

Computational Investigation of Increased Virulence and Pathogenesis of SARS-CoV-2 Lineage B.1.1.7

N. Arul Murugan[†], Prashanth S. Javali^γ, Chitra Jeyaraj Pandian[‡], Muhammad Akhtar Ali^ϑ, Vaibhav Srivastava^ϑ, Jeyakanthan Jeyaraman^γ

[†] Department of Theoretical Chemistry and Biology, School of Engineering Sciences in Chemistry, Biotechnology and Health, KTH Royal Institute of Technology, S-106 91, Stockholm, Sweden

[‡] Department of Biotechnology, Dr. Umayal Ramanathan College for Women, Karaikudi, India

^ϑ Division of Glycoscience, Department of Chemistry, School of Chemistry, Biotechnology and Health, KTH Royal Institute of Technology, Stockholm, Sweden

^γ Department of Bioinformatics, Alagappa University, Karaikudi, Tamilnadu, India

Abstract

New variants of SARS-CoV-2 are being reported worldwide. More specifically, the variants reported in South Africa (501Y.V2) and United Kingdom (B.1.1.7) were found to be more contagious than the wild type. There are also speculations that the variants might evade the host immune responses induced by currently available vaccines and develop resistance to drugs under consideration. The first step of viral infection in COVID-19, occurs through the interaction of receptor binding domain (RBD) of the spike protein with peptidase domain of the human ACE-2 (hACE-2) receptor. So, possibly the mutations in the RBD domain of spike protein in the new variants could modulate the protein-protein interaction with hACE-2 receptor leading to the increased virulence. In this study, we aim to get molecular level understanding into the mechanism behind the increased infection rate due to such mutations in these variants. We have computationally studied the interaction of the spike protein in both wild-type and B.1.1.7 variant with hACE-2 receptor using combined molecular dynamics and binding free energy calculations using molecular mechanics-Generalized Born surface area (MM-GBSA) approach. The binding free energies computed using configurations from minimization run and low temperature simulation show that mutant variant of spike protein has increased binding affinity for hACE-2 receptor (i.e. $\Delta\Delta G(N501Y,A570D)$ is in the range -20.4 to -21.4 kcal/mol) The residue-wise decomposition analysis and intermolecular hydrogen bond analysis evidenced that the N501Y mutation has increased interaction between RBD of spike protein with ACE-2 receptor. We have also carried out calculations using density functional theory and the results evidenced the increased interaction between three pairs of residues (TYR449 (spike)-ASP38 (ACE-2), TYR453-HIE34 and TYR501-LYS353) in the variant that could be attributed to its

increased virulence. The free energies of wild-type and mutant variants of the spike protein computed from MM-GBSA approach suggests that latter variant is stable by about -10.4 kcal/mol when compared to wild type suggesting that it will be retained in the evolution due to increased stability. We demonstrate that with the use of the state-of-the art of computational approaches, we can in advance predict the more virulent nature of variants of SARS-CoV-2 and alert the world health-care system.

Keywords SARS-CoV-2; B117 variant; Spike protein:ACE-2 receptor complex; COVID-19; Molecular Mechanics-Generalized Born Surface Area approach; QM fragmentation scheme.

1 Introduction

A distinct phylogenetic cluster of Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) referred as VUI 202012/01 (Variant Under Investigation, year 2020, month 12, variant 01) belonging to the lineage B.1.1.7 was recently identified in the United Kingdom through viral genome sequencing.¹ It was found that this variant has unusual multiple mutations especially in the spike protein namely N501Y, A570D mutations in receptor binding domain (RBD) and H69/V70, 144/145 deletion mutations in S1 sub unit besides P681H, T716I, S982A, D1118H mutations in S2 sub unit.² Driven by the increased infectiousness of this new strain, about 50,000 individuals per day had been infected in the United Kingdom till first week of January 2021 with the first case being identified on 20th September, 2020.³ VUI 202012/01 cluster varies by 29 nucleotide substitutions from the Wuhan strain that is much higher than the recent molecular clock predictions of about 2 substitutions per month.¹ The non-synonymous mutations in the spike protein for VUI 202012/01 variant is higher than the expected random mutations and about 27 % of 22 nucleotide substitutions in the S-gene are acquired from common ancestor next strain clade 20B.⁴ Phylogenetic analysis of this variant evidenced 14 non-synonymous mutations involving amino acid alterations; 6 synonymous non-amino acid altering mutations and 3 deletion mutations.¹ The six synonymous mutations include five mutations in ORF1ab namely C913T, C16176T, C14676T, C5986T, C15279T and one mutation in M gene viz. T26801C.¹ Three deletion mutations in VUI 202012/01 variant comprise of i) several spontaneous double deletions (69/70 deletion) that possibly cause conformational change in the spike protein; ii) spontaneous multiple mutation (P681H) near high variability S1/S2 furin cleavage site and iii) open reading frame 8 (ORF8) stop codon (Q27stop) mutation.⁵ The N501Y mutation occurring in the RDB of spike protein is proposed to increase the binding affinity of the spike protein to murine and human angiotensin-converting enzyme-2 (ACE-2) receptors.^{6,7} Accumulation of high mutation rates within short time periods suggest that this variant might not have emerged from gradual accumulation of mutations but could have arisen by selection pressure through inpatient virus genetic diversity after convalescent plasma treatment.⁸ The probable sources that has given rise to this variant with multiple mutations can be explained by three possible processes namely i) prolonged SARS-CoV-2 infection in immune-deficient/suppressed patients leading to accumulation of immune evading mutations at higher rate; ii) adaptation process of SARS-CoV-2 as in the case of Y453F mutation or 69/70 deletions that have occurred in different animal species followed by zoonotic transmission to human host and iii) antibody-mediated selective pressure creating multiple genetic changes in SARS-CoV-2 through direct selection or by genetic hitchhiking.^{9,10} Experimental studies have suggested

that N501Y mutation in the RBD region of spike protein at position 501 can increase its affinity to the ACE-2 receptor.¹¹ Animal studies using mouse model have also evidenced increased infectivity and virulence of N501Y mutation in SARS-CoV-2.¹² Moreover, mutation in the residue P681H that creates furin cleavage among S1 and S2 sub-units of the spike protein resulted in enhanced SARS-CoV-2 variant transmission and entry into host respiratory epithelial cells in animal models.^{13,14} The double deletion mutation involving amino acids at sites 69-70 in N terminal domain of the spike protein is a recurrent mutation that has been recorded in multiple lineages linked to various RDB mutations.¹⁵ High frequency of the recurrent deletion mutations in RDB region of the spike protein were evidenced in Y453F mutation in Denmark's mink-associated Y453F mutant outbreak and N439K mutation associated infections in human.¹⁶ Other significant mutation in B.1.1.7 variant occurs in ORF8 Q27 stop codon that truncate and inactivates ORF8 protein favouring further downstream mutations.¹⁷ For instance, the Singaporean strain had double deletions mutations leading to ablated ORF8 and truncated ORF7b expression besides 382nt deletion associated with reduced post-infection inflammation with milder clinical infection.¹⁷ Several studies have also reported that ORF8 deletion mutation has only modest effect on the replication of SARS-CoV-2 in the primary airway cells of human when compared to virus without this ORF8 deletion.¹⁸ Therefore, mutations in the critical residues of the RBD region plays a very important role in enhancing the interactions between the spike protein of SARS-CoV-2 and the ACE-2 receptor.¹⁹ Presence of F486 in RBD site instead of usual I472 can result in the formation of strong aromatic-aromatic interactions.²⁰ This aromatic interaction is formed between Y83 of ACE2 and E484 in CTD, instead of ionic interactions between P470 in RBD and K31 which provide higher affinity for receptor binding.²¹ Though there is no evidence that B.1.1.7 variant could produce severe illness compared to other SARS-CoV-2 strains, the following possible implications has been suggested for the new variant. The UK reports that the new variant has increased transmissible rate of about 70 % besides increase in reproductive number by 0.4.²² The UK variant was also found to have impact on diagnosis of SARS-CoV-2 as the deletion mutations in 69-70 can cause negative result from S-gene in RT-PCR assay.²³⁻²⁶ Mutations in the RBD of spike protein or other surface structures can alter the antigenic property of new SARS-CoV-2 variant that could lead to reduction in neutralization activity of antibodies and thereby resulting in higher risk of reinfection or decreased effectiveness of vaccines.²⁷ Therefore, there is an urgent need to study the mutation pattern of SARS-CoV-2 and mechanism of virulence or pathogenesis of new variant B.1.1.7 to develop effective therapeutics. Based on the above considerations, in the present study, we have comprehensively investigated the multiple mutations in the genome of new

SARS-CoV-2 variant B.1.1.7 besides its potential role in viral entry, replication or pathogenesis in comparison with the wild-type Wuhan strain and less contagious variants. This study also includes mutation pattern analysis, phylogenetic study and the possible sources of new variant B.1.1.7 through genomic analysis of SARS-CoV-2 sequences retrieved using Nextstrain data from GISAID for different geographical locations like India, Nigeria, United Kingdom and United Arab Emirates. Here, molecular dynamics simulation, free energy calculations and electronic structure theory based analysis of interactions of the residues from RBD domain of the spike protein and the ACE-2 receptor were performed to explore the stabilizing interactions in the formation of protein-protein complex and the effect of mutations (occurring in the RBD domain of the spike protein) in modulating complex stability. The findings of the study will enhance the understanding of genomic properties of new variant B.1.1.7 and the possible implications of its mutations on the viral protein structure that will offer a clue for futuristic development of novel therapeutics or potent vaccine against SARS-CoV-2.

2 Results and Discussion

The binding free energies based on MM-GBSA approach^{28,29} have been presented for configurations corresponding to minimization run and low temperature simulation (30 K and 1 atm pressure). The results corresponding to low temperature have been used since the entropic contributions are less significant in this condition. Further we will be able to compare the results from force-field methods to interaction energies computed using approaches based on quantum mechanics (QM). The binding free energies for the RBD domain of the spike protein with the ACE-2 receptor³⁰ are given in Table 1. In particular, two cases namely wild type and mutant variants of the spike proteins were considered as receptors. As we see, the binding free energy for mutant (using the configuration from the minimization) are lower by about 20 kcal/mol when compared to wild type spike protein (will be referred as set-1). The trend remains the same even in the case of calculations using trajectories corresponding to low temperature simulation. Now the difference between the binding free energies for mutant when compared to wild type with ACE-2 receptor is about -21.0 kcal/mol (will be referred as set-2). Interestingly, the binding free energies obtained in this set of calculations are quantitatively lower when compared to results from minimization run. It has to be attributed that the low temperature simulation allows the system to sample configurational phase space which corresponds to minimum free energies. The binding free energies have the contributions from van der Waals and electrostatic interactions between spike protein-ACE-2

receptor and the polar and non-polar solvation free energy contributions.^{28,29} The latter two terms account for the change in free energies associated with protein-protein complex formation in aqueous solvent like environment. The effect of solvent is to screen the charges on the amino acids and so the polar solvation free energy contributions generally are positive.³¹ The electrostatic interactions and van der Waals interactions are contributing to the stabilization of the complex formation. The overall, electrostatic interactions are against the complex formation as the sum of electrostatic interaction and polar solvation free energies are positive. The values are respectively 102 and 77 kcal/mol for wild type and mutant variant of the spike protein with ACE-2 receptor (we refer to the results corresponding to set-2) and the complexes will be referred as wt-spike:ACE-2 and mt-spike:ACE-2 respectively. Therefore, the protein-protein complexation is driven by hydrophobic interactions between the residues in the spike protein with residues in the ACE-2 receptor. The van der Waals interactions are as much as -146 and -142 kcal/mol for the wild type and mutant variant of the spike protein with ACE-2 receptor. Interestingly, the N501Y mutation has lowered the van der Waals interaction (by about 4 kcal/mol while benefited due to increased electrostatic interaction by about 25 kcal/mol). The non-polar solvation contribution to the binding free energies are almost comparable in both cases even though its overall effect is to stabilize the complex formation.

We have analysed the residue-wise contributions to the binding free energies and this analysis has been carried out for both the spike protein and the ACE-2 receptor. Since our focus is to investigate the effect of mutations on the binding free energies in the complex formation, we have only presented the results for the spike protein. As we discussed earlier, we have only considered two mutations occurring in the RBD domain namely N501Y and A570D. The latter mutation occurs in the N-terminal domain which is located far away from the interfacial region involved in the protein-protein interaction (Refer to Figure 1a). So, it is expected that the latter mutation has no or less significant effect on the protein-protein interaction and the former one is supposed to have major contribution to difference in binding free energies between the wild-type and mutant variants of the spike protein with ACE-2 receptor. Figure 2a shows the residue wise contributions to the binding free energies from the residues of the spike protein and both wild type and mutant variants were considered. The numbering of the residues has been done as if they are in the whole spike protein even though we have studied only the RBD domain of this protein (322-580). In order to see the residues that are dominantly contributing to difference in binding free energies, we have computed $\Delta\Delta E_{\text{Res}(\text{spike})}(\text{wt} \rightarrow \text{mt})$. One would expect major contribution from 501 residue as this corresponds to mutation site located in the interfacial region. However, the

contributions from many other residues are not less significant which has to be attributed to the fact that the mutation occurring in 501 position also alters the distance between the other residues of the spike protein and ACE-2 receptor residues and so the interactions are altered significantly. However the major changes are observed for two residues 417 and 501. As expected the $\Delta\Delta E_{\text{Res(spike)}}(\text{wt} \rightarrow \text{mt})$ has stabilizing contribution (i.e. negative in magnitude and amounts to -7.7 kcal/mol) to the complex formation. It is also worth noticing that in the wild type variant, the 501 residue has destabilizing contribution to the complex formation (refer to the positive value for $\Delta E_{\text{Res(spike)}}$ for this residue in the case of wild type in Figure 2a). The contribution due to residue 417 is rather positive which suggests that this residue contributes to destabilization of complex formation in the case of N501Y mutant. In addition to stabilizing contributions from residue 501, there are other residues such as 486, 489, 493, 496, 500, 502 are stabilizing the mutant N501Y with a contribution < -1 kcal/mol.

Table 1: Various contributions to the binding free energies of wild-type and mutant variants of the spike protein with the ACE-2 receptor. The free energies were computed using MM-GBSA approach and the energies are in kcal/mol.

System	ΔE_{vdw}	ΔE_{elec}	ΔG_{GB}	ΔG_{SA}	$\Delta G_{binding}$
Minimization					
wt-spike	-83.55	-1042.27	1092.09	-12.66	-46.39
B.1.1.7 spike	-91.70	-979.95	1018.80	-13.98	-66.83
$\Delta\Delta G(\text{N501Y,A570D})$					-20.44
Low temperature simulation					
wt-spike	-145.57	-1118.60	1220.65	-18.99	-62.51
B.1.1.7 spike	-142.18	-998.37	1075.55	-18.89	-83.90
$\Delta\Delta G(\text{N501Y,A570D})$					-21.39

Further to address the stability of the two protein-protein complexes (i.e wt-spike:ACE-2 and mt-spike:ACE-2), root mean square displacement (RMSD), root mean square fluctuation (RMFS) and hydrogen bond analysis were carried out using the trajectories corresponding to 50 ns production run. The results for RMSD are given in Figure 3a and 3c while the RMFS results are given in 3b and 3d. The first configuration in the production run has been used as the reference for computing RMSD and the values are presented for the spike protein and the ACE-2 protein separately. The RMSD values are shown to fluctuate over a larger range for the spike protein when compared to ACE-2 receptor which indicates the

former biomolecule undergoes larger conformational flexibility than the latter. It is expected behaviour as the RBD domain of the spike protein only has been studied in this work and the conformational stability (or structural stability) is reduced as we do not include the whole spike protein in trimeric form.³² Another interesting observation is that both the spike protein and the ACE-2 receptor in the case of mutant variant are associated with the larger RMSD values suggesting that conformational fluctuations are increased due to mutation N501Y. We have seen that this mutation is associated with lowering of the binding free energies but it is interesting to see it also contributes to larger conformational flexibility. The increased conformational flexibility has been attributed to the increased ability for the spike protein to escape from the neutralizing potential of antibodies.³³ RMSF values shown for the spike protein suggest that the such conformational flexibility occurs rather over a small range of residues namely 527-580. In the case of the ACE-2 protein bound to mutant variant of the spike protein larger RMSF values were observed signifying that the mutation occurring in spike protein can induce changes in structure and dynamics of ACE-2 protein.

We have also analysed the intermolecular hydrogen bonds in the spike protein:ACE-2 complex for the two variants namely wild type and B.1.1.7 variant. The results are presented in Table 2. The residues involved in hydrogen bonds (hbonds), fraction of configurations the hbonds survive and average hydrogen bond geometries are provided in the Table. We have only listed H bonds that have survival percentage greater than 40%. This means these hbonds were observed at least in 40% configurations in whole trajectories. The hbonds having larger survival time are the most stable. As can be observed from Table 2, the number of hbonds with survival fraction of 0.4 are 6 in the case of wt-spike:ACE-2 complex. In the case of mutant variant the number of such hbonds increased to 8 which explains why the protein-protein complex is more stable in this case. It is interesting to note that the most stable hydrogen bond (with larger survival fraction) is not the same in the case of both wt-spike:ACE-2 and mt-spike:ACE-2 complexes. In the case of wt-spike:ACE-2 complex, the most stable hydrogen bond is formed between the TYR449 of the spike protein and ASP38 of ACE-2 receptor while for mt-spike:ACE-2 complex, it is established between ASN487 (of spike protein) and TYR83 (of ACE-2 receptor). Even though the residue in mutation site is not involved in hydrogen bonding, the residues in the vicinity of the mutation site such as THR500, GLY502, GLN498 are involved in hydrogen bonding. This explains why the $\Delta\Delta E_{\text{Res}}(\text{wt} - \text{spike} - \text{mspike})$ were non-zero for the residues located in the vicinity of mutation site. As can be seen the hydrogen bonding formed by these residues in the wt-spike protein with ACE-2 are significantly affected by the mutation. There were two hydrogen bonds formed by THR500 in wild type spike protein which is reduced to single hydrogen

bond in the case of mutant variant. Further the hydrogen bond formed by GLN498 in wild type is lost in the case of mutant. In the latter case, new hbonds are formed. Single mutation occurring at 501 site significantly alters the intermolecular hydrogen bonding pattern which stabilizes the protein-protein complex formation for N501Y mutant. Further, the N501Y leads to increased number of hbonds having larger survival time and this has to be attributed to the increased interaction between the spike protein and ACE-2 receptor.

Table 2: Hydrogen bond analysis carried out for the spike protein:ACE-2 complex. Only those hbonds with survival fraction greater than 0.4 are shown here. Also the average hydrogen bond geometries are shown.

No	Acceptor	Donor	Fraction	R_{H-bond}	θ_{H-bond}
wt-spike:ACE-2 receptor complex					
1	ASP38 (ACE-2)	TYR449 (spike)	0.9378	2.6814	164.2380
2	HIE34 (ACE-2)	TYR453 (spike)	0.7929	2.8201	162.9223
3	ASP38 (ACE-2)	GLN498 (spike)	0.7187	2.8274	164.3970
4	LYS353 (ACE-2)	GLY502 (spike)	0.6827	2.8685	161.4827
5	ASP355 (ACE-2)	THR500 (spike)	0.4387	2.6902	163.6439
6	TYR41 (ACE-2)	THR500 (spike)	0.4089	2.7926	161.8899
mt-spike:ACE-2 receptor complex					
1	ASN487 (spike)	TYR83 (ACE-2)	0.9338	2.7036	159.1010
2	TYR83 (ACE-2)	TYR489 (spike)	0.8733	2.7876	159.7186
3	HIE34 (ACE-2)	TYR453 (spike)	0.7182	2.8324	162.9428
4	LYS353 (ACE-2)	GLY502 (spike)	0.5964	2.8780	162.8891
5	TYR41 (ACE-2)	THR500 (spike)	0.5867	2.8052	159.6009
6	GLN24 (ACE-2)	ASN487 (spike)	0.5507	2.8531	155.7901
7	GLU35 (ACE-2)	GLN493 (spike)	0.4276	2.8268	163.8225
8	ALA475 (spike)	SER19 (ACE-2)	0.4142	2.7294	160.5896

Finally we have analysed the interaction energies of the key residues using electronic structure theory to further validate the results from force-field based approaches. We have identified the pair of residues from the spike protein and the ACE-2 receptor appearing within a cut-off distance of 8 Å . The pair of residues separated by distances larger than this may not contribute to the stabilization of the protein-protein complex. So, they have not been considered for this analysis. In the case of configuration corresponding to minimization runs there

were about 53 such pairs and the interaction energies between them and the total sum were obtained using two different approaches. In the first method, the energies of the three systems namely residues from spike protein, ACE-2 receptor and their complexes were obtained in aqueous solvent environment and the difference in energies of the complex and subsystems (i.e. spike protein and ACE-2 receptor centered residues) gives the interaction energies of the pairs of residues of protein-protein complex in aqueous environment. The interaction energies between the pairs of residues located on the spike protein and ACE-2 receptor were obtained from the expression as shown in equation 1. In particular, we employed M06-2X/6-31+G* level of theory³⁴ and for solvent description SMD solvent model³⁵ has been adopted. The interaction energies of the protein-protein complex were computed as sum over the interaction energies between the pairs of residues and the values are presented in Table 3. The sum over interaction energies for wt-spike:ACE-2 complex is -67.5 kcal/mol while for the mt-spike:ACE-2 complex is -86.4. The $\Delta\Delta E_{\text{spike:ACE-2}}(\text{wt} - \text{spike} \rightarrow \text{mt} - \text{spike})$ corresponding to N501Y mutation is about -19.0 kcal/mol which explains why the B.1.1.7 mutation is more contagious when compared to wild type. In addition to the energetics of the subsystems and complexes in aqueous environment, this set of calculations also provide the solvation energies of these systems. These energies refer to the change in energies corresponding to the change of the environment from vacuum (or gas-like) to aqueous condition. These solvation free energies are referred as $\Delta G_{\text{Solv}_{\text{aqueous}}}$ and the difference in solvation free energies of the complex to the subsystem solvation free energies are given in the Table 3.

In the second method, the interaction energies were obtained using counterpoise correction method by employing M06-2X/6-31+G* level of theory. The values obtained using this method are presented in Table 3 as $\Delta E_{\text{BSSE}(\text{raw})}$ and $\Delta E_{\text{BSSE}(\text{cor})}$. In particular, the former one gives the ΔE as the mere energy difference between the complex and subsystems. However, the latter case the energies are corrected for basis sets superposition error^{36,37} using counterpoise correction method.³⁸ Moreover, M06-2X level of theory has been often used to estimate the interaction energies in even weak intermolecular complexes where the major contributions are due to van der Waals type interaction.³⁹ Even density functional theory with dispersion correction is known to perform reasonably well to describe such interactions.³⁹ However, we have used M06-2X level of theory to describe the interactions between the pairs of residues in the interfacial region of protein-protein complex. All the electronic structure theory calculations were carried out using Gaussian09 software.⁴⁰

$$\Delta E_{\text{aqueous}} = E(A_{(\text{spike})} : B_{(\text{ACE-2})})_{(\text{aqueous})} - E(A_{(\text{spike})})_{(\text{aqueous})} - E(B_{(\text{ACE-2})})_{(\text{aqueous})} \quad (1)$$

Since this approach, considers vacuum like environment to estimate the interactions be-

tween the residues, it tends to overestimate when the residues are oppositely charged or when a single residue is charged while the other residue is neutral. Using this approach, the sum over interaction energies computed for wt-spike:ACE-2 complex is -243.7 kcal/mol while for the mt-spike:ACE-2 complex is -304.3 kcal/mol (we are referring to BSSE corrected values). The $\Delta\Delta E_{\text{spike:ACE-2}}$ corresponding to N501Y mutation using the counterpoise correction method is about -60.6 kcal/mol which also explains why the B.1.1.7 mutation is more contagious when compared to wild type. As we mentioned this approach underestimates the interaction energies as the solvation effect is not accounted for. However, one can add to this the solvation free energies (i.e. $\Delta\text{GSolv}_{\text{aqueous}}$) which are computed in the first step and the sum has been presented as $\Delta\Delta E_{1\text{aqueous}}$ in Table 3. The change in interaction energy for the N501Y mutation is about -15.6 kcal/mol which is closer in magnitude to the values presented in the column $\Delta\Delta E_{\text{aqueous}}$. Two approaches namely the interaction energies obtained using the expression 1 and the interaction energies computed from the counterpoise method added together with solvation free energies ($\Delta\text{GSolv}_{\text{aqueous}}$) provide comparable results for the binding free energies for spike protein:ACE-2 receptor. Further both method shows that due to N501Y mutation the binding free energies are lowered by -19 and -15.6 kcal/mol. These results are consistent with the observation that B.1.1.7 mutant is 70% more contagious which has to be attributed to the increased interaction between the spike protein :ACE-2 receptor which is the first step in the infection.

We have also analysed interactions between the pairs of residues that majorly contribute to the total interaction energies. The pairs that contribute with interaction energies <-4.0 kcal/mol in the case of wt-spike:ACE-2 receptor are LYS417-ASP30, TYR505-GLU37 and TYR505-LYS353 and the $\Delta E_{(\text{spike}):(\text{ACE-2})}$ for these pairs of residues are respectively, 6.8, -6.4 and -4.7 kcal/mol. In the case of mt-spike:ACE-2 complex, interestingly there are six such pairs: LYS417-ASP30, TYR449-ASP38, TYR453-HIE34, TYR501-LYS353, TYR505-GLU37 and TYR505-LYS353. Further the contributions to the total interaction energies from these residues are respectively, -6.3, -7.9, -4.5, -5.2, -6.3 and -4.7 kcal/mol. As can be seen the increased affinity for ACE-2 receptor in the case of mutant variant of spike protein is caused by the pairs, TYR449-ASP38, TYR453-HIE34 and TYR501-LYS353. Even though the mutation occurs in site 501 only, the effect is also propagated to other residues and this is an very interesting observation in this study.

Table 3: Interaction energies ($\Delta E_{\text{aqueous}}$) computed by different approaches using electronic structure theory. The solvation energies ($\Delta \text{Solv}_{\text{aqueous}}$) associated with complex formation has been computed as the difference in solvation free energies for the complex and for the subsystems.

System	$\Delta E_{\text{aqueous}}$	$\Delta \text{Solv}_{\text{aqueous}}$	$\Delta E_{\text{BSSE}(\text{raw})}$	$\Delta E_{\text{BSSE}(\text{cor})}$	$\Delta E1_{\text{aqueous}}$
wt-spike	-67.51	191.71	-259.09	-243.69	-51.98
B.1.1.7 spike	-86.51	236.72	-323.22	-304.29	-67.57
$\Delta \Delta E(\text{N501Y, A570D})$	-19.00	45.01	-64.13	-60.60	-15.60

We have also analysed whether the increased stability of the mutant may be a cause for the retainment and propagation of this specific variant. The free energy of the spike protein also will be able to shed light into this aspect. So, we have also calculated the free energy of the both wt-spike and mt-spike proteins and listed in Table 4. Both total intramolecular energy (total energy of the protein due to bonded and non-bonded interactions between its residues) and solvation energies (both polar and non-polar contributions) are also shown. As can be evidenced the results corresponding to minimization run, the stability of mutant variant is increased by about 4.7 kcal/mol (i.e. the free energy of the mutant is lowered by this much value). The results from low temperature run also shows the same trend and suggest that mutant variant is more stable when compared to wild type by about -10.4 kcal/mol. The study also demonstrates one should be able to estimate the relative stability of spike proteins of different mutants to analyse whether it will be retained as the virus evolves.

3 Computational Methods

3.1 Mutants preparation

The B.1.1.7 is reported to have as many as 23 mutations in the viral genome. However, only 8 of them occurred in the spike protein.¹⁻³ Among these two are deletions occurring at positions HV 69-70, Y144 and others are N501Y, A570D, P681H, T716I, S982A and D1118H. As can be seen only one mutation is occurring in the RBD domain (N501Y) and another mutation is in the linker region between the RBD domain and NTD domain (A570D). Since the RBD domain of the spike protein is responsible for binding to the ACE-2 receptor, we have only studied the effect of these two mutations on the spike protein binding to the ACE-2

receptor. We have used PyMOL software⁴¹ to propose the structure for the mutant variant of the spike protein RBD domain.

3.2 Molecular Dynamics Simulation and MM-GBSA free energy calculations

Two systems of protein-protein complexes were considered in this study. (i) RBD domain of wild spike (will be referred as wt-spike) protein (322-580) with the ACE-2 receptor (19-614) (ii) double muted variant of spike protein (will be referred as mt-spike) with the ACE-2 receptor. As we mentioned above, the mutant spike protein differ from wild type in terms of the two mutations N501Y and A570D. The positions of the two mutations are shown in Figure 1. As can be seen the former mutation occurs in the RBD domain which is in direct contact with residues in the ACE-2 receptor (key residues are TYR41, GLY352, LYS353, GLY354, ASP355). The second mutation occurs in the NTD domain of the spike protein and does not have any role in dictating the binding affinity towards ACE-2 receptor. The three dimensional structure for the wt-spike:ACE-2 complex is based on the homology models build using the template structure reported for the complex in the PDB ID 7A91.⁴² The structure for the complex was build using SWISSMODEL server.⁴³ There are structures with better resolution reported for the ACE-2 receptor however, we chose this as template file as we wanted to propose the structure for the spike protein with residues in the range 322-580. The structure obtained from homology modeling has been used for preparing the input files for carrying out molecular dynamics simulations. The complex has been solvated with sufficient number of water molecules and neutralized with counter ions. The solvated complex structure for the wt-spike:ACE-2 complex has been energy minimized and followed by this low temperature simulations (at 30 K and 1 atm pressure) were carried out. Finally, MD simulations in ambient conditions (300 K and 1atm) were carried out. A similar protocol has been followed for mt-spike:ACE-2 complex except that before processing the PDB structure for the MD input file preparation, the two aforementioned mutations were introduced in the spike protein using Mutagenesis module of PyMOL software. All the finite temperature simulations were carried out isothermal-isobaric ensemble by using AMBER16 software.⁴⁴ The time step for solving Newton's equation of motion was set to 2fs. The time scale for the production runs for the wt-spike:ACE-2 and mt-spike:ACE-2 complexes was set to 50 ns. This set of simulations were carried out to study the stability of the complexes during a long time scale. In addition, the trajectories were used for computing various properties like RMSD, RMSF and hydrogen bond analysis. In particular, we were only analysed the intermolecular hydrogen bonds between the spike protein and ACE-2 receptor. The low

temperature simulation and the structure corresponding to minimization run were used for computing the binding free energies using molecular mechanics-Generalized Born surface area approach (MM-GBSA). Even for the trajectory corresponding to finite temperature simulation at 300 K and 1 atm pressure the binding free energies were computed and the trends in them reflected the results corresponding to above two sets (even though quantitatively the values were in the higher side). The reason for using the configurations from the low temperature simulation is to avoid the need for computing the entropic contributions and in addition it allows us to directly compare to the energetics obtained from subsequent QM calculations. For the QM calculations, only those pairs of residues located in the interfacial region and are separated by a distance of 8 Å were considered. Further, the residues were capped with hydrogens to meet with the valency requirements for carbon and nitrogen atoms after cutting along the peptide bonds.

Table 4: Various contributions to the free energies of wild and mutant variants of the spike protein in aqueous solvent environment. The free energies were computed using MM-GBSA approach and the energies are in kcal/mol.

System	E_{intra}	G_{polar}	$G_{non-polar}$	G_{spike}
Minimization				
wt-spike	-6552.25	-2394.82	105.24	-8841.83
B.1.1.7 spike	-6527.80	-2427.62	108.89	-8846.53
$\Delta G_{spike}(N501Y, A570D)$				-4.70
Low temperature simulation				
wt-spike	-6249.64	-2362.61	94.04	-8518.21
B.1.1.7 spike	-6263.04	-2361.27	95.70	-8528.61
$\Delta G_{spike}(N501Y, A570D)$				-10.40

4 Conclusions

Currently, a major attention is given to identify the location and spread of highly contagious variants of SARS-CoV-2 as it is feared that certain variants can escape the vaccination or develop resistance to drugs^{45,46} based therapy. Along this line the variant reported in South Africa and in United Kingdom (referred as B.1.1.7 variant) are reported to be more contagious and are reported to spread at 70% higher rate than the wild variant. It is not clearly established why these variants are highly contagious and we have undertaken a

computational investigation to identify the mechanism behind the increased virulence and pathogenesis. The infection occurs through the first step that the spike protein binds to the ACE-2 receptor in human cell. So, naturally the mutations occurring in the RBD domain should give us details about the increased virulence of these variants. Therefore, we have studied the interaction between the wt-spike with ACE-2 receptor and mt-spike interaction with the same receptor using computational approaches. The calculated binding free energies clearly showed that the interaction between the mt-spike to ACE-2 receptor was increased by about -20 to -21 kcal/mol. We have also computed the change in interaction energies due to N501Y mutation using electronic structure theory based calculations and it is consistently found that the binding free energies for ACE-2 receptor are lowered by < -17 to 19 kcal/mol for the mutant variant of spike protein. The study clearly explains why the variants showed increased virulence and we attribute this to the more stable complex formation between the spike protein:ACE-2 receptor. We have also found the increased interaction between the certain pairs of residues (such as TYR449-ASP38, TYR453-HIE34 and TYR501-LYS353) and increased number of intermolecular hydrogen bonds to be responsible for this. We hereby, demonstrate that the computational modeling can provide valuable insight on how the mutations can modulate the protein-protein interaction in the spike protein:ACE-2 complex. Finally, we have also shown the mutant variant of the spike protein is stable by -4. to -10 kcal/mol when compared to wild type suggesting that it may be retained in the virus evolution.

5 Acknowledgments

This work was supported by the grants from the Swedish Infrastructure Committee (SNIC) for the projects “In-silico Diagnostic Probes Design” (snic2020-5-2). JJ gratefully acknowledges the MHRD-RUSA 2.0 [F. 24/51/2014-U, Policy (TNMulti-Gen), Dept. of Edn. Govt. of India] for the infrastructure facilities provided to the Department of Bioinformatics, Alagappa University. JJ thanks DST INDO-TAIWAN (GITA/DST/TWN/P-86/2019 dated: 04/03/2020) for the funding.

6 Author contributions statement

NAM and JJ designed the project. MD, binding free energy and QM calculations and analysis were carried out by NAM. Genomics analysis was carried out by MAA, VS, PSJ, CJP and JJ. Manuscript was written by NAM, CJP and JJ with contributions from all

authors. All authors participated in the scientific discussion.

7 Supporting information

The supporting information includes details about genomics analysis and various mutations in B.1.1.7 variant when compared to wild type SARS-CoV-2. Further the sequences of wild type and sequence information for different mutants are provided.

References

- [1] Tang, J. W.; Tambyah, P. A.; Hui, D. S. Emergence of a new SARS-CoV-2 variant in the UK. *Journal of Infection* **2020**, (DOI :10.1016/j.jinf.2020.12.024)
- [2] Kemp, S.; Harvey, W.; Datir, R.; Collier, D.; Ferreira, I.; Carabelii, A.; Robertson, D. L.; Gupta, R. K. Recurrent emergence and transmission of a SARS-CoV-2 Spike deletion Δ H69/V70. *bioRxiv* **2020**, (DOI: 10.1101/2020.12.14.422555)
- [3] Kirby, T. New variant of SARS-CoV-2 in UK causes surge of COVID-19. *The Lancet Respiratory Medicine*
- [4] Rambaut, A.; Holmes, E. C.; Hill, V.; OToole, A.; McCrone, J.; Ruis, C.; du Plessis, L.; Pybus, O. A dynamic nomenclature proposal for SARS-CoV-2 to assist genomic epidemiology. *Nature* **2020**, *5*, 1403—1407
- [5] Pereira, F. Evolutionary dynamics of the SARS-CoV-2 ORF8 accessory gene. *Infection, Genetics and Evolution* **2020**, *85*, 104525.
- [6] Mathavan, S.; Kumar, S. Evaluation of the Effect of D614G, N501Y and S477N Mutation in SARS-CoV-2 through Computational Approach. **2020**, (DOI : 10.20944/preprints202012.0710.v1)
- [7] Ahmed, W.; Phillip, A. M.; Biswas, K. H. Stable Interaction Of The UK B. 1.1. 7 lineage SARS-CoV-2 S1 Spike N501Y Mutant With ACE2 Revealed By Molecular Dynamics Simulation. *bioRxiv* 2021–01.
- [8] Choi, B.; Choudhary, M. C.; Regan, J.; Sparks, J. A.; Padera, R. F.; Qiu, X.; Solomon, I. H.; Kuo, H.-H.; Boucau, J.; Bowman, K., et al. Persistence and evolution of SARS-CoV-2 in an immunocompromised host. *New England Journal of Medicine* **2020**, *383*, 2291–2293.
- [9] Chaw, S.-M.; Tai, J.-H.; Chen, S.-L.; Hsieh, C.-H.; Chang, S.-Y.; Yeh, S.-H.; Yang, W.-S.; Chen, P.-J.; Wang, H.-Y. The origin and underlying driving forces of the SARS-CoV-2 outbreak. *Journal of biomedical science* **2020**, *27*, 1–12.
- [10] Munnink, B. B. O.; Sikkema, R. S.; Nieuwenhuijse, D. F.; Molenaar, R. J.; Munger, E.; Molenkamp, R.; Van Der Spek, A.; Tolsma, P.; Rietveld, A.; Brouwer, M., et al. Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. *Science* **2020**, *371*, 172–177.

- [11] Starr, T. N.; Greaney, A. J.; Hilton, S. K.; Ellis, D.; Crawford, K. H.; Dingens, A. S.; Navarro, M. J.; Bowen, J. E.; Tortorici, M. A.; Walls, A. C., et al. Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell* **2020**, *182*, 1295–1310.
- [12] Gu, H.; Chen, Q.; Yang, G.; He, L.; Fan, H.; Deng, Y.-Q.; Wang, Y.; Teng, Y.; Zhao, Z.; Cui, Y., et al. Adaptation of SARS-CoV-2 in BALB/c mice for testing vaccine efficacy. *Science* **2020**, *369*, 1603–1607.
- [13] Hoffmann, M.; Kleine-Weber, H.; Pöhlmann, S. A multibasic cleavage site in the spike protein of SARS-CoV-2 is essential for infection of human lung cells. *Molecular Cell* **2020**, *78(4)*, 779–784.e5
- [14] Peacock, T. P.; Goldhill, D. H.; Zhou, J.; Baillon, L.; Frise, R.; Swann, O. C.; Kugathasan, R.; Penn, R.; Brown, J. C.; Sanchez-David, R. Y., et al. The furin cleavage site of SARS-CoV-2 spike protein is a key determinant for transmission due to enhanced replication in airway cells. *bioRxiv* **2020**, (DOI: 10.1101/2020.09.30.318311)
- [15] McCarthy, K. R.; Rennick, L. J.; Nambulli, S.; Robinson-McCarthy, L. R.; Bain, W. G.; Haidar, G.; Duprex, W. P. Natural deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. *bioRxiv* **2020**, (DOI : 10.1101/2020.11.19.389916)
- [16] Kemp, S.; Collier, D.; Datir, R.; Ferreira, I.; Gayed, S.; Jahun, A.; Hosmillo, M.; Rees-Spear, C.; Mlcochova, P.; Lumb, I., et al. Neutralising antibodies drive Spike mediated SARS-CoV-2 evasion (medRxiv). *bioRxiv* **2020**, (DOI : 10.1101/2020.12.05.20241927)
- [17] Young, B. E.; Fong, S.-W.; Chan, Y.-H.; Mak, T.-M.; Ang, L. W.; Anderson, D. E.; Lee, C. Y.-P.; Amrun, S. N.; Lee, B.; Goh, Y. S., et al. Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: an observational cohort study. *The Lancet* **2020**, *396*, 603–611.
- [18] Gamage, A. M.; Tan, K. S.; Chan, W. O.; Liu, J.; Tan, C. W.; Ong, Y. K.; Thong, M.; Andiappan, A. K.; Anderson, D. E.; Wang, L.-F., et al. Infection of human Nasal Epithelial Cells with SARS-CoV-2 and a 382-nt deletion isolate lacking ORF8 reveals similar viral kinetics and host transcriptional profiles. *PLoS pathogens* **2020**, *16*, e1009130.
- [19] 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*, 1260–1263.

- [20] Walls, A. C.; Park, Y.-J.; Tortorici, M. A.; Wall, A.; McGuire, A. T.; Velesler, D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* **2020**, *181*(2), 281–292.e6
- [21] Wang, Q.; Zhang, Y.; Wu, L.; Niu, S.; Song, C.; Zhang, Z.; Lu, G.; Qiao, C.; Hu, Y.; Yuen, K.-Y., et al. Structural and functional basis of SARS-CoV-2 entry by using human ACE2. *Cell* **2020**,
- [22] Santos, J. C.; Passos, G. A. The high infectivity of SARS-CoV-2 B. 1.1. 7 is associated with increased interaction force between Spike-ACE2 caused by the viral N501Y mutation. *bioRxiv* **2021**, 2020–12.
- [23] Wang, R.; Hozumi, Y.; Yin, C.; Wei, G.-W. Mutations on COVID-19 diagnostic targets. *Genomics* **2020**, *112*, 5204–5213.
- [24] Khan, K. A.; Cheung, P. Presence of mismatches between diagnostic PCR assays and coronavirus SARS-CoV-2 genome. *Royal Society open science* **2020**, *7*, 200636.
- [25] Rahimi, A.; Mirzazadeh, A.; Tavakolpour, S. Genetics and genomics of SARS-CoV-2: A review of the literature with the special focus on genetic diversity and SARS-CoV-2 genome detection. *Genomics* **2020**, *113*, 1221–1232
- [26] Tahamtan, A.; Ardebili, A. Real-time RT-PCR in COVID-19 detection: issues affecting the results. *Expert review of molecular diagnostics* **2020**, *20*, 453–454.
- [27] Thomson, E. C.; Rosen, L. E.; Shepherd, J. G.; Spreafico, R.; da Silva Filipe, A.; Wojcechowskyj, J. A.; Davis, C.; Piccoli, L.; Pascall, D. J.; Dillen, J., et al. The circulating SARS-CoV-2 spike variant N439K maintains fitness while evading antibody-mediated immunity. *bioRxiv* **2020**, (DOI: 10.1101/2020.11.04.355842)
- [28] Wang, E.; Sun, H.; Wang, J.; Wang, Z.; Liu, H.; Zhang, J. Z.; Hou, T. End-point binding free energy calculation with MM/PBSA and MM/GBSA: strategies and applications in drug design. *Chemical reviews* **2019**, *119*, 9478–9508.
- [29] Rastelli, G.; Rio, A. D.; Degliesposti, G.; Sgobba, M. Fast and Accurate Predictions of Binding Free Energies Using MM-PBSA and MM-GBSA. *J. Comput. Chem.* **2010**, *31*, 797–810.
- [30] Zhao, P.; Praissman, J. L.; Grant, O. C.; Cai, Y.; Xiao, T.; Rosenbalm, K. E.; Aoki, K.; Kellman, B. P.; Bridger, R.; Barouch, D. H., et al. Virus-receptor interactions of glyco-

- sylated SARS-CoV-2 spike and human ACE2 receptor. *Cell host & microbe* **2020**, *28*, 586–601.
- [31] Gohlke, H.; Case, D. A. Converging free energy estimates: MM-PB (GB) SA studies on the protein–protein complex Ras–Raf. *Journal of computational chemistry* **2004**, *25*, 238–250.
- [32] Benton, D. J.; Wrobel, A. G.; Xu, P.; Roustan, C.; Martin, S. R.; Rosenthal, P. B.; Skehel, J. J.; Gamblin, S. J. Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion. *Nature* **2020**, *588*, 327–330.
- [33] Papageorgiou, A. C.; Mohsin, I. The SARS-CoV-2 Spike Glycoprotein as a Drug and Vaccine Target: Structural Insights into Its Complexes with ACE2 and Antibodies. *Cells* **2020**, *9*, 2343.
- [34] Hohenstein, E. G.; Chill, S. T.; Sherrill, C. D. Assessment of the Performance of the M05-2X and M06-2X Exchange–Correlation Functionals for Noncovalent Interactions in Biomolecules. *J. Chem. Theory Comput.* **2008**, *4*, 1996–2000.
- [35] Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *The Journal of Physical Chemistry B* **2009**, *113*, 6378–6396.
- [36] Turi, L.; Dannenberg, J. Correcting for basis set superposition error in aggregates containing more than two molecules: ambiguities in the calculation of the counterpoise correction. *The Journal of Physical Chemistry* **1993**, *97*, 2488–2490.
- [37] Richard, R. M.; Bakr, B. W.; Sherrill, C. D. Understanding the many-body basis set superposition error: Beyond Boys and Bernardi. *Journal of chemical theory and computation* **2018**, *14*, 2386–2400.
- [38] Boys, S. F.; Bernardi, F. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Molecular Physics* **1970**, *19*, 553–566.
- [39] Marom, N.; Tkatchenko, A.; Rossi, M.; Gobre, V. V.; Hod, O.; Scheffler, M.; Kronik, L. Dispersion interactions with density-functional theory: Benchmarking semiempirical and interatomic pairwise corrected density functionals. *Journal of Chemical Theory and Computation* **2011**, *7*, 3944–3951.

- [40] Frisch, M. J. et al. Gaussian 09, Revision A.02. Gaussian, Inc., Wallingford, CT, 2009.
- [41] PyMOL: The PyMOL Molecular Graphics System, Version 1.8, Schrödinger, LLC. **2015**,
- [42] Benton, D. J.; Wrobel, A. G.; Xu, P.; Roustan, C.; Martin, S. R.; Rosenthal, P. B.; Skehel, J. J.; Gamblin, S. J. Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion. *Nature* **2020**, *588*, 327–330.
- [43] Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L., et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, *46*, W296–W303.
- [44] Case, D. et al. Amber 16, University of California, San Francisco. 2016.
- [45] Murugan, N. A.; Kumar, S.; Jeyakanthan, J.; Srivastava, V. Searching for target-specific and multi-targeting organics for Covid-19 in the Drugbank database with a double scoring approach. *Scientific reports* **2020**, *10*, 1–16.
- [46] Panda, P. K.; Arul, M. N.; Patel, P.; Verma, S. K.; Luo, W.; Rubahn, H.-G.; Mishra, Y. K.; Suar, M.; Ahuja, R. Structure-based drug designing and immunoinformatics approach for SARS-CoV-2. *Science advances* **2020**, *6*, eabb8097.

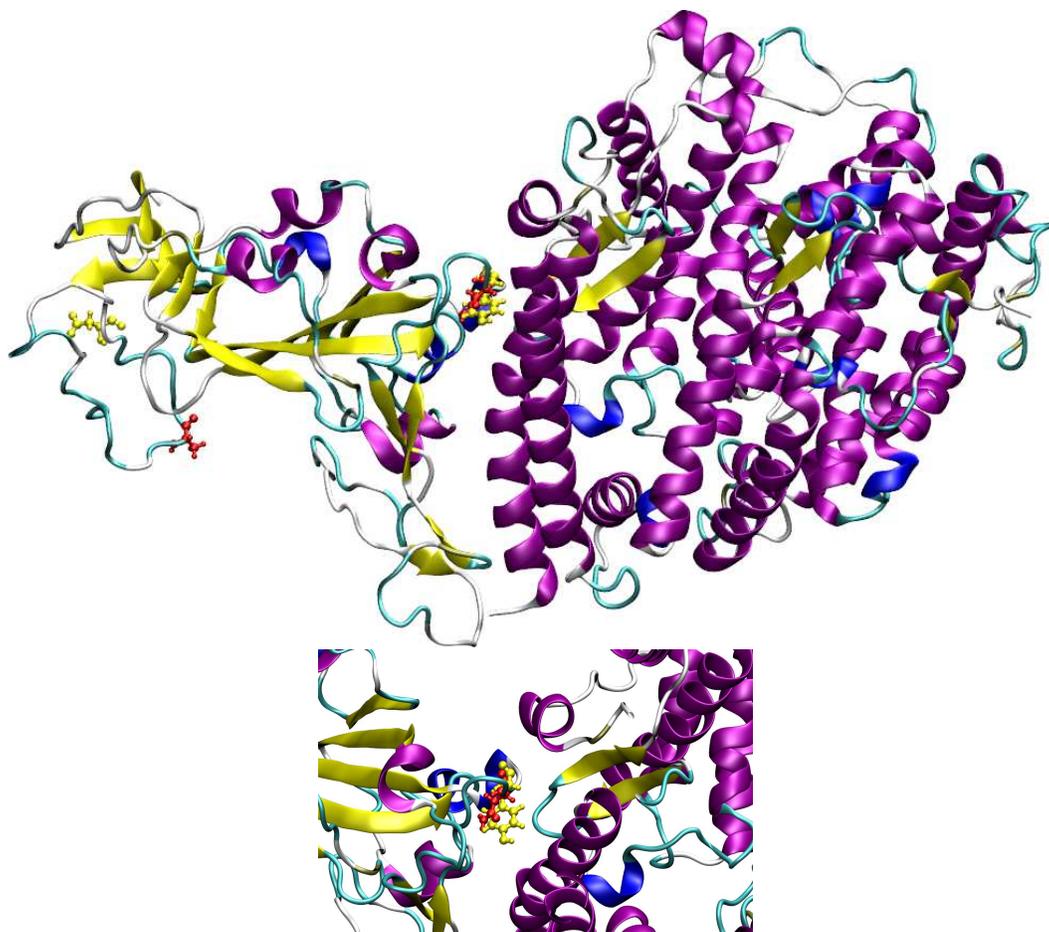


Figure 1: Three dimensional structure the spike protein:hACE-2 complex. The residues that are mutated in B.1.1.7 variant are shown in yellow color.

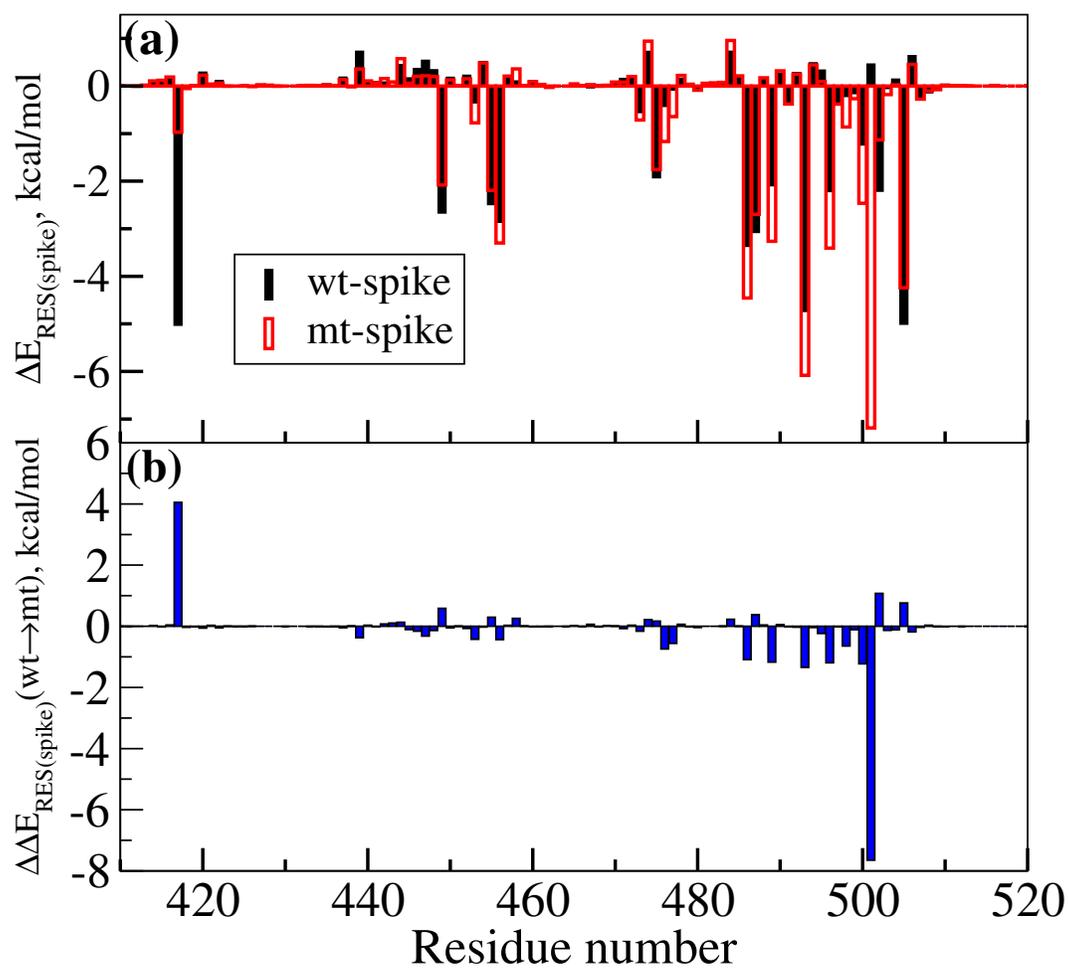


Figure 2: The spike protein residue wise contributions to the binding free energies with ACE-2 receptor. Results are shown for both wild type and B.1.1.7 variants of SARS-CoV-2 in (a) and the differences between these two values are shown in (b)

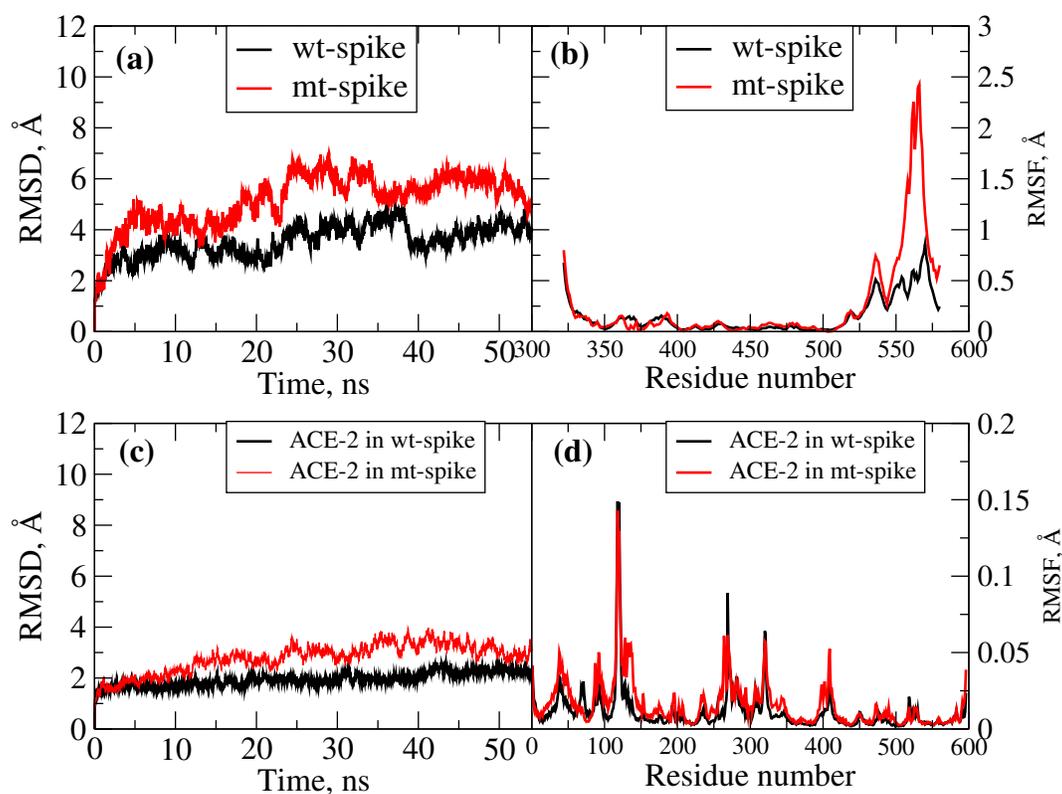


Figure 3: (a) RMSD as a function of time computed for wild type and mutant variant of the RBD domain of spike protein when it is bound to the ACE-2 receptor. (b) RMSF for wild type and mutant variant of the RBD domain of spike protein when it is bound to the ACE-2 receptor. (c) RMSD as a function of time computed for the ACE-2 receptor when it is bound to wt-spike and mt-spike proteins. (d) RMSF for ACE-2 receptor when it is bound to wt-spike and mt-spike proteins.