Binding Profile Assessment of N501Y: a More Infectious Mutation on the Receptor-Binding Domain of SARS-CoV-2 Spike Protein

Yuzhao Zhang, Xibing He, Viet Hoang Man, Jingchen Zhai, Beihong Ji, Junmei Wang*

Department of Pharmaceutical Sciences and Computational Chemical Genomics Screening Center, School of Pharmacy, University of Pittsburgh, Pittsburgh, PA 15261, USA

* Corresponding Author: Email: junmei.wang@pitt.edu

Abstract

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was first reported in December 2019 and has accumulated nearly a hundred million reported infections thereafter. This highly transmissible and pathogenic coronavirus has caused a pandemic of acute respiratory disease, coronavirus disease 2019 (COVID-19), which has caught extensive attention and greatly changed people's lifestyles all over the world. As an RNA virus, SARS-CoV-2 mutates rapidly as the virus replicates. The world health organization is now closely monitoring the emergence of a new variant, N501Y, on the spike protein. This N501Y variant is found to have higher transmission ability and infectivity, and is believed to be related to the rapid increase of COVID-19 cases in December 2020 in the UK. It was recently reported that the N501Y variants reduce neutralization sensitivity to convalescent sera and monoclonal antibodies. The Tyr mutation at 501 is located at the receptor binding domain (RBD) of the spike protein, the area that directly contacts human ACE2 (hACE2).

It’s urgent to figure out the driving force of the new mutant’s enhanced infectivity. Thus, a computational aided binding profile prediction is made to investigate the binding affinity alteration and potential structural change of the N501Y mutant. The resulting structures of N501Y mutant from MD simulations could be used to develop drug inhibitors against hACE2/RBD binding.

Keywords:

N501Y; SARS-CoV-2; spike protein; RBD; COVID-19; MD simulation; binding affinity; ACE2

Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a highly pathogenic virus, started a worldwide pandemic since December 2019.[1] The disease caused by the virus, known as COVID-19, presents a wide range of symptoms including dry cough, fever, headache, dyspnea, and pneumonia, to name a few[2]. The overall estimated mortality of COVID-19 is 2~5%.[3, 4] After more than one year of spreading, numerous variants of SARS-CoV-2 have been discovered and caught great attention, especially the variants on viral proteins that are structurally and functionally crucial to the virus.

A new mutant with an abnormal spreading rate, N501Y, has been detected in numerous countries including the UK, the US, and Canada, and is thought to be 75% more infectious. The new mutant was officially reported by the UK where a rapid increase of COVID-19 cases occurred in December 2020.[5] According to the report, the new mutant accounts for about 60% of new infections in London, suggesting that the new variant is highly transmissible.[6] After the report of N501Y, UK imposed a harsh lockdown policy to prevent the new variant from spreading. Now, the world is closely monitoring this new mutant. As of Jan 8th, 2021; 63 N501Y infections were

reported in the US.[7] Though the variant has not been thought to cause more severe symptoms yet, it must be taken under control due to its enhanced infectivity.

At the structural level, the mutant probably binds more tightly to the human angiotensin-converting enzyme 2 (hACE2) so that it may have better chance to infect people than other types do. Although it is not conclusive, the N501Y variants can reduce neutralization sensitivity to convalescent sera and monoclonal antibodies according a report by Hu et al.[8] As the crystalized structure of the new mutant hasn't been discovered yet, it's our virtue to apply all the resources to uncover the mystery of the new mutant. In this work, a model was built for the N501Y mutant receptor binding domain (RBD) based on the prototype crystal structure. Extensive molecular dynamic (MD) simulations were performed to study the binding characteristics of the mutant to hACE2. The binding affinity of the N501Y mutant was then compared to the prototype RBD with hACE2 applying a series of computational tools. The aim is to demonstrate the distinct structural features of the mutant and its potential effect on the vaccines, to elucidate the binding free energy change at the residue level of the RBD, and to provide reasonable structures of N501Y mutant to community for the sake of structure-based drug design.

Results & Discussion

Five independent MD simulation runs were performed to assess the structural stabilities for both the prototype RBD and the N501Y mutant RBD/hACE2 complexes using AMBER18.[9] MD snapshots were collected for binding free energy calculation and binding free energy decompositions. Each MD run lasts 100 nanoseconds and the collected snapshots were applied to characterize RBD/hACE2 binding.

Structural stability and binding free energy

An average N501Y complex structure was computed using snapshots selected from equilibrium simulation phase (20ns to 120ns, see **Figure S1** in Supporting Information) when both the simulation and root-mean-square deviation (RMSD) are stable. Then the representative structure which has the smallest RMSD value compared with the average structure was selected to represent the MD ensemble. The RMSD vs. time plots exhibiting the stability of MD simulation are shown in **Figure S1**. The representative N501Y RBD/hACE2 structure was compared to the prototype RBD/hACE2 complex crystal structure, and it was found that the representative MD structure is very similar to the crystal structure, with some slight movements around the binding area (**Figure 1A**). **Figure 1B** shows the change around the 501 position in the two structures. These results suggest that the efficacy of the vaccines may not be greatly affected by the N501Y mutation itself. To better compare the model of N501Y and prototype, fluctuation of solvent accessible surface (SAS) versus time plots are shown in **Figure S2**. The figure showed the SAS change during the stable simulation phase (20ns to 120ns) at the binding surface (**Figure S2A** and **2B**) and the complex level (**Figure S2C** and **2D**). No significant SAS change is detected at the whole complex nor the binding surface level, indicating that the model change before and after the mutation is not significant. Thus, we suspected that the mutation may not a threat to vaccine afficacy from a viewpoint of shape complementarity between RBD and antibodies.

To calculate the binding free energy of the two complexes, we applied molecular mechanics Poisson Boltzmann surface area (MM-PBSA) method, an end-point approach for free energy calculation with the solvent free energy being calculated with the PBSA method[10-15] and the conformational entropy being estimated with the WASA method.[16] The free energy is calculated using equations below for the molecule in the solvent:

$$\Delta G_{MM-PBSA-WSAS} = \Delta H - T\Delta S$$

$$= \Delta E_{int} + \Delta E_{vdw} + \Delta E_{eel} + \Delta G_p^{sol} + \Delta G_{np}^{sol} - T\Delta S$$

ΔE_{int} which represents the internal energy contribution was canceled out due to the use of "Single Trajectories" sampling protocol[17]. ΔE_{vdw} and ΔE_{eel} are the van der Waals and gas phase electrostatic energies, respectively; $\Delta G_p^{sol}$and ΔG_{np}^{sol} stand for the polar and nonpolar components of the solvation free energy, respectively; ΔG_p^{sol} is calculated by solving the Poisson-Boltzmann equations using the Delphi program[13]. ΔG_{np}^{sol} is estimated using solvent accessible surface area with the surface tension coefficient of 0.00542 kcal/(mol·Å^2) and a constant of 0.92 kcal/mol[18].We found out that the N501Y mutant has a significantly lower binding free energy (-24.48 kcal/mol) compared to the prototype (-16.04 kcal/mol). The contributions of individual energy terms are listed in **Table 1**. Examination on this table, we found that van der Waals (ΔE_{vdw}) and electrostation ($\Delta E_{eel} + \Delta G_p^{sol}$) interactions make roughly equale contribution in lowering the binding free energy. The MM-PBSA-WSAS binding free energies were calculated using 1000 evenly collected MD snapshots.

Binding profile analysis

To better analyze the binding profile of the mutant, we performed energy decomposition to show the individual residue contribution to the binding free energy and the interaction between the hACE2 residues and RBD residues using 50,000 evenly collected MD snapshots. The residues with high contribution to binding free energy are potentially important for the hACE2/RBD binding affinity. The residues selected with binding free energy contribution < -0.1 kcal/mol (the lower the binding free energy is, the greater the contribution) are shown in the heatmap (**Figure 2A and 2B**) and the original data for all the residues is listed in **Table 2**. After the mutation from

Asn to Tyr, the residue at the position 501 had a significant increase in binding free energy contribution, indicating the residue may hold greater binding and structural importance.

We also calculated the interaction energy of each RBD residue with every hACE2 residue and vice versa. The most signficantly interactive residues were selected using an interaction energy cutoff of -10.0 kcal/mol (**Figure 2C and 2D**). Not surprisingly, Y501 was selected in the N501Y model but N501 was not selected in the prototype model, indicating that the strengthened interaction between Y501 with the hACE2 residues. There were also several RBD residues or residue clusters within blue rectangles have significantly increased interactions with hACE2 after the N501Y mutation, such as Ser477, and Gln493 and its neigbouring residues (**Table 2**). For the binding profile of hACE2, there are also several residules or residue clusters have distinct interaction patterns between the prototype and the mutant as shown in **Table 2** and **Figure 2A** and **2B**. The total interaction energies of residues in the green rectangles are roughly the same, while the residues in red rectangules and in blue rectangles have much more favorable interactions with RBD in the prototype and the N501Y mutant, respectively. As illustrated in **Figure 2C** and **2D**, there are much more polar contacts formed in the protein-protein binding interface of the N501Y system. These changes all contribute to the much stronger interaction between RBD and hACE2 for the N501Y mutant.

Conclusion

Through a series of computational methods, we found that the N501Y mutant had a higher binding affinity to hACE2, which may increase its infectivity; and the mutated Tyr at 501 position has a higher binding free energy contribution compared to the counterpart in the prototype. Furthermore, our energy decomposition results indicate that Y501 has stronger interaction with

the hACE2 protein, in agreement with the binding affinity enhancement. Besides N/Y501, S477 also has a much stronger interaction with hACE2 protein in the mutant. Though the N501Y mutant is found to have stronger ability to alter binding affinity and inner interaction, the overall interaction pattern of the N501Y/hACE2 complex seems comparable to that of the prototype/hACE2 complex, which may ease the concern that the overwhelming spread of this new mutant may alter the efficacy of vaccines.

Supporting Information

Figure S1 shows the time courses of the RMSD of main cahin atoms of five MD trajectories for both prototype and N501Y mutant; Figure S2 shows the fluctuations of the total solvent accessible surface areas (SAS) and the SAS change upon protein-protein binding along MD simulation time for both systems; representative MD structure of N501Y is provided in pdb format.

Acknowledgments

This work was supported by the funds from the National Science Foundation (1955260) and National Institutes of Health (P30DA035778). The authors also thank the computing resources provided by the Center for Research Computing (CRC) at the University of Pittsburgh.

Conflict of Interests

The authors declare no competing financial interest.

Reference

1. Li, Q.; Guan, X.; Wu, P.; Wang, X.; Zhou, L.; Tong, Y.; Ren, R.; Leung, K. S. M.; Lau, E. H. Y.; Wong, J. Y.; Xing, X.; Xiang, N.; Wu, Y.; Li, C.; Chen, Q.; Li, D.; Liu, T.; Zhao, J.; Liu, M.; Tu, W.; Chen, C.; Jin, L.; Yang, R.; Wang, Q.; Zhou, S.; Wang,

R.; Liu, H.; Luo, Y.; Liu, Y.; Shao, G.; Li, H.; Tao, Z.; Yang, Y.; Deng, Z.; Liu, B.; Ma, Z.; Zhang, Y.; Shi, G.; Lam, T. T. Y.; Wu, J. T.; Gao, G. F.; Cowling, B. J.; Yang, B.; Leung, G. M.; Feng, Z., Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *N Engl J Med* **2020,** *382* (13), 1199-1207.

2. Huang, C.; Wang, Y.; Li, X.; Ren, L.; Zhao, J.; Hu, Y.; Zhang, L.; Fan, G.; Xu, J.; Gu, X.; Cheng, Z.; Yu, T.; Xia, J.; Wei, Y.; Wu, W.; Xie, X.; Yin, W.; Li, H.; Liu, M.; Xiao, Y.; Gao, H.; Guo, L.; Xie, J.; Wang, G.; Jiang, R.; Gao, Z.; Jin, Q.; Wang, J.; Cao, B., Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **2020,** *395* (10223), 497-506.
3. Wang, W.; Tang, J.; Wei, F., Updated understanding of the outbreak of 2019 novel coronavirus (2019-nCoV) in Wuhan, China. *J Med Virol* **2020,** *92* (4), 441-447.
4. Wu, Y. C.; Chen, C. S.; Chan, Y. J., The outbreak of COVID-19: An overview. *J Chin Med Assoc* **2020,** *83* (3), 217-220.
5. Interim: Implications of the Emerging SARS-CoV-2 Variant VOC 202012/01. **2021**.
6. Kupferschmidt, K., Mutant coronavirus in the United Kingdom sets off alarms, but its importance remains unclear. *European News* **2020**.
7. National Center for Immunization and Respiratory Diseases (NCIRD), D. o. V. D., US COVID-19 Cases Caused by Variants. **2021**.
8. Hu, J.; Peng, P.; Wang, K.; Liu, B.-Z.; Fang, L.; Luo, F.-Y.; Jin, A.-S.; Tang, N.; Huang, A.-L., Emerging SARS-CoV-2 variants reduce neutralization sensitivity to convalescent sera and monoclonal antibodies. *bioRxiv* **2021**.
9. Case, D. A.; Cheatham III, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz Jr., K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., The Amber biomolecular simulation programs. *Journal of Computational Chemistry* **2005,** *26* (16), 1668-1688.
10. Gohlke, H.; Case, D. A., Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf. *J Comput Chem* **2004,** *25* (2), 238-50.
11. Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., 3rd, Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Accounts of chemical research* **2000,** *33* (12), 889-97.
12. Hou, T.; Wang, J.; Li, Y.; Wang, W., Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *Journal of chemical information and modeling* **2011,** *51* (1), 69-82.
13. Rocchia, W.; Alexov, E.; Honig, B., Extending the Applicability of the Nonlinear Poisson−Boltzmann Equation: Multiple Dielectric Constants and Multivalent Ions. *The Journal of Physical Chemistry B* **2001,** *105* (28), 6507-6514.
14. Kuhn, B.; Gerber, P.; Schulz-Gasch, T.; Stahl, M., Validation and use of the MM-PBSA approach for drug discovery. *Journal of medicinal chemistry* **2005,** *48* (12), 4040-4048.
15. Swanson, J. M. J.; Henchman, R. H.; McCammon, J. A., Revisiting Free Energy Calculations: A Theoretical Connection to MM/PBSA and Direct Calculation of the Association Free Energy. *Biophysical Journal* **2004,** *86* (1), 67-74.
16. Wang, J.; Hou, T., Develop and test a solvent accessible surface area-based model in conformational entropy calculations. *Journal of chemical information and modeling* **2012,** *52* (5), 1199-212.

17. Junmei, W.; Tingjun, H.; Xiaojie, X., Recent Advances in Free Energy Calculations with a Combination of Molecular Mechanics and Continuum Models. *Current Computer-Aided Drug Design* **2006,** *2* (3), 287-306.
18. Knight, J. L.; Brooks Iii, C. L., Surveying implicit solvent models for estimating small molecule absolute hydration free energies. *Journal of Computational Chemistry* **2011,** *32* (13), 2909-2923.

Table 1. The binding free energy calculation results using MM/PBSA.

	ΔE_{vdw}	ΔE_{eel}	ΔG_p^{sol}	ΔG_{np}^{sol}	$T\Delta S$	$\Delta G_{MM-PBSA-WSAS}$
N501Y	-96.43	-638.99	686.94	-10.18	-34.17	-24.48
Prototype	-91.54	-615.23	667.81	-10.1	-33.02	-16.04

Table 2. Results of binding free energy decomposition of N501Y and the prototype systems. Results for the hACE2 residues and the RBD residues are shown on the top and bottom of the table, respectively. A residue's $\Delta\Delta G_{inter}$ is colored red if $\Delta\Delta G_{inter}^{WT} - \Delta\Delta G_{inter}^{N501Y} \leq -1.0\ kcal/mol$, and blue if $\Delta\Delta G_{inter}^{WT} - \Delta\Delta G_{inter}^{N501Y} \geq 1.0\ kcal/mol$. The total $\Delta\Delta G_{inter}$ for residues in a green rectangle is roughly same for WT and N501Y mutant, while the total $\Delta\Delta G_{inter}$ for residues in a red rectangle is much lower in WT than in N501Y mutant, and the total $\Delta\Delta G_{inter}$ for residues in a blue rectangle is much lower in N501Y mutant than in WT.

hACE2						RBD					
Residue ID	$\Delta\Delta G_{inter}$		Residue ID	$\Delta\Delta G_{inter}$		Residue ID	$\Delta\Delta G_{inter}$		Residue ID	$\Delta\Delta G_{inter}$	
	WT	N501Y		WT	N501Y		WT	N501Y		WT	N501Y
SER19	-5.44	-1.86	LEU79	-2.65	-3.10	LYS403	-2.14	-1.66	GLY485	-0.65	-0.56
THR20	-0.12	-3.67	ALA80	-0.12	-0.12	ASP405	0.00	-0.20	PHE486	-6.86	-7.61
ILE21	-0.24	-0.45	MET82	-2.44	-2.41	ARG408	-0.28	-0.33	ASN487	-6.55	-5.65
GLU22	-0.13	-0.17	TYR83	-5.46	-5.64	VAL417	-5.24	-5.96	CYS488	-0.22	-0.19
GLU23	-1.26	-2.88	ASN90	-0.24	-0.36	TYR421	-0.25	-0.29	TYR489	-7.55	-7.61
GLN24	-7.01	-9.30	ASN322	-0.16	0.00	ARG439	-0.10	-0.11	PHE490	-1.51	-2.21
ALA25	-0.16	-0.17	THR324	-0.89	-0.70	THR444	-0.16	-0.14	PRO491	-0.20	-0.25
LYS26	-0.19	-0.11	GLN325	-0.53	-0.84	SER445	-0.33	-0.15	LEU492	-1.22	-1.32
THR27	-6.60	-6.26	GLY326	-0.35	-0.42	THR446	-0.42	-0.12	GLN493	-9.98	-10.95
PHE28	-2.86	-3.22	PHE327	-0.10	0.00	GLY447	-0.15	0.00	SER494	-0.88	-0.99
LEU29	-0.17	-0.17	GLU329	-0.19	-0.22	TYR449	-4.77	-1.40	TYR495	-2.42	-0.84
ASP30	-7.33	-8.45	ASN330	-1.88	-2.09	TYR453	-1.90	-1.96	GLY496	-1.73	-0.57
LYS31	-12.35	-15.43	LEU351	-0.23	-0.25	ARG454	-0.11	-0.12	PHE497	-0.42	-0.24
PHE32	-0.24	-0.24	GLY352	-0.73	-0.63	LEU455	-3.77	-3.95	GLN498	-4.21	-2.55
ASN33	-0.20	-0.22	LYS353	-15.42	-13.79	PHE456	-4.79	-5.21	PRO499	-0.49	-0.48
HIS34	-7.10	-8.01	GLY354	-2.61	-2.48	ARG457	-0.11	-0.16	THR500	-9.23	-9.14
GLU35	-4.32	-3.61	ASP355	-9.38	-8.65	LYS458	-0.15	-0.52	ASN/TYR501	-7.18	-11.21
GLU37	-4.21	-5.57	PHE356	-0.48	-0.46	TYR473	-1.89	-2.41	GLY502	-3.51	-3.66
ASP38	-5.00	-2.01	ARG357	0.00	-0.23	GLN474	-0.22	0.00	VAL503	-1.15	-1.26
LEU39	-0.13	0.00	MET383	-0.11	0.00	ALA475	-6.20	-5.53	GLY504	-0.23	-0.24
TYR41	-2.81	-3.53	ALA386	-0.70	-0.57	GLY476	-2.70	-3.06	TYR505	-10.75	-11.84
GLN42	-0.84	-0.47	ALA387	-0.34	-0.47	SER477	-0.97	-5.11	GLN506	-0.47	-0.47
LEU45	-0.57	-0.50	PHE390	-0.12	-0.12	THR478	-0.58	-0.55			
GLU75	0.00	-0.13	ARG393	-0.86	-0.86	PRO479	0.00	-0.17			
GLN76	-0.23	-0.24				GLU484	-1.73	-3.21			

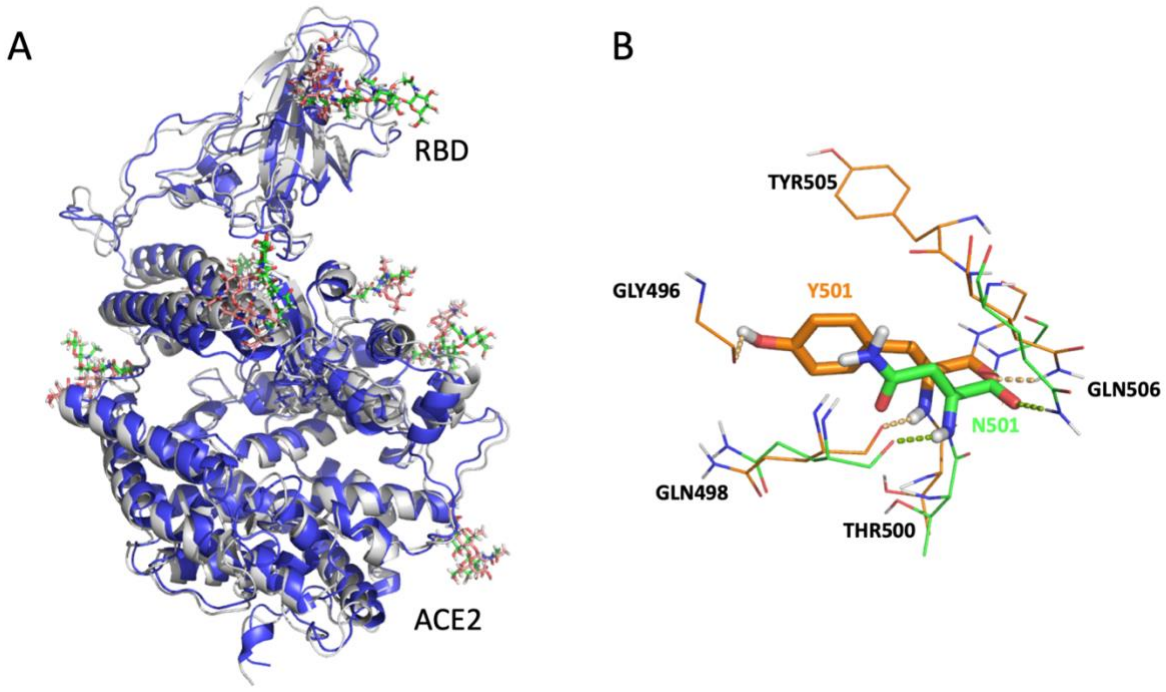

Figure 1. Panel A shows an overlay of the N501Y RBD/hACE2 model (blue) with the prototype RBD/hACE2 crystal structure (grey). Glycosylated residues are shown in green sticks for N501Y model and orange for the prototype model. Panel B shows an overlay of Y501 mutant with the N501 prototype and the interaction with the Y/N501 around 2.5 Å.

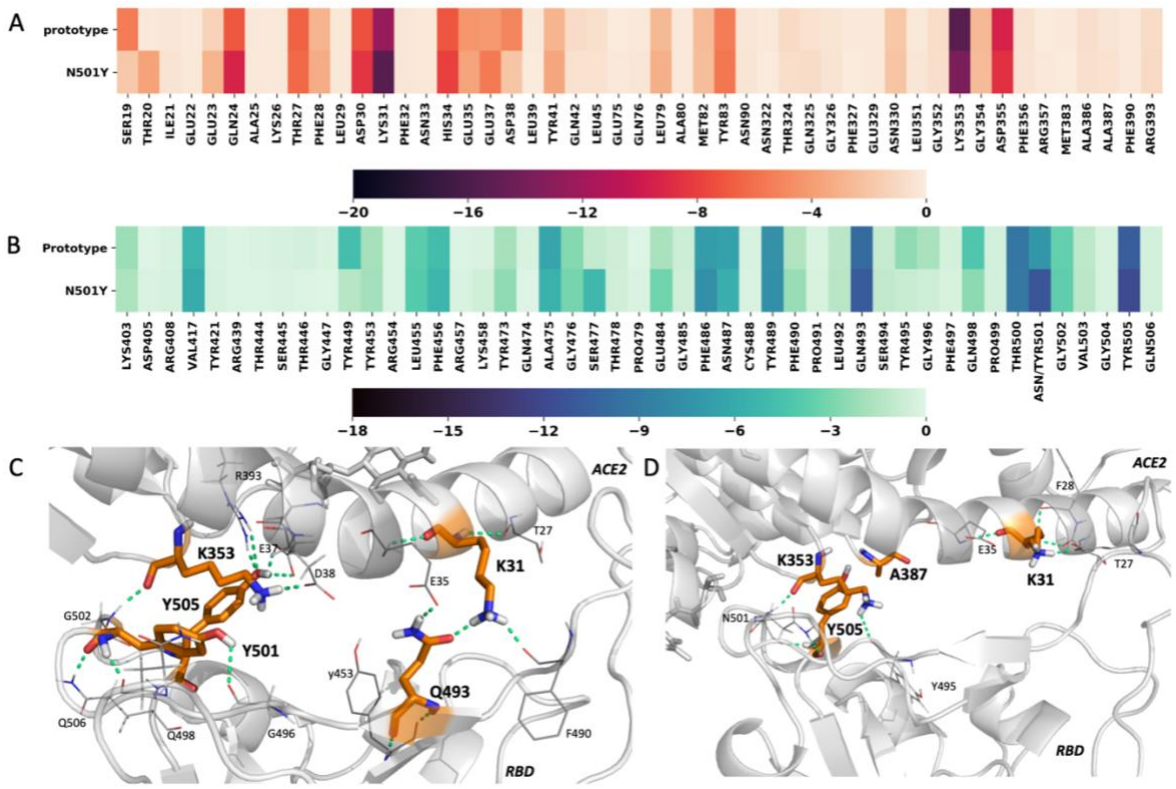

Figure 2. Panels A and B are the heatmaps that show the binding free energy contributions for selected residues from hACE2 and RBD, respectively. Panel C shows the interface between RBD and hACE2 in N501Y model. Panel D shows the interface between RBD and hACE2 in prototype model. Hydrogen bonds formed between hACE2 and RBD residues are shown as green dashed lines. The location of N/Y501 in RBD is highlighted by a red rectangle.