10.1371/journal.pone.0063906>
- Norman RA, Ambrosetti F, Bonvin AMJJ, Colwell LJ, Kelm S, Kumar S, Krawczyk K (2019) Computational approaches to therapeutic antibody design: established methods and emerging trends. *Briefings in Bioinformatics* bbz095. <https://doi.org/10.1093/bib/bbz095>
- Pettersen E, Goddard T, Huang C, Couch G, Greenblatt D, Meng E, Ferrin T (2004) UCSF Chimera? A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* 25 (13): 1605-1612. <https://doi.org/10.1002/jcc.20084>
- Prabakaran P, Gan J, Feng Y, Zhu Z, Choudhry V, Xiao X, Ji X, Dimitrov D (2006) Structure of severe acute respiratory syndrome coronavirus receptor-binding domain

- complexed with neutralizing antibody. *Journal of Biological Chemistry* 281 (23): 15829-15836. <https://doi.org/10.1074/jbc.m600697200>
- Sircar A, Gray J (2010) SnugDock: Paratope structural optimization during antibody-antigen docking compensates for errors in antibody homology models. *PLOS Computational Biology* 6 (1): e1000644. <https://doi.org/10.1371/journal.pcbi.1000644>
 - Song W, Gui M, Wang X, Xiang Y (2018) Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. *PLOS Pathogens* 14 (8): e1007236. <https://doi.org/10.1371/journal.ppat.1007236>
 - Sui J, Li W, Murakami A, Tamin A, Matthews LJ, Wong SK, Moore MJ, Tallarico ASC, Olurinde M, Choe H, Anderson LJ, Bellini WJ, Farzan M, Marasco WA (2004) Potent neutralization of severe acute respiratory syndrome (SARS) coronavirus by a human mAb to S1 protein that blocks receptor association. *Proceedings of the National Academy of Sciences* 101 (8): 2536-2541. <https://doi.org/10.1073/pnas.0307140101>
 - Wan Y, Shang J, Graham R, Baric R, Li F (2020) Receptor recognition by the novel coronavirus from Wuhan: an Analysis based on decade-long structural studies of SARS coronavirus. *Journal of Virology* 94 (7). <https://doi.org/10.1128/jvi.00127-20>
 - Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TA, Rempfer C, Bordoli L, Lepore R, Schwede T (2018) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research* 46 (W1): W296-W303. <https://doi.org/10.1093/nar/gky427>
 - Weitzner BD, Jeliakov JR, Lyskov S, Marze N, Kuroda D, Frick R, Adolf-Bryfogle J, Biswas N, Dunbrack RL, Gray JJ (2017) Modeling and docking of antibody structures with Rosetta. *Nature Protocols* 12 (2): 401-416. <https://doi.org/10.1038/nprot.2016.180>
 - World Health Organization (2020) Coronavirus disease 2019 (COVID-19) Situation Report – 78. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/>. Accessed on: 2020-4-15.
 - Wrapp D, Wang N, Corbett K, Goldsmith J, Hsieh C, Abiona O, Graham B, McLellan J (2020) Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* 367 (6483): 1260-1263. <https://doi.org/10.1126/science.abb2507>
 - Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q (2020) Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* 367 (6485): 1444-1448. <https://doi.org/10.1126/science.abb2762>
 - Yuan M, Wu N, Zhu X, Lee C, So RY, Lv H, Mok CP, Wilson I (2020) A highly conserved cryptic epitope in the receptor-binding domains of SARS-CoV-2 and SARS-CoV. *Science* <https://doi.org/10.1126/science.abb7269>
 - Zhao J, Nussinov R, Wu W-J, Ma B (2018) In Silico methods in antibody design. *Antibodies* 7 (3): 22. <https://doi.org/10.3390/antib7030022>