bioRxiv preprint doi: https://doi.org/10.1101/2021.04.09.439181; this version posted April 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

RESEARCH ARTICLE

# Theoretical causes of the Brazilian P.1 and P.2 lineages of the SARS-CoV-2 virus through molecular dynamics

Micael D. L. Oliveira [a], Kelson M. T. Oliveira [a], Jonathas N. Silva [a], Clarice Santos[b], João Bessa[b], and Rosiane de Freitas [b]

[a] Laboratory of Theoretical and Computational Chemistry. Department of Chemistry, Federal University of Amazonas, 69077-000, Manaus, AM, Brazil;

[b] Institute of Computing (IComp), Federal University of Amazonas, 69077-000, Manaus, AM, Brazil.

**ARTICLE HISTORY**

Compiled April 9, 2021

**ABSTRACT**

The novel $\beta$-coronavirus has caused sad losses worldwide and the emergence of new variants has been causing great concern. Thus throughout this research the lineage B.1.1.28 of clade P.2 (K417N, N501Y, E484K) that emerged in Brazil was studied but also in a less depth the P.1 lineage, where through simulations of molecular dynamics in the NAMD 3 algorithm in the $8ns$ interval it was possible to understand the thermodynamic impacts in the interaction of the ACE2-RBD complex and the neutralizing antibody RBD-IgG. From the molecular dynamics, we noticed that the RMSF averages in the P.2 strain were more expressive in comparison to the ACE2-RBD wild-type and consequently some regions have undergone more expressive conformational changes although, in general, a greater stabilization of the complex was perceived. In addition, was an increase in the average number of Hydrogen bonds generating a lower RMSD and greater system compaction measured by Radius of Gyration (Rg). The change in native contacts was also important where the decrease from $0.992 \pm 0.002$ to $0.988 \pm 0.002$ reflects structural changes, which could reflect in greater transmissibility and difficulty in recognizing neutralizing antibodies. Through the MM-PBSA decomposition, we found that Van der

CONTACT micaeloliveira@ufam.edu.br

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.09.439181; this version posted April 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Waals interactions predominated and were more favorable when the structure has mutations of the P.2 lineage. Therefore we believe that the greater stabilization of the ACE2-RBD complex may be a plausible explanation of why some mutations are converging in different lineages such as E484K and N501Y.

**KEYWORDS**

SARS-CoV-2, molecular dynamics, mutagenesis, P.1/P.2 lineages.

## 1. Introduction

In January 2021, Brazilian researchers identified a novel variant of the coronavirus, called P.1 (20J/501Y.V3), during a period of unprecedented increase in hospitalizations and incidence in cases of reinfection in the city of Manaus, where it was shown to be correlated with the emergence of novel lineage. From the genetic sequencing, it consists of multiple mutations, among the most worrying are 3 (three) that affected the RBD region of Spike glycoprotein: K417N, E484K and N501Y (Naveca et al., 2021). The P.1 strain identified has many similarities with the discovery in the United Kingdom B.1.351 (20H/501Y.V2 or VOC-202012/01). Since there has been an increase in transmissibility by 1.4-2.2 times, in addition to the greater probability of lethality by 1.1-1.8 times (Coutinho et al., 2021; Faria et al., 2021). According to data from the Epidemiological Surveillance System in Brazil in relation to the first three months of 2021, it was noticed that in the age groups from 30 to 39 years old there was an increase in cases by 565.08%, while in the age group from 40 to 49 years for a rate of 626%. Finally, when analyzing the death rate, the data are less expressive. Consequently, there is a shift towards younger age groups while the number of occurrences at older ages has been reduced (Castro, 2021).

The P.2 lineage evolved independently in Rio de Janeiro without being directly related to the P.1 lineage (Voloch et al., 2021). We should note that the P.2 and South African lineage have exactly the same mutations in common in the RBD region of the Spike protein. In South Africa, lineage B.1.351 emerged whose studies report greater difficulty in the recognition of antibodies (Moyo-Gwete et al., 2021; P. Wang et al., 2021). It is then constituted by eight mutations in the protein Spike, being 3 (three) of these: K417N, E484K and N501Y. It is noticeable the convergence of some

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.09.439181; this version posted April 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

mutations, specifically in position E484. Thus, several variants of the coronavirus, whether identified in South Africa, Brazil or the United Kingdom, carry this same mutation.

First of all, we have to highlight that theoretical and computational chemistry techniques are essential to see what is often inaccessible experimentally or even to provide new insights into the most intrinsic molecular mechanisms of a given phenomenon. Thus, molecular dynamics simulations are an essential tool for understanding the mechanisms involved in biomolecular interactions. Although, it is known that the precision of the results depends essentially on the elapsed time of simulation in addition to the precision of the adopted force field (Durrant & McCammon, 2011). In this research we studied the P.2 variant impact using simulations of atomistic molecular dynamics (all-atom) for the ACE2-RBD and antibody-antigen complexes to find theoretical explanations of the causes of the increase in viral transmissibility or even severity of symptoms. In the end, we analyzed the RMSF atomic fluctuations, RMSD stability, formed Hydrogen bonds and SASA value of solvent exposure in the ACE2-RBD interaction. In addition, the non-covalent terms of molecular mechanics $\langle E_{MM} \rangle$ were also estimated in addition to the total solvation energy by decomposing MM-PBSA energy.

## 2. Materials and methods

### *2.1. Mutagenesis in silico*

As a result of the unavailability of crystallographic structures containing the mutations of the P.2 lineage (K417N, E484K and N501Y) that arose in Brazil, first it was necessary to use reference structures from the RCSB Protein Data Bank database (`https://www.rcsb.org/`) (Berman et al., 2000) where amino acid substitutions would be applied. The *in silico* mutation was performed in the PyMOL 2.3 software (Schrödinger, LLC, 2015) with the module "Mutagenesis" (see Figure 1) having the rotamer with lowest steric tension automatically defined by the software. In addition, the standard rotamer library was adopted, being dependent on the $\phi$ and $\psi$ angles of the backbone. The prediction of new conformation of ACE2-RBD (PDB ID: 6M0J) (X. Wang et al., 2020) and antibody-antigen (PDB ID: 7BWJ) (Ju et al., 2020)

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.09.439181; this version posted April 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

with P.2 mutations was performed with Branch-and-Prune implementation algorithm. After the insertion of mutations, it was necessary to remove steric clashes between the atoms and thus was performed molecular dynamics simulations to minimize and equilibrate the resulting structure.

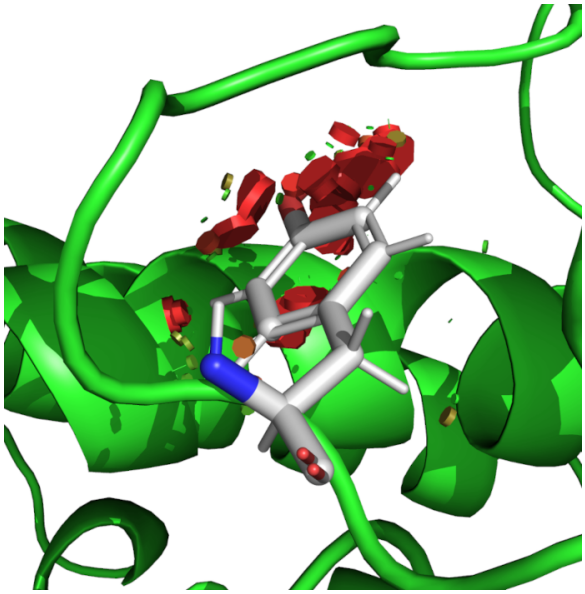

**Figure 1.** Insertion of the N501Y mutation is presented, belonging to the P.1/P.2 strains and applied to the ACE2-RBD complex (PDB ID: 6M0J). The software PyMOL (Schrödinger, LLC, 2015) was used for mutagenesis and presented a probability of tension at the best rotamer with a score of 31.18. The small green disks denote atoms that are almost in a steric or subtly overlapping conflict. On the other hand, the large red discs indicate a significant overlap of the Van der Waals radius.

In parallel, the I-Mutant 3.0 (Capriotti, Fariselli, & Casadio, 2005) tool was used, which implements the SVM machine-learning approach to predict $\Delta\Delta G$ as a result of the emerging variants in the ACE2-Spike complex (PDB ID: 7DF4) and neutralizing antibodies RBD-Ty1 (PDB ID: 6ZXN) where we adopt the physiological condition at $pH = 7.4$, temperature of $25°C$ and SVM3 ternary classification algorithm where it denotes stability ($\Delta\Delta G \geq +0.5 kcal{\cdot}mol^{-1}$), destabilization ($\Delta\Delta G \leq -0.5 kcal{\cdot}mol^{-1}$) and neutrality ($-0.5 kcal{\cdot}mol^{-1} \leq \Delta\Delta G \leq +0.5 kcal{\cdot}mol^{-1}$). It is noted that although not all authors denote the positive sign as stabilization, it was agreed in this work that: $+$ (Stabilization); $-$ (destabilization) as shown in the equation below:

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.09.439181; this version posted April 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

$$\Delta\Delta G_{binding} = \Delta G_{mutation} - \Delta G_{wild-type} \quad (1)$$

### *2.2. Parameters of molecular dynamics*

In this step, atomistic simulations (all-atom) of molecular dynamics were performed in the ACE2-RBD complex (PDB ID: 6M0J) (X. Wang et al., 2020) in front of the lineage B.1.1.28/P.2 (E484K, K417N, N501Y). We also studied the impact of P.2 on the interaction with specific neutralizing antibodies for RBD (PDB ID: 7BWJ) (Ju et al., 2020) that were derived from B lymphocytes from patients infected with SARS-CoV-2. All proteins were previously prepared with the software in its version for academic purposes Schrödinger Maestro 2020-4 using the "Protein Preparation" module, a step that included adjustments in ionization states, removal of water molecules, cofactors, assignment of partial charges and addition of Hydrogen atoms to the structure. The generation of all input and configuration files was done using the QwikMD 1.3 (Ribeiro et al., 2016) plugin implementing in the VMD 1.9.4.48a (Humphrey, Dalke, & Schulten, 1996) graphical interface. The complex was immersed in a solvation box (see Figure 2) with cubic geometry and applying PBC (periodic boundary conditions) containing water molecules described by the TIP3P model at a standard distance of 12.0 Å in relation to the frontiers of the box, as well as the addition of $Na^+$ and $Cl^-$ counterions to neutralize the system at a physiological molar concentration of 0.15 $mol \cdot L^{-1}$. Meanwhile, all topology files were generated with the CHARMM36 (Huang & MacKerell Jr, 2013) force field.

The system was minimized using 1000 steps with conjugate gradient approach. Shortly thereafter, there was gradual heating of 60-300K at the rate of $1K \cdot ps^{-1}$ over $1ns$ under the NPT ensemble and then the system was equilibrated in the NPT ensemble with steepest descent algorithm. Finally, the trajectory calculation step used the integration of Verlet in an NPT thermodynamic cycle with the elapsed time of $8ns$ and recording every $20ps$. Hoover's stochastic piston was used for pressure control similar to a barostat under the standard conditions of $1.01325Bar$, period of $200fs$ and decay time with $100fs$. While a Langevin dynamic thermostat was adopted with a coupling

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.09.439181; this version posted April 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

coefficient of $\tau_p = 1ps^{-1}$. In addition, long-range electrostatic interactions were estimated according to the Particle-Mesh-Ewald (PME) formalism while the SHAKE (Kräutler, van Gunsteren, & Hünenberger, 2001) algorithm was used to restrict covalent bonds between hydrogens along of the simulation. Finally, the NAMD3 (Nanoscale Molecular Dynamics) (Phillips et al., 2020) algorithm was used to run all simulations with acceleration of the Nvidia GTX 1050 2 GB GPU with 640 CUDA cores at a time-step of $2fs$. As a result of computational limitations, studies of mutations were restricted to the RBD region (229 amino acids) instead of the entire Spike receptor (1273 amino acids).

In the end, the conformational changes in the macromolecule as a result of the mutations were measured, through the temporal evolution RMSD and RMSF of the atomic displacement in relation to the $C\alpha$ of the ACE2-RBD complex where frame 0 was adopted as a reference. The analysis of trajectories and construction of RMSD and RMSF graphics were possible with the MDAnalysis (Michaud-Agrawal, Denning, Woolf, & Beckstein, 2011) and Matplotlib libraries implemented in the Python 3 programming language with the Shrake and Rupley algorithm (Shrake & Rupley, 1973). The fraction of native contacts was calculated based on the first frame of the MD simulations as a native conformation. Throughout the analysis of native contacts, the MDAnalysis tool where the function $radius_cut_q(r, r0, radius)$ was adopted, with the cut radius $< 4.5$Å. The calculation of the hydrogen bonds formed during the simulation was made using a plugin present in the VMD 1.9.4a.51 software where the donor-acceptor cut distance was 3.0Å while the cut angle was of 20°. Throughout the identification of saline bridges, a plugin was also used in this same software with an oxygen-nitrogen cutoff distance of 3.2Å. All SASA values were estimated with MDTraj (McGibbon et al., 2015) library with the Shrake and Rupley algorithm (Shrake & Rupley, 1973).

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.09.439181; this version posted April 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

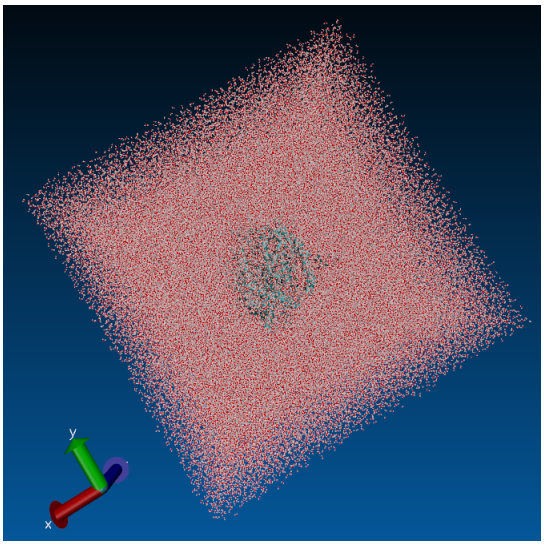


**Figure 2.** Cubic solvation box whose ACE2-RBD complex was immersed 12.0Å away from the boundaries with a volume of $(163.57Å)^3$, and whose visualization was in the software VMD 1.9.4a48 (Humphrey et al., 1996).

The estimation of the error in the RMSD, RMSF, SASA and Hydrogen bonds values were based on the standard deviation of the measurements over time. We used the Scipy library in Python for all statistical analysis with the command $ttest_ind(a, b, equal_var = False)$, where $(a, b)$ denote two matrices containing the molecular dynamics RMS data. We analyzed the statistical significance of the comparative results using the two-way Student t-test to confirm the hypothesis of the difference between two means $(\mu_1 \neq \mu_2)$ of the RMSD and RMSF values obtained from a set of statistically independent data and variance $\sigma_1^2 \neq \sigma_2^2$. We considered the value $p < 0.05$ for a statistically significant difference for a confidence interval of 95%. Regarding the impact of the P.2 strain on the recognition of antibodies, one complex antigen-antibody was subjected to molecular dynamics simulations in the range of $12ns$ and later the mean values of RMSD, RMSF, SASA, in addition to formed hydrogen bonds.

### *2.3. MM-PBSA energy decomposition*

The decomposition of the energy values $\langle E_{MM} \rangle$, $\langle E_{elec} \rangle$, and $\langle E_{VdW} \rangle$ refer to the average of all molecular dynamics frames. All components of molecular mechanics and

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.09.439181; this version posted April 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

solvation terms were estimated for the ACE2-RBD complex (PDB ID: 6M0J) in the presence and absence of the P.2 strain, in which the explicit water molecules and ions were previously removed for analysis. We used the NAMDEnergy plugin to calculate $E_{MM}$ while the term $\langle SASA \rangle$ was previously obtained by the MDTraj library. From the average solvation area it was possible to calculate the term $\langle E^{apolar}_{solvation} \rangle$ where the empirical values $\gamma \approx 0.00542\ kcal \cdot mol^{-1} \cdot \text{Å}^{-2}$ and $\beta \approx 0.92 kcal \cdot mol^{-1}$ were adopted. The $G_{PB} = E^{polar}_{solvation}$ component of Poisson-Boltzmann was estimated for the last frame by the finite difference method using the DelPhi v8.4.5 algorithm (Li et al., 2012) with a 0.0001 convergence criterion, in a $0.10M$ solution having a dielectric constant for the solute with $\epsilon = 2.0$ and for the solvent the water dielectric constant was $\epsilon = 80.0$. It is important to note that the polar solvation term was considered implicit continuum solvation by the DelPhi algorithm. As a result of the high computational complexity and the inherent uncertainty in the estimation of the entropic term $T \cdot \Delta S$, we did not include the MM-PB(SA) decomposition of this research.

The equation to estimate the polar contribution corresponds to the finite difference of the solution of the Poisson-Boltzmann equation (PB) (Homeyer & Gohlke, 2012; C. Wang, Greene, Xiao, Qi, & Luo, 2018):

$$-\nabla \cdot \epsilon(\vec{r}) \nabla \phi(\vec{r}) + \overline{\lambda}(\vec{r}) \phi(\vec{r}) = 4\pi \rho_f(\vec{r}) \tag{2}$$

The entropic term is generally estimated only by normal or quasi-harmonic analysis and is the most complex component to obtain. The nonpolar solvation free energy stems from Van der Waals interactions between the solute and the solvent. The value of $SASA$ denotes the surface area accessible to the solvent:

$$\langle G^{apolar}_{solvation} \rangle = \gamma \times \langle SASA \rangle + \beta \tag{3}$$

The solvation free energy is then estimated by the sum of $\Delta_{PB}$ associated with the Poisson-Boltzmann term and by the term $\Delta G^{apolar}_{solvation}$, making up respectively the

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.09.439181; this version posted April 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

polar and nonpolar contributions to the solvation free energies. The polar contribution is typically obtained by solving the PB equation, while the term apolar is estimated from a linear relationship with the surface area accessible to the solvent (SASA).

$$\langle \Delta G_{solvation} \rangle = \langle \Delta G_{PB/GB} \rangle + \langle \Delta G^{apolar}_{solvation} \rangle \tag{4}$$

The term $E_{MM}$ corresponds to the total molecular mechanics (MM) energy referring to the sum of the covalent interactions $E_{int}$ with the non-covalent $E_{VdW}$ and $E_{elec}$. The internal energy term $E_{int}$ consists of the sum of the energies associated with the chemical bonds $E_{bond}$, the angles $E_{angles}$ and the dihedral torsions $E_{torsions}$.

$$\langle E_{MM} \rangle = \langle E_{internal} \rangle + \langle E_{elec} \rangle + \langle E_{VdW} \rangle \tag{5}$$

$$\Delta H_{gas} \approx \langle \Delta E_{gas} \rangle = \langle \Delta E_{MM} \rangle \tag{6}$$

In the MM-PBSA method, the free energy of binding results from the sum of the interaction energy of the complex associated with the formation in its gaseous phase (vacuum), the term associated with total solvation and finally the entropic penalty:

$$\Delta G_{binding} = \Delta H - T \cdot \Delta S = \langle \Delta E_{MM} \rangle + \langle \Delta G^{total}_{solvation} \rangle - T \cdot \Delta S \tag{7}$$

## 3. Results and Discussions

### *3.1. Study of the Brazilian lineage B.1.1.28 of clades P.1 and P.2*

First of all, the structural reconstruction of the Spike protein containing the P.1 and P.2 mutations was possible by implementing an implicit enumeration algorithm called

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.09.439181; this version posted April 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Branch-and-Prune (BP). It is therefore a method that lists all possible atoms and positions while discarding invalid ones. After mutagenesis and protein reconstruction, it was finally possible to perform the various analyzes of chemical interactions, molecular dynamics and MM-PBSA decomposition.

Through an analysis with the aid of the I-Mutant 3.0 tool (Capriotti et al., 2005), the global value of $\Delta\Delta G$ in the ACE2-RBD structure (PDB ID: 7DF4) as a result of the 3 (three) mutations that constitute the Amazonian lineage B.1.1.28/P.1 was $-1.85 kcal \cdot mol^{-1}$ where: N501Y ($-0.08 kcal \cdot mol^{-1}$); K417T ($-1.50 kcal \cdot mol^{-1}$); E484K ($-0.27 kcal \cdot mol^{-1}$). Thus we can see that the N501Y mutation that appeared in the UK a priori would not significantly affect the ACE2-RBD interaction. The same conclusion refers to the E484K mutation in South Africa that did not reach the critical threshold of $-0.5 kcal \cdot mol^{-1}$. On the other hand, K417T that emerged in Amazonas generated an unexpected destabilization in the ACE2-RBD recognition, which although it can decrease the probability of cell invasion could make it difficult to recognize neutralizing antibodies induced by vaccines. In addition, the critical destabilization due to K417T could have a correlation with the unprecedented increase in January, 2021 in infections and deaths in Amazonas due to COVID-19. The P.2 clade lineage is distinguished by the K417N mutation that generates less significant destabilization although still critical of $-1.35 kcal \cdot mol^{-1}$. It is noted that mutations where there is destabilization tend to disappear due to evolutionary pressure since they cause a negative impact on the protein function (Ancien, Pucci, Godfroid, & Rooman, 2018). We must note that these instability results are only predictions, and only with molecular dynamics can we be more convinced of the conclusions. As reported in the literature, the phenomenon of stability is proportional to a lower conformational entropy, but it is not possible to infer that greater stability is necessarily related to greater severity of the disease.

In order to have greater confidence in the stability results of the ACE2-RBD complex, we predicted $\Delta\Delta G$ using the mCSM (Pires, Ascher, & Blundell, 2013) tool where: N501Y ($-0.757 kcal \cdot mol^{-1}$); K417T ($-1.594\ kcal \cdot mol^{-1}$); E484K ($-0.081\ kcal \cdot mol^{-1}$). Again, the K417T mutation of the P.1 lineage showed a critical destabilization, as occurred in the I-Mutant 3.0 tool. When the P.2 lineage was analyzed, the

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.09.439181; this version posted April 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

K417T mutation generated a critical destabilization of $-1.582\ kcal \cdot mol^{-1}$ providing preliminary evidences of changes in residue Lys417 in the Spike protein may be the main responsible for the increase in the transmissibility of the virus in the Amazonas. Through the tool DynaMut2 (Rodrigues, Pires, & Ascher, 2021) it was possible to understand the impact of mutations on chemical bonds (see Figure 3). The K417N mutation belonging to the P.2 lineage is considered destabilizing as a result of the loss of 1 (one) Hydrogen bond, 3 (three) hydrophobic contacts in addition to 2 (two) polar interactions at the ACE2-RBD interface. The same pattern was observed in the P.1 lineage with the K417T mutation, although with greater stability as a result of having lost fewer interactions with the disappearance of 1 (one) hydrogen bond, 1 (one) hydrophobic contact and 1 (one) interaction polar.

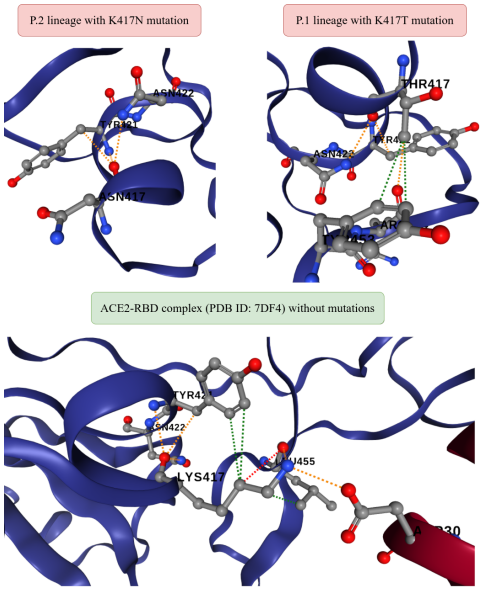


**Figure 3.** Comparison between the chemical interactions formed in the ACE2-RBD complex (PDB ID: 6M0J) as a result of the K417T and K417N mutations related to the P.1 and P.2 lineages, respectively. All diagrams were generated on the DynaMut2 (Rodrigues et al., 2021) platform. The dashed lines in green represent the hydrophobic contacts, the lines in red correspond to the Hydrogen bonds and finally the color orange refers to the polar interactions.

Through the aid of the PDBePISA platform (`https://www.ebi.ac.uk/msd-srv/prot_int/pistart.html`) (Krissinel, 2010), a worrying mutation in the B.1.1.28/P.1 and P.2 lineages would be the same that affected the United Kingdom. In this, the N501Y mutation resulted in the formation of a hydrogen bond ($-OH$) between the

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.09.439181; this version posted April 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Tyr501-Lys353 residues at a distance of 2.90Å which would explain the $\Delta_i G \approx -6.5$ $kcal \cdot mol^{-1}$ compared to the wild-type structure where $\Delta_i G \approx -5.9$ $kcal \cdot mol^{-1}$. This was perhaps the cause of the greatest transmissibility in the British lineage B.1.1.7, but it would also indicate a priori that all lineages emerged in Brazil would also be characterized by an increase in transmissibility because it shares similar mutations to other strains around the world. Exposure to water molecules in the Spike protein containing the P.2 strain increased due to a SASA of 934.7Å comparatively higher compared to the wild-type structure with 880.2Å$^2$ as predicted in the PDBePISA tool through the last frame resulting from the MD simulations. Consequently, the greater probability of hydrophobic contacts in P.2 provides an increase in stability and therefore causes an increase in the affinity of interaction with the ACE2 receptor. In addition, the P.2 strain would make the Spike protein more soluble due to a higher SASA value, which could facilitate interaction with the ACE2 receptor. It is important to note that these results refer only to the last frame, and that is why it was essential that we also evaluate the average value $\langle SASA \rangle$ later with molecular dynamics on (see Table 1).

When analyzing the RBD-Ty1 interaction (PDB ID: 6ZXN) with the mutations present in the P.1 strain, a destabilizing behavior was found, which reflects a less interaction between the Spike glycoprotein and monoclonal antibodies so that: N501Y ($0.35 kcal \cdot mol^{-1}$); K417T ($-0.90 kcal \cdot mol^{-1}$); E484K ($-0.50 kcal \cdot mol^{-1}$). Among all the results, the most worrying is precisely the K417T mutation that emerged in South Africa, which presents a critical destabilization and may a priori hamper the immune response through the administration of vaccines. Regarding the P.2 strain, the K417N mutation showed lower stability compared to the K417T, resulting in $-0.82 kcal \cdot mol^{-1}$. Lastly, when we performed the protein-protein docking on the HDock platform (`http://hdock.phys.hust.edu.cn/`) (Yan, Zhang, Zhou, Li, & Huang, 2017) for the ACE2-RBD complex (PDB ID: 6M0J) containing the lineage P.2, it was noticeable a greater affinity with a relative score of $-311.75$ while in the absence of mutations it was $-310.19$. These results corroborate the hypothesis that P.2 really makes Spike interaction with the ACE2 cell receptor more expressive. On the PatchDock platform (`http://bioinfo3d.cs.tau.ac.il/PatchDock/`) (Schneidman-

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.09.439181; this version posted April 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Duhovny, Inbar, Nussinov, & Wolfson, 2005) P.2 mutations also reflected an increase in affinity in the ACE2-RBD complex where the score increased from 15206 to 15356 compared to the reference structure. When analyzing all the mutations that constitute P.1/P.2, we realized that the N501Y mutation was the only one that directly affected the antibody-antigen interface although the value of $\Delta\Delta G$ was not significant, because otherwise, the harm to patients could be even greater than it currently is.

#### *3.1.1. Impact of P.2 through MD analysis*

It should be noted that the mutations that constitute the South African strain affected the same positions in the RBD region of the Spike protein in relation to the Brazilian strain B.1.1.28/P.2. Thus, the simulations presented here will present similar conclusions for both strains. Although molecular dynamics is a rigorous approach to study the interaction of biomolecules, it is necessary to calculate as many trajectories as possible to minimize statistical errors. Thus although the results presented here have a relatively short time of $8ns$ for ACE2-RBD and $12ns$ for RBD-IgG, we believe that the results are consistent and relevant. Despite the widespread use of MD simulations, little statistical approach is used to attest to the significance of the results and to minimize subjectivity in the conclusions. This may be due mainly to the large sampling of data resulting from molecular dynamics, which tends to make statistically significant even very small variations. A similar approach is found in the literature that used ANOVA to study the impacts on the activation of the GPCR protein (Bruzzese, Dalton, & Giraldo, 2020). For this reason, we used the Student's t-test to confirm whether the differences in the conformational fluctuations of the P.2 strain are really significant. Our criterion for quantifying the influence of structural flexibility in the presence and absence of the P.2 strain was the value of RMSD and also for the analysis as a function of amino acids by RMSF. This difference corresponds to the variation of the average values RMSD and RMSF between the structure with the mutations of P.2 and the wild-type structure, which was important for the tests of statistical significance (see *Supplementary Material*).

We performed the superposition of the last frame obtained from the molecular dynamics referring to the P.2 lineage and the respective original crystallographic struc-

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.09.439181; this version posted April 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

ture of the ACE2-RBD complex (PDB ID: 6M0J) in order to understand the structural impacts of the mutations. Thus, using the Schrödinger Maestro 2020-4 software, the alignment resulted in an atomic displacement with RMSD of 0.8299Å which therefore reflects few structural changes at the global level. This result was also corroborated by the PDBeFold (Krissinel & Henrick, 2004) platform with a RMSD of 0.7540Å. The most notable comparison occurred when we aligned the last frame containing the P.2 with the last frame of wild-type structure resulting in the end an RMSD of 1.7778Å.

Through RMSF analyzes it is possible to know the fluctuation variation around a specific residue (see Figure 4), thus allowing to understand the theoretical impacts of the mutations that make up the B.1.1.28/P.2 lineage. The K417N mutation indicated RMSF fluctuations in the Spike protein around 0.725Å, while N501Y in the amount of 0.816Å and finally E484K with 1.08Å. In terms of fluctuations, the E484K residue that emerged in South Africa was the most worrisome because it induced greater conformational changes in the Spike protein, although it may be only because it is a naturally more unstable position and perhaps that is why it is more susceptible to mutations. The greater stabilization in the ACE2-RBD complex as a result of certain mutations may be an explanation of why the virus has been following a convergent evolution as already reported in some experimental studies (Bobay, O'Donnell, & Ochman, 2020; Hodcroft et al., 2021; Kemp et al., 2021; Lam et al., 2020), although the causes until still unknown. Through the simulations of this work, we conclude that there is a tendency towards greater structural stability in the most frequent mutations. In other words, the mutations that have been recurring in several strains such as E484K tend to have greater stability.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.09.439181; this version posted April 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

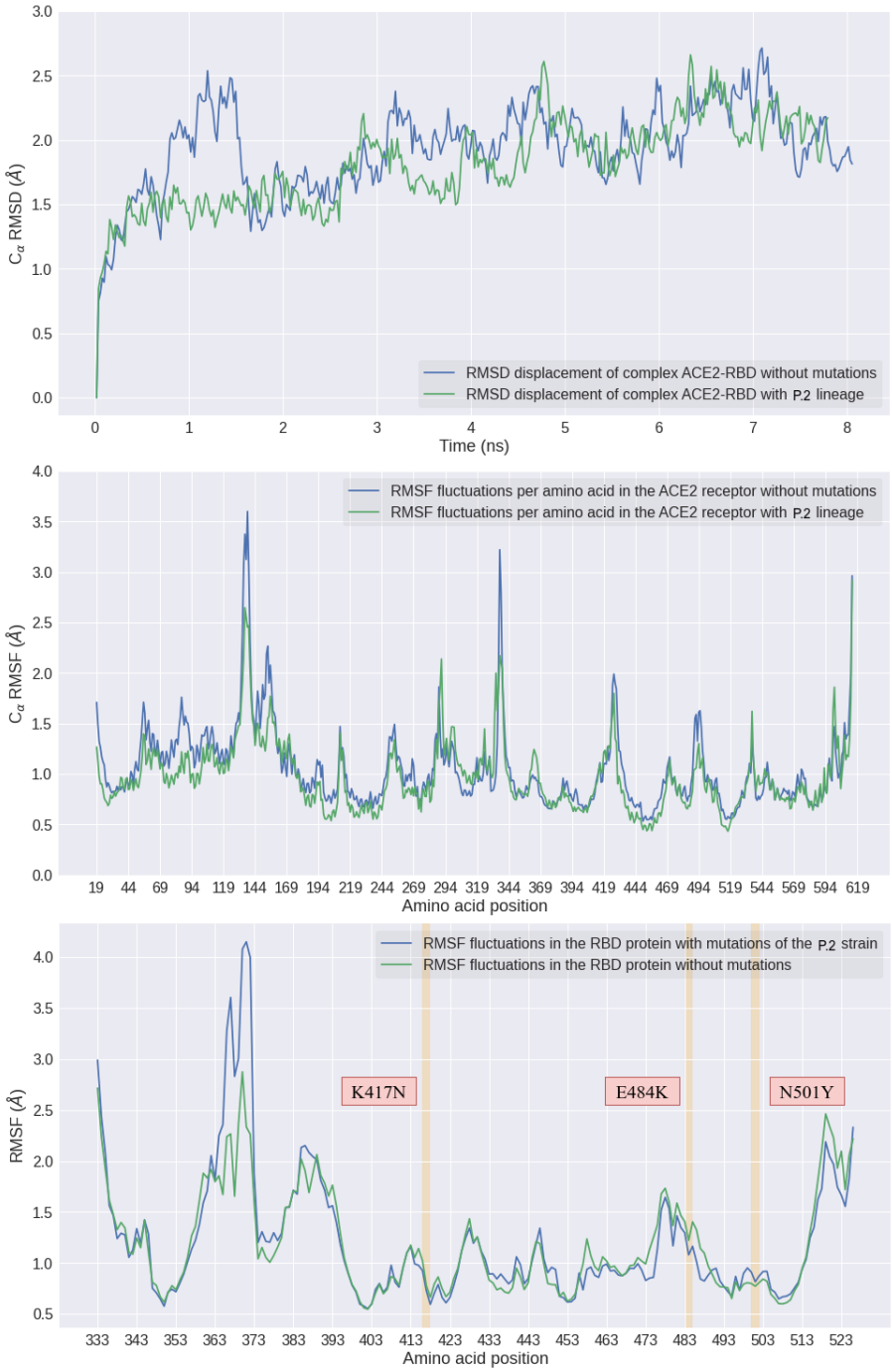


**Figure 4.** Comparative analysis between RMSD and RMSF fluctuations in relation to $C\alpha$ with and without the Brazilian lineage B.1.1.28/P.2.

When analyzing the ACE2-RBD complex without any mutations, we realized that the RMSF fluctuations for residue K417 was approximately 0.776Å, compared to residue N501 it was 0.773Å and lastly the amino acid E484 was 1.22Å. In the absence

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.09.439181; this version posted April 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

of mutations, the RMSF peak was around 2.87Å, but when the P.2 strain was analyzed there was an increase in the maximum fluctuation to 4.15Å. In general, although the fluctuation differences may be partly a consequence of a new velocity distribution associated with each simulation performed, the mutations indeed contributed to an increase in conformational changes in some very specific regions. This is because, when analyzing only the average of fluctuations (see Table 1), a lower RMSD is noticeable in P.2 and therefore greater stability. Consequently, our hypothesis is that the increase in transmissibility it may be correlated with greater affinity in the interaction of the ACE2-RBD complex as a consequence of the highest stability. It seems paradoxical that although P.2 presents comparatively larger peaks of fluctuations, the average RMSD was lower compared to the structure absent from mutations. Despite everything, it is consolidated in the literature that in the protein-protein interaction, greater stabilization is the main factor for increasing affinity. Therefore, the possible explanation for the highest peak is an indication that some very specific regions tend to suffer more from the impact of mutations, although at the global level there are fewer fluctuations. Finally, at the same time that the Spike protein becomes more flexible in some regions as a result of P.2, it may be indicative of greater exposure to neutralizing antibodies, although there may be a decrease in the affinity of interaction due to the RMSF instability of the system.

**Table 1.** Comparison between the average values of RMSD, RMSF and Hydrogen bonds in the interaction of RBD region of Spike protein with ACE2 receptor. It should be noted that the estimate of the SASA value for the RBD-IgG complex (PDB ID: 7BWJ) structure was only at the last frame, as a result of computational limitations, and the algorithm used was DelPhi instead of the MDTraj library.

| | ACE2-RBD (PDB ID: 6M0J) | | RBD-IgG (PDB ID: 7BWJ) | |
|---|---|---|---|---|
| Analysis | Wild-type | Lineage P.2 | Wild-type | Lineage P.2 |
| Average RMSD | $(1.93 \pm 0.34)$Å | $(1.81 \pm 0.34)$Å | $(3.75 \pm 0.85)$Å | $(2.40 \pm 0.56)$Å |
| Average RMSF | $(1.18 \pm 0.50)$Å | $(1.33 \pm 0.76)$Å | $(1.59 \pm 0.71)$Å | $(1.75 \pm 0.56)$Å |
| Hydrogen bonds | $201 \pm 12$ | $204 \pm 11$ | $160 \pm 10$ | $155 \pm 9$ |
| Average SASA | $(221.9 \pm 5.4)nm^2$ | $(221.8 \pm 5.3)nm^2$ | $(305.6)nm^2$ | $(317.5)nm^2$ |
| Native contacts | $0.992 \pm 0.002$ | $0.988 \pm 0.002$ | $0.981 \pm 0.003$ | $0.980 \pm 0.003$ |
| Average Rg | $(31.63 \pm 0.21)$Å | $(31.48 \pm 0.18)$Å | $(33.60 \pm 0.42)$Å | $(33.51 \pm 0.30)$Å |

Through analyzes of the average RMSD value (see Table 1), the absence of mu-

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.09.439181; this version posted April 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

tations in the ACE2-RBD complex induced approximately $(1.93 \pm 0.34)$Å in atomic displacement and average RMSF fluctuations of $(1.18 \pm 0.50)$Å. On the other hand, in the presence of P.2 lineage mutations, the average RMSD decreased to $(1.81 \pm 0.34)$Å, while the average RMSF value increased to $(1.33 \pm 0.76)$Å. Thus, the higher RMSD in the absence of mutations can be associated with a greater probability of detachment of the Spike protein with the ACE2 receptor interface. In addition, $\alpha$-helices for being present in naturally unstable positions were the ones that presented the highest conformational fluctuations, and that also suffered the direct effects of the mutations. After identification by the VMD software, we noticed that the Asn370 residue in the RBD protein showed an RMSF fluctuation of 4.08Å in the structure containing the P.2 lineage, but in the absence there was an increase in stability to 2.87Å. It is noteworthy that this region was the one that showed the greatest fluctuation among all amino acids regardless of the presence of mutations. Therefore, these data possibly reflect an increase in the ACE2-RBD interaction due to lower RMSD displacements as a result of the multiple mutations that make up the P.2 strain.

Regarding the analysis of the formation of Hydrogen bonds over the ACE2-RBD simulation (see Figure 5), it was found that the P.2 strain had an average (see Table 1) of $203 \pm 11 \cdots H$ bonds and in the absence there was a decrease to $200 \pm 12$ on average bonds. When analyzing the formation of saline bridges, it was noticed that only the wild-type structure formed 2 (two) interactions in the ACE2-RBD interface between the residues: RBD: Glu484 $\rightarrow$ ACE2: Lys31 and RBD: Lys417 $\rightarrow$ ACE2: Asp30. On the other hand, the structure containing P.2 has lost these interactions. Consequently, we can conclude that Hydrogen bonds play a fundamental role in the stabilization of P.2 to the detriment of saline bridges. Therefore, the formation of more chemical interactions, especially hydrogen bonds, therefore seems to increase the structural stability of P.2. In summary, there seems to be a favoring of this lineage so that there is a greater ACE2-RBD interaction.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.09.439181; this version posted April 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

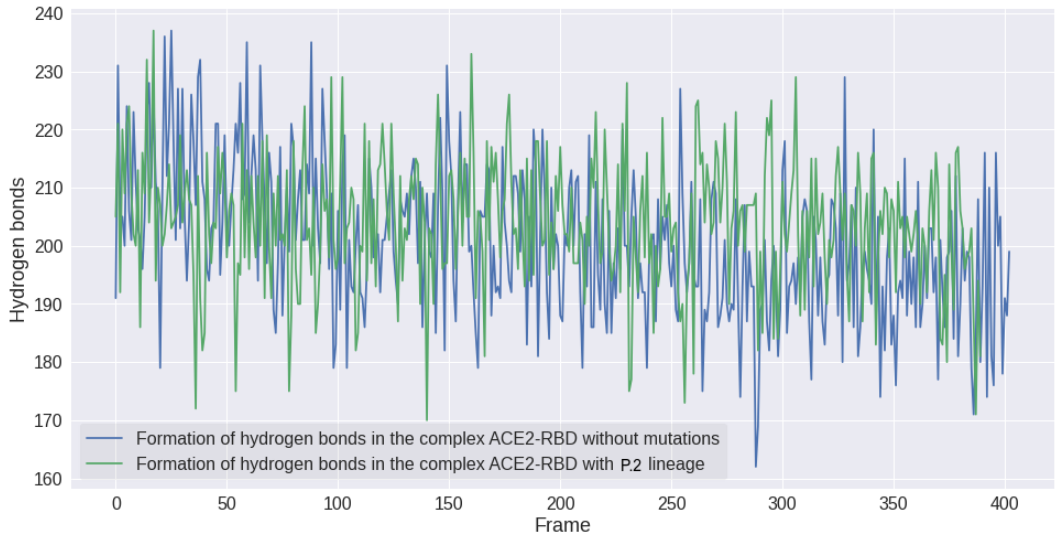


**Figure 5.** Impact on the formation of hydrogen bonds in the ACE2-RBD complex as a result of the P.2 lineage.

When analyzing the total SASA value over time (see Figure 6) the differences were not statistically significant where P.2 presented an average equal to $(2218.7 \pm 5.3)nm^2$ differing only $(+0.4nm^2)$ compared to the wild-type structure. Therefore this an indication that exposure to the solvent was not affected by the mutations that make up the P.2 strain. From the joint analysis of RMSD, RMSF, SASA and Hydrogen bonds, in general, it can be seen that mutations of the P.2 strain have stabilized the ACE2-RBD structure, although in some central residues an increase in structural flexibility has been noted according to RMSF analysis. Finally, it is important to highlight that 3 (three) analyzes corroborate the hypothesis of greater stability ACE2-RBD as a result of P.2, among them: Average values of RMSD, Hydrogen bonding and Radius of Gyration (Rg). While only 2 (two) analyzes reflect the hypothesis of greater instability, these being: Average values of RMSF and native contacts. However, we must remember that the average RMSF value is natural to be more unstable in face of more expressive peak values as a result of the mutations, although in general we can conclude that stability is in fact the main conclusion for the transmissibility and virulence changes of P.2.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.09.439181; this version posted April 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

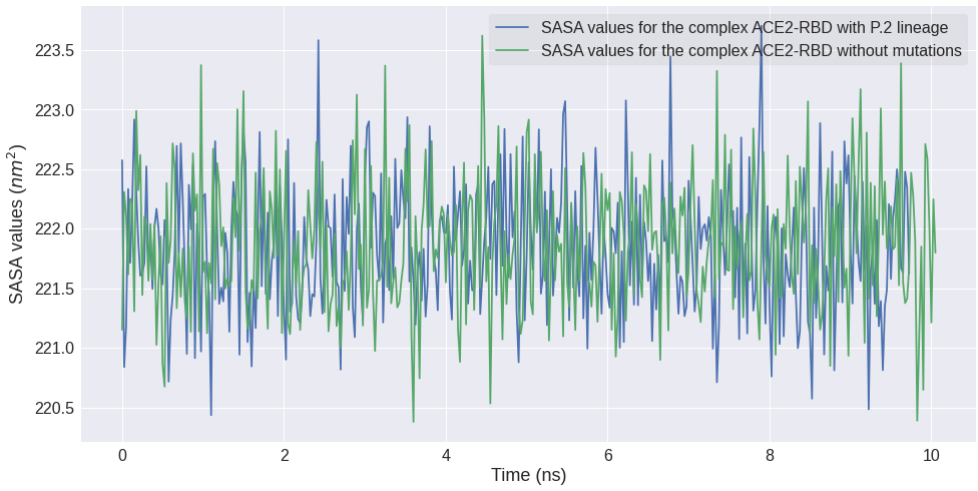


**Figure 6.** Comparison of the SASA values for exposure to the solvent of the ACE2-RBD complex (PDB ID: 6M0J) in view of the mutations that make up the P.2 strain.

Firstly as a result of the ACE2-RBD interaction being a temperature-dependent process as discovered in the research (He, Tao, Yan, Huang, & Xiao, 2020) so the premises $\Delta H < 0$ and $\Delta S < 0$ are possibly true. Furthermore, it is known that Coronaviruses tend to spread with a higher incidence, even more negative values of $\Delta G < 0$, due to the seasonality in winter. Consequently, we can conclude that the P.2 lineage, being more stable, will have less difficulty in overcoming the entropic penalty because greater affinity leads to a reduction in conformational entropy. This stems from the equation that describes the free binding energy between two proteins: $\Delta G = \Delta H - T \cdot \Delta S$. As the term $-T \cdot \Delta S$ becomes even more positive as a result of less structural stability, the free energy of binding tends to be less spontaneous and the affinity of interaction becomes less expressive with the increase of temperature. In other words, the entropic penalty of free energy decreases in proportion to structural stability. The increased compaction of the ACE2-RBD complex as a result of the P.2 strain could also make the enthalpy variation more negative a hypothesis verified by the MM-PBSA decomposition, which will lead to a greater propensity for proteolytic activation of the virus.

A greater deviation from native contacts as a result of the P.2 lineage was noticeable in the ACE2-RBD complex (see Figure 7), decreasing from $0.992 \pm 0.002$ to $0.988 \pm 0.002$. This therefore indicates notable and statistically significant structural changes

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.09.439181; this version posted April 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

($p < 0.05$). Thus, these modifications may have a correlation to the appearance of a phenotypic characteristic where there is greater viral transmissibility in similar strains such as P.1 and B.1.1.351 as reported in an important study (Faria et al., 2021).

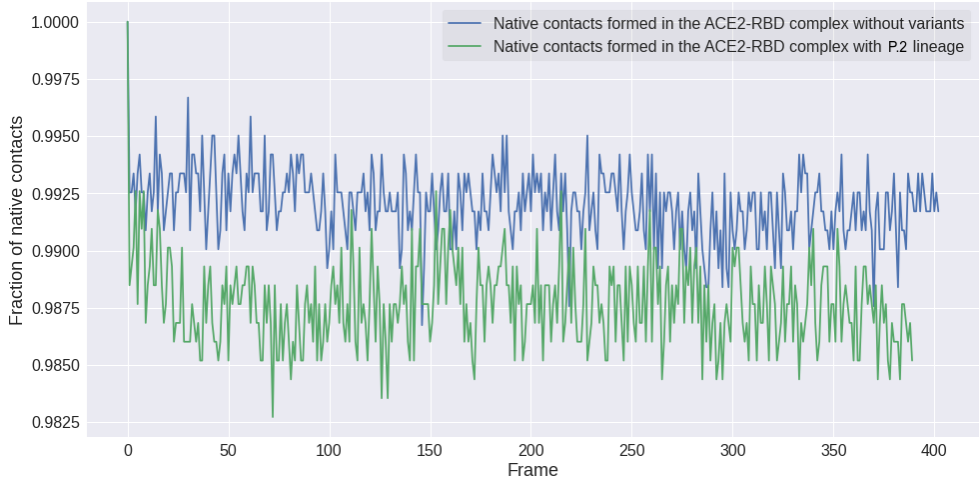


**Figure 7.** Comparison between the fractions of native $Q(t)$ contacts in the ACE2-RBD complex (PDB ID: 6M0J) as a result of the mutations that constitute the P.2 lineage.

In terms of compaction of the ACE2-RBD complex (see Figure 8) we can see that the P.2 lineage induced greater packaging along the interaction described by the average Radius of Gyration (Rg) of 31.48Å $\pm$ 0.18 when analyzing all Cartesian axes. On the other hand, in the absence of mutations, there was an increase in Rg to 31.63Å $\pm$ 0.21. When performing a statistical comparison using the Student's t-test, the Pearson's coefficient value was $p < 0.05$ within the 95 % confidence interval, thus showing a statistically significant difference between the groups analyzed. Although the statistical approach was important to minimize subjectivity in the interpretation of the results, it is important to note that the $p$ coefficient cannot be seen as something absolute or that the differences are necessarily statistically significant (Wasserstein, Schirm, & Lazar, 2019). In spite of everything, the current paradigm is that strains similar to P.2, mainly P.1 lineage, really generated an increase in transmissibility and lethality (Faria et al., 2021), and consequently we seek to see this in the results through the greater ACE2-RBD affinity, compaction and system stabilization as a consequence of P.2 lineage.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.09.439181; this version posted April 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

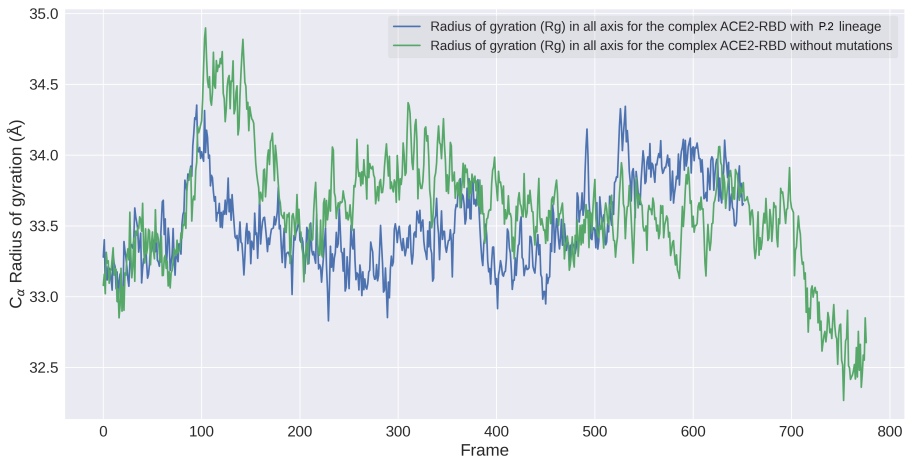


**Figure 8.** Comparative plot between the radius of gyration (Rg) values in the presence and absence of P.2 lineage.

Through the simulations of molecular dynamics in the NAMD 3 algorithm, it was possible to understand the impacts of the P.2 strain on the effectiveness of neutralizing antibody in the interaction with the Spike protein (RBD). From the individual analysis of RMSF fluctuations associated with each residue, the impact of the P.2 strain was realized so that the E484K mutation in the RBD-IgG complex (PDB ID: 7BWJ) generated an increase in instability of 1.262Å $\rightarrow$ 2.490Å. As for the K417N mutation, there was the same behavior with the transition from 1.139Å $\rightarrow$ 1.543Å. Again these results were repeated for N501Y where the change was 1.653Å $\rightarrow$ 2,154Å. In this way, it seems that a characteristic of P.2 is precisely to induce less stability in the interaction with neutralizing antibodies, which reflects in less affinity of the interaction. Furthermore, it was noticed a decrease of 1 (one) saline bridge during the MD simulations, where the RBD-IgG complex (PDB ID: 7BWJ) lost the interactions between the residues RBD: Glu484 $\rightarrow$ Heavy Chain: Arg112 and RBD: Glu484 $\rightarrow$ Light Chain: Lys55 although it formed RBD: Lys484 $\rightarrow$ Light Chain: Glu52. This therefore reflects a possible decrease in the interaction of neutralizing antibodies due to greater instability in the antibody-antigen complex. Lastly, the alignment of the crystallographic structure of the RBD-IgG complex (PDB ID: 7BWJ) with the last frame resulting from the molecular dynamics with P.2, resulted in an RMSD value of 2.5636Å, which indicates substantial conformational changes. All graphical comparative analysis for RMSD, RMSF, Hydrogen bonds and others, for the RBD-IgG

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.09.439181; this version posted April 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

complex (PDB ID: 7BWJ), are found as *Supplementary Material*.

In general, almost all comparisons were statistically significant, with the exception of the SASA value. In order to justify the validity of the t-test approach employed, the Shapiro-Wilk normality test was performed for the number of hydrogen bonds, as a test example, with a minimum confidence interval of 87 %. Levene's test was applied to verify equality between variances where $p > 0.05$, and therefore the inequality had to be considered throughout the hypothesis test. Consequently, we believe that the T-Test formalism proposal for the analysis of MD simulations is justified. Regarding the RMSD and RMSF data set, the confidence interval of 95% was adopted. Finally, we must note that the results of any computational simulation must always be seen with the perspective that the improvement of the force fields and algorithms is continuous, and therefore these results are provisional until corroborated by more in-depth theoretical studies or by means of experimental techniques.

### *3.1.2. MM-PBSA decomposition results*

The Van der Waals interaction and electrostatic energy were estimated in addition to the solvation energy. Thus, will help us to better understand the theoretical causes of the greater stability of the ACE2-RBD complex as a result of P.2 mutations. Throughout the discussion of the results (see Figure 3.1.2), we focused on the terms that make up the $\langle E_{MM} \rangle$ energy between the RBD chain and the ACE2 receptor, instead of the $\Delta G_{binding}$ because it would require accurate estimation of the entropic term. We consider that the internal energy component of molecular mechanics did not make a significant contribution, therefore approaching zero. The favorable formation and greater stability of the ACE2-RBD complex as a function of P.2 may be correlated mainly with even smaller terms of the energy of Van der Waals $\langle E_{VdW} \rangle$. Apparently, Van der Waals contribution (see Figure 9) has been playing a central role in a greater formation of intermolecular bonds in the complex containing P.2.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.09.439181; this version posted April 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

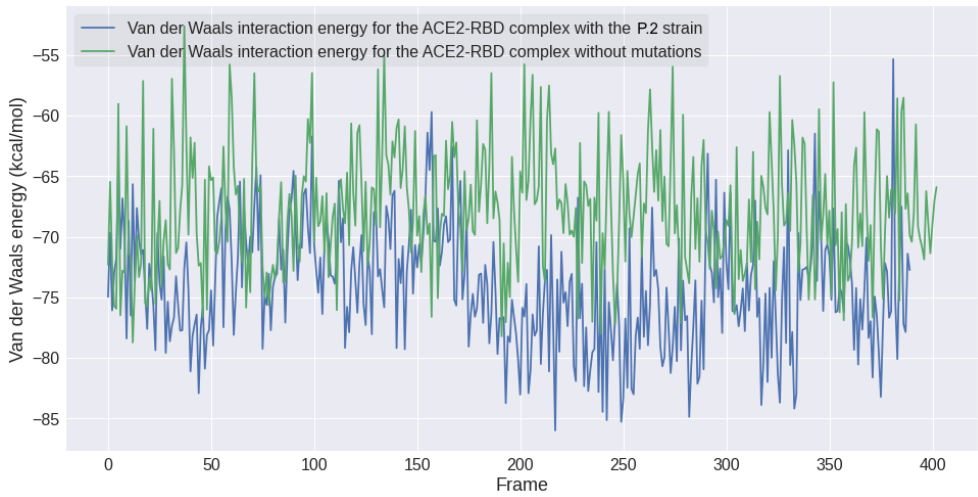


**Figure 9.** Study of the P.2 strain on the energy of Van der Waals $\langle E_{VdW} \rangle$ on the interaction with the ACE2-RBD complex (PDB ID: 6M0J).

In terms of polar solvation, the impact of the P.2 strain was more notable compared to non-polar solvation where the difference was not significant. However, the polar solvation free energy $G_{PB}$ favored the interaction of the wild-type complex much more, which could explain the greater instability and greater interaction with the surrounding environment instead of the ACE2 cell receptor. The unexpected results for electrostatic interactions may be a consequence of the numerous fluctuations in the energy $\langle E_{elec} \rangle$ that characterized the wild-type structure ranging from $-335.89\ kcal \cdot mol^{-1}$ up to $+176.07\ kcal \cdot mol^{-1}$, and therefore increasing the instability of the ACE2-RBD system. Therefore, Van der Waals interactions are the most prevalent when comparing energy data between the complex containing and absent P.2 mutations. After estimating the molecular energy and solvation, it is finally possible to obtain the enthalpy $H$ associated with the system. Thus, we realized that the P.2 strain unexpectedly had the less favorable enthalpic term compared to the reference structure, suggesting that greater ACE2-RBD interaction is correlated to less negative enthalpy but that it will possibly reflect in a more favorable and therefore greater $\Delta H$ and spontaneity of $\Delta G_{binding}$. Finally, we must highlight that the energy values presented are very dependent on the methodology used. Consequently, numerical values should only be seen in a qualitative way for the impact of P.2 instead of absolute quantitative.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.09.439181; this version posted April 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

**Table 2.** All MM-PB(SA) energy decomposition values refer to the ACE2-RBD complex (PDB ID: 6M0J) and RBD-IgG (PDB ID: 7BWJ). Throughout the analyzes for the antibody-antigen complex, only the light chain (L) of the antibody was considered when interacting with the RBD. The energy values are in the unit $kcal \cdot mol^{-1}$. As a result of computational limitations, the nonpolar solvation term for the RBD-IgG complex was calculated from a SASA value referring only to the last frame of the molecular dynamics and using the DelPhi algorithm instead of the MDTraj library. The term for $E_{int}$ covalent interactions was considered to be approximately zero in the analyzes.

| | ACE2-RBD (PDB ID: 6M0J) | | RBD-IgG (PDB ID: 7BWJ) | |
|---|---|---|---|---|
| Term | Wild-type | Lineage P.2 | Wild-type | Lineage P.2 |
| $\langle E_{elec} \rangle$ | $-182.57 \pm 26.52$ | $-45.91 \pm 25.62$ | $-10.67 \pm 17.23$ | $+4.43 \pm 17.75$ |
| $\langle E_{VdW} \rangle$ | $-67.80 \pm 4.95$ | $-74.15 \pm 4.91$ | $-8.37 \pm 2.53$ | $-3.82 \pm 4.07$ |
| $\langle E_{MM} \rangle$ | $-250.37 \pm 26.00$ | $-120.06 \pm 26.43$ | $-19.05 \pm 16.67$ | $+0.62 \pm 18.78$ |
| $G^{polar}_{solvation}$ | $-2651.31$ | $-2578.58$ | $-1319.65$ | $-1367.02$ |
| $\langle G^{apolar}_{solvation} \rangle$ | $+120.16$ | $+120.15$ | $+166.57$ | $+173.02$ |
| $G^{total}_{solvation}$ | $-2531.15$ | $-2458.43$ | $-1153.08$ | $-1194.0$ |
| $Enthalpy$ | $-2781.52$ | $-2578.49$ | $-1172.13$ | $-1193.38$ |

### *3.1.3. Impacts on electrostatic distribution*

Based on the physico-chemical properties inherent to each amino acid (Nelson & Cox, 2017), we realized that the transition from the Lys417 residue ($pI \approx 9.47$) positively charged, of a polar character and basic pH to the amino acid Asn417 ($pI \approx 5.41$) also polar although electrically neutral would increase the likelihood of hydrogen bonding at the expense of saline bridges. The substitution of the amino acid Asn501 for Tyr501 ($pI \approx 5.63$) apparently would not induce significant differences since both are polar and electrically neutral, therefore having similar probabilities in the formation of hydrogen bonds, except that Tyrosine has an amphipathic character. Finally, the amino acid Glu484 ($pI \approx 3.08$) of a nonpolar character, acidic in addition to being negatively charged when changed by Lys484 with a positive net electrical charge and basic character could therefore alter the electrostatic distribution around it and increase the probability of solubility both in acidic and alkaline conditions where $pI > pH$. From the electrostatic distribution surface (see Figure 10) calculated by the APBS (Adap-

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.09.439181; this version posted April 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

tive Poisson-Boltzmann Solver) algorithm (Jurrus et al., 2018), we can see that the P.2 strain contributed to there being regions with a more negative potential ($-597.134$ to $+549.126$) possibly as a result of the E484K mutation. On the other hand, the wild-type structure showed a slightly more positive distribution ($-587.879$ to $+501.106$). In the qualitative aspect, we noticed that the P.2 strain made the $\beta$-sheet formed between the K417N and N501Y mutations with a shorter length compared to the structure absent from mutations.

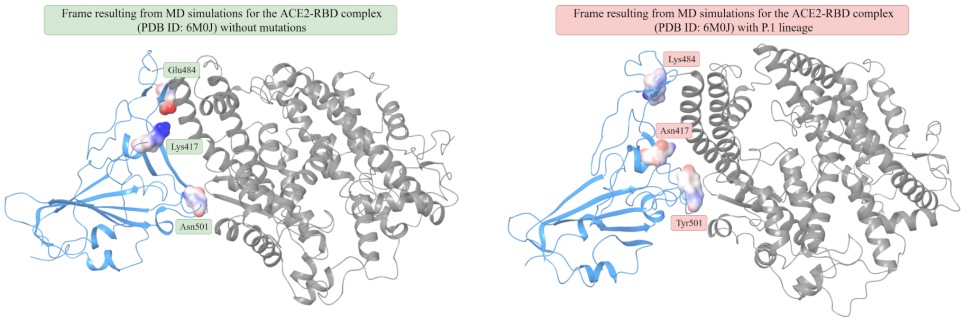


**Figure 10.** The electrostatic distribution surfaces were generated with the APBS algorithm from the last frames of the respective molecular dynamics simulations for the ACE2-RBD complex (PDB ID: 6M0J). The visualization was using Schrödinger Maestro 2020-4 software.

## 4. Conclusions

In summary, the RMSF averages of $(1.33 \pm 0.76)$Å were more expressive in the P.2 lineage compared to the ACE2-RBD wild-type complex. These results reflect less entropic penalty and therefore greater susceptibility to seasonal influences as a result of the term $-T \cdot \Delta S$ in addition to the increase in the interaction affinity and greater probability of cell entry. In the P.2 lineage we concluded that N501Y mutation showed a formation of an additional hydrogen bond between RBD:Tyr501 $\rightarrow$ ACE2: Lys353. While at the global level, P.2 generated an average of $204 \pm 11$ while without mutations this number decreased to $201 \pm 12$. This can therefore be correlated with the greater transmissibility of this strain, where the ACE2-RBD interaction becomes more spontaneous. In terms of native contacts $Q(t)$ the P.2 lineage induced a decrease from $0.992 \pm 0.002$ to $0.988 \pm 0.002$ indicating a more notable structural change, although still stable. As for the compaction aspects, the Radius of Gyration (Rg) value de-

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.09.439181; this version posted April 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

creased from $(31.63 \pm 0.21)$Å to $(31.48 \pm 0.18)$Å, and therefore P.2 generated greater compaction of the ACE2-RBD complex. In general, the P.2 strain provided stabilization of the ACE2-RBD complex and consequently, a priori, it would not disappear due to evolutionary pressure since it benefits the virus, and perhaps that is why there is a convergence of certain mutations such as E484K and N501Y. Finally, we hope with these results to help in the theoretical understanding of the impacts of the emerging lineages in Brazil. For this reason, it is essential to update the vaccines as soon as possible in order to minimize the effects of current and future non-synonymous mutations specifically on the Spike glycoprotein.

**Acknowledgements**

We have sincere gratitude for all the people who came together to face this difficult period of COVID-19. We would like to thank National Council for Scientific and Technological Development (CNPq) for supplying this research with grant 136222/2020-0.

**Disclosure statement**

The authors declare no conflict of interests in this research.

**Supplementary material**

All supplementary material for this research can be found at (`https://figshare.com/`).

**Author contributions statement**

*Kelson M. T. Oliveira:* led the project, review and article writing; *Micael D. L. Oliveira:* developed simulations in molecular dynamics/thermodynamics stability, and thus takes responsibility for data integrity and analysis accuracy; *Jonathas N. Silva:* validated all the results and discussions carried out in the research; *Rosiane de Freitas:* writing-review, editing and development of reconstruction algorithm; *Clarice Santos:*

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.09.439181; this version posted April 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

developed the protein reconstruction algorithm. *João Bessa:* Helped scientific discussions throughout the project.

## ORCID

*Micael D. L. Oliveira* https://orcid.org/0000-0002-1832-0542

*Kelson M.T. Oliveira* https://orcid.org/0000-0002-8578-6967

*Jonathas N. Silva* https://orcid.org/0000-0001-6283-3831

*Rosiane F. Rodrigues* https://orcid.org/0000-0002-7608-2052

*Clarice dos Santos*

*João Bessa*

## References

Ancien, F., Pucci, F., Godfroid, M., & Rooman, M. (2018). Prediction and interpretation of deleterious coding variants in terms of protein structural stability. *Scientific Reports*, *8*(1).

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000). The Protein Data Bank. *Nucleic acids research*, *28*(1), 235-242.

Bobay, L.-M., O'Donnell, A. C., & Ochman, H. (2020, 12). Recombination events are concentrated in the spike protein region of betacoronaviruses. *PLOS Genetics*, *16*(12), 1-14.

Bruzzese, A., Dalton, J. A. R., & Giraldo, J. (2020). Statistics for the analysis of molecular dynamics simulations: providing p values for agonist-dependent gpcr activation. *Scientific Reports*, *10*(1).

Capriotti, E., Fariselli, P., & Casadio, R. (2005, 07). I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Research*, *33*(2), 306-310.

Castro, R. (2021). *Fiocruz alerta para rejuvenescimento da pandemia no Brasil.* Retrieved from https://portal.fiocruz.br/noticia/observatorio-covid-19-fiocruz-alerta-para-rejuvenescimento-da-pandemia-no-brasil (Acessado em 27 de março de 2021)

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.09.439181; this version posted April 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Coutinho, R. M., Marquitti, F. M. D., Ferreira, L. S., Borges, M. E., da Silva, R. L. P., Canton, O., ... Prado, P. I. (2021). Model-based estimation of transmissibility and reinfection of sars-cov-2 p.1 variant. *medRxiv*.

Durrant, J. D., & McCammon, J. A. (2011). Molecular dynamics simulations and drug discovery. *BMC Biology*, *9*(71).

Faria, N. R., Mellan, T. A., Whittaker, C., Claro, I. M., Candido, D. d. S., Mishra, S., ... Sabino, E. C. (2021). Genomics and epidemiology of a novel sars-cov-2 lineage in manaus, brazil. *medRxiv*.

He, J., Tao, H., Yan, Y., Huang, S.-Y., & Xiao, Y. (2020). Molecular mechanism of evolution and human infection with sars-cov-2. *Viruses*, *12*(4).

Hodcroft, E. B., Domman, D. B., Snyder, D. J., Oguntuyo, K. Y., Van Diest, M., Densmore, K. H., ... Kamil, J. P. (2021). Emergence in late 2020 of multiple lineages of sars-cov-2 spike protein variants affecting amino acid position 677. *medRxiv*.

Homeyer, N., & Gohlke, H. (2012). Free energy calculations by the molecular mechanics poisson-boltzmann surface area method. *Molecular Informatics*, *31*(2), 114-122.

Huang, J., & MacKerell Jr, A. D. (2013). Charmm36 all-atom additive protein force field: Validation based on comparison to nmr data. *Journal of Computational Chemistry*, *34*(25), 2135-2145.

Humphrey, W., Dalke, A., & Schulten, K. (1996). Vmd: Visual molecular dynamics. *Journal of Molecular Graphics*, *14*(1), 33-38.

Ju, B., Zhang, Q., Ge, J., Wang, R., Sun, J., Ge, X., et al. (2020). Human neutralizing antibodies elicited by SARS-CoV-2 infection. *Nature*, *584*(7819), 115-119.

Jurrus, E., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L. E., ... Baker, N. A. (2018). Improvements to the APBS biomolecular solvation software suite. *Protein Science*, *27*(1), 112-128.

Kemp, S., Harvey, W., Lytras, S., Carabelli, A., Robertson, D., & Gupta, R. (2021). Recurrent emergence and transmission of a sars-cov-2 spike deletion h69/v70. *bioRxiv*.

Krissinel, E. (2010). Crystal contacts as nature's docking solutions. *Journal of Computational Chemistry*, *31*(1), 133-143. Retrieved from `https://doi.org/10.1002/jcc.21303`

Krissinel, E., & Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallographica Section D*, *60*(12), 2256-2268. Retrieved from `https://doi.org/10.1107/S0907444904026460`

Kräutler, V., van Gunsteren, W. F., & Hünenberger, P. H. (2001). A fast shake algorithm to

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.09.439181; this version posted April 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

solve distance constraint equations for small molecules in molecular dynamics simulations. *Journal of Computational Chemistry*, *22*(5), 501-508.

Lam, T. T.-Y., Jia, N., Zhang, Y.-W., Shum, M. H.-H., Jiang, J.-F., Zhu, H.-C., ... others (2020). Identifying sars-cov-2-related coronaviruses in malayan pangolins. *Nature*, *583*(7815), 282-285.

Li, L., Li, C., Sarkar, S., Zhang, J., Witham, S., Zhang, Z., et al. (2012). DelPhi: a comprehensive suite for DelPhi software and associated resources. *BMC Biophysics*, *5*(1).

McGibbon, R. T., Beauchamp, K. A., Harrigan, M. P., Klein, C., Swails, J. M., Hernández, C. X., ... Pande, V. S. (2015). Mdtraj: A modern open library for the analysis of molecular dynamics trajectories. *Biophysical Journal*, *109*(8), 1528-1532.

Michaud-Agrawal, N., Denning, E. J., Woolf, T. B., & Beckstein, O. (2011). MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry*, *32*(10), 2319-2327.

Moyo-Gwete, T., Madzivhandila, M., Makhado, Z., Ayres, F., Mhlanga, D., Oosthuysen, B., ... Moore, P. L. (2021). Sars-cov-2 501y.v2 (b.1.351) elicits cross-reactive neutralizing antibodies. *bioRxiv*.

Naveca, F., Nascimento, V., Souza, V., Corado, A., Nascimento, F., Silva, G., ... others (2021). Phylogenetic relationship of SARS-CoV-2 sequences from Amazonas with emerging Brazilian variants harboring mutations E484K and N501Y in the Spike protein. *Virological*. Retrieved from `https://virological.org/t/phylogenetic-relationship-of-sars-cov-2-sequences-from-amazonas-with-emerging-brazilian-variants-harboring-mutations-e484k-and-n501y-in-the-spike-protein/585`

Nelson, D. L., & Cox, M. M. (2017). Lehninger Principles of Biochemistry. In (7th ed., p. 119). Macmillan Higher Education.

Phillips, J. C., Hardy, D. J., Maia, J. D. C., Stone, J. E., Ribeiro, J. V., Bernardi, R. C., ... others (2020). Scalable molecular dynamics on CPU and GPU architectures with NAMD. *The Journal of Chemical Physics*, *153*(4), 044130.

Pires, D. E. V., Ascher, D. B., & Blundell, T. L. (2013, 11). mCSM: predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics*, *30*(3), 335-342.

Ribeiro, J. V., Bernardi, R. C., Rudack, T., Stone, J. E., Phillips, J. C., Freddolino, P. L., & Schulten, K. (2016). Qwikmd-integrative molecular dynamics toolkit for novices and experts. *Scientific Reports*, *6*(1).

Rodrigues, C. H., Pires, D. E., & Ascher, D. B. (2021). Dynamut2: Assessing changes in

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.09.439181; this version posted April 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

stability and flexibility upon single and multiple point missense mutations. *Protein Science*, *30*(1), 60-69.

Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., & Wolfson, H. J. (2005, 07). PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Research*, *33*(2), 363-367.

Schrödinger, LLC. (2015, November). *The PyMOL molecular graphics system, version 2.3.0.*

Shrake, A., & Rupley, J. (1973). Environment and exposure to solvent of protein atoms. lysozyme and insulin. *Journal of Molecular Biology*, *79*(2), 351-371.

Voloch, C. M., da Silva Francisco, R., de Almeida, L. G. P., Cardoso, C. C., Brustolini, O. J., Gerber, A. L., ... de Vasconcelos, A. T. R. (2021). Genomic characterization of a novel sars-cov-2 lineage from rio de janeiro, brazil. *Journal of Virology*.

Wang, C., Greene, D., Xiao, L., Qi, R., & Luo, R. (2018). Recent developments and applications of the mmpbsa method. *Frontiers in Molecular Biosciences*, *4*, 87.

Wang, P., Nair, M. S., Liu, L., Iketani, S., Luo, Y., Guo, Y., et al. (2021). Antibody resistance of sars-cov-2 variants b.1.351 and b.1.1.7.

Wang, X., Lan, J., Ge, J., Yu, J., Shan, S., et al. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*, *581*(7807), 215-220.

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond "p¡0.05". *The American Statistician*, *73*(sup1), 1-19.

Yan, Y., Zhang, D., Zhou, P., Li, B., & Huang, S.-Y. (2017, 05). HDOCK: a web server for protein–protein and protein–DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Research*, *45*(W1), W365-W373.