

Supplementary Materials for the article:

Plausible blockers of Spike RBD in SARS-CoV2 – Molecular design and underlying interaction dynamics from high-level structural descriptors

Sankar Basu*, Devlina Chakravarty, Dhananjay Bhattacharyya, Pampa Saha, Hirak K Patra

Figure S1. Pre- and Post-fusion forms of the SARS-CoV2 Spike protein. To construct the physical body of the pre-fusion form (on the left of the schematic) three cryo-EM structures with PDB ID: 6VXX (shown in blue cartoon), 6CRZ (red), 6XR8 (yellow) have been superposed. The post-fusion form is displayed using the PDB ID: 6XRA. Different important structural regions have been demarcated as in the literature [1]. Proteolytic cleavages occur at two S1/S2 junction points in the polypeptide chain (cleavage sites) situated along the horizontal direction (i.e., at either ends of the S1/S2 boundary [2]: shown as the thick dashed line) of the displayed pre-fusion structure. The cleavage is concomitantly followed by a conformational transition (longitudinal shrinkage) of the resultant post-fusion form constrained by the embedding membrane matrix environment. Image constructed in PyMol [3].

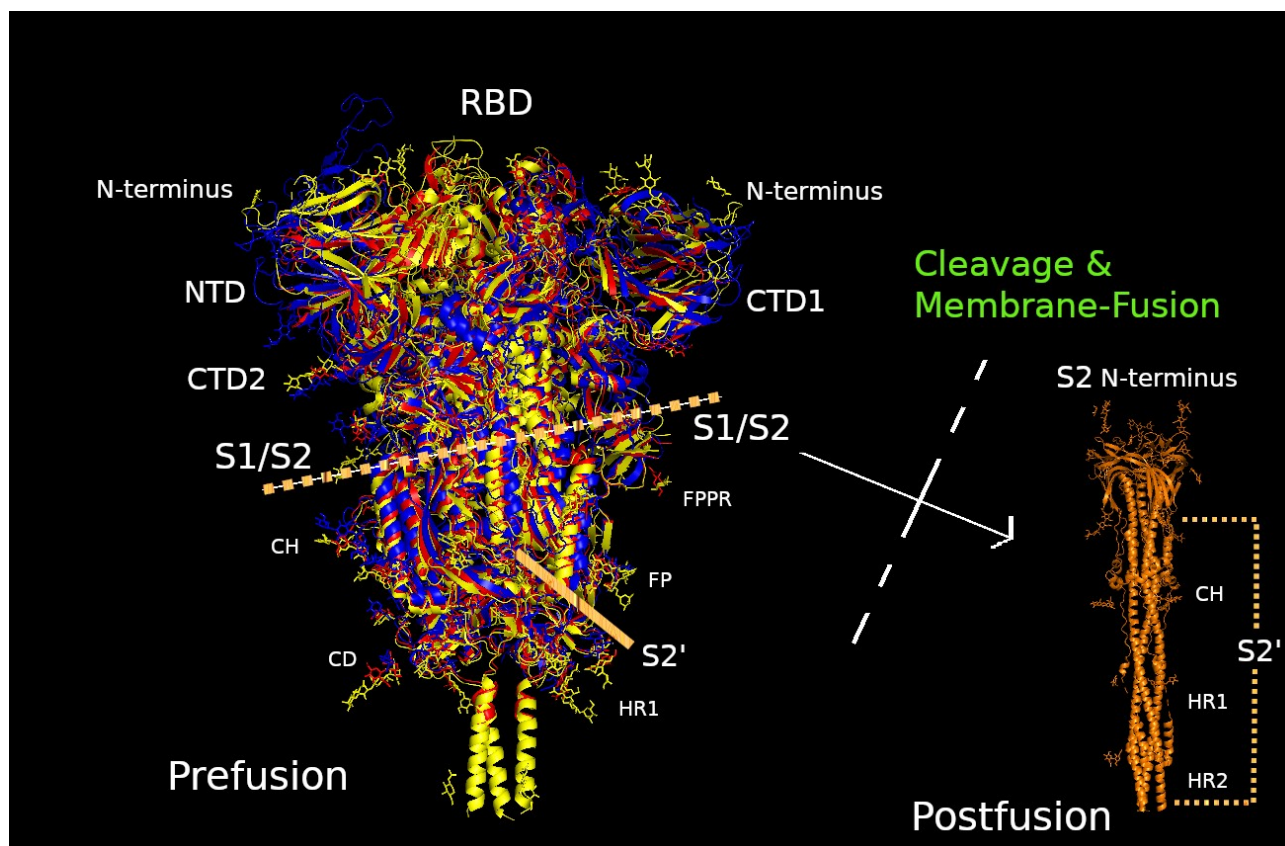


Figure S2. The Spike-RBD-hotspot. In the Figure (constructed from PDB ID: 6VW1), the ACE2 (chain A) and RBD_{Spike} (chain E) are colored in light magenta and green respectively. The hotspot region is colored in 'red' which corresponds to a 51 amino acid stretch (residues 455-505 of the ligand chain) consisting of the key mutations in the CoV-2 RBD_{Spike}.

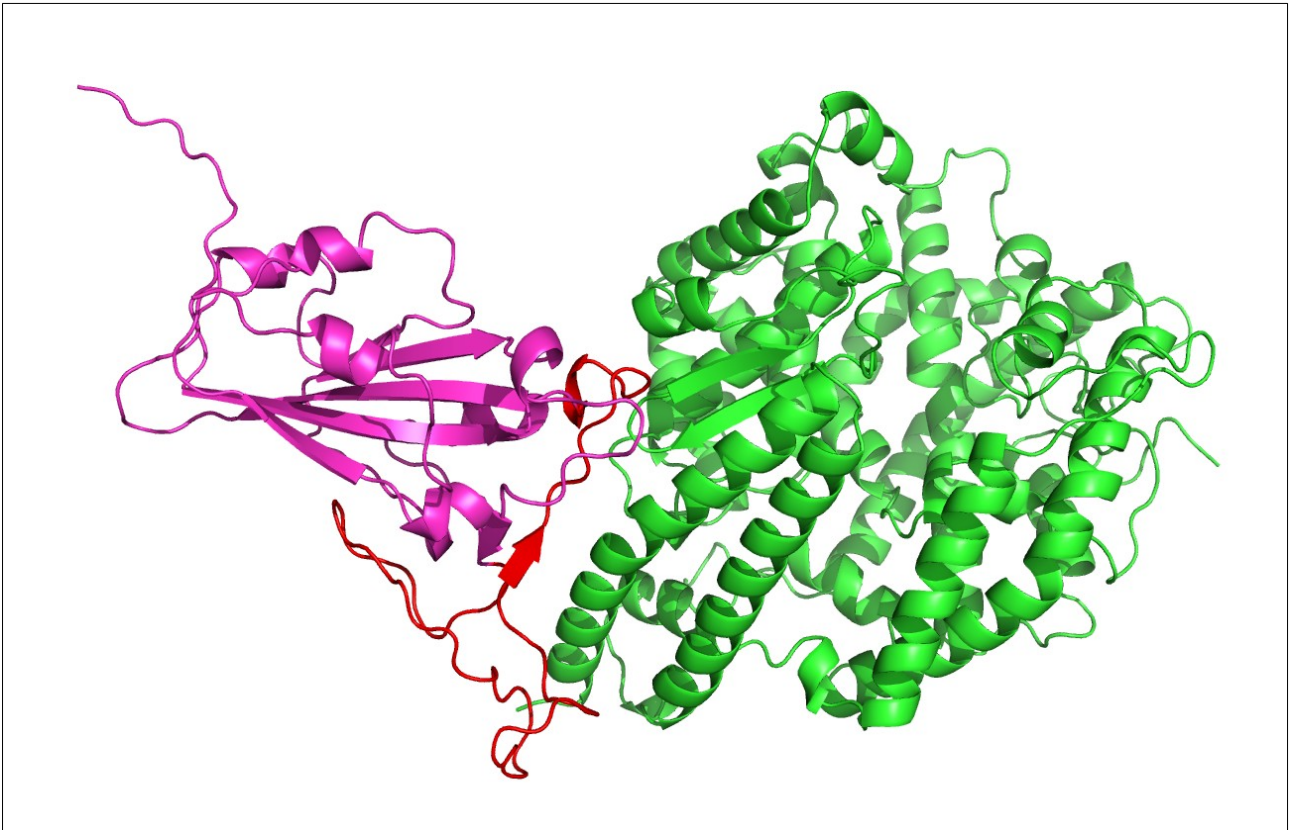


Figure S3. The local neighborhood of RBD_{Down} in the full-length Spike protein. In the Figure, the RBD_{down} unit as well as its local neighborhood in the rest of the Spike protein are both displayed as solid molecular (Connolly) surfaces, colored in green and blue respectively. The local neighborhood of RBD_{down} was delineated by collecting those residues (from the ‘rest of the Spike protein’) which were found within a C^α-C^α cut-off distance of 12Å from any residue in RBD_{down}. The calculation was also repeated at a 15 Å cut-off which returned the same Sc_{RBD_{down}}. The three chains (cartoons) in the trimeric native Spike protein (PDB ID: 6XR8) are drawn in different colors.

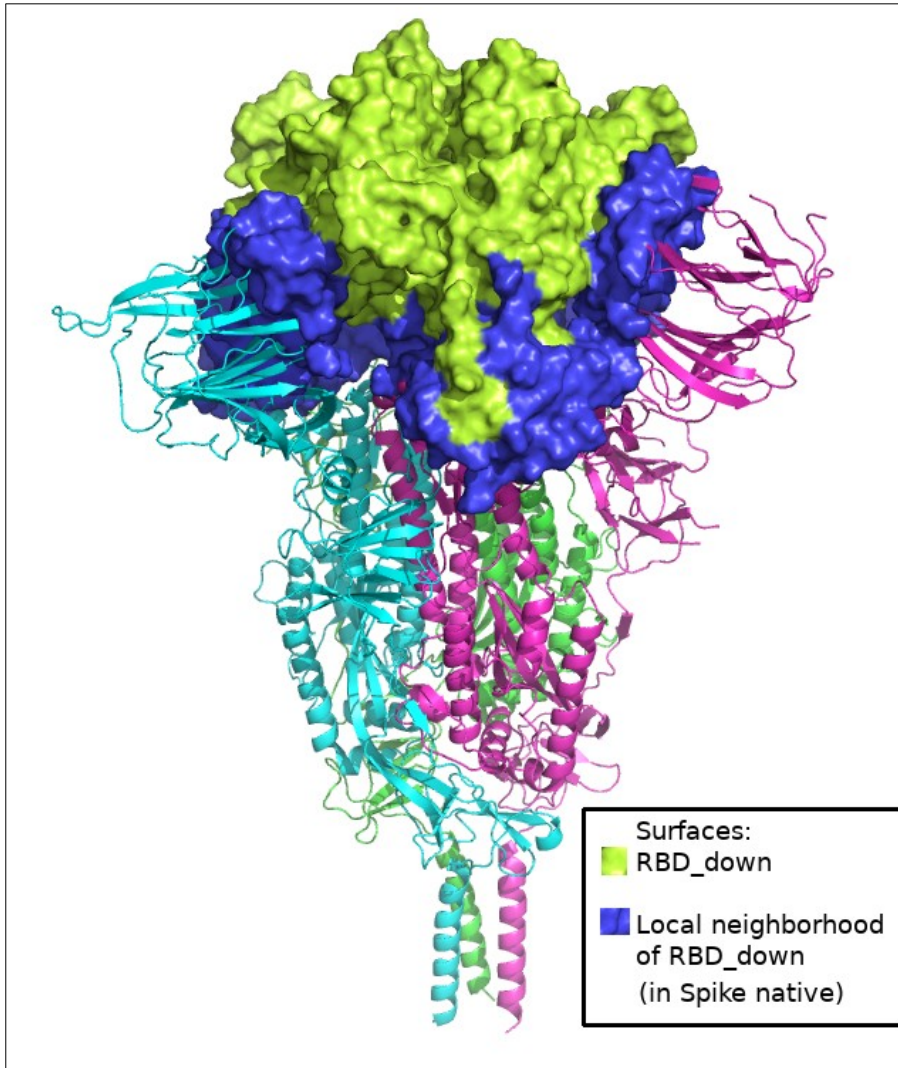


Table S1. Comparative stability of the RBD_{Spike} conformers in terms of the complementarity and other structural measures. The essential molecular details of the calculation(s) are tabulated in rows 1-4 of the Table while the numbers for the attempted measures (Sc, EC, BSA) are given in the rows 5-7.

	RBD _{down}	RBD _{up}
PDB ID	6XR8	6VW1
Molecular Source	The native Spike protein	The RBD _{Spike} -ACE2 complex
Target	RBD _{down}	RBD _{up}
Neighbor	The 'rest of the Spike protein'	ACE2
Sc	0.617	0.566
EC	0.254	0.055
BSA (Å ²)	6306.1	875.3

Figure S4. Electrostatic surface of RBD_{Down} when embedded in the native Spike protein. Panel A and B respectively map the electrostatic potential surface of RBD_{Down} due to the electric fields coming from its own atoms (i.e., self-potentials) and that from the partner atoms distributed throughout the ‘rest of the Spike protein’ (partner-potentials). Atomic coordinates of the native Spike protein are taken from PDB ID: 6XR8. In each panel, the thick arrows indicate whether the surface potentials are due to ‘self’ (panel: A) or ‘partner’ (panel: B). Panels C and D display the same colored surfaces in isolation as in panels A and B respectively, to present a greater clarity of the surface coloring. The electrostatic surface coloring was done in Chimera [4] using Delphi [5] electrostatic focusing files (.cube) with a color scale set to -10 kT/e for ‘pure blue’ to +10 kT/e for ‘pure red’. As can be seen from the colors, the anti-correlation in surface electrostatic potential is appreciably better than RBD_{Up} (portrayed in **Figure 3**, Panels A and B, part of Main Manuscript), leading to the resultant EC_{1,2} value of 0.254 (where, 1 and 2 in the subscripts of EC refer to the RBD_{Down} and the ‘rest of the Spike protein’ respectively). For purpose of comparison, the EC_{1,2} obtained for RBD_{Up} (referred to as the ‘ligand’ in **Figure 3**) is 0.055.

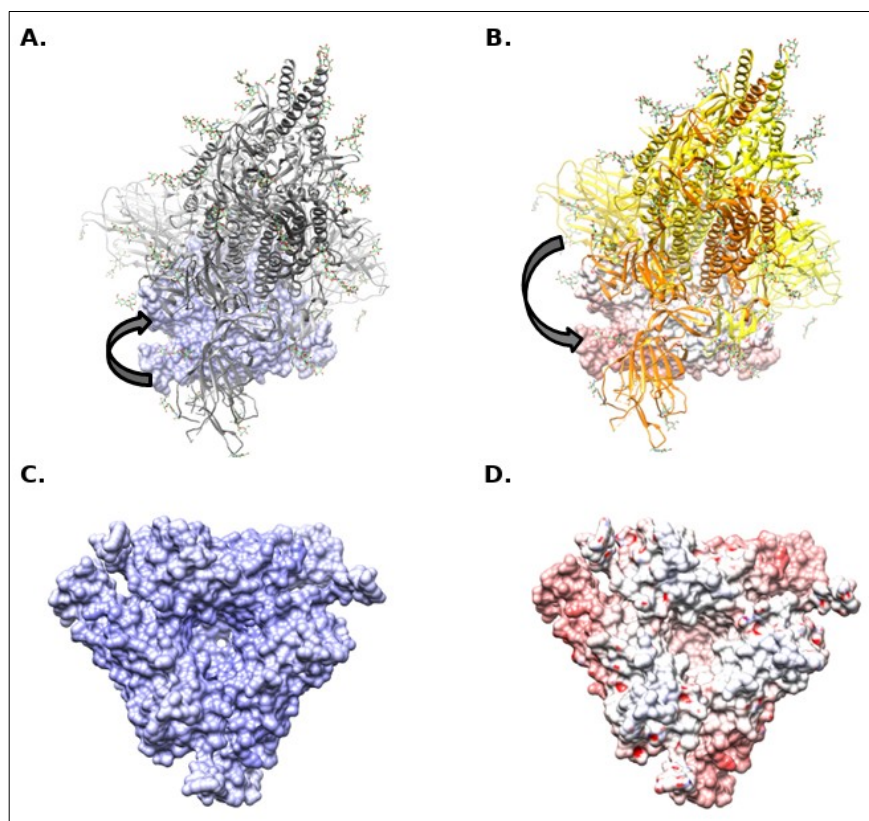
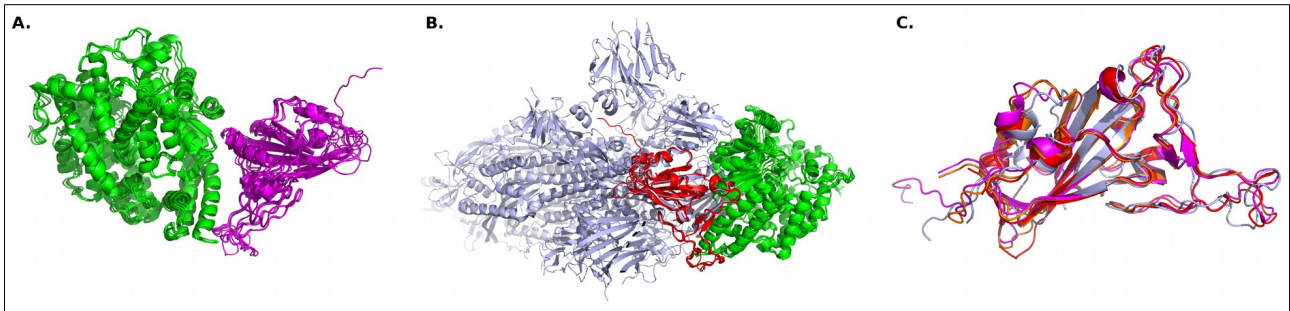


Table S2. Contact map of the native interface pertaining to the SARS-CoV-2 RBD_{Spike}-ACE2 interaction (6VW1). Receptor (6VW1_A) and ligand (6VW1_E) residues involved in the contacts are tabulated as column 1 and 2 with their contact strength (number of side-chain atomic contacts) given in column 3. Contacts calculated at a cutoff inter-atomic distance of 4 Å.

Interfacial Contact		Number of side-chain atomic contacts
Receptor partner (chain A)	Ligand partner (chain E)	
24-GLN-A	487-ASN-E	6
28-PHE-A	489-TYR-E	3
31-LYS-A	456-PHE-E	1
31-LYS-A	484-GLU-E	1
31-LYS-A	489-TYR-E	4
34-HIS-A	453-TYR-E	3
34-HIS-A	455-LEU-E	4
34-HIS-A	493-GLN-E	11
37-GLU-A	505-TYR-E	1
38-ASP-A	449-TYR-E	1
41-TYR-A	498-GLN-E	5
41-TYR-A	500-THR-E	6
41-TYR-A	501-ASN-E	5
42-GLN-A	498-GLN-E	3
45-LEU-A	498-GLN-E	3
79-LEU-A	486-PHE-E	4
82-MET-A	486-PHE-E	4
83-TYR-A	486-PHE-E	15
83-TYR-A	489-TYR-E	1
353-LYS-A	501-ASN-E	4
353-LYS-A	505-TYR-E	8
355-ASP-A	500-THR-E	1
357-ARG-A	500-THR-E	2

Figure S5. The RBD_{Spike}–ACE2 complex in coronavirus: Homologous structure superposed. Panel A shows the superimposed binary PPI complexes while panel B shows them further being superposed onto the RBD_{Spike} of the native Spike protein (PDB ID: 6VXX). Panel C shows the superimposed RBD_{Spike} domains taken from 6VXX (orange), 6VW1 (red), 6CRZ (light blue) & 6XR8 (magenta).



Dataset S1. Full length sequences of a set of representative designed RBD mimics. These best predicted designed mimics appears to have the potential to serve as plausible competitive inhibitor(s) of the native RBD_{Spike}-ACE2 interaction in SARS-CoV2. Sequences are given in FASTA format and their names as per referred in the main-text.

>>**HM0**

```
NLCPFGEVFNATKFPSVYAWERKKISNCVADYSVLYNSTFFSTFKCYGVSATKLNLDLCSNV
YADSFVVKGDDVRQIAPGQTGVIADYNYKLPDDFMGCVLAWNTRNIDATSTGNINYKYR
YLRHGKLRPFERDISNVPFSPDGKPCPTVPAPNCYWPLRGYGFYTTTGIGYQPYRVVLSFE
LLNAPATVCGPKLSTDLIK
```

>>**HM3**

```
NLCPFGEVFNATKFPSVYAWERKKISNCVADYSVLYNSTFFSTFKCYGVSATKLNLDLCSNV
YADSFVVKGDDVRQIAPGQTGVIADYNYKLPDDFMGCVLAWNTRNIDATSTGNINYKYR
YLRHGKLRPFERDISNVPFSPDGKPCPTVPAPNCYWPLNGYGFYTTTGIGYQPYRVVLSFE
LLNAPATVCGPKLSTDLIK
```

>>**HM5**

```
NLCPFGEVFNATKFPSVYAWERKKISNCVADYSVLYNSTFFSTFKCYGVSATKLNLDLCSNV
YADSFVVKGDDVRQIAPGQTGVIADYNYKLPDDFMGCVLAWNTRNIDATSTGNINYKYR
YLRHGKLRPFERDISNVPFSPDGKPCPTVPAPNCYWPLNGYGFYTTTGIGHQPYRVVLSFE
LLNAPATVCGPKLSTDLIK
```

>>**HM19**

```
NLCPFGEVFNATKFPSVYAWERKKISNCVADYSVLYNSTFFSTFKCYGVSATKLNLDLCSNV
YADSFVVKGDDVRQIAPGQTGVIADYNYKLPDDFMGCVLAWNTRNIDATSTGNINYKYR
YLRHGKLRPFERDISNVPFSPYGKPCPTVPAPNAYWPLNGYGFYTYTGIGEOPYRVVLSFE
LLNAPATVCGPKLSTDLIK
```

>>**HM21**

```
NLCPFGEVFNATKFPSVYAWERKKISNCVADYSVLYNSTFFSTFKCYGVSATKLNLDLCSNV
YADSFVVKGDDVRQIAPGQTGVIADYNYKLPDDFMGCVLAWNTRNIDATSTGNINYKYR
YLRHGKLRPFERDISNVPFSPYGKPCPTVPAPNAYWPLNGYGFYTYTGIGNQPYRVVLSFE
LLNAPATVCGPKLSTDLIK
```


Figure S6. Time series plots of C^α-RMS deviations for the selected designed ACE2-complexes. RMS deviations were computed for each sampled snapshot for the respective ACE2-complexes (see section 3.7) considering both ligand and receptor chains, subsequent to their superposition on the corresponding initial ACE2-complex (i.e., starting structures of the MD simulation) by TM-align [6]. The superposed coordinates (of the snapshots) were built from translational and rotational parameters returned by TM-align. The X-axes represents the time frames (1 to 2000) corresponding to the full 200 ns MD simulation trajectories (sampled at 100 ps interval).

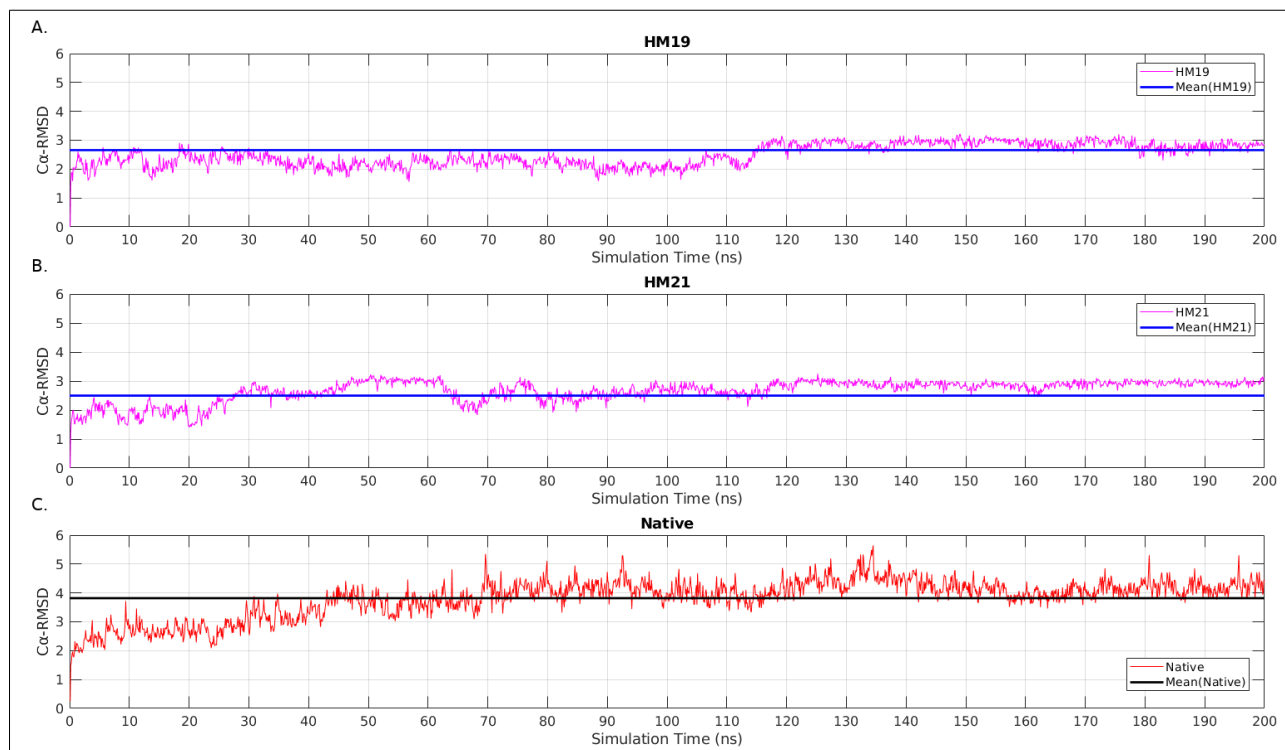


Figure S7. Electrostatic surface representation of one of the best predicted designed binary complexes (for HM21). Panel A-D represent the electrostatic surface map of the snapshot (picked up from its 200 ns MD simulation trajectory) with the highest attained EC value for HM21 (section 3.8). Rest of the figure may be described likewise to that of **Figure 3** (part of Main Manuscript). Briefly, panels A, C represent ‘self-potentials’ while B, D represent ‘partner-potentials’ realized on the ligand and receptor surface respectively for HM21. Self- and partner-potentials are as defined in the legend of **Figure 3**. Arrows indicate whether the surface potentials are due to ‘self’ (panels A, C) or ‘partner’ (panels B, D). Coloring of ‘cartoon’s are as in **Figure 3**. A direct comparison with **Figure 3** clearly shows that the match in counter-colors (red and blue’s) improves appreciably between corresponding patches on the contact surfaces (due to their respective self- and partner-potentials) with respect to that of the native ACE2-complex (see section 3.7).

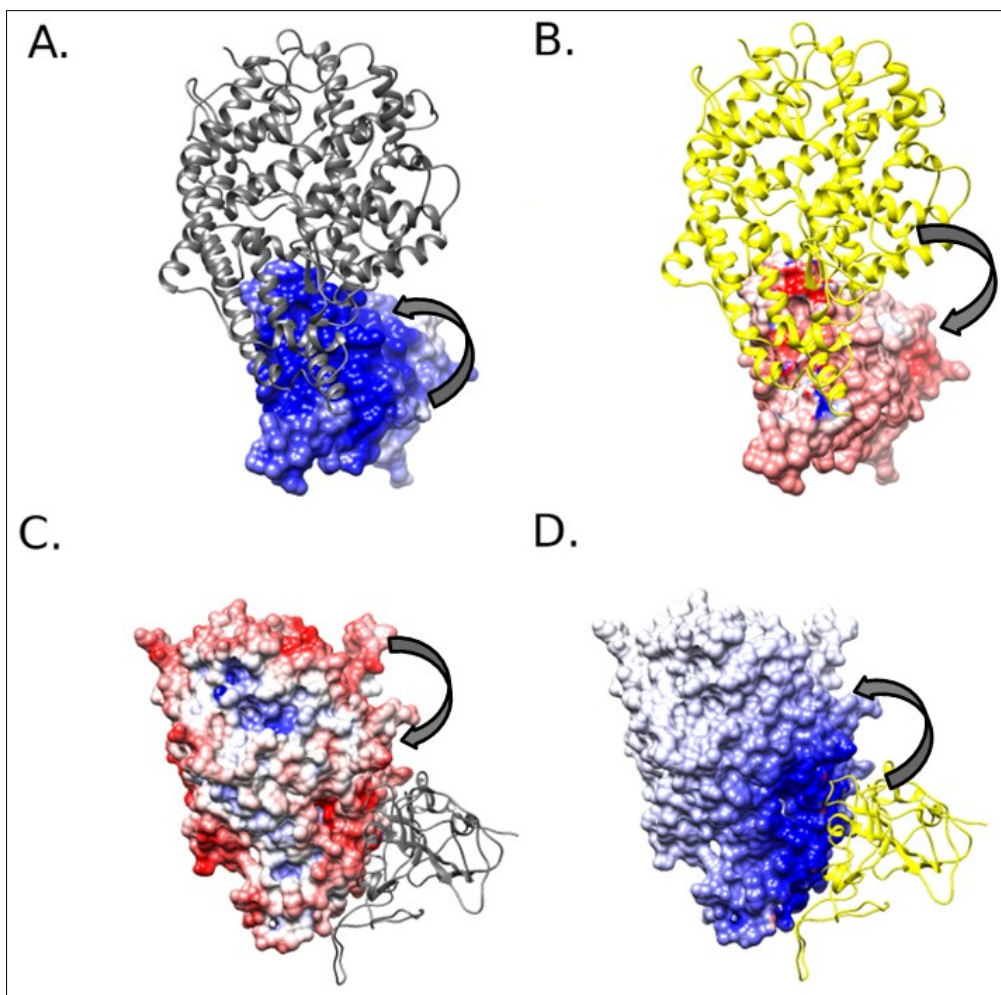


Figure S8. Time-series plots of E2d for the selected designed structural mimics in comparison to the native. Time-evolved E2d values for the ACE2-complexes (see section 3.7) pertaining to HM19, HM21 and the native has been plotted with different colors as given in the legend-box. The corresponding time-series averages are shown as dashed lines (- -) with different colors, also, as given in the legend-box and their values (along with standard deviations in parenthesis) displayed in the embedded text box. E2d is a non-negative 2d metric (i.e., a distance measure) in Euclidean space without having a defined range. The X-axis represents the simulation time (in units of ns).

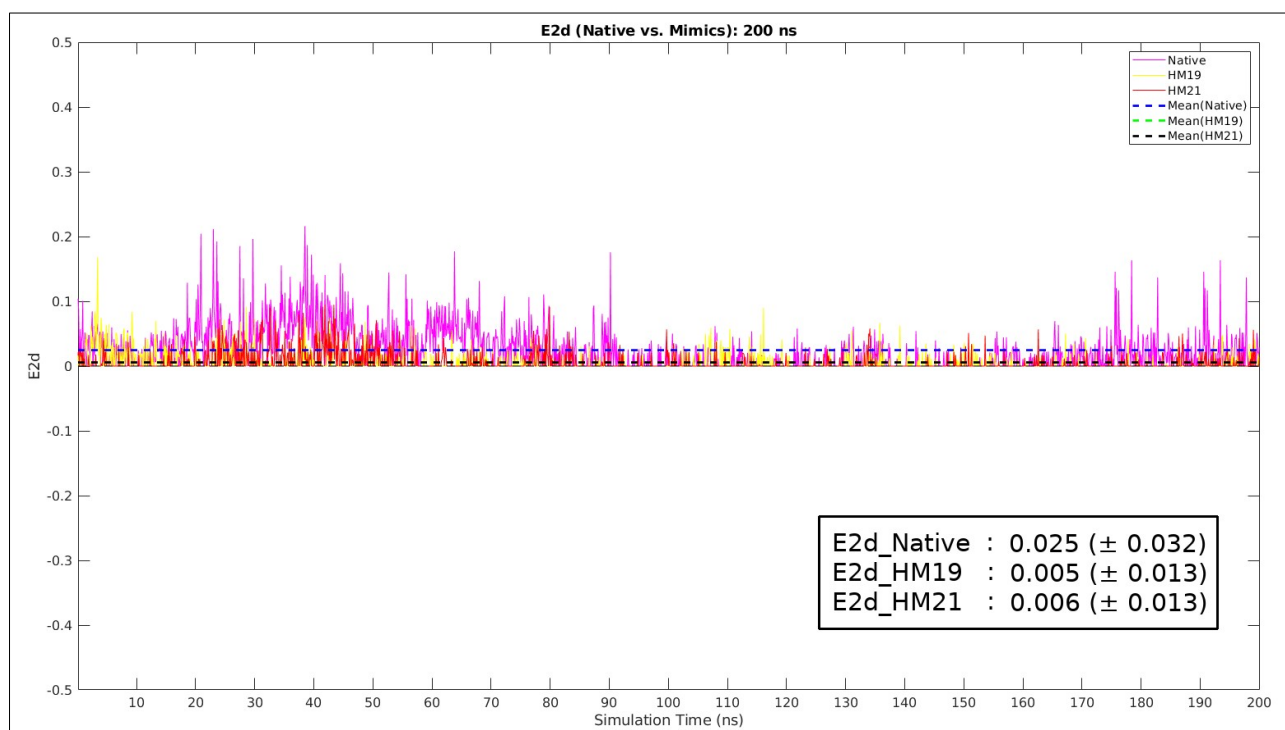


Figure S9. Population density plots for the simulated trajectories corresponding to the selected designed ACE2-complexes. These plots may also be described as the three-dimensional versions of CP_{dock} , wherein, the third dimension (i.e., the height: Z-axis) represents the normalized frequency distribution (i.e., discrete probabilities) of the points spanning the 2D {Sc, EC} map. Colorbars given beside each plot represent the raw frequencies corresponding to each X-Y square-grid of the 2D plot. The 2D grid-contours on the X-Y plane serve to demarcate the ‘optimal’ (‘probable’ + ‘less probable’) regions are drawn in ‘magenta’ and ‘blue’ respectively for ‘probable’ and ‘less probable’ regions. The surface plots have been constructed with enough transparency so that the 2D grid contours (on the X-Y plane) does not get obscured.

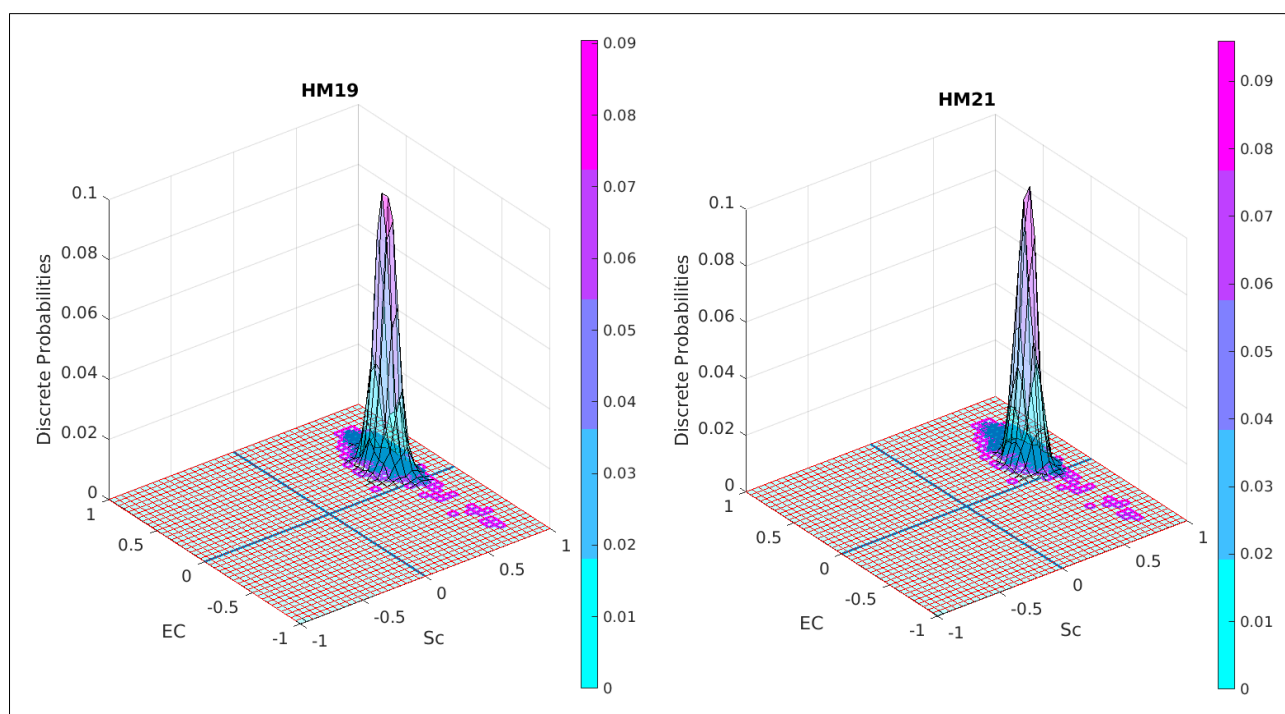


Figure S10. Structural analysis of the ACE2 receptor and angiotensin II to test the possibility of conflicting with the proposed inhibitor binding. In the top panel of the composite figure, from left to right are displayed respectively the NMR structure (1N9V) of the octa-peptide angiotensin-II, the same molecule superposed onto the ACE2 binding site in RBD_{Spike} and the corresponding CLUSTAL-OMEGA [7] pairwise sequence alignment based on which the superposition was performed. The displayed RMSD (3.45 Å) is computed for a stretch of just 8 mapped amino acids. The bottom panel displays the superposed structure of 1N9V (shown as mesh) onto the RBD_{Spike} (6VW1_E) in complexation with ACE2. The superposition yields an RMSD of 4.28 Å again for a stretch of just 8 mapped amino acids. The superposed angiotensin II is clearly away from the ACE2 binding site which is displayed as solid surface. For greater visual clarity, the mesh and the surface was colored according to two different atomtype based coloring schemes in PyMol.

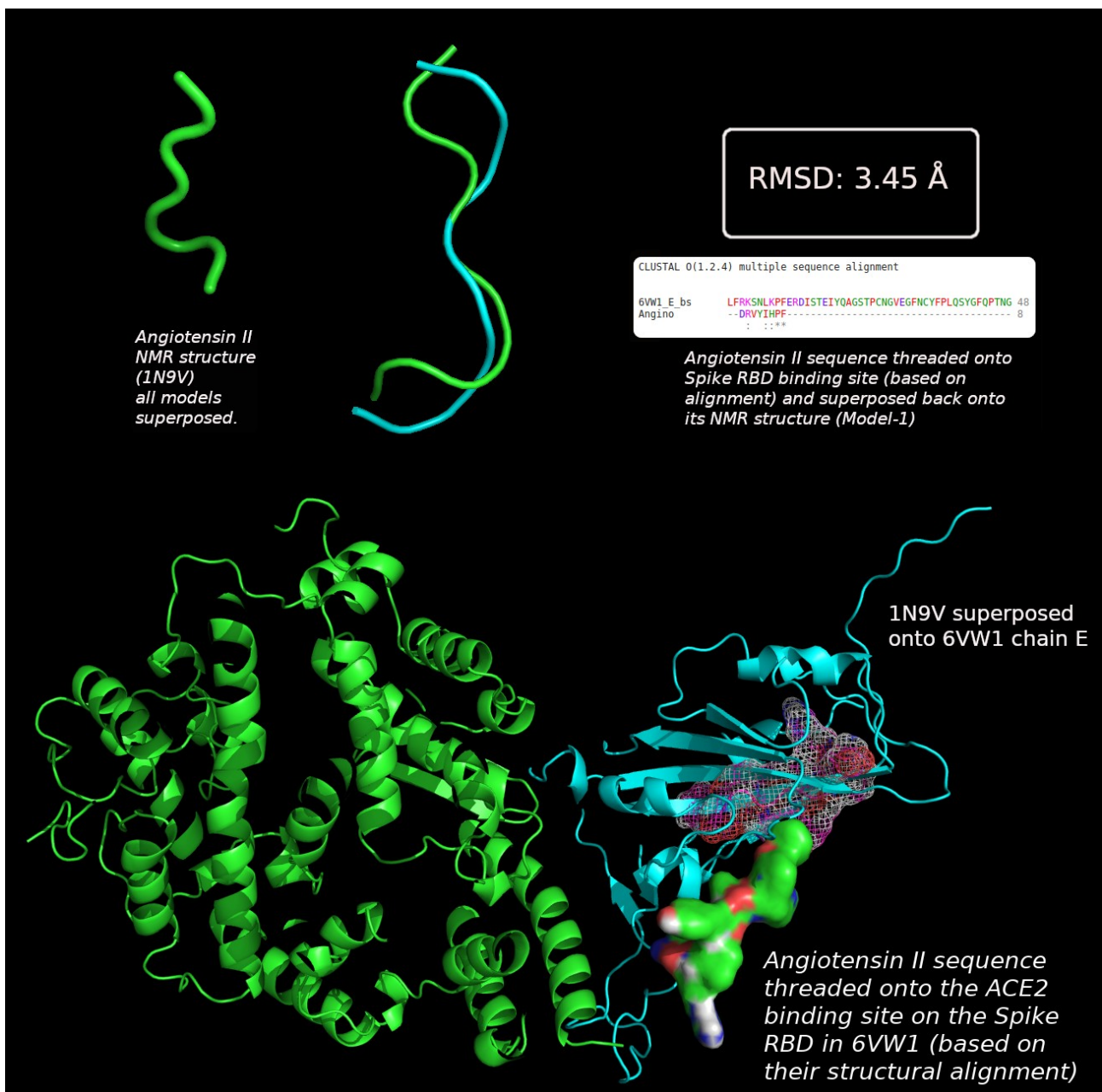


Figure S11. BRANEart visual outputs for ACE2 and Angiotensin II. Panels A and B portray the BRANEart visual outputs respectively for ACE2 (PDB ID: 6VW1, chain A) and Angiotensin II (1N9V, MODEL 1) in their free-forms. The component figures are rebuilt in PyMol from their corresponding BRANEart .pml out-files. In consistency with **Figure 11** (part of Main Manuscript) angiotensin II is displayed both as cartoon and dots (surface points). Coloring of structural regions follow the coloring scheme specified in the colorbar: blue: hydrophilic, white: neutral, red: hydrophobic (see section 3.9).

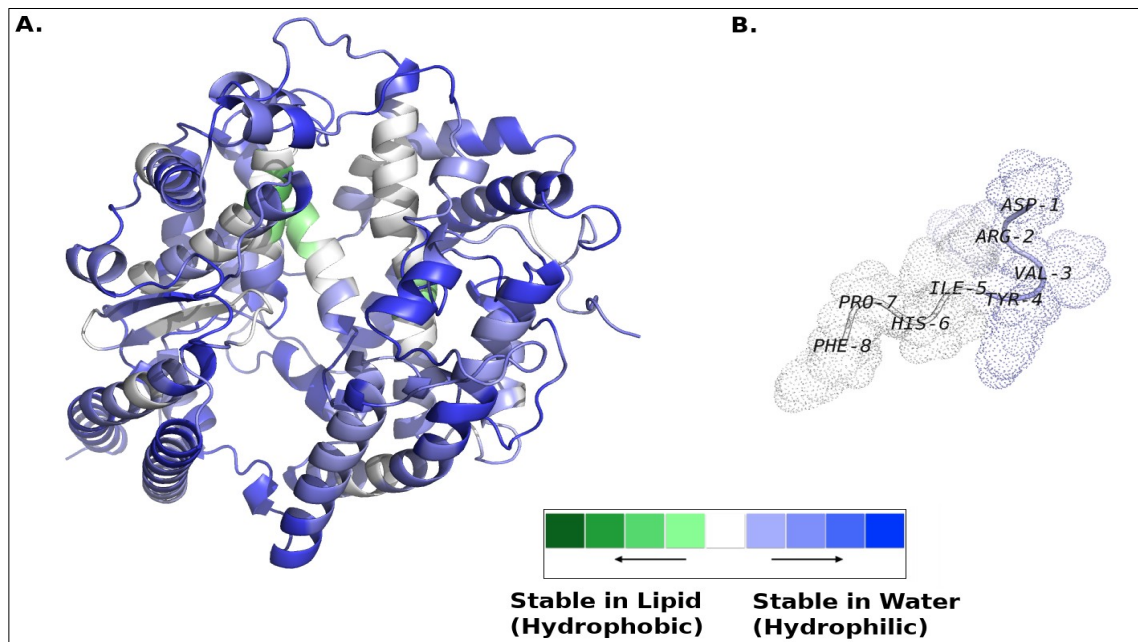
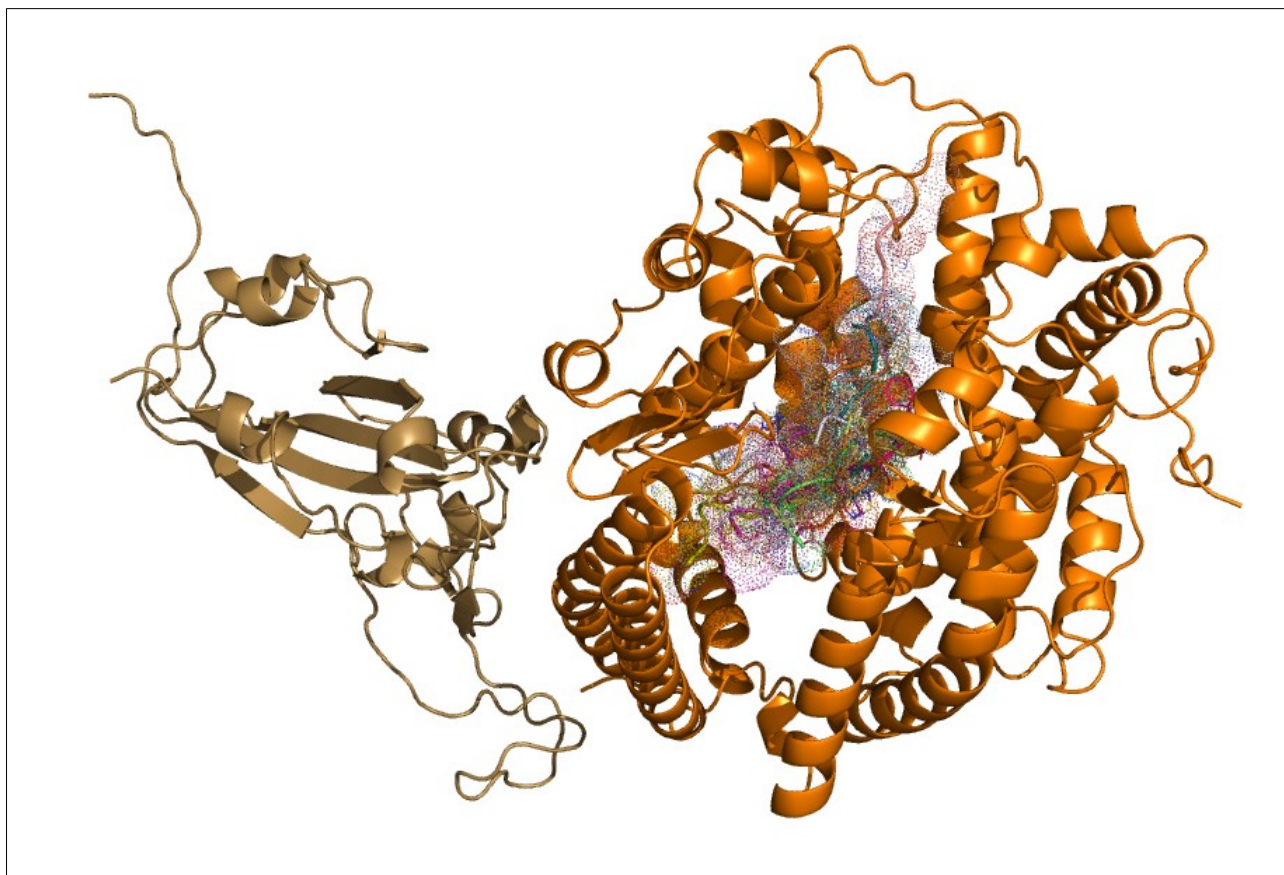


Figure S12. Cluspro docking results for Angiotensin II and the ACE2–RBD_{Spike} co-complex. As in **Figure 11** (part of the Main Manuscript), the 10 top-ranked docked poses of the ligand (angiotensin II; PDB ID: 1N9V, MODEL 1) are displayed both as cartoon and dots. The whole binary PPI complex (ACE2–RBD_{Spike}: 6VW1) is used as the receptor in the performed docking while its own receptor (ACE2) and ligand (RBD_{Spike}) chains are displayed in orange and sand (cartoons) respectively.



References

- [1] Cai Y, Zhang J, Xiao T, Peng H, Sterling SM, Walsh RM, et al. Distinct conformational states of SARS-CoV-2 spike protein. *Science* 2020. <https://doi.org/10.1126/science.abd4251>.
- [2] Shang J, Wan Y, Luo C, Ye G, Geng Q, Auerbach A, et al. Cell entry mechanisms of SARS-CoV-2. *PNAS* 2020;117:11727–34. <https://doi.org/10.1073/pnas.2003138117>.
- [3] PyMOLWiki n.d. https://pymolwiki.org/index.php/Main_Page (accessed August 4, 2020).
- [4] Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 2004;25:1605–12. <https://doi.org/10.1002/jcc.20084>.
- [5] DelPhi Suite: New Developments and Review of Functionalities - Li - 2019 - *Journal of Computational Chemistry* - Wiley Online Library n.d. <https://onlinelibrary.wiley.com/doi/full/10.1002/jcc.26006> (accessed May 25, 2020).
- [6] Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* 2005;33:2302–9. <https://doi.org/10.1093/nar/gki524>.
- [7] Sievers F, Higgins DG. Clustal Omega for making accurate alignments of many protein sequences: Clustal Omega for Many Protein Sequences. *Protein Science* 2018;27:135–45. <https://doi.org/10.1002/pro.3290>.