Transition structures between unbound and bound SARS-CoV-2 spike protein to human ACE2: potential targets for drug design

Authors: N. Zanna[1]*

Affiliations:

[1]Independent Researcher, Italy

*Correspondence to: zanna.nicola1@gmail.com

Abstract:

The recent outbreak of COVID-19 caused by SARS-CoV-2 led to a race in finding a cure. Different drug targets were recognized, but in most cases the main target has been identified in the virus spike protein because it is crucial for the virus to gain entry into human cells. The virus spike protein undergoes large dynamic changes in order to bind to the entry point in human cells, which is a surface protein known as Angiotensin-Converting Enzyme 2 (ACE2). The spike – ACE2 interaction represents the major target for vaccines and antiviral drugs. Incidentally, all intermediate structures in the folding pathways could become potential drug targets. This study reports the simulation of the transition pathway of the spike protein and includes its animation that can help also non-experts to visually understand how the infection starts.

Introduction

The ongoing outbreak of COVID-19 caused by a novel coronavirus, Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), led to, as of 27 May 2020, up to 5 million confirmed cases and about 350,000 deaths according to the World Health Organization (WHO). As reported in detail by Wadman *et al*.(*1*) this coronavirus can attack almost anything in the body with devastating consequences, which explain the high mortality. This is because the virus is able to enter cells through the cell-surface receptor called Angiotensin-Converting Enzyme 2 (ACE2) (*2*, *3*) which is present in almost all cells and has an important role in blood pressure regulation.

One of the possible methods to stop the spread of infection into the body is to reduce the possibility of the virus to bind to the ACE2 receptors.

After the first reported genome sequence of SARS-CoV-2,(*4*) the main challenge was to obtain the 3D structures of all the key proteins involved in the infection, which could help to understand the infection mechanism and design new drugs. SARS-CoV-2 makes use of a homotrimeric spike glycoprotein (S) to gain entry into human cells by binding to the ACE2 receptors. For this reason, the main challenge has been the discovery of its 3D structure. Different approaches were used to obtain its 3D structure: either through experimental techniques (after the synthesis of a copy of the spike protein) or through computer simulations. The cryo-EM structure of the SARS-CoV-2 spike trimer has just been reported in two independent studies. (*5*, *6*) Like SARS-CoV-1, this protein exists in two states, referred as "up" and "down". The up state is the one involved in the interaction with ACE2, as in SARS-COV-1. (*7*)

The first studies devoted to rapidly finding a treatment were performed as target-based virtual ligand screening by using computational docking of a library of FDA approved drugs and a database of commercially available compounds to the receptor-binding domain (RBD) of the S protein. (8)

While most of these screens were done by using the spike protein in its up conformation as a target, it is important to note that there is a large conformational change between the two states. Intermediate structures between both conformations could potentially be targets for computational virtual ligand screening. For this reason, obtaining the intermediate structures by passing from the down to the up state is of great importance.

Results and Discussion

There are a lot of different computational techniques able to reproduce the large-scale conformational changes that occur in proteins. (9–11) The most used methods able to reproduce transition pathways between two different states make use of Molecular Dynamics (MD), Coarse Grained (CG) Modeling, Normal Mode Analysis (NMA), Elastic Network Model (ENM), and linear interpolation in Cartesian coordinates. In this list, sorted in ascending order of complexity and accuracy, it is important to remember that the first method is too computationally expensive for large structures, while the last one produces pathways that do not correspond at all to realistic transitions. The other methods, sometimes used in combination, are able to give accurate results at a reasonable simulation time. Several software packages using these techniques are readily accessible over web servers. In this study, intermediate states between the up and down conformations of SARS-CoV-2 spike protein were generated by targeted simulations with the NMSim approach. (12, 13) NMSim makes use of a three-step protocol including CG, NMA, and ENM methods, providing realistic intermediates at a reasonable simulation time. (14, 15)

First of all, an appreciable 3D structure is needed for both the up and down states. The only experimental structure of the entire spike protein, unfortunately, does not represent the up conformation well, as reported also by Lan *et al. (16)* The reason could be ascribed to the cryo-EM technique, which usually captures snapshots of the conformational intermediates that lie in the transition path from up to down states. Other structures of the spike protein were obtained by homology modelling. Some of these structures (the up state, in particular) were compared with the recently reported crystallographic structure (6M0J) containing the synthetic RBD of S interacting with the ACE2 receptor. (16) The best result was obtained with the structure predicted with the C-I-TASSER server by the Zhang group, which perfectly reproduces the interaction between S and ACE2 while superimposing the two structures. (17–19) It it important to note that the perfect superposition could be ascribed to the fact that homology modelling takes as reference structures with similar amino acidic sequence deposited in PDB databases as reference. Since 6M0J is the only structure available at this moment, it was probably used to reproduce the RBD interaction with ACE2.

To reduce the computational cost for the pathway simulation, two of the three chains forming the trimeric spike protein were removed from both up and down full S structures, leaving the only chain that undergoes conformational changes. This simplification does not affect the results because the other chains do not change conformations as demonstrated by the cryo-EM structures. To simplify the simulation we also remove the tails from the spike proteins starting from Glycine 1124 since this region is not functional and only serve to link with the viral envelope. The simulation with the NMSim web server was run with almost all the default parameters (full details in supporting materials), modifying only the number of intermediate

structures between the two states. After different attempts, it was found that despite the large displacement between the two structures 120 is the best number of intermediates necessary to capture all the different changes in conformation.

For a better visualization of the transition pathway movies have been created by using all the computed intermediate structures (Movie S1). The same was done to visualize the motion of the full S protein binding to ACE2, in this case the other two chains of the homotrimeric structure and the interacting chain of ACE2 were added and repacked without design. (Movie S2)

In both cases it is possible to visualize how the RBD, which in the down state is angled closer to the central cavity of the trimer, undergoes not only an "up" movement away from the cavity but also a torsion of about 45 degrees, exposing the amino acids interacting with ACE2. (Figure 1) The pathway can be observed also by overlaying the intermediate conformations as shown in Figure 2, indicating that the transition does not only displace the bulk RBD, but is also affects the amino acids binding to ACE2.

It is possible that some of these transition structures have higher binding energies with some known drugs. If so, a treatment that prevents the interaction between S and ACE2 could be designed. This treatment could be readily available, with its dosage and side effects being known as well.

In conclusion, using virtual screening as the methods reported, but using these intermediate structures as targets, it is possible to screen drugs that are able to block the first step of infection, and thus stop the spread of SARS-CoV-2.

In addition, the generated movies are useful to illustrate to a broad audience how the virus enters human cells to begin infection.
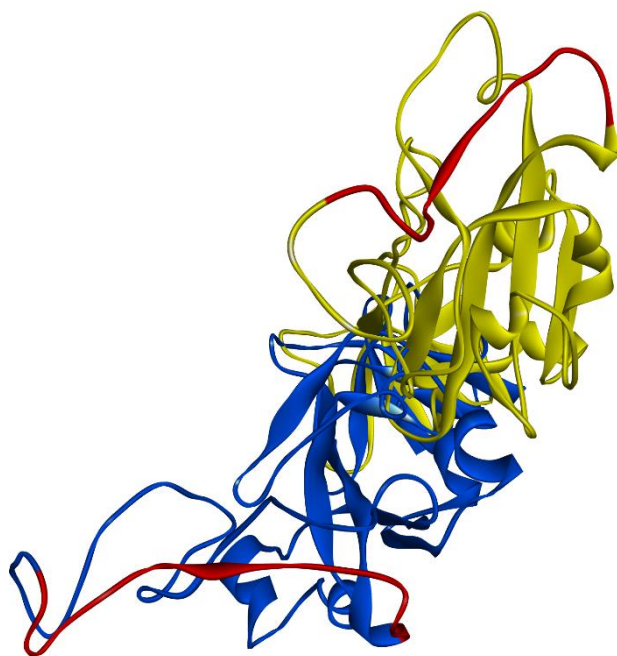

Fig. 1. Superimposed up (yellow) and down (blue) conformations of a single chain of the homotrimeric spike protein (top view). The ACE2-interacting RBD was depicted in red in both conformations, showing a torsion of about 45°.
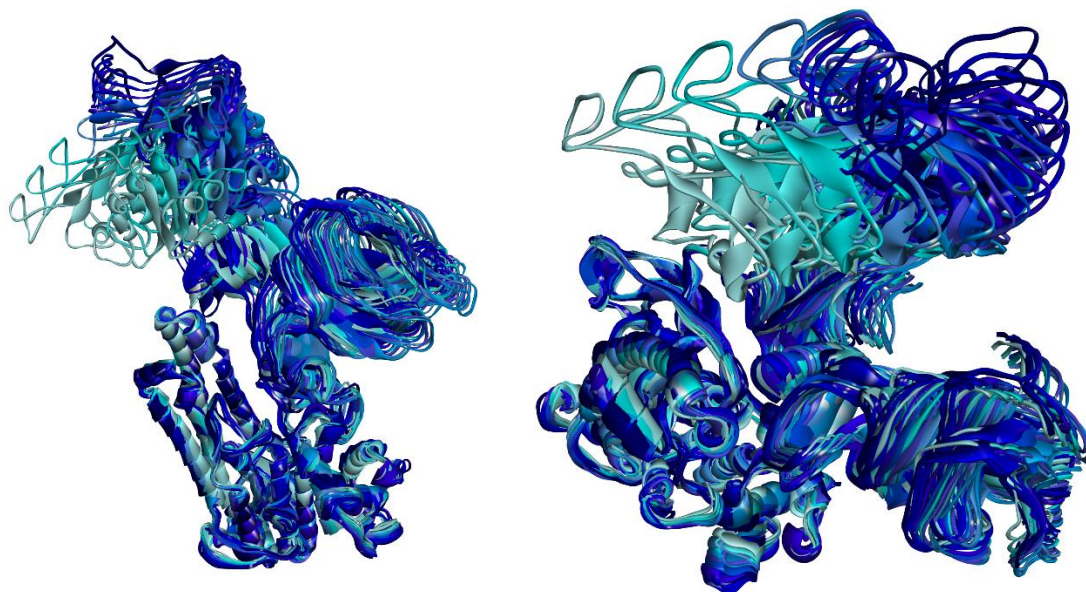
Fig. 2. Overlay of computed intermediates in the transition pathway. For clarity, only one in every ten of the 120 intermediate models were used. Light blue: down conformation; deep blue: up conformation. Left: side view; right: top view.

References and Notes:

1. M. Wadman, C.-F. Jennifer, K. Jocelyn, M. Catherine, How does coronavirus kill? Clinicians trace a ferocious rampage through the body, from brain to toes. *Science (80-.).* (2020), doi:10.1126/science.abc3208.

2. Y. Wan, J. Shang, R. Graham, R. S. Baric, F. Li, Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. *J. Virol.* **94**, 1–9 (2020).

3. X. Xu, P. Chen, J. Wang, J. Feng, H. Zhou, X. Li, W. Zhong, P. Hao, Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. *Sci. China Life Sci.* **63**, 457–460 (2020).

4. F. Wu, S. Zhao, B. Yu, Y. M. Chen, W. Wang, Z. G. Song, Y. Hu, Z. W. Tao, J. H. Tian, Y. Y. Pei, M. L. Yuan, Y. L. Zhang, F. H. Dai, Y. Liu, Q. M. Wang, J. J. Zheng, L. Xu, E. C. Holmes, Y. Z. Zhang, A new coronavirus associated with human respiratory disease in China. *Nature*. **579**, 265–269 (2020).

5. A. C. Walls, Y. J. Park, M. A. Tortorici, A. Wall, A. T. McGuire, D. Veesler, Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*. **181**, 281-292.e6 (2020).

6. D. Wrapp, N. Wang, K. S. Corbett, J. A. Goldsmith, C. L. Hsieh, O. Abiona, B. S. Graham, J. S. McLellan, Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science (80-.)*. **367**, 1260–1263 (2020).

7. R. N. Kirchdoerfer, N. Wang, J. Pallesen, D. Wrapp, H. L. Turner, C. A. Cottrell, K. S. Corbett, B. S. Graham, J. S. McLellan, A. B. Ward, Stabilized coronavirus spikes are resistant to conformational changes induced by receptor recognition or proteolysis. *Sci. Rep.* **8**, 1–11 (2018).

8. C. Wu, Y. Liu, Y. Yang, P. Zhang, W. Zhong, Y. Wang, Q. Wang, Y. Xu, M. Li, X. Li, M. Zheng, L. Chen, H. Li, Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. *Acta Pharm. Sin. B* (2020), doi:10.1016/j.apsb.2020.02.008.

9. L. Skjærven, N. Reuter, A. Martinez, Dynamics, flexibility and ligand-induced conformational changes in biological macromolecules: A computational approach. *Future Med. Chem.* **3**, 2079–2100 (2011).

10. S. Kmiecik, M. Kouza, A. E. Badaczewska-Dawid, A. Kloczkowski, A. Kolinski, Modeling of protein structural flexibility and large-scale dynamics: Coarse-grained simulations and elastic network models. *Int. J. Mol. Sci.* **19** (2018), doi:10.3390/ijms19113496.

11. L. Orellana, Large-Scale Conformational Changes and Protein Function: Breaking the in silico Barrier. *Front. Mol. Biosci.* **6** (2019), doi:10.3389/fmolb.2019.00117.

12. D. M. Krüger, A. Ahmed, H. Gohlke, NMSim Web Server: integrated approach for normal mode-based geometric simulations of biologically relevant conformational transitions in proteins. *Nucleic Acids Res.* **40**, W310–W316 (2012).

13. A. Ahmed, F. Rippmann, G. Barnickel, H. Gohlke, A Normal Mode-Based Geometric Simulation Approach for Exploring Biologically Relevant Conformational Transitions in Proteins. *J. Chem. Inf. Model.* **51**, 1604–1622 (2011).

14. D. Ciupka, H. Gohlke, On the potential alternate binding change mechanism in a dimeric structure of Pyruvate Phosphate Dikinase. *Sci. Rep.* **7**, 1–11 (2017).

15. A. Minges, D. Ciupka, C. Winkler, A. Höppner, H. Gohlke, G. Groth, Structural intermediates and directionality of the swiveling motion of Pyruvate Phosphate Dikinase. *Sci. Rep.* **7**, 1–14 (2017).

16. J. Lan, J. Ge, J. Yu, S. Shan, H. Zhou, S. Fan, Q. Zhang, X. Shi, Q. Wang, L. Zhang, X. Wang, Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* (2020), doi:10.1038/s41586-020-2180-5.

17. B. He, S. M. Mortuza, Y. Wang, H. Bin Shen, Y. Zhang, NeBcon: Protein contact map prediction using neural network training coupled with naïve Bayes classifiers. *Bioinformatics*. **33**, 2296–2306 (2017).

18. W. Zheng, Y. Li, C. Zhang, R. Pearce, S. M. Mortuza, Y. Zhang, Deep-learning contact-map guided protein structure prediction in CASP13. *Proteins Struct. Funct. Bioinforma.* **87**, 1149–1164 (2019).

19. C. Zhang, S. M. Mortuza, B. He, Y. Wang, Y. Zhang, Template-based and free modeling of I-TASSER and QUARK pipelines using predicted contact maps in CASP12. *Proteins Struct. Funct. Bioinforma.* **86**, 136–151 (2018).

Supplementary Materials:

Materials and Methods

Movies S1-S2

Data S1-S3