Supporting information for

"Critical differences between the binding features of the spike proteins of SARS-CoV-2 and SARS-CoV"

Chen Bai and Arieh Warshel

Department of Chemistry, University of Southern California, 418 SGM Building, 3620 McClintock Avenue, Los Angeles, CA 90089-1062

Email: warshel@usc.edu

Constructing the Assemblies

The protein complex construction was done by utilizing all available experimental structures, where all the structure used in this work are the combination of elements from cryo-EM or X-ray diffraction (XRD) structures. For receptor ACE2, the structure comes from cryo-EM ACE2-SARS-CoV-2 complex (PDB ID: 6VSB)[1]. For the protein body of the novel virus SARS-CoV-2, we utilized two structures, where one is the virus body from PDB 6VSB, However, this structure misses a loop near the binding domain. Another available structure is that of the XRD complex of ACE2 and the receptor binding domain of the SARS-CoV-2. In this structure the loop near the RBD is complete (PDB ID 6M0J)[2]. We combined the protein body of PDB 6VSB and the loops near RBD from 6M0J to obtain a complete SARS-CoV-2 protein body. The entire ACE2-SARS-CoV complex structure was taken from the cryo-EM structure (PDB ID: 6CS2)[3]. For the SARS-CoV antibody m396 complex , we used the XRD structure (PDB ID: 2DD8)[4]. The binding patterns of ACE2-SARS-CoV-2 and ACE2-SARS-CoV to m396-SARS-CoV are directly obtained from the above mentioned PDB structures. The binding pattern between m396 and SARS-CoV-2 was obtained by mimicking the binding pattern between m396 and SARS-CoV. The above homology-modelling is done by MODELLER[5].

After obtaining the structures, we trim them into coarse-grained (CG) representation, and perform steepest decent energy minimization and MD relaxation to eliminate unreasonable closed interactions, until the potential energy is converged. Next, we use a Monte Carlo Proton Transfer (MCPT) procedure to obtain the CG free energies (see below). All simulations and calculations were done by the MOLARIS-XG package software[6].

The binding energy was calculated by the following equations:

For the ACE2/virus complexes:

$$\Delta G_{binding} = G_{complex} - G_{ACE2} - G_{virus}$$

For the m396/virus complexes:

$$\Delta G_{binding} = G_{complex} - G_{m396} - G_{virus}$$

The energetics of the CG protein model

The development of realistic CG models in the field has been slow, partly due to the difficulties of obtaining unique points for calibration. We have used a relatively extensive benchmark of folding experiments and membrane insertion experiments. The CG model that has been continuously developed and refined in our group is based on solvation model of ionizable residues, which emphasis the key-role of electrostatic effects of the protein[7].

In our CG model, the entire side chain is represented by a CB (Fig S1). The total CG free energy is given by:

$$\Delta G_{fold}^{CG} = \Delta G_{main}^{CG} + \Delta G_{side}^{CG} + \Delta G_{main-side}^{CG}$$

The total CG folding free energy is taken relative to the free energy of the unfolded system in water at zero allied potential. The two terms represent the main chain and side chain contributions, while the third term accounts for the total protein and side chain flexibility in estimating the overall conformational entropy.

The main chain energy is given by the contributions of backbone solvation and hydrogen bonds:

$$\Delta G_{main}^{CG} = c_2 \Delta G_{solv}^{CG} + c_3 \Delta G_{HB}^{TOTAL}$$

Where c_2 and c_3 are the scaling coefficients (0.25 and 0.15, respectively).

The side chain term is given by:

$$\Delta G_{side}^{CG} = \Delta G_{side}^{elec} + \Delta G_{side}^{polar} + \Delta G_{side}^{hyd} + c_1 \Delta G_{side}^{vdw}$$

Where the first three terms are electrostatic, polar, and hydrophobic components, respectively. The last term is the van der Waals component for side chain interactions and c_1 is a scaling coefficient (0.10). The focus is placed on the electrostatic term, which is computed as a sum of change in free energy associated with charge-charge interactions between ionizable side chains $\Delta\Delta G_{QQ}^{w \to p}$, and the change in solvation free energy of those residues in their specific environment ,$\Delta\Delta G_{self}^{w \to p}$, inside the protein and in water.

$$\Delta G_{side}^{elec} = \Delta\Delta G_{QQ}^{w \to p} + \Delta\Delta G_{self}^{w \to p}$$

In case of membrane proteins we represent the membrane by a grid of unified atom. The membrane grid has a regular spacing between the membrane particles. The width of such CG membrane grid is equivalent to a hydrophobic thickness of lipid bilayer or the membrane protein under investigate. The self-energy term now is a function of the number of membrane grid points neighboring the ionized residue. However, membranes is not presented in the level of the current system.

The main chain/side chain coupling term consists of two parts, the electrostatic and the van der Waals parts:

$$\Delta G_{main-side}^{CG} = \Delta G_{main-side}^{elec} + \Delta G_{main-side}^{vdw}$$

The electrostatic part is treated with the same electrostatic interaction form as in side chain electrostatic term but with a different ε_{eff}. The van der Waals for main-side interactions consists of two parts: (a) the one where the side chain is a regular protein side chain $\Delta G^{vdw}_{main-side\ protein}$ and (b) the one where the side chain is a membrane grid atom $\Delta G^{vdw}_{main-side\ mem}$.[7b]

Finally, if electrodes and electrolytes are presented, then another term is added to the total energy:

$$\Delta G = \Delta G^{CG}_{fold} + \Delta G^{fold}_{lyte-voltage}(V_{ext})$$

The last term is the CG representation of the effect of the external potential.

For more details see ref. 7a and 7b.

Before the CG energy evaluation, we use a Monte Carlo Proton Transfer (MCPT) algorithm to obtain the charge configuration of the system[7a]. In the MCPT approach, the MC controls proton transfer between ionizable residues or between one ionizable residue and the bulk. The acceptance possibility of the move is determined by standard Metropolis criteria.

The actual MCPT can be used for time dependent study of proton transport processes as described in ref 7b. However, here we use this approach just to obtained equilibrated ionization states.

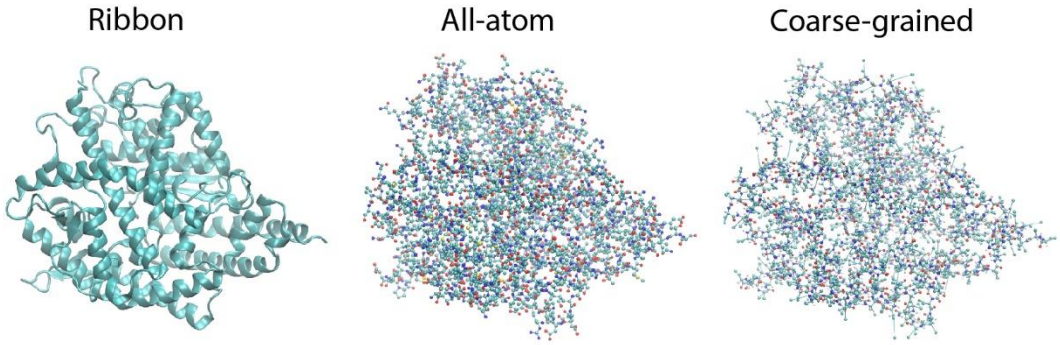

Fig. S1. A visual presentation of the CG model used in this study. The human receptor ACE2 is shown in ribbon representation on the left, in all-atom CPK representation in the middle, and in CG CPK representation on the right. In the CG model, the CB represent an entire sidechain of a residue in the CG model (see text for details). The atoms are colored cyan, red, blue and yellow for carbon (and CB), oxygen, nitrogen and sulfur, respectively.

References

1. Wrapp, D.; Wang, N.; Corbett, K. S.; Goldsmith, J. A.; Hsieh, C.-L.; Abiona, O.; Graham, B. S.; McLellan, J. S., Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **2020,** *367* (6483), 1260-1263.
2. Lan, J.; Ge, J.; Yu, J.; Shan, S.; Zhou, H.; Fan, S.; Zhang, Q.; Shi, X.; Wang, Q.; Zhang, L.; Wang, X., Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **2020**.
3. Kirchdoerfer, R. N.; Wang, N.; Pallesen, J.; Wrapp, D.; Turner, H. L.; Cottrell, C. A.; Corbett, K. S.; Graham, B. S.; McLellan, J. S.; Ward, A. B., Stabilized coronavirus spikes are resistant to conformational changes induced by receptor recognition or proteolysis. *Scientific Reports* **2018,** *8* (1), 15701.
4. Prabakaran, P.; Gan, J.; Feng, Y.; Zhu, Z.; Choudhry, V.; Xiao, X.; Ji, X.; Dimitrov, D. S., Structure of Severe Acute Respiratory Syndrome Coronavirus Receptor-binding Domain Complexed with Neutralizing Antibody. *Journal of Biological Chemistry* **2006,** *281* (23), 15829-15836.
5. Šali, A.; Blundell, T. L., Comparative Protein Modelling by Satisfaction of Spatial Restraints. *Journal of Molecular Biology* **1993,** *234* (3), 779-815.
6. (a) Lee, F. S.; Chu, Z. T.; Warshel, A., Microscopic and semimicroscopic calculations of electrostatic energies in proteins by the POLARIS and ENZYMIX programs. *Journal of Computational Chemistry* **1993,** *14* (2), 161-185; (b) Kamerlin, S. C. L.; Vicatos, S.; Dryga, A.; Warshel, A., Coarse-Grained

(Multiscale) Simulations in Studies of Biophysical and Chemical Systems. *Annual Review of Physical Chemistry* **2011,** *62* (1), 41-64.
7. (a) Vorobyov, I.; Kim, I.; Chu, Z. T.; Warshel, A., Refining the treatment of membrane proteins by coarse-grained models. *Proteins: Structure, Function, and Bioinformatics* **2016,** *84* (1), 92-117; (b) Vicatos, S.; Rychkova, A.; Mukherjee, S.; Warshel, A., An effective coarse-grained model for biological simulations: recent refinements and validations. *Proteins* **2014,** *82* (7), 1168-1185.