

Support information for

“Predicting Mutational Effects on Receptor Binding of the Spike Protein of SARS-CoV-2 Variants”

Chen Bai^{1*}, Junlin Wang¹, Geng Chen¹, Honghui Zhang¹, Ke An¹, Peiyi Xu¹, Yang Du^{1*}, Richard D Ye^{1*}, Arjun Saha², Aoxuan Zhang², Arieh Warshel^{2*}

Warshel Institute for Computational Biology, School of Life and Health Sciences, The Chinese University of Hong Kong, Shenzhen 518172, China

Department of Chemistry, University of Southern California, 3620 McClintock Avenue, Los Angeles, California 90089, United States

Email: baichen@cuhk.edu.cn, yangdu@cuhk.edu.cn, richardye@cuhk.edu.cn, warshel@usc.edu

Constructing the complex

We assembled the protein complex by incorporating all available experimental structures, where all the structural elements were taken from cryo-EM or XRD structures. The structure of receptor ACE2 was obtained from cryo-EM ACE2-SARS-CoV-2 complex (PDB ID: 6VSB)¹. The structure of the novel SARS-Cov-2 virus, bound with ACE2 was based on PDB ID 6VSB, whereas the structure of the missing loop near the receptor binding domain (RBD) was based on PDB ID 6M0J². A Cryo-EM structure was used to model the ACE2-SARS-CoV complex (PDB ID: 6CS2)³. A XRD structure was used for the SARS-CoV antibody m396 complex (PDB ID 2DD8)⁴. The binding patterns can be directly obtained from the PDB structures for ACE2-SARS-Cov-2, ACE2-SARS-CoV and m396-SARS-CoV. As for m396-SARS-CoV-2, the binding pattern was similar to that of m396-SARS-CoV. The homology models were constructed by MODELLER⁵ in the above examples.

In the next step, the obtained structures were trimmed into coarse-grained representations⁶ and the steepest descent energy minimization was carried out, followed by MD relaxation to reduce the unreasonable closed interactions, until the potential energy was converged. Then, we estimated the free energies of the CGs with a Monte Carlo Proton Transfer (MCPT) process (see below). MOLARIS-XG software⁷ was used for all simulations and calculations. The binding energy of those complexes was calculated as mentioned in the main text.

The energetics of the CG protein model

Developing realistic CG models has been a long and arduous process, in part because the unique calibration points are difficult to obtain. The relatively extensive benchmark of folding experiments as well as membrane insertion experiments have been used. The CG model has been constantly upgraded within our team using the ionizable residues salvation model, which emphasizes the importance of electrostatic interaction in proteins^{6, 8}.

The entire side chain is represented by a CB in our CG model (Fig S1). To calculate the total CG free energy, use:

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

The total CG folding free energy is weighted by the unfolded system's free energy in water at zero allied potential. The first two terms represent the contribution of the main chains and the side chains, while the third term describes the flexibility of the total protein and side chain in estimating the overall conformational entropy.

The backbone solvation and hydrogen bonds contribute to the main chain energy.:

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

Here the scaling coefficients C_2 and C_3 have the values of a 0.25 and 0.15, respectively.

The side chain term can be expressed by:

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

The first three terms of the right-hand side, $\Delta G_{side}^{elec}, \Delta G_{side}^{polar}, \Delta G_{side}^{hyd}$, are electrostatic, polar, and hydrophobic interactions, respectively. $c_1 \Delta G_{side}^{vdw}$ represents the effective van der Waals interaction for simplified side chains and the scaling constant c_1 equals 0.10. The emphasis is on the electrostatic term, which is calculated as the sum of charge-charge interactions between ionizable side chains $\Delta \Delta G_{QQ}^{w \rightarrow P}$ in addition to the solvation free energy of those residues within their special environment $\Delta \Delta G_{self}^{w \rightarrow P}$, inside the protein and in water.

$$\Delta G_{side}^{elec} = \Delta \Delta G_{QQ}^{w \rightarrow P} + \Delta \Delta G_{self}^{w \rightarrow P}$$

Protein membranes are represented as a grid of unified atoms. In membranes, membrane particles are spaced regularly. CG membrane grids are approximately the same width as the lipid bilayer or membrane protein. In this model, the self-energy term is determined by the number of membrane grid points surrounding the ionized residue. In the current system, membranes are not presented.

The main chain/side chain coupling term, consisting of the electrostatic and the van der Waals parts, is given by:

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

Where the $\Delta G_{main-side}^{elec}$ is of the same electrostatic interaction form as in side chain electrostatic term but has a different ϵ_{eff} (10.0). The van der Waals forces for main-side interactions can be divided two parts, marked as $\Delta G_{main-side}^{vdw, protein}$ (where the side chain is a regular protein side chain) and $\Delta G_{main-side}^{vdw, mem}$ (where the side chain is a membrane grid atom), respectively⁶.

Additionally, the total energy then increases if electrodes and electrolytes are also present:

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

Where V_{ext} represents the effect of the external potential in the CG model. More details can be

found in ref. 6 and 8.

To evaluate the CG energy, we first calculated the reliable charges for the protein ionized groups using the Monte Carlo Proton Transfer algorithm (MCPT)⁸. The MC procedure involves the proton transfer of ionizable residue pairs or the ionizable residue along with the bulk. The moves of residues were accepted and employed based on the Metropolis algorithm⁹. The time-dependent processes of proton transfer can be obtained by MCPT method according to ref. 6. Even though, the purpose of this method here is merely to get the equilibrated ionization states.

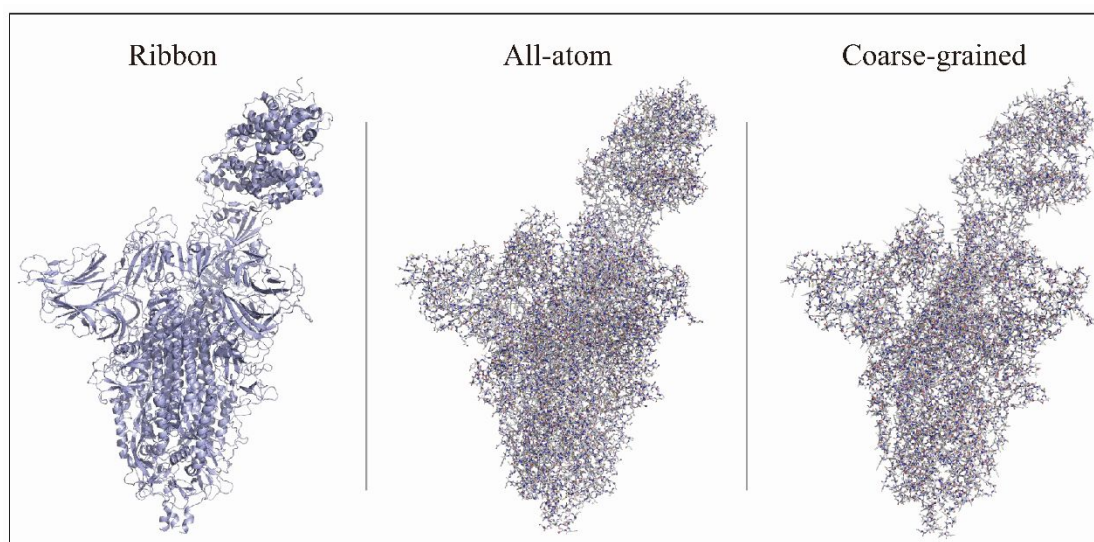


Fig. S1. CG model representation of the SARS-CoV-2 S protein bound to the ACE2 receptor. The complex model is shown in ribbon representation on the left, in all-atom stick and ball representation in the middle, and in CG stick and ball representation on the right. As mentioned afore, the entire sidechain of a residue in the CG model is unified as a CB atom. The oxygen, nitrogen, carbon (CB atoms also included), sulfur atoms are shown in red, blue, gray, and yellow, respectively.

Table S1. Term contributions for the $\Delta G1$ and $\Delta G2$ defined in main text (Unit: kcal/mol).

(A). $\Delta G1$ of ACE2-Spike complex. (B). $\Delta G1$ of m396-Spike complex. (C). $\Delta G2$ of Spike protein. The largest contribution is annotated by the red box for each row.

(A) ACE2-Spike complex							
	G_a	G_b	G_c	G_d	G_e	G_f	$\Delta G1$
Wild-type	0.0	0.0	0.0	0.0	0.0	0.0	0.0
K417N	-1.2	0.0	0.2	0.0	-0.1	-0.2	-1.3
E484K	-6.3	0.1	0.5	0.0	0.0	0.0	-5.7
N501Y	-7.3	0.0	0.1	0.0	0.0	0.1	-7.1
D614G	-1.6	-0.1	-0.1	0.0	0.1	0.1	-1.6
Africa	0.3	0.0	-7.0	0.7	-0.1	-0.2	-6.3
UK	-6.0	-0.3	4.6	-0.1	-0.1	-0.1	-2.0
Delta	-2.0	0.0	2.7	0.0	0.0	-0.2	0.5

(B) m396-Spike complex							
	G_a	G_b	G_c	G_d	G_e	G_f	$\Delta G1$
Wild-type	0.0	0.0	0.0	0.0	0.0	0.0	0.0
K417N	18.4	0.0	-7.8	-1.8	0.1	0.5	9.4
E484K	17.4	0.1	-2.4	-2.6	0.0	0.4	12.9
N501Y	0.0	0.0	0.0	0.0	0.0	0.1	0.1
D614G	0.8	-0.1	-0.1	0.0	0.0	0.1	0.7
Africa	10.1	0.0	4.1	-1.0	-0.1	0.3	13.4
UK	18.2	-0.3	0.1	-2.5	-0.2	0.7	16.0
Delta	27.5	0.0	-2.9	-1.3	0.0	0.2	23.5

(C) Spike protein							
	G_a	G_b	G_c	G_d	G_e	G_f	$\Delta G2$
Wild-type	0.0	0.0	0.0	0.0	0.0	0.0	0.0
K417N	-2.5	0.0	-0.5	0.0	0.1	0.1	-2.8
E484K	-0.8	-0.1	-0.2	0.0	0.0	0.0	-1.1
N501Y	4.4	0.0	-2.7	0.1	0.0	-0.1	1.7
D614G	-3.7	0.1	-6.0	0.0	-0.1	-0.1	-9.8
Africa	-4.6	0.0	-0.4	0.0	0.1	0.1	-4.8
UK	3.8	0.3	-7.9	0.0	0.1	0.2	-3.5
Delta	-1.6	0.2	-7.8	0.0	0.1	0.2	-8.9

G_a : the electrostatic energy term determined by whole residue charges.

G_b : the empirical term that takes into account the effect of protein size when calculating the folding free energy.

G_c : the scaled hydrophobic energy term.

G_d : the scaled van der Waals energy term.

G_e : the negative of a scaled charge-charge energy estimate of an unfolded protein.

G_f : the polar interaction contribution.

Table S2. The distance changes (Unit: Å) between the mutation sites and the nearby

interacting residues. The positively charged residues are colored by blue. The negatively charged residues are colored by red. The hydrophobic residues are colored by gray.

	Residue	The distance in the wildtype structure	The distance in the mutation structure	The change of distance
K417N ACE2-Spike	K424	10.2	10.3	0.1
	R454	6.5	6.7	0.2
	D30	10.2	10.1	-0.1
	D405	10.5	10.4	-0.1
K417N Spike	K424	10.2	10.4	0.2
	R454	6.6	6.8	0.2
	D405	10.3	10.2	-0.1
K417N m396-Spike	E406	8.4	8.6	0.2
	K424	10.3	9.4	-0.9
	R454	6.9	6.7	-0.2
E484K ACE2-Spike	D30	15.4	16.1	0.7
	D38	19.2	19.8	0.7
	K31	12.1	12.9	0.8
	H34	16.0	16.7	0.7
	R454	16.4	17.2	0.8
E484K m396-Spike	K417	20.7	21.4	0.7
	E471	10.6	10.8	0.2
	D467	19.6	20.0	0.4
	K417	20.9	21.8	0.9
	R454	16.6	17.5	0.9
N501Y ACE2-Spike	R457	14.6	14.4	-0.2
	E37	11.6	11.8	0.2
	D355	4.2	4.0	-0.2
	D382	12.8	12.7	-0.1
	R357	9.9	9.8	-0.1
N501Y Spike	K444	11.9	12.0	0.1
	E406	15.5	15.4	-0.1
	D405	13.0	12.9	-0.1
	D442	14.4	14.3	-0.1
	R454	22.0	21.9	-0.1
	R509	17.5	17.4	-0.1
D614G Spike	R403	11.8	11.7	-0.1
	F324	9.4	9.3	-0.1
	F592	5.6	5.4	-0.2
	V595	9.1	8.9	-0.2
	A628	7.0	6.9	-0.1

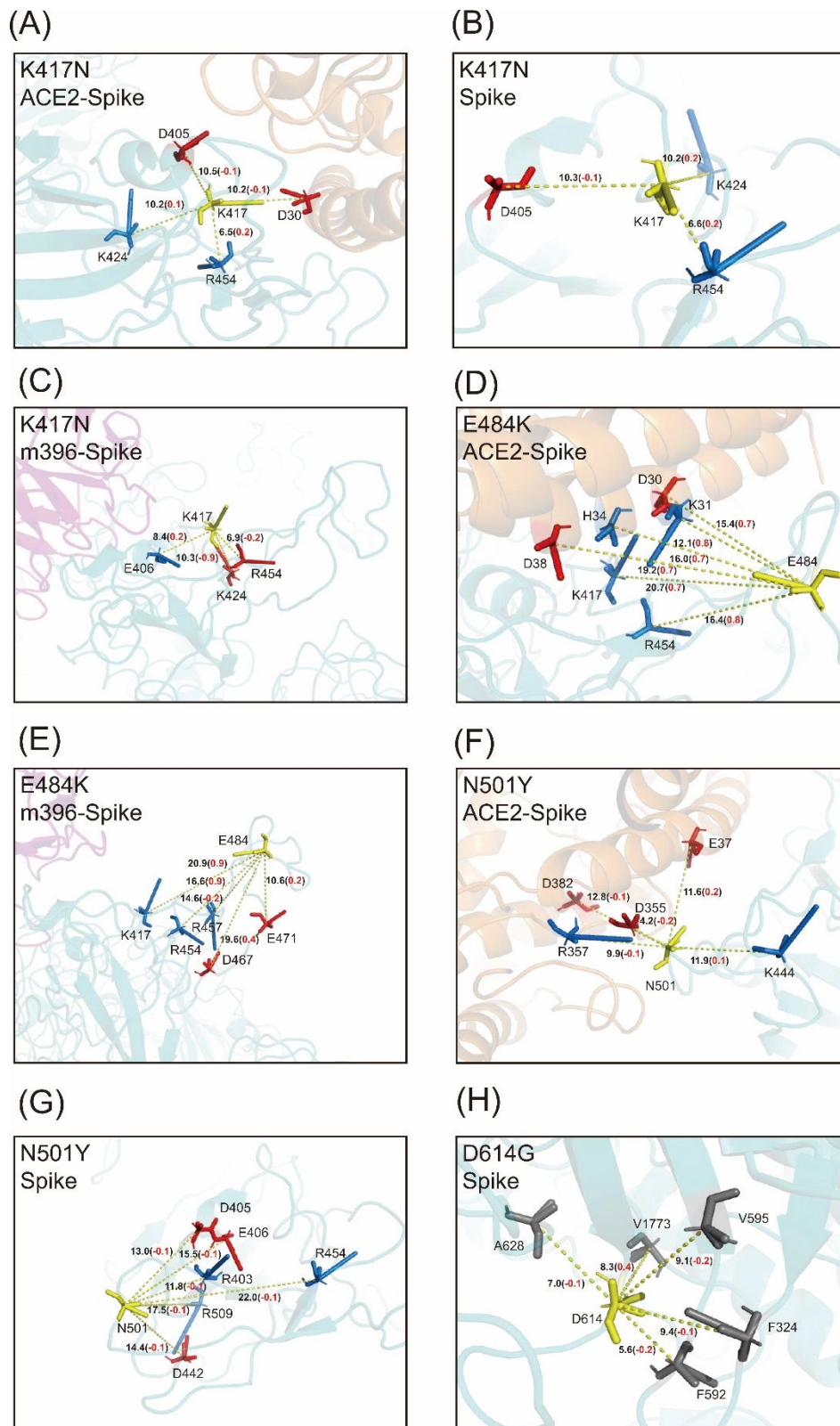


Fig. S2. The distance changes between the mutation sites and the residues listed in Table S2. Yellow: the mutation sites; Blue: positively charged residues; Red: negatively charged residues;

Gray: hydrophobic residues; Orange: ACE2 receptor; Magenta: m396 antibody; Cyan: SARS-CoV-2 spike protein.

Table S3. Root-mean-square-displacement (RMSD, Unit: Å) of different domains after mutation.

	domain	RMSD
K417N	ACE2	0.0089
	RBD	0.0298
	non-RBD	0.0005
	all	0.0079
E484K	ACE2	0.014
	RBD	0.1883
	non-RBD	0.0007
	all	0.0448
N501Y	ACE2	0.123
	RBD	0.1256
	non-RBD	0.126
	all	0.1256

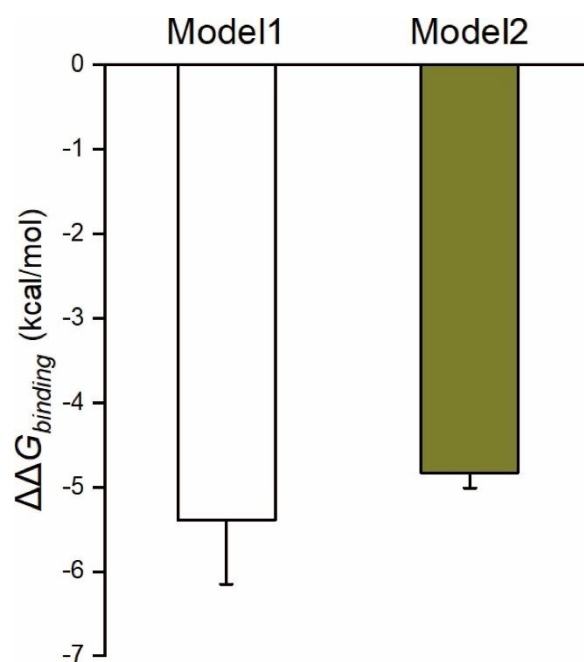


Fig. S3. $\Delta\Delta G_{binding}$ of two different models of SARS-CoV-2 N501Y mutation. Error bars represent standard error of mean. Model 1 was built based on PDB 6VSB (the current work), and model 2 was constructed using PDB 7DF3.

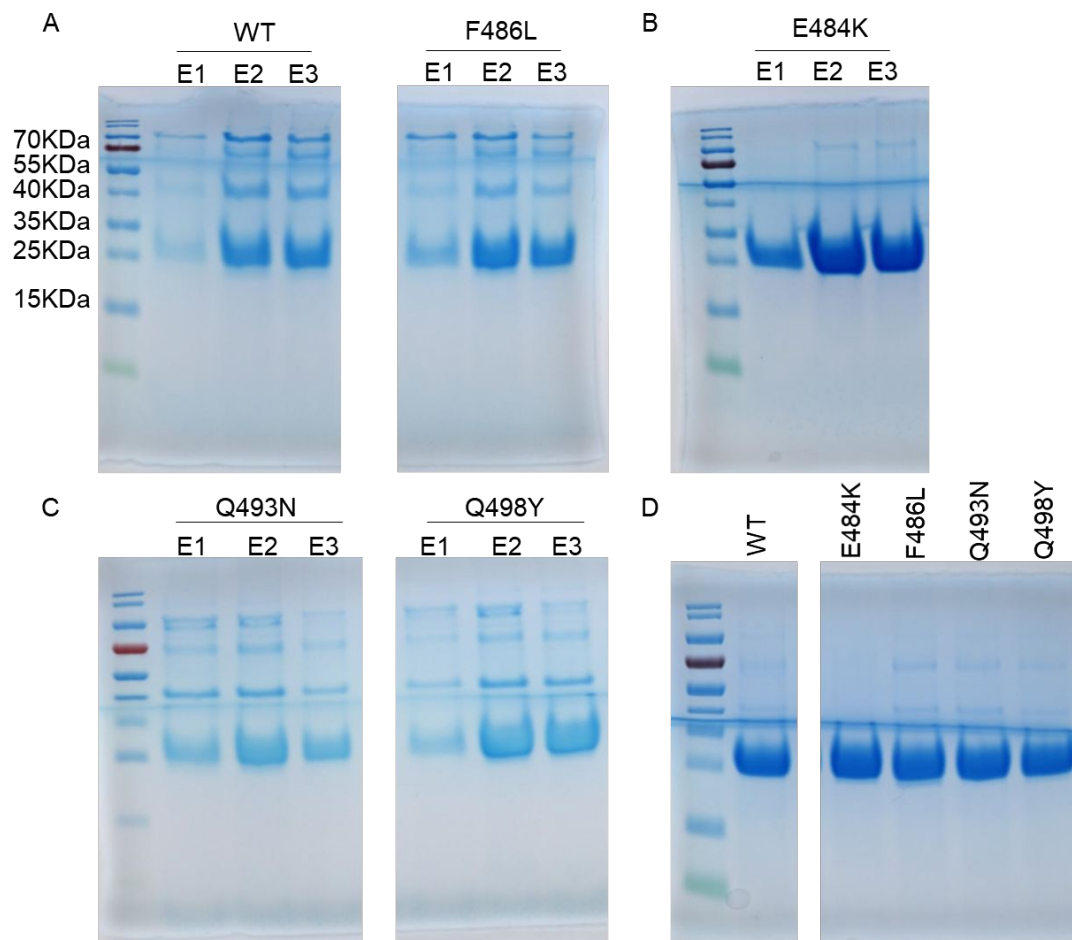


Fig. S4. SDS-PAGE of s protein WT and mutants. (A) Different Elution of s protein WT, F486L, (B) s protein E484K and (C) s protein Q493N, Q498Y from a Ni-NTA purification, E1-E3: fractions eluted with 250mM imidazole. (D) S protein WT and mutant fractions between 16mL-21mL from size exclusion chromatography.

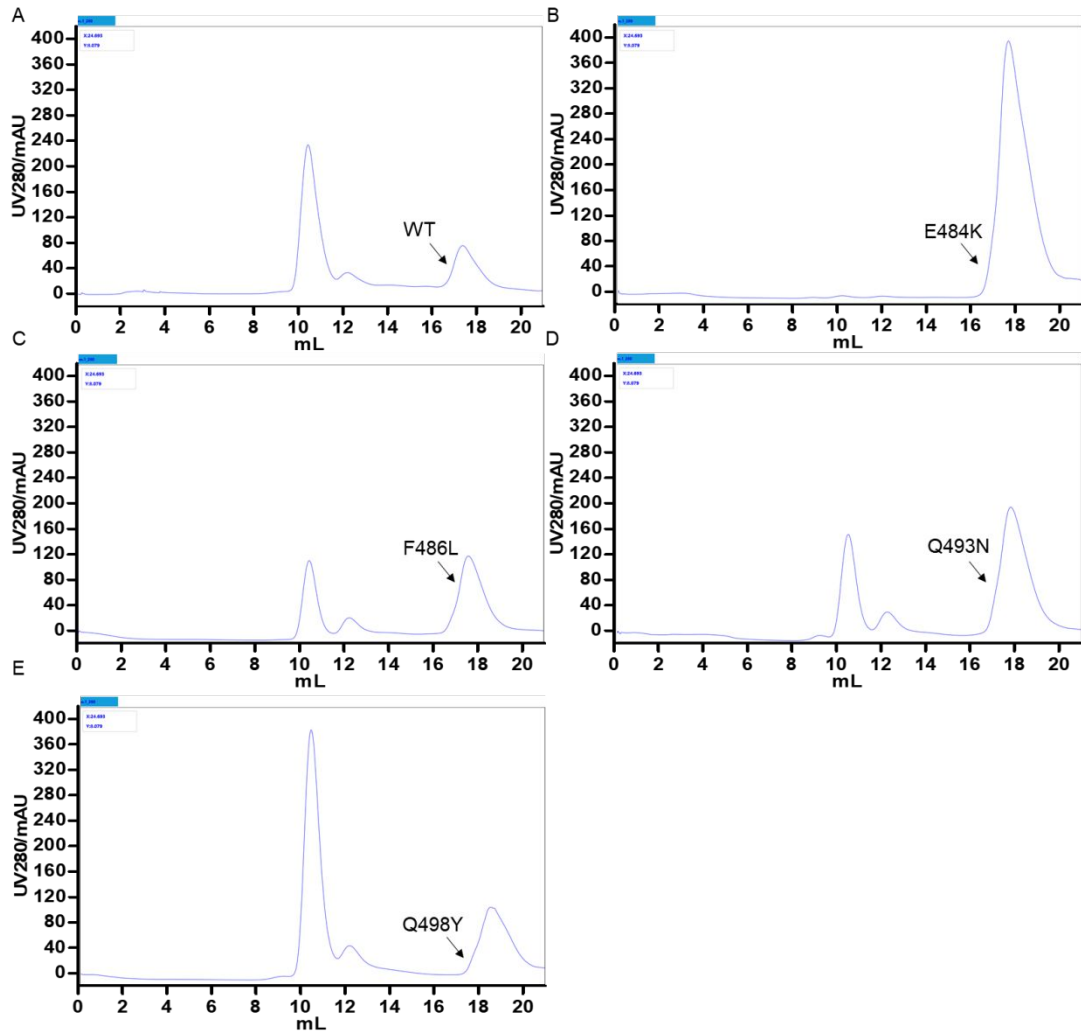


Fig. S5. Size exclusion chromatography profile of the S protein and mutants. The peak of S protein WT and mutants elution is marked with an arrow. Fractions between 16mL-21mL were pooled and concentrated for S protein/ACE2 binding experiments.

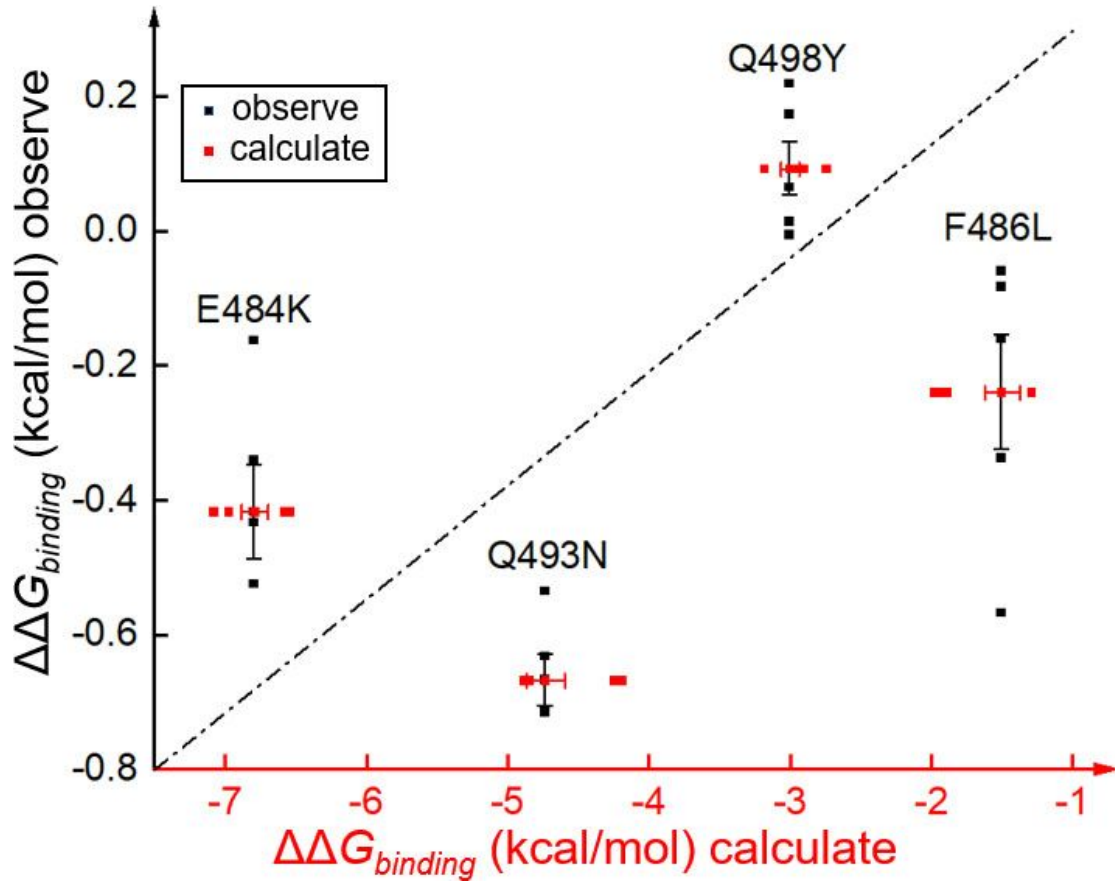


Fig. S6. Scatter plot of experimental and computational results for spike protein mutants.

The “observe” points represent the experimental results, while the “calculate” represent the computational outcomes. Error bars show the standard error of mean (S.E.M.).

Table S4. Kinetic constants and fitting parameters

	Koff (1/s)	Kon (1/Ms)	KD (M)	R ²
S protein RBD WT/ACE2	6.96x10 ⁻³	3.40x10 ⁵	2.05x10 ⁻⁸	0.990
	6.29x10 ⁻³	3.71x10 ⁵	1.69x10 ⁻⁸	0.993
	7.12x10 ⁻³	3.81x10 ⁵	1.87x10 ⁻⁸	0.994
	4.09x10 ⁻³	2.53x10 ⁵	1.62x10 ⁻⁸	0.995
	4.45 x10 ⁻³	2.96 x10 ⁵	1.51 x10 ⁻⁸	0.996
S protein RBD E484K/ACE2	4.84x10 ⁻³	3.64x10 ⁵	1.33x10 ⁻⁸	0.994
	4.93x10 ⁻³	5.01x10 ⁵	9.85x10 ⁻⁹	0.989
	4.61x10 ⁻³	6.38x10 ⁵	7.22x10 ⁻⁹	0.988
	3.10x10 ⁻³	3.68x10 ⁵	8.43 x10 ⁻⁹	0.990
	3.42 x10 ⁻³	3.94 x10 ⁵	8.66 x10 ⁻⁹	0.994
S protein RBD F486L/ACE2	6.62x10 ⁻³	4.36x10 ⁵	1.52x10 ⁻⁸	0.971
	4.35x10 ⁻³	4.39x10 ⁵	9.90x10 ⁻⁹	0.976
	3.27x10 ⁻³	4.87x10 ⁵	6.72x10 ⁻⁹	0.974
	5.68x10 ⁻³	4.16x10 ⁵	1.36x10 ⁻⁸	0.976
	5.50 x10 ⁻³	3.48 x10 ⁵	1.58 x10 ⁻⁸	0.984
S protein RBD Q493N/ACE2	3.14x10 ⁻³	4.43x10 ⁵	7.09x10 ⁻⁹	0.995
	2.93x10 ⁻³	5.60x10 ⁵	5.23x10 ⁻⁹	0.989
	3.38x10 ⁻³	6.40x10 ⁵	5.28x10 ⁻⁹	0.987
	2.17x10 ⁻³	3.83x10 ⁵	5.68x10 ⁻⁹	0.988
	2.10 x10 ⁻³	3.48 x10 ⁵	6.03 x10 ⁻⁹	0.985
S protein RBD Q498Y/ACE2	6.98x10 ⁻³	4.04x10 ⁵	1.73x10 ⁻⁸	0.984
	6.89x10 ⁻³	3.84x10 ⁵	1.79x10 ⁻⁸	0.984
	8.06x10 ⁻³	4.12x10 ⁵	1.95x10 ⁻⁸	0.968
	7.31x10 ⁻³	3.13x10 ⁵	2.34x10 ⁻⁸	0.994
	7.18 x10 ⁻³	2.84 x10 ⁵	2.53 x10 ⁻⁸	0.992

Table S5. Primer information

Primer name	Primer sequence
1COVS-F-Q493N-overlap	CCTTTAAATTCATATGGTTTCCAACCCACT
2COVS-R-Q493N-overlap	ACCATATGAATTTAAAGGAAAGTAACAATT
3COVS-F-Q498Y-overlap	GGTTTCTATCCCACTAATGGTGTTGGTTAC
4COVS-R-Q498Y-overlap	ATTAGTGGGATAGAAACCATATGATTGTAA
5COVS-F-F486L-overlap	GAAGGTCTCAATTGTTACTTTCCTTTACAA
6COVS-R-F486L-overlap	GTAACAATTGAGACCTTCAACACCATTACA
7COVS-F-E484K-overlap	GGTGTTAAAGGTTTTAATTGTTACTTTCCT

8COVS-R-E484K-overlap	ATTAAAACCTTTAACACCATTACAAGGTGT
9COVS-F	ATAAGAATGCGGCCGCATGCTACTAGTAAATCAGT CAC
10COVS-R	CCGCTCGAGTCAATGGTGATGGTGGTGATGGTGAT GTTTAGGTCCACAAACAGTTGCT

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.
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**, *581* (7807), 215-220.
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. *Sci Rep* **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. Webb, B.; Sali, A., Comparative Protein Structure Modeling Using MODELLER. *Current Protocols in Bioinformatics* **2016**, *54* (1), 5.6.1-5.6.37.
6. Vicatos, S.; Rychkova, A.; Mukherjee, S.; Warshel, A., An effective Coarse-grained model for biological simulations: Recent refinements and validations. *Proteins: Structure, Function, and Bioinformatics* **2014**, *82* (7), 1168-1185.
7. Lee, F. S.; Chu, Z. T.; Warshel, A., Microscopic and semimicroscopic calculations of electrostatic energies in proteins by the POLARIS and ENZYMIK programs. *Journal of Computational Chemistry* **1993**, *14* (2), 161-185.
8. Vorobyov, I.; Kim, I.; Chu, Z. T.; Warshel, A., Refining the treatment of membrane proteins by coarse-grained models. *Proteins* **2016**, *84* (1), 92-117.
9. Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E., Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics* **1953**, *21* (6), 1087-1092.