

pH and Receptor Induced Conformational Changes- Implications Towards S1 Dissociation of SARS-CoV2 Spike Glycoprotein

Jesu E. Castin, Daniel A. Gideon, Karthik S. Sudarsha and Sherlin A. Rosita

Department of Biotechnology and Bioinformatics, Bishop Heber College (Autonomous),
Tiruchirappalli, India

Abstract:

Viruses, being obligate intracellular parasites, must first attach themselves and gain entry into host cells. Viral fusion machinery is the central player in the viral attachment process in almost every viral disease. Viruses have incorporated an array of efficient fusion proteins on their surfaces to bind efficiently to host cell receptors. They make use of the host proteolytic enzymes to rearrange their surface protein(s) into the form which facilitates their binding to host- cell membrane proteins and subsequently, fusion. This stage of viral entry is very critical and has many therapeutic implications. The current global pandemic of COVID-19 has sparked severe health crisis and economic shutdowns. SARS-CoV2, the etiological agent of the disease has led to millions of deaths and brought the scientific community together in an attempt to understand the mechanisms of SARS-CoV2 pathogenesis and mortality. Like other viral fusion machinery, CoV2 spike (S) glycoprotein- ‘The Demogorgon’ poses the same questions about viral-host cell fusion. The intermediate stages of S protein-mediated viral fusion are unclear owing to the lack of structural insights and concrete biochemical evidence. The mechanism of conformational transition is still unclear. S protein binding and fusion with host cell receptors, Eg., angiotensin-converting enzyme-2 (ACE2) is accompanied by cleavage of S1/S2 subunits. To track the key events of viral-host cell fusion, we have identified (*in silico*) that low pH-induced conformational change and ACE-2 binding events promote S1 dissociation. Deciphering key mechanistic insights of SARS-CoV2 fusion will further our understanding of other class- I fusion proteins

Background:

Viruses constitute a vast majority of space under the infectious disease category. In the history of pandemics, viruses along with bacteria have conferred a major threat to world health as well as the economy. Unlike bacteria, they have higher mutation rates which make them easy to evolve into a better construct. HIV-I genome showed the highest mutation rate of $(4.1 \pm 1.7) \times 10^{-3}$ per base per cell [1]. The infectivity and lethality can vary from benign to acute. Although there are specific characteristics of each virus that set each of them apart are being crudely focused, there is some common key machinery that helps them evade the host cellular environment. One of such machinery is viral-host membrane fusion mediated by viral glycoproteins and host proteolytic cleavage. Viral glycoproteins also known as fusion proteins, are extremely capable of transitioning into different structurally distinct prefusion/post-fusion states [2-4]. Based on these common features of transitions, fusion proteins are classified into three classes. Despite the knowledge about two states, the intermediate events for any viral fusion are poorly understood [5]. The novel construct of Coronavirus genome- SARS- CoV2 poses the same inadequacy of knowledge about fusion machinery. Like other fusion proteins, it undergoes spontaneous transition driven by factors like i) low pH [6], ii) Host Receptor Binding [7], iii) Host-mediated proteolytic cleavage [8] and iv) host membrane composition (In case of Ebola) [9]. Although there are many biochemical analyses [10-14] for the intermediates, many structural insights still remain unclear. Spike glycoprotein of SARS-CoV2 is the class I fusion protein with trimeric assembly [15]. Deciphering the intermediate stages of the pre-fusion to post-fusion conformational transition can change the whole knowledge about viral fusion proteins. This could lead us to develop effective therapeutics which can block the transition as well as the successful fusion.

Introduction:

The spike glycoprotein of SARS-CoV2 is the large trimeric structure crucial for viral-cell fusion with the molecular weight of about 424 kDa. The structure has the three major topological domains which include Head, Stalk and Cytoplasmic Tail (CT). Each monomeric unit of the protein has two subunits namely S1 and S2 [16]. S1 has the regions responsible for host receptor binding- N-terminal Domain (NTD) and Receptor Binding domain (RBD). Most of the S2 regions serve as the fusion machinery- Fusion Peptide (FP), Central Helix (CH), Heptad Repeat (HR), etc (Figure 1.1, Figure 1.2.A). Although these regions are mostly conserved (compared to older strains of Coronaviridae family), few mutations are seen in the Receptor Binding Domain (RBD) and it is also noteworthy that this new construct has polybasic cleavage site [18] which is not present in the previously encountered viral strains. This new region increases the rate of S1/S2 cleavage and thereby, elevates efficiency of viral entry into the host cells. Only for the Head domain, the experimental structure is available and the rest of the topological domains have been successfully modelled in accordance with many experimental data [19]. The stalk domain is responsible for global protein flexibility and allows the glycoprotein to bend in different angles [20]. It is the best example of a mechanistic protein *i.e.*, it has exhibited many states which are shown in the Figure 1.2 B.

Followed by Angiotensin Collecting Enzyme-2 (ACE-2) Receptor binding in open form/ RBD 'Up' form (active form), the fusion events are initiated. The cleavage is likely to be mediated by host furin or TRPMSS2 at two distinct sites- S1/S2 and S2/S2' respectively [21]. As stated earlier, the polybasic cleavage site at the S1/S2 interface brings about effective proteolytic cleavage, thereby promoting the mechanistic tendency of the spike glycoprotein to mature into the post-fusion state. The steps of such transitions are unclear because very few biochemical evidence has been found for the intermediates.

But it is well known that S1/S2 cleavage (Figure 1.3. A) promotes S1 dissociation whereas S2/S2' (Figure 1.3. B) cleavage paves the way for S2 maturation (into the postfusion state) because Fusion peptide carrying the membrane fusing segment- 'FPPR' is released along with Heptad Repeat (HR) which holds the FP straight towards the host membrane. Although the cleaved states for the Stabilized CoV1 did not show any immediate conformational changes in the Cryo-EM experiments [22], our analysis on CoV2 Spike showed that the post cleavage state is much more disordered. Right after the binding event with ACE-2, the structural changes like decreased S1/S2 contact start to occur [23]. Our preliminary analysis confirmed this conformational change and yielded deeper insights.

The key role of glycans in protein shielding has come to the fore recently. Due to the recent reports, several glycan associated analyses have become prominent and repositories (such as protein data bank) have focused on easier retrieval of glycans which are associated with proteins. A recent study on SARS-CoV2 spike by Casalino et al., 2020 has shed more light on the role of glycans beyond just protein shielding. In that study, the group identified that the glycans in N165 and N234 stabilize the RBD 'Up' state [24]. And another study depicted its importance in maintaining contact with the ACE-2 Receptor [25]. Recently, the same glycans involved in RBD 'Up' state are known to form a cryptic pocket during the intermediate's opening pathway [26].

Results and Discussion:

i) Low pH creates unfavourable charge distribution in S1/S2 Interface and thereby promotes S1 Dissociation:

S1 dissociation is a spontaneous step which occurs after S1/S2 Cleavage and the ACE-2 Binding event. S1 and S2 have two putative contact interfaces C1 and C2 (Supplementary Figure 2.1). Of the two, Amaro Lab ‘Closed’ state model (depicting postcleaved S1/S2 state) showed that C2 has a complex hydrogen-bonding network (Supplementary Figure 2.1). Also, a strong electrostatic complementarity was seen in the C1 and C2 (Supplementary Figure 2.2). But in the dynamical events, the external physical conditions were found to play a vital role in the conformational arrangement of proteins. Hence, we hypothesised that the low pH would create an unfavourable charge distribution. The hypothesis is successfully proven by the preliminary results from ProPka. The pKa values for the titratable residues are calculated as follows:

$$pK_a = pK_{model} + pK_{HB} + pK_{DS} + pK_{CC}$$

where pK_{model} is the reference pKa value for each titratable amino acid and pK_{HB} , pK_{DS} , pK_{CC} are the pKa values decreased in the event of hydrogen bonding interaction, desolvation and coulombic contact, respectively.

On analysing the titratable residues in the S1/S2 Interface (Figure 1.4.A) of ‘Closed’ and ‘ACE-2 Bound’ experimental structures (PDB ID: 6VXX and 6A98 respectively), it becomes clear that the interface residues of the ‘ACE-2 Bound’ state exhibited higher pKa values than the interface residues of the ‘Closed’ state (Figure 1.4.B). This implies that the hydrogen bonding network and Coulombic contacts begin to diminish at the low pH receptor binding event, which would result in increased pK_{HB} , pK_{CC} values. Also, charge calculations across the pH

spectrum proved that low pH induces unfavourable positive/neutral charge distribution in the S1/S2 interface (Figure 1.5).

We obtained a different result from Amaro lab model of ‘Open’ state (Supplementary figure 2.4.A). Unlike in ‘ACE-2 Bound’ state, the pKa values of the interface residues are decreased, providing us with the insights on the differences between ‘Open’ and ‘ACE-2 Bound’ states. The charge distribution again showed increasing values across decreasing pH (Supplementary figure 2.4.B).

ii) S1/S2 Cleavage Elevates Local Disorderliness in S1/S2 Interface Environment:

We compared the uncleaved (precleaved) (CHARMM Model) and postcleaved (Amaro Lab Model) S1/S2 models (Figure 1.6) with Normal Mode Analysis incorporated with Anisotropic Network Model (ANM). NMA is the approximation technique of Molecular Dynamic Simulation that uses a network model and calculates projection of each atom assuming a harmonic oscillation. Calculating Hessian Matrix for each atomic projection can provide us with insights about fluctuations and other dynamical properties. Applying NMA-ANM for the postcleaved as well as precleaved states, we came to find that postcleaved S1/S2 state showed increased RMSF (especially in S1/S2 interface) than that of precleaved S1/S2 state (Figure 1.7). It provided the preliminary insight that host-mediated proteolytic cleavage elevates relative disorderliness in S1/S2 interface and further promotes S1 dissociation. It is noteworthy that neither experimental structures nor models are capable of illustrating S2/S2’ cleavage, which can be studied only through rigorous MD approaches.

iii) Experimental Evidence Proves Host-Receptor Induced Conformational Changes:

We also generated distant maps for the experimental structures in ‘Closed’ and ‘ACE-2 Bound’ states (Figure 1.8). It became clear that the contact distance in S1/S2 interface of the ‘ACE-2 Bound’ state appeared to have diminished. We could observe that the NTD and CTD-2 regions

are fluctuating in nature whereas HR of S2 remained stable (Figure 1.9). This confirms that Host-Receptor binding also influences S1 dissociation.

Materials and Methods:

i) pH and Charge Calculations:

To observe pH-induced changes, we used PropKa 3.0 [27]. The pKa calculation was done for all the titratable residues of the different states of S protein (Chain A). Initially, we performed the calculations for the experimental S protein structures in Closed and Three RBD bound states (PDB ID: 6VXX and 6A98 respectively). To also assess the nature of S1/S2 interface of the full-length model of open and closed states, we used Amaro Lab model which is available as open-source. The pKa values of the titratable interface residues were then taken for further consideration, as our interest fits into the S1/S2 interface where possible initial events of S1 dissociation occur. Then the charge of the titratable residues upon decreasing pH *i.e.*, 7 to 3 was calculated for the experimental and modelled structures to find the effect of pH in order to prove our hypothesis about unfavourable charge distribution. The charge values were obtained with the pH plot profiling module available at VMD 1.9.3. The titratable residues of S1/S2 interface were subjected for the charge profiling with the decreasing pH in the units of 0.01. The data points were then retrieved for comprehensive plotting with GNUPLOT 5.2.

ii) Comparing Precleaved/Postcleaved States with ANM-NMA:

For better understanding the roles of S1/S2 cleavage, we need to compare the S1/S2 interface environment in both the states. We chose the models which clearly depict the precleaved (CHARMM-Model)/postcleaved (Amaro Lab Model) states. For both structures, backbone Anisotropic Network Model-Normal Mode Analysis (ANM-NMA) was performed for 3 modes due to the reason that 3 modes NMA calculation is optimum for observing global fluctuations. The square fluctuations for both states were then calculated and resulting differential

fluctuation values were compared. All these steps were performed with the Python-based tool-ProDy 1.10.11 [28].

iii) Distance Map Generation for the Receptor Bound/Unbound States:

For deriving a conclusive proof for the receptor-induced conformational changes in the S1/S2 interface, we built S1/S2 distance maps for the experimental structures in closed as well as in Receptor bound states (PDB ID: 6VXX and 6A98 respectively). Also, generating the distance map may help us understand how receptor binding influences the S1 dissociation. With the Protein Contact Map generator [29], inter-residue distances were calculated and then the final correlation map was generated.

Conclusion:

Viral glycoproteins are the enhanced fusion machinery exhibiting the diverse dynamical and structural nature of biomacromolecules. This makes them pivotal to delineate their function. The key events during viral-host cell fusion are still enigmatic and there are several unsolved questions. One such event is the dissociation of protein subunits S1 and S2 in SARS-CoV2 S protein. Understanding this spontaneous but substantial event with the aid of unconventional MD approaches which when validated by using molecular biological techniques would give us a crystal-clear perspective of how class- I fusion proteins function. Our preliminary results on key conformational changes provide an insight towards the importance of further extending our knowledge about fusion proteins and their dissociation events. Delineating fusion proteins like this can impact Structural Virology as well as promiscuous drug development.

References:

- [1] Cuevas, J. M., Geller, R., Garijo, R., López-Aldeguer, J., & Sanjuán, R. (2015). Extremely High Mutation Rate of HIV-1 In Vivo. *PLoS Biology*, 13(9). <https://doi.org/10.1371/journal.pbio.1002251>
- [2] Modis, Y., Ogata, S., Clements, D. et al. Structure of the dengue virus envelope protein after membrane fusion. *Nature* 427, 313–319 (2004). <https://doi.org/10.1038/nature02165>
- [3] White, J. M., Delos, S. E., Brecher, M., & Schornberg, K. (2008). Structures and Mechanisms of Viral Membrane Fusion Proteins: Multiple Variations on a Common Theme. *Critical Reviews in Biochemistry and Molecular Biology*, 43(3), 189–219. <https://doi.org/10.1080/10409230802058320>
- [4] Benton, D.J., Gamblin, S.J., Rosenthal, P.B. et al. Structural transitions in influenza haemagglutinin at membrane fusion pH. *Nature* 583, 150–153 (2020). <https://doi.org/10.1038/s41586-020-2333-6>
- [5] Barfoot, S., Poger, D., & Mark, A. E. (2020). Understanding the Activated Form of a Class-I Fusion Protein: Modeling the Interaction of the Ebola Virus Glycoprotein 2 with a Lipid Bilayer. *Biochemistry*, 59(41), 4051–4058. <https://doi.org/10.1021/acs.biochem.0c00527>
- [6] Korber, B., Fischer, W. M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., ... Montefiori, D. C. (2020). Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell*, 182(4), 812-827.e19. <https://doi.org/10.1016/j.cell.2020.06.043>
- [7] Gaudin, Y., Ruigrok, R. W. H., & Brunner, J. (1995, July 1). Low-pH induced conformational changes in viral fusion proteins: Implications for the fusion mechanism. *Journal of General Virology*, Vol. 76, pp. 1541–1556. <https://doi.org/10.1099/0022-1317-76-7-154>
- [8] Matsuyama, S., & Taguchi, F. (2002). Receptor-Induced Conformational Changes of Murine Coronavirus Spike Protein. *Journal of Virology*, 76(23), 11819–11826. <https://doi.org/10.1128/jvi.76.23.11819-11826.2002>
- [9] Meulendyke, K. A., Wurth, M. A., McCann, R. O., & Dutch, R. E. (2005). Endocytosis Plays a Critical Role in Proteolytic Processing of the Hendra Virus Fusion Protein. *Journal of Virology*, 79(20), 12643–12649. <https://doi.org/10.1128/jvi.79.20.12643-12649.2005>
- [10] Carette, J., Raaben, M., Wong, A. et al. Ebola virus entry requires the cholesterol transporter Niemann–Pick C1. *Nature* 477, 340–343 (2011). <https://doi.org/10.1038/nature10348>
- [11] Kawase, M., Kataoka, M., Shirato, K., & Matsuyama, S. (2019). Biochemical Analysis of Coronavirus Spike Glycoprotein Conformational Intermediates during Membrane Fusion. *Journal of Virology*, 93(19). <https://doi.org/10.1128/jvi.00785-19>
- [12] Park, J. E., Li, K., Barlan, A., Fehr, A. R., Perlman, S., McCray, P. B., & Gallagher, T. (2016). Proteolytic processing of middle east respiratory syndrome coronavirus spikes expands

virus tropism. *Proceedings of the National Academy of Sciences of the United States of America*, 113(43), 12262–12267. <https://doi.org/10.1073/pnas.1608147113>

[13] Simmons, G., Reeves, J. D., Rennekamp, A. J., Amberg, S. M., Piefer, A. J., & Bates, P. (2004). Characterization of severe acute respiratory syndrome-associated coronavirus (SARS-CoV) spike glycoprotein-mediated viral entry. *Proceedings of the National Academy of Sciences of the United States of America*, 101(12), 4240–4245. <https://doi.org/10.1073/pnas.0306446101>

[14] Mille, J. K., & Whittaker, G. R. (2014). Host cell entry of Middle East respiratory syndrome coronavirus after two-step, furin-mediated activation of the spike protein. *Proceedings of the National Academy of Sciences of the United States of America*, 111(42), 15214–15219. <https://doi.org/10.1073/pnas.1407087111>

[15] 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 of the United States of America*, 106(14), 5871–5876. <https://doi.org/10.1073/pnas.0809524106>

[16] Delmas, B., & Laude, H. (1990). Assembly of coronavirus spike protein into trimers and its role in epitope expression. *Journal of Virology*, 64(11), 5367–5375. <https://doi.org/10.1128/jvi.64.11.5367-5375.1990>

[17] Fan, X., Cao, D., Kong, L. et al. Cryo-EM analysis of the post-fusion structure of the SARS-CoV spike glycoprotein. *Nat Commun* 11, 3618 (2020). <https://doi.org/10.1038/s41467-020-17371-6>

[18] Hoffmann, M., Kleine-Weber, H., & Pöhlmann, S. (2020). A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells. *Molecular Cell*, 78(4), 779-784.e5. <https://doi.org/10.1016/j.molcel.2020.04.022>

[19] Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T., & Veerler, D. (2020). Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*, 181(2), 281-292.e6. <https://doi.org/10.1016/j.cell.2020.02.058>

[20] Kyo Choi, Y., Cao, Y., Frank, M., Woo, H., Park, S.-J., Sun Yeom, M., ... Im, W. (2020). Structure, Dynamics, Receptor Binding, and Antibody Binding of Fully-glycosylated Full-length SARS-CoV-2 Spike Protein in a Viral Membrane. *BioRxiv*, 2020.10.18.343715. <https://doi.org/10.1101/2020.10.18.343715>

[21] Jaimes, J. A., Millet, J. K., & Whittaker, G. R. (2020). Proteolytic Cleavage of the SARS-CoV-2 Spike Protein and the Role of the Novel S1/S2 Site. *IScience*, 23(6), 101212. <https://doi.org/10.1016/j.isci.2020.101212>

[22] Kirchdoerfer, R. N., Wang, N., Pallesen, J., Wrapp, D., Turner, H. L., Cottrell, C. A., ... Ward, A. B. (2018). Stabilized coronavirus spikes are resistant to conformational changes induced by receptor recognition or proteolysis. *Scientific Reports*, 8(1), 1–11. <https://doi.org/10.1038/s41598-018-34171-7>

[23] Benton, D. J., Wrobel, A. G., Xu, P., Roustan, C., Martin, S. R., Rosenthal, P. B., ... Gamblin, S. J. (2020). Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion. *Nature*. <https://doi.org/10.1038/s41586-020-2772-0>

- [24] Casalino, L., Gaieb, Z., Goldsmith, J. A., Hjorth, C. K., Dommer, A. C., Harbison, A. M., ... Amaro, R. E. (2020). Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Central Science*. <https://doi.org/10.1021/acscentsci.0c01056>
- [25] Zhao, P., Praissman, J. L., Grant, O. C., Cai, Y., Xiao, T., Rosenbalm, K. E., ... Wells, L. (2020). Virus-Receptor Interactions of Glycosylated SARS-CoV-2 Spike and Human ACE2 Receptor. *Cell Host and Microbe*, 28(4), 586-601.e6. <https://doi.org/10.1016/j.chom.2020.08.004>
- [26] Brotzakis, Z. F., Lohr, T., & Vendruscolo, M. (2020). *Determination of Intermediate State Structures in the Opening Pathway of SARS-CoV-2 Spike Using Cryo-Electron Microscopy*. <https://doi.org/10.26434/CHEMRXIV.13222073.V1>
- [27] Olsson, M. H. M., SØndergaard, C. R., Rostkowski, M., & Jensen, J. H. (2011). PROPKA3: Consistent treatment of internal and surface residues in empirical pKa predictions. *Journal of Chemical Theory and Computation*, 7(2), 525–537. <https://doi.org/10.1021/ct100578z>
- [28] Ahmet Bakan, Lidio M. Meireles, Ivet Bahar, ProDy: Protein Dynamics Inferred from Theory and Experiments, *Bioinformatics*, Volume 27, Issue 11, 1 June 2011, Pages 1575–1577, <https://doi.org/10.1093/bioinformatics/btr168>
- [29] Rafferty, B., Flohr, Z. C., & Martini, A. (2010). *Protein Contact Maps*. <https://doi.org/10.21981/02DQ-MT84>

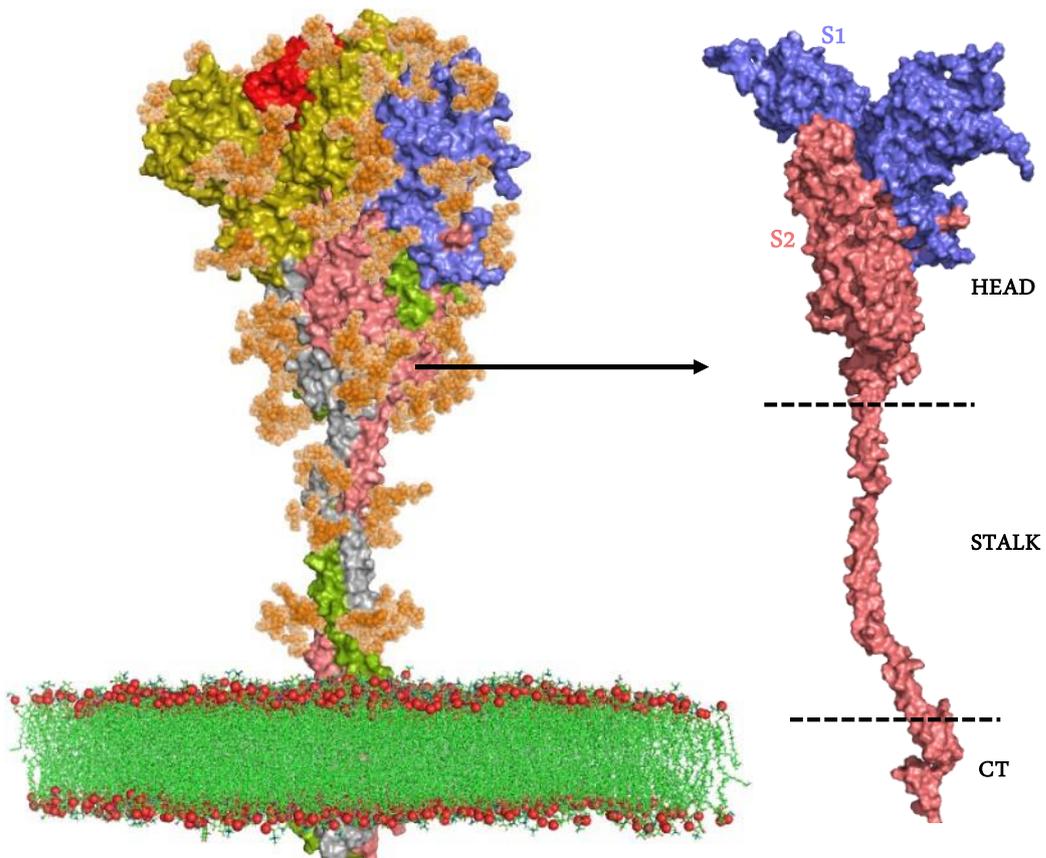


Figure1.1- Full Length Glycosylated Spike Model

SARS-CoV2 spike glycoprotein forms a trimeric assembly in which each monomeric unit has two subunits S1 and S2. The main topological domains include head, stalk and cytoplasmic tail (CT).

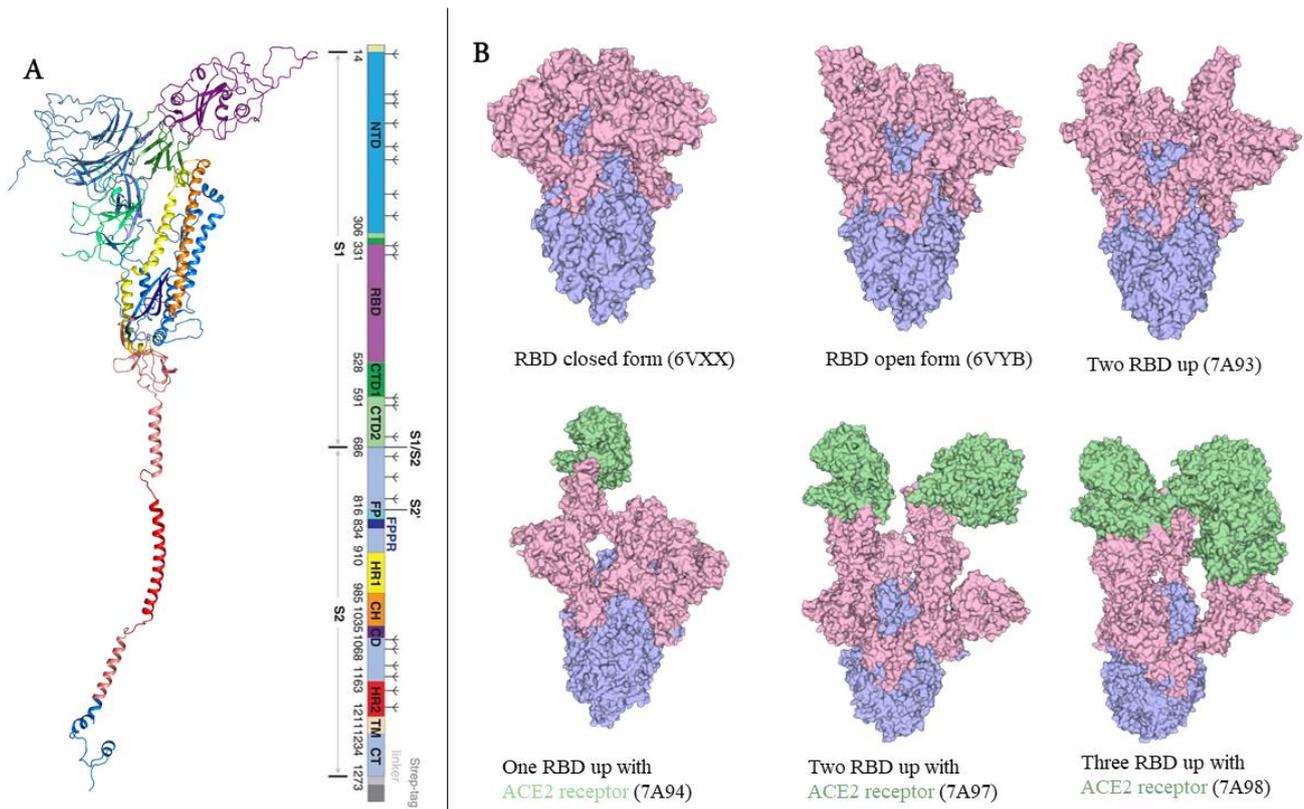


Figure 1.2 A- Characteristic regions of the spike glycoprotein (Chain A).

Spike has several key regions which are responsible for various functions that promote the viral entry into the host cell.

Each region is coloured with respective to the one-dimensional legend.

Figure 1.2 B- Spike Exhibiting Multiple States (with the PDB ID)

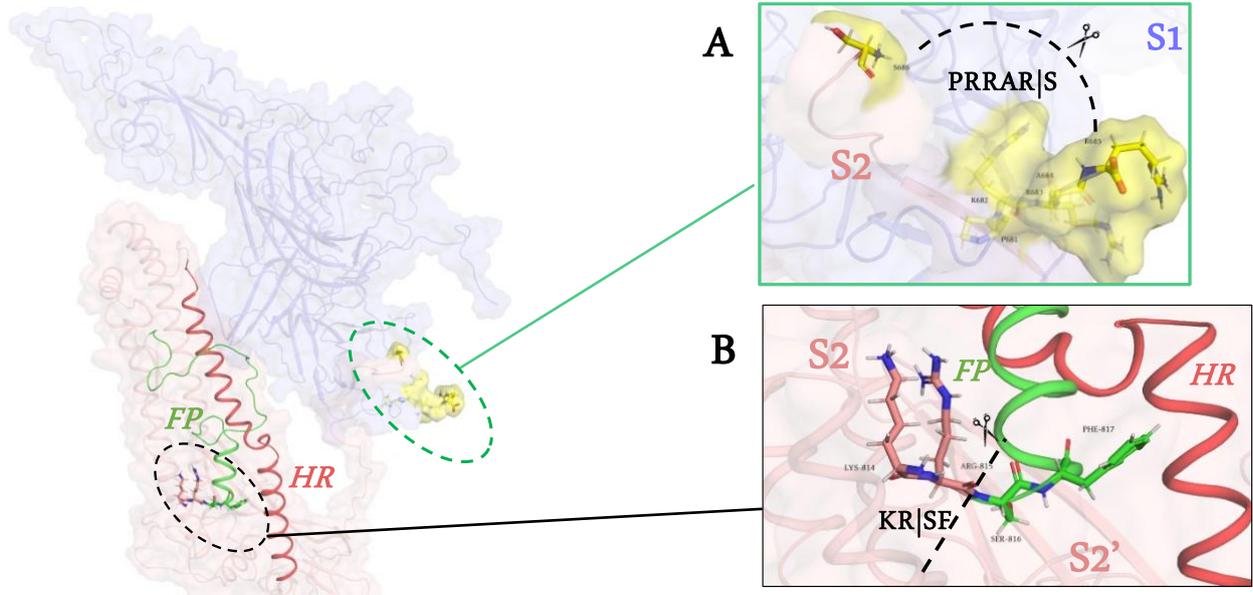


Figure 1.3- S1/S2 and S2/S2' cleavage sites

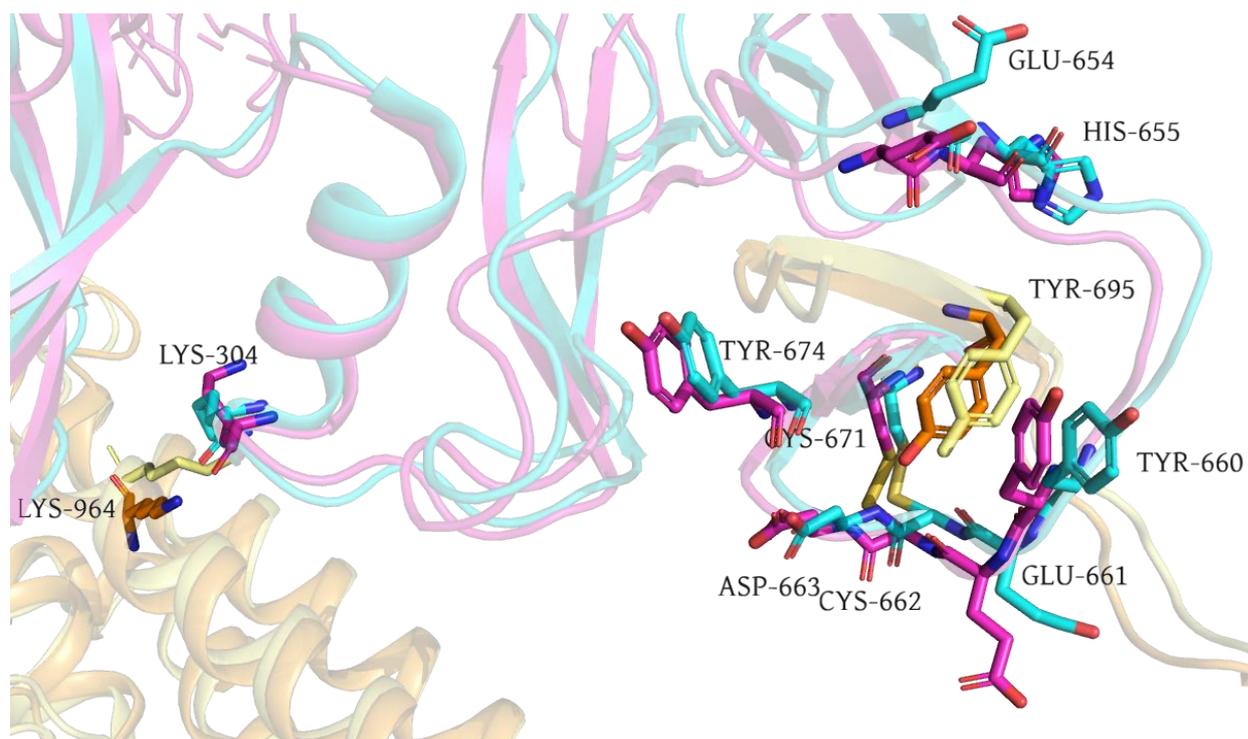


Figure 1.4.A- Titratable residues at the S1/S2 interface of Closed (PDB: 6VXX) and ACE-2 Bound states (PDB: 6A98)

Figure 1.4.B- pKa values of S1/S2 interface titratable residues.

The pKa values of the interface residues in 'ACE-2 Bound' state (shown in arrows) are increased due to the diminishing polar and coulombic contact

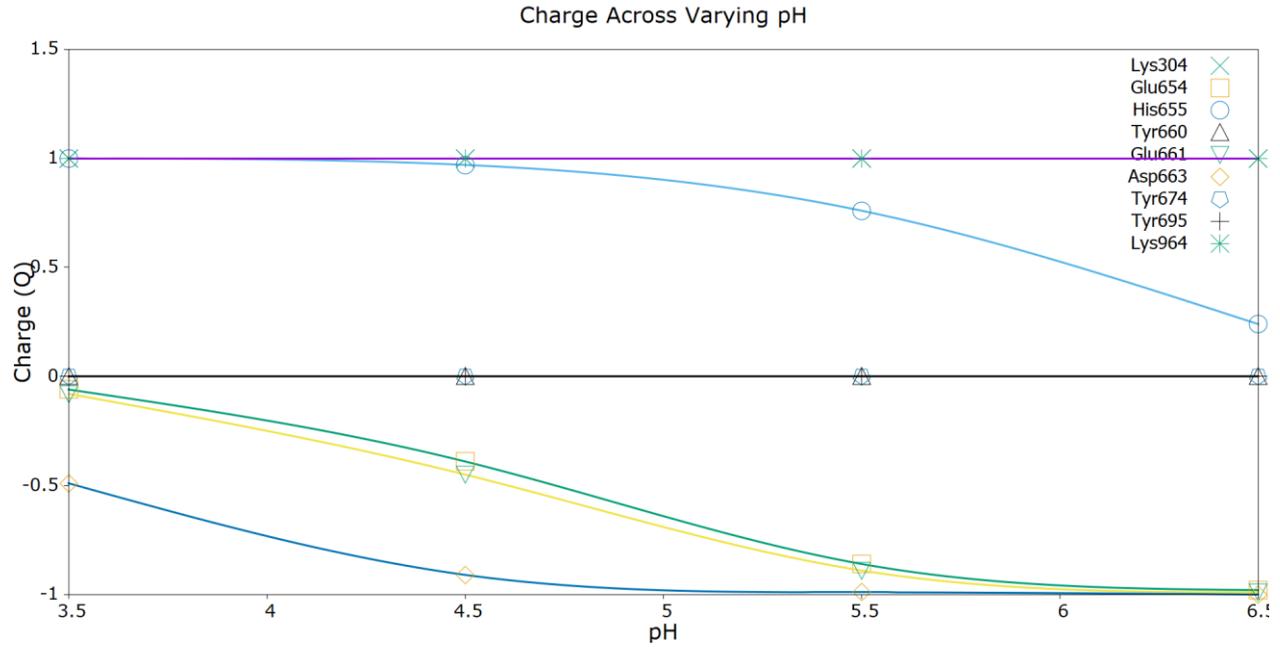


Figure 1.5- Charge calculation across the decreasing pH

The graph implies that most of the interface residues show increasing positive/neutral charge distribution when the pH decreases (other than lysine and tyrosine residues)

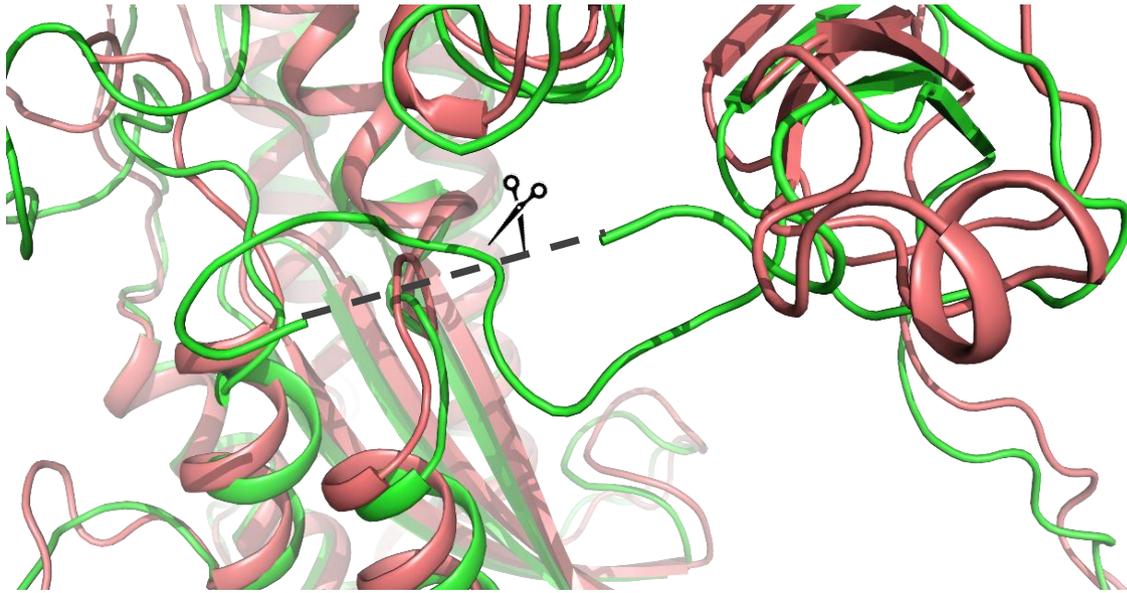


Figure 1.6- Superimposed **precleaved** (CHARMM Model) and **postcleaved** (Amaro Lab Model) states

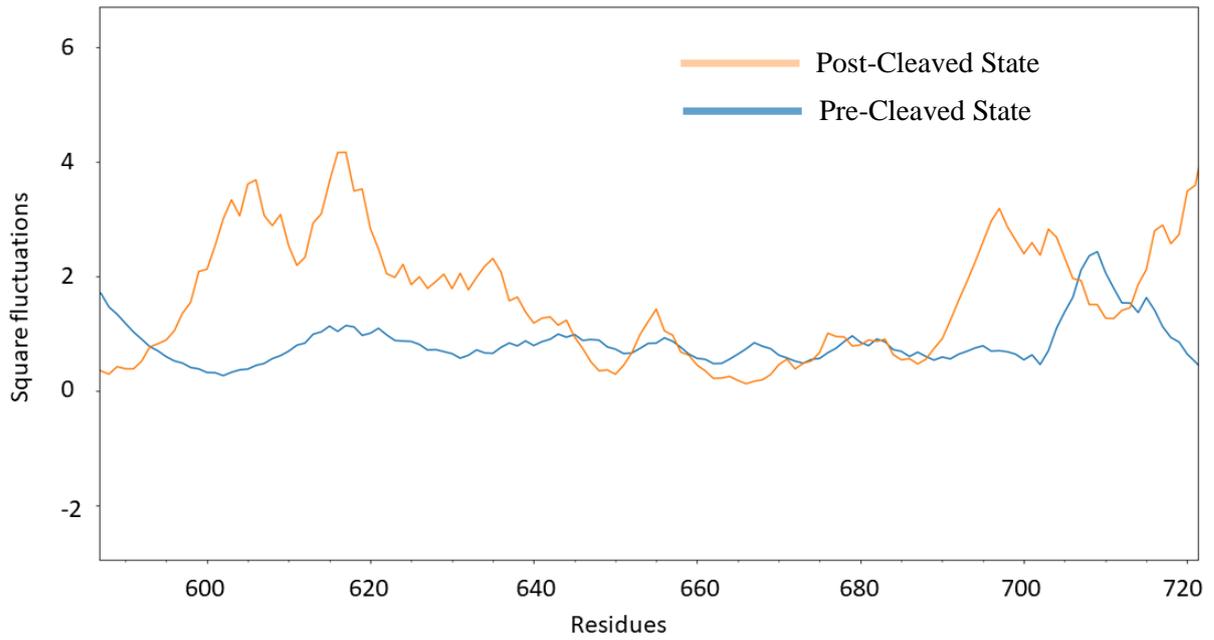


Figure 1.7- RMSF plot shows that the S1/S2 interface in post-cleaved state is highly fluctuating in nature

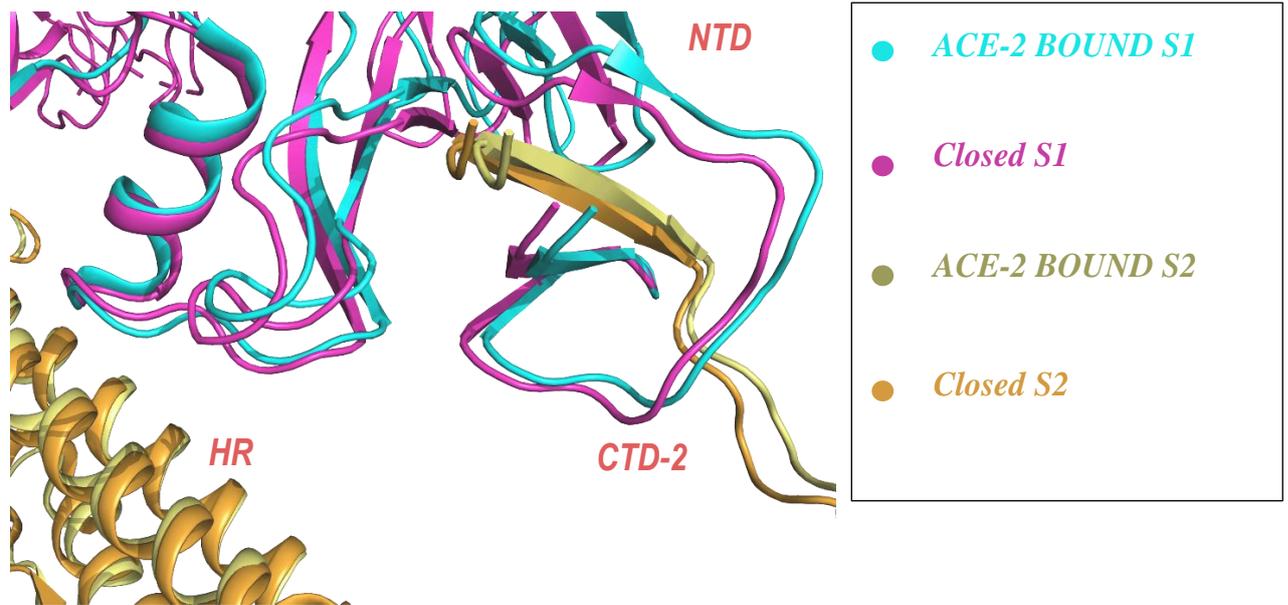


Figure 1.8- Superimposed S1/S2 interface at ‘Closed’ and ‘ACE-2 Bound’ states

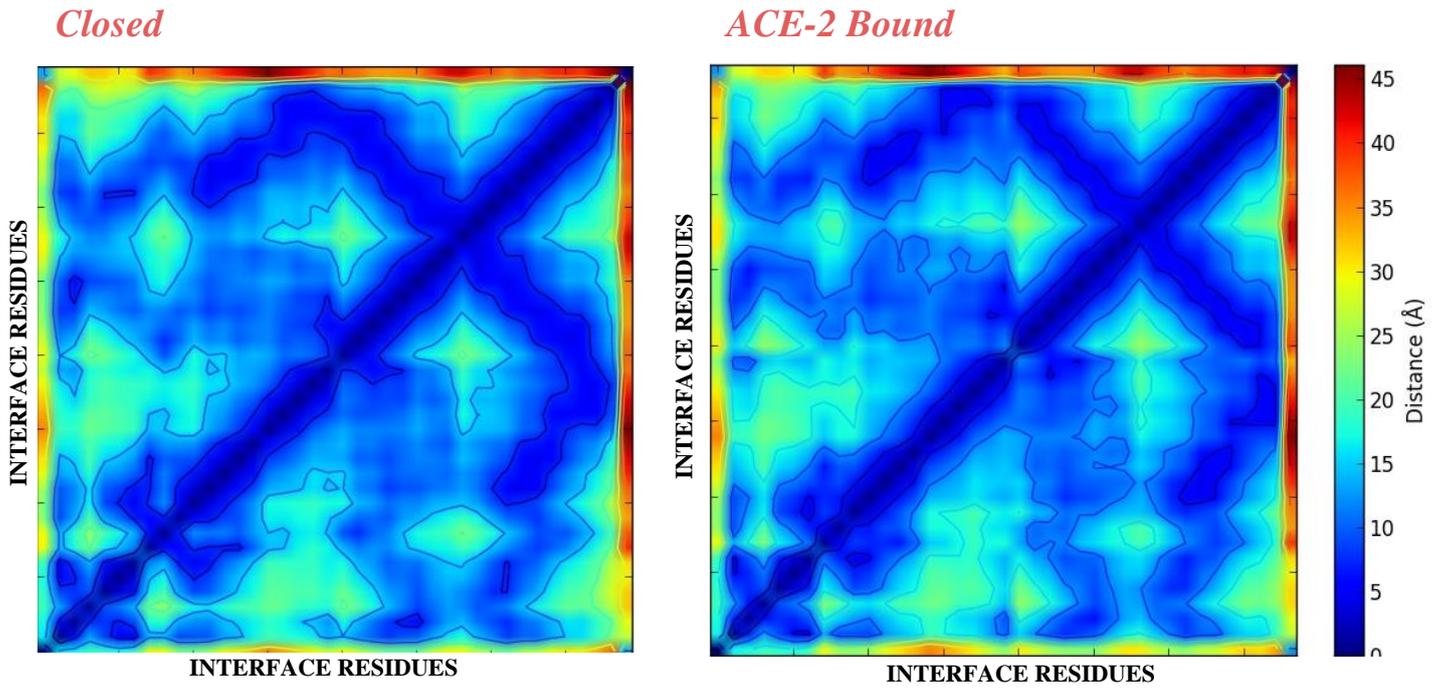


Figure 1.9- S1/S2 Distance Map confirms the fact that the relative distance between S1/S2 interface residues start diminishing in the event of ACE-2 binding.