Energetic and structural basis for the differences in infectivity between the wild-type and mutant spike proteins of SARS-CoV-2 in the Mexican population

Martiniano Bello [a,*], Md Kamrul Hasan [b], Nazmul Hussain [b]

[a] Laboratorio de Diseño y Desarrollo de Nuevos Fármacos e Innovación Biotecnológica de la Escuela Superior de Medicina, Instituto Politécnico Nacional, Mexico
[b] Department of Biochemistry and Molecular Biology, Tejgaon College, National University, Gazipur, 1704, Bangladesh

A B S T R A C T

SARS-CoV-2 is the causative agent of the ongoing viral pandemic of COVID-19. After the emergence of this virus, it became a global public health concern and quickly evolved into a pandemic. Mexico is currently in the third position in the number of deaths due to SARS-CoV-2. To date, there have been several lineages of SARS-CoV-2 worldwide; in the Mexican population, two variants of the spike protein (S-protein) are found, localized at H49Y and D614G, which have been related to increased infectivity with respect to the wild-type S-protein. To understand how these differences impact the structural behavior of the S-protein of SARS-CoV-2, as well as binding with ACE2, we performed MD simulations combined with the molecular mechanics generalized Born/Poisson-Boltzmann surface area (MMGB(PB)SA) approach starting from X-ray crystallography data. Energetic and structural analysis showed that the differences in infectivity can be explained by differences in affinity of the protein-protein interface between the wild-type and mutant S-protein with ACE2. Conformational analysis showed that molecular recognition between the S-protein and ACE2 is linked to a decrease in the conformational flexibility of wild-type and mutant S-protein; however, an increase in the conformational mobility of ACE2 could also contribute to the binding affinity observed using the MMGB(PB)SA method.

1. Introduction

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which is responsible for the ongoing viral pandemic COVID-19, can be considered as the most catastrophic viral pathogen discovered in this decade. It was first discovered in Wuhan, China after an extreme outbreak of pneumonia-like symptoms [1,2]. While it is the seventh discovered coronavirus species that affect humans, if we compare its infection speed and fatality rate, it's undoubtedly the first among its peers [3,4]. Even though the World Health Organization (WHO) declared the SARS-CoV-2, a global pandemic due to the extreme viral propagation and the global authorities made no delay in taking immediate countermeasures (including travel bans) afterward, the viral spread is hardly showing any signs of slowing down. According to the Worldometer by April 27th, new infections are being reported still above 800,000 daily globally, with the death toll being no less than 14,000 daily (https://www.worldometers.info/coronavirus/) [5]. While some countries, after a strict quarantine period, have become free of

COVID-19, the situation in some others have relapsed. SARS-CoV-2 infection generally causes fever, coughing, sneezing, and pneumonia-like breathing difficulties in most cases [1], although recent studies have also reported the occurrence of kidney dysfunction and myocardial injury due to viral proliferation [6–8].

SARS-CoV-2 is an enveloped, positive-sense, single-stranded RNA virus with four structural proteins, including N (nucleocapsid protein), M (membrane protein), E (envelope protein), and S (spike protein) along with 14 open reading frames (ORF) containing the viral genome. It belongs to the *Betacoronavirus* order of the Coronaviridae family under the Nidovirales order, and evidence suggests that the virus originated from bats, not unlike that of its brethren, 2002 SARS-CoV and 2012 MERS-CoV. Genetic analysis of SARS-CoV-2 showed 96% similarity with bat coronavirus and 80% sequence similarity with 2002 SARS-CoV [9]. As the majority of coronaviruses that infect humans share common zoonotic origins, researchers and health officials share common concerns in terms of understanding their natural source and evolutionary mechanism to better prepare prevention mechanisms for future outbreaks.

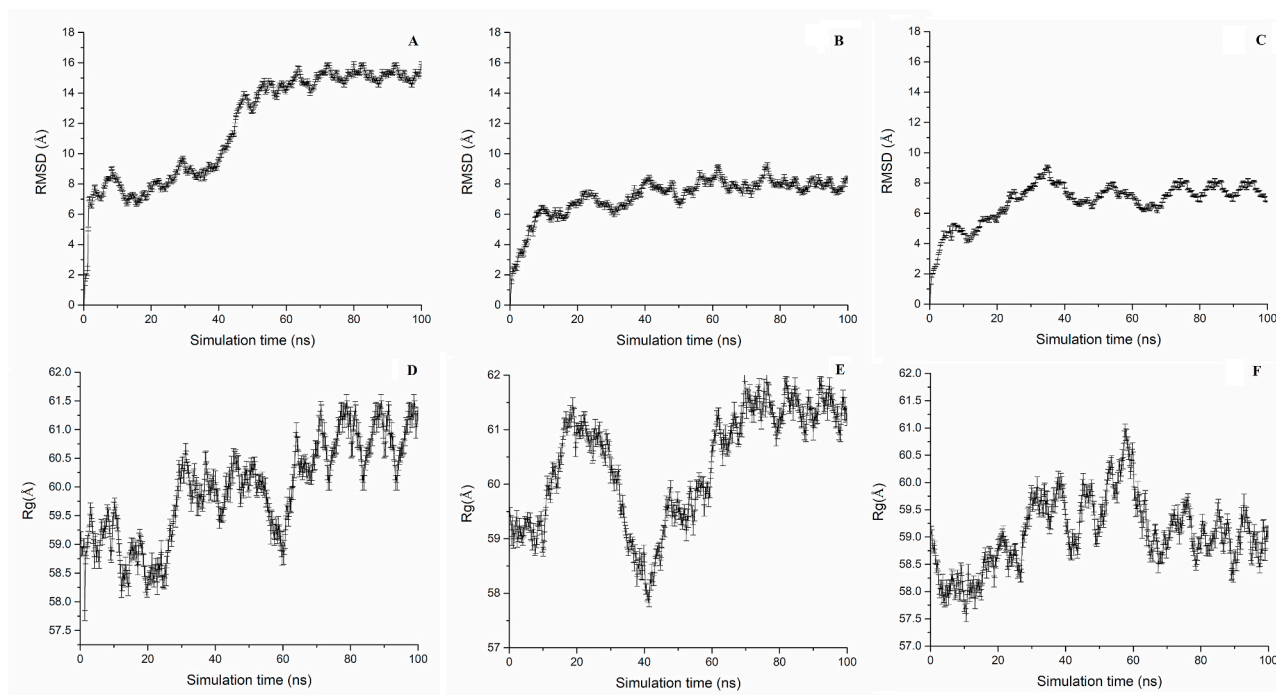**Fig. 1.** RMSD and Rg values of S-protein$_{WT}$-ACE2, S-protein$_{H49Y}$-ACE2 and S-protein$_{D614G}$-ACE2 complexes through MD simulations. RMSD values of A) S-protein$_{WT}$-ACE2, B) S-protein$_{H49Y}$-ACE2 and C) S-protein$_{D614G}$-ACE2 complexes. Rg values of D) S-protein$_{WT}$-ACE2, E) S-protein$_{H49Y}$-ACE2 and F) S-protein$_{D614G}$-ACE2 complexes. Error bars represent the standard deviation of triplicate simulations.

Amongst the 14 ORFs of SARS-CoV-2 ORF, 1a/b plays the most vital role in viral replication as it consists of two overlapping polyproteins, pp1a and pp1ab, which are cleaved into 16 major non-structural proteins. The remaining one-third of the viral genome of SARS-CoV-2 translates into the major structural proteins, including N (nucleo-capsid protein), M (membrane protein), E (envelope protein), and spike glycoprotein (S-protein) along with other accessory proteins [10–13]. The S-protein of SARS-CoV-2 plays the most substantial role in infecting humans. The S-protein is built in a trimeric conformation in which each monomer consists of two subunits. S-protein monomers are built of an N-terminal, S1 (residues 13–685) subunit that mediates binding to the human angiotensin-converting enzyme ACE2, including to the receptor-binding domain (RBD, residues 319–541) and the S2 subunit (residue 686–1273) that mediates fusion with the host cell, thus mediating host invasion in humans [14]. SARS-CoV-2 and SARS-CoV share 79% sequence similarities [15], and both employ ACE2 as their cellular receptors. As the S-protein plays the most vital role in viral propagation, it bears considerable importance in understanding the virus's evolutionary mechanisms which led to its increased infectivity against humans [16]. Studies suggest any change or mutation in the S-protein of SARS-CoV-2 has a substantial impact on the respective strain's infectivity [17]. To date, the major mutations detected worldwide have been localized on the viral S-protein [18–21]; in the Mexican population, two of three viral lineages have been reported [22], however, structural insight into how these mutations impact the S-protein of SARS-CoV-2 and its molecular recognition of ACE2 is still missing. In a recent computational study, it was identified that H49Y, D614G, and T573I mutations in the Mexican population differently impact the affinity of some potential inhibitors of the S-protein [23]. In addition, it has been reported that the H49Y mutation is linked to a reduction in total free energy, while the D614G substitution has a more stabilizing effect [21]. In this study, these two mutations (H49Y and D614G) of the S-protein of SARS-CoV-2, discovered among the Mexican population for which increased infectivity has been observed, were studied using a computational approach that combines molecular dynamics simulations and free energy calculation using the MMGB(PB)SA approach, here we

attempt to discover the potential variations in their infectivity.

2. Methods

2.1. Preparation of systems

The complex between monomeric SARS-CoV-2 S-protein and ACE2 (S-protein$_{WT}$-ACE2) was retrieved from the trimeric spike protein complex co-crystallized with ACE2 available in the Protein Data Bank (PDB entry 6ACG). Free S-protein$_{H49Y}$ and S-protein$_{D614G}$ systems and S-protein$_{H49Y}$-ACE2 and S-protein$_{D614G}$-ACE2 mutations were done on PyMOL [24], using S-protein$_{WT}$.

2.2. MD simulations

The stability of the wild-type and mutants complexes between monomeric SARS-CoV-2 spike protein and ACE2 was evaluated through MD simulations using AMBER16 software [25] and the ff14SB force field [26]. The systems were placed in a dodecadic box of 15.0 Å and solvated with a TIP3P model of water [27]. After that, the charge of the systems was neutralized with a 0.15 M concentration of Na+ and Cl-counter ions to mimic physiological conditions. After the systems were minimized and equilibrated, they were submitted to MD simulations for a period of 100 ns with triplicate experiments under an NPT ensemble at 310 K. The electrostatic forces were fixed using the PME method [28], whereas a 10 Å cutoff was selected for the van der Waals forces. The SHAKE algorithm [29] was chosen to set bond lengths. The pressure and temperature of the complexes were maintained by a weak-coupling algorithm [30]. The root means square deviation (RMSD) values of the α-carbon atoms, the radius of gyration (Rg) values of the α-carbon atoms, and the root means squared fluctuation (RMSF) was performed over the last 40 ns with AmberTools16 [25]. Hydrogen bonds were identified using a maxima distance of 2.8 Å, a minimum acceptor angle of 90.0° and a minimum donor angle of 120°. Representative conformations were obtained through a cluster analysis utilizing the kclust algorithm in the MMTSB toolset [31], and the images were obtained using Pymol [24].
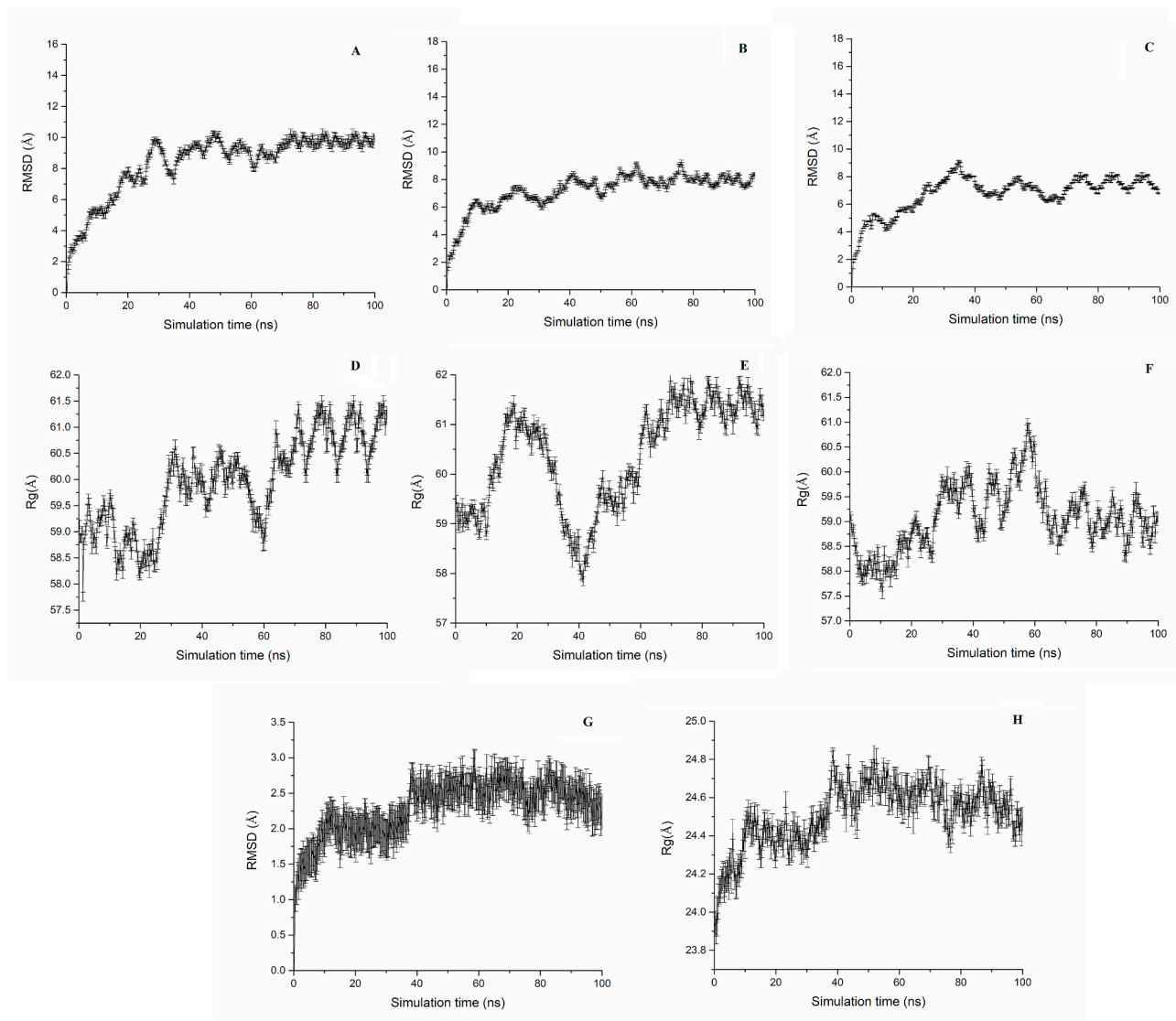
Fig. 2. RMSD and Rg values of S-protein$_{WT}$, S-protein$_{H49Y}$, S-protein$_{D614G}$ and ACE2 through MD simulations. RMSD values of A) S-protein$_{WT}$, B) S-protein$_{H49Y}$ and C) S-protein$_{D614G}$ systems. Rg values of D) S-protein$_{WT}$, E) S-protein$_{H49Y}$ and F) S-protein$_{D614G}$ systems. G) RMSD and H) Rg values of ACE2.

2.3. Principal component analysis

Principal component (PC) analysis was carried out to observe the collective motion of atoms. To this, a covariance matrix was constructed over the last 40 ns to identify the number of eigenvectors and eigenvalues that contain the total mobility of the system. The essential subspace was observed by projecting the systems onto the most important eigenvectors from the analysis using AmberTools16 [25].

2.4. Binding free-energy and per-residue decomposition calculations

The MMGB(PB)SA [32,33] approach was chosen to measure the binding free energy (ΔG_{bind}) for the different systems and to evaluate the per-residue decomposition free energy. In total, 400 snapshots at time intervals of 100 ps were taken over the last 40 ns (over the equilibrated simulation time). All counterions and water molecules were removed previous to the ΔG_{bind} evaluation and a salt concentration of 0.15 M of NaCl with the implicit solvation model [34] were considered. ΔG_{bind} and per-residue free energy decomposition evaluation was calculated as reported elsewhere [35].

3. Results and discussion

3.1. Convergence of MD simulations

Evaluation of convergence was performed to MD simulations to evaluate the simulation time where systems reached equilibrated values. Analysis for the bound systems: S-protein$_{WT}$-ACE2, S-protein$_{H49Y}$-ACE2, and S-protein$_{D614G}$-ACE2 complexes showed that they reached stable RMSD and Rg values between 40 and 60 ns (Fig. 1). Similarly, the free systems: S-protein$_{WT}$, S-protein$_{H49Y}$, S-protein$_{D614G}$, and ACE2 showed that they reached convergence between 40 and 60 ns (Fig. 2). Based on this result, the first 60 ns were excluded for further analyses.

3.2. RMSF for S-protein$_{WT}$-ACE2, S-protein$_{H49Y}$-ACE2 and S-protein$_{D614G}$-ACE2 complexes

Evaluation of the RMSF for S-protein$_{WT}$-ACE2 complex shows that the regions with the highest mobility in S-protein$_{WT}$, excluding the N and C-terminal regions, was localized in residues along S1 subunit (residues 65–82, 118–180, 242–264, and 556–574) and S2 subunit (residues 700–717, 736–763, 835–852, 969–100 and 1072–1113) (Fig. 3A). The highest mobility regions for ACE2 were residues 80–97,
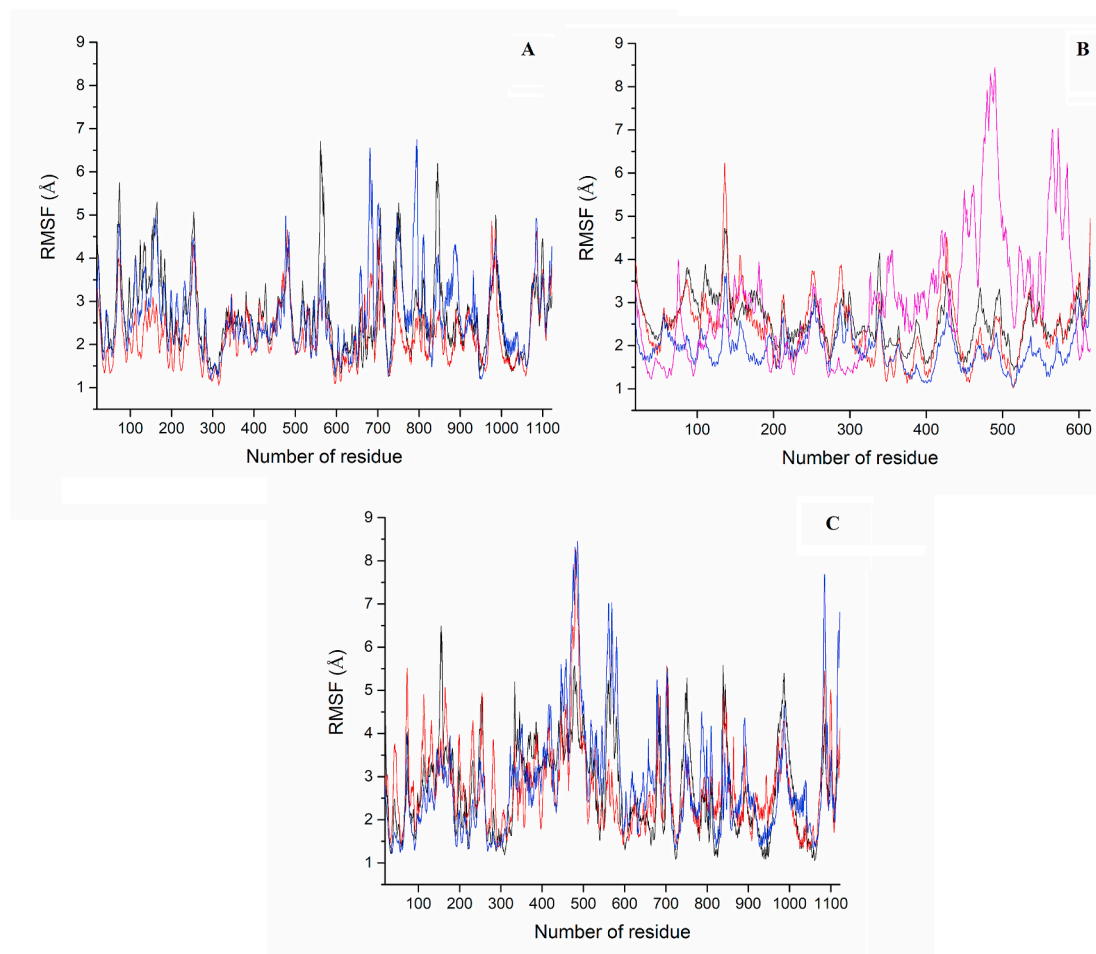
Fig. 3. RMSF of free and bound S-protein in wild-type and mutant S-protein systems through MD simulations. A) S-protein in wild-type and mutant S-protein-ACE2 complexes: S-protein$_{WT}$-ACE2 (black line), S-protein$_{H49Y}$-ACE2 (blue line) and S-protein$_{D614G}$-ACE2 (red line) complexes. B) ACE2 in wild-type and mutant S-protein-ACE2 complexes: ACE2 in S-protein$_{WT}$-ACE2 (black line), ACE2 in S-protein$_{H49Y}$-ACE2 (blue line), ACE2 in S-protein$_{D614G}$-ACE2 (red line) complexes and free ACE2 (magenta line) systems. C) free wild-type and mutant S-protein systems: S-protein$_{WT}$ (black line), S-protein$_{H49Y}$ (blue line) and S-protein$_{D614G}$ (red line) systems. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

108–122, 132–142, 163–180, 246–261, 284–304, 334–343, 383–394, 414–437, 462–499, 529–550 and 593–603 (Fig. 3B). Among these residues, some (residues 80–97, 334–343, 383–394) were localized within a close distance to some residues forming the protein-protein interface with S-protein$_{WT}$.

In the S-protein$_{H49Y}$-ACE2 complex, the highest mobility was localized in residues along S1 subunit (residues 67–80, 106–170, 246–259 and 469–491) and S2 subunit (residues 655–713, 739–760, 784–798, 813–816, 839–851, 870–900, 968–1003 and 1072–1096) (Fig. 3A). For ACE2, the most flexible regions were at the following residues: 51–90, 101–115, 132–142, 150–162, 188–200, 209–221, 244–261, 282–305, 325–343, 412–441, 462–492, 530–550, 569–582 and 596–604; some of these regions (residues 51–90 and 325–343) are close to residues involved in forming the complex with S-protein$_{H49Y}$ (Fig. 3B).

In the S-protein$_{D614G}$-ACE2 complex, the highest mobility was localized in residues along with the S1 subunit (residues 64–83, 242–264, and 450–492) and S2 subunit (residues 665–717, 969–995, and 1063–1095) (Fig. 3A). In the case of ACE2, the most dynamic regions are localized in the following residues: 75–91, 131–142, 151–163, 198–218, 241–265, 278–297, 334–343, 410–444, 465–498, 526–553 and 589–603, among these residues 75–90 and 334–343 were within a close distance to residues involved in forming the complex with S-protein$_{D614G}$ (Fig. 3B). In general, this analysis among the wild-type and mutant S-protein-ACE2 complexes point out that the more flexible regions are far from the RBD (residues 319–541), whereas, in the case of

ACE2, only some regions are within a close distance to regions of ACE2 binding site (residues 19, 23, 24, 27, 28, 30, 31, 34, 35, 37, 38, 41, 42, 83, 330, 353–355, 357 and 393) which are involved in forming the protein-protein interface of the S-protein$_{WT}$-ACE2 complex.

3.3. RMSF of S-protein, S-protein$_{H49Y}$, S-protein$_{D614G}$ and ACE2 systems

The S-protein$_{WT}$ shows that the highest mobility was localized in residues along the S1 subunit (residues 126–184, 244–259, 331–504, and 554–586) and S2 subunit (residues 675–716, 735–768, 836–854 and 965–1010). In the S-protein$_{H49Y}$, the highest flexibility was placed in residues along with the S1 subunit (residues 467–495 and 553–585) and S2 subunit (residues 683–706, 783–814, 835–855, 880–904, 972–999, and 1077–1090). In the S-protein$_{D614G}$ system, the highest mobility was localized in residues along S1 subunit (residues 69–85, 106–135, 156–174, 225–238, 247–261 and 468–494) and S2 subunit (residues 678–713, 739–762, 829–856, 889–894, 965–997 and 1077–1099) (Fig. 3C). This analysis indicates that only S-protein$_{WT}$ exhibited high flexibility of residues (residues 331–504) involved in forming the RBD (residues 319–541), indicating that both mutations contributed to a decrease in mobility at this region involved in the molecular recognition. This lower mobility at RBD for mutants may be responsible for the differences in affinity for potential S-protein inhibitors as previously reported [23]. For free ACE2, it was observed that high mobility was localized along with residues 71–86, 146–186,
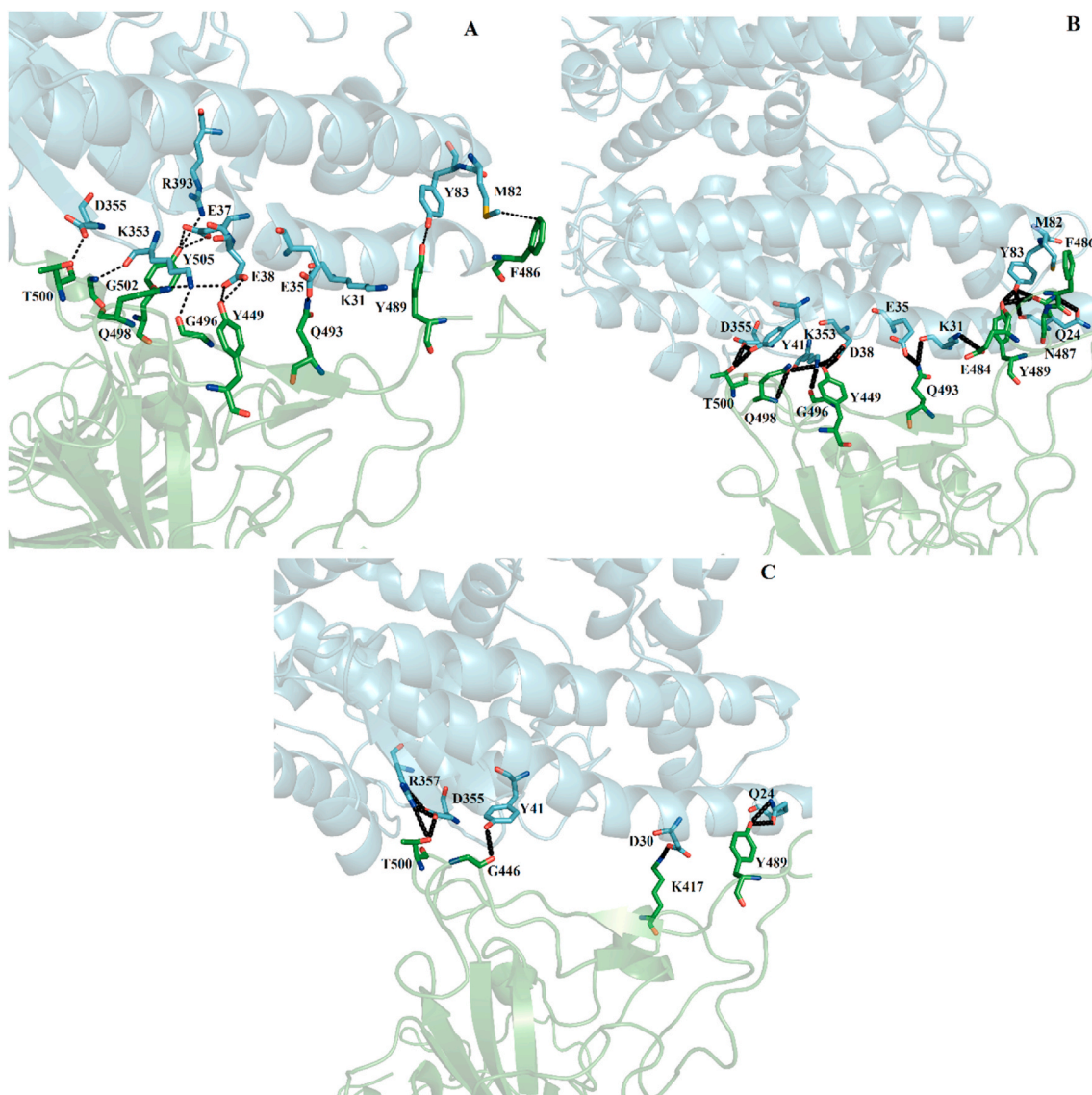
Fig. 4. Hydrogen bonds of wild-type and mutant S-protein-ACE 2 complexes. A) S-protein$_{WT}$-ACE2, B) S-protein$_{H49Y}$-ACE2 and C) S-protein$_{D614G}$-ACE2 complexes. Representative conformations were obtained during the last 40 ns of MD simulation with the kclust algorithm (see Methods).

246–266, 330–370, 404–431, 447–470, 471–505, and 557–587 (Fig. 3B), of which one region (residues 330–370) is directly involved in forming protein-protein interactions with the S-protein. Comparison of the free and bound ACE2 states suggest that a decrease in the conformational mobility is present in residues 330–370 upon complex formation with wild-type and mutant S-protein systems.

3.4. Hydrogen bonds for S-protein$_{WT}$-ACE2, S-protein$_{H49Y}$-ACE2, and S-protein$_{D614G}$-ACE2 complexes

Analysis of the most populated conformation of S-protein$_{WT}$-ACE2 obtained along the equilibrated simulation time (over the last 40 ns) showed that the protein-protein interface is coordinated by 12 H-bonds involving eight residues of the RBD: Y449, Y453, A475, G476, Q493, T500, G502 and Y505 and 11 residues of ACE2: S19, E23, K31, H34, E35, E37, D38, Y41, K353, D355 and R393 (Fig. 4A). For S-protein$_{H49Y}$-ACE2, it was observed that the protein-protein interface is also maintained by 12 H-bonds, involving nine residues of S-protein$_{H49Y}$: Y449, F486, Y489, Q493, G496, Q498, T500, G502 and Y505, and 10 residues of ACE2: K31, E35, E37, D38, M82, Y83, K353, D355 and R393 (Fig. 4B). For S-protein$_{D614G}$-ACE2, it was found that seven H-bonds maintained

the protein-protein interface through six residues of S-protein$_{D614G}$ (K417, Y489, Q493, T500, G502, and T500) and six residues of ACE2 (Q24, D30, H34, K353, D355, and R357) (Fig. 4C).

Analysis of the map of interactions among the three systems shows that primarily D614G impacted the map of interactions at the protein-protein interface compared with S-protein$_{WT}$-ACE2 and S-protein$_{H49Y}$-ACE2 complexes. Henceforth, altered SARS-CoV-2 spike conformation and enhanced protease cleavage at the S1/S2 Junction was found due to D614G mutation that results in increased fitness and transmissibility of D614G isolates [36]. Whereas, a high number of residues participating in the stabilization of the protein-protein interface are shared between the S-protein$_{WT}$-ACE2 and S-protein$_{H49Y}$-ACE2 complexes. Comparison among the number of interactions for the three systems suggests a higher affinity for the S-protein$_{WT}$-ACE2 and S-protein$_{H49Y}$-ACE2 complexes compared with the S-protein$_{D614G}$-ACE2 complex.

3.5. Binding free-energy determination

The differences in the binding free energy (ΔG_{bind}) were evaluated using the molecular mechanics generalized Poisson-Boltzmann surface area (MM/PBSA) and the molecular mechanics generalized Born surface

Table 1

Binding free energy components of S-protein$_{WT}$-ACE2, S-protein$_{H49Y}$-ACE2 and S-protein$_{D614G}$-ACE2 complexes using MM/GBSA and MM/PBSA approaches (in units of kcal/mol).

System	ΔE_{vdw}	ΔE_{ele}	$\Delta G_{ele,sol}$	$\Delta G_{npol,sol}$	DG_{mmgbsa}
MM/GBSA					
S-protein$_{WT}$-ACE2	−79.4 ± 5.60	−752.3 ± 48.00	817.0 ± 47.00	−11.3 ± 0.70	−26.0 ± 5.90
S-protein$_{H49Y}$-ACE2	−90.3 ± 8.00	−786.9 ± 52.00	846.6 ± 51.00	−13.3 ± 1.10	−43.9 ± 10.00
S-protein$_{D614G}$-ACE2	−86.2 ± 5.00	−882.0 ± 56.00	948.8 ± 55.00	−12.1 ± 0.90	−31.5 ± 7.00
MM/PBSA					
S-protein$_{WT}$-ACE2	−79.4 ± 5.60	−752.3 ± 48.00	790.7 ± 46.00	−9.5 ± 0.450	−50.5 ± 8.70
S-protein$_{H49Y}$-ACE2	−90.3 ± 8.00	−786.9 ± 52.00	822.3 ± 53.00	−10.5 ± 0.80	−65.4 ± 11.20
S-protein$_{D614G}$-ACE2	−86.2 ± 5.00	−882.0 ± 56.00	923.3 ± 55.00	−10.2 ± 0.50	−55.1 ± 7.00

Table 2

Per-residue free energy for of S-protein$_{WT}$-ACE2, S-protein$_{H49Y}$-ACE2 and S-protein$_{D614G}$-ACE2 complexes (values kcal/mol).

	S-protein$_{WT}$-ACE2	S-protein$_{H49Y}$-ACE2	S-protein$_{D614G}$-ACE2
R403		−0.2	
K417	−1.07	−0.6	−0.9
I418		−0.2	−0.6
Y449	−0.1	−1.0	
L455	−2.1	−2.6	−2.0
F456	−2.2	−0.6	−2.2
Y473	−0.2	−1.0	
A475	−2.8	−0.8	−1.8
G476	−1.0	−0.6	−0.2
T478	−0.3		−0.5
P479			−0.3
F486	−0.6	−0.4	−0.1
N487	−0.6		−0.4
Y489	−1.6	−1.8	−1.6
F490			−0.40
P491		−0.3	−0.2
L492		−0.1	−0.3
Q493	−4.4	−2.1	−4.0
Y495		−2.4	−0.2
G496		−1.2	−0.9
F497		−0.5	−0.3
Q498	−1.0	−5.0	−1.2
P499		−0.1	
T500	−1.0	−0.7	−1.1
N501	−1.8	−3.4	−1.3
G502	−1.5	−1.6	−1.5
V503	−0.4	−0.4	−0.4
Y505	−3.4	−3.2	−2.8
P507		−0.1	

3.6. Per-residue free energy decomposition

Per-residue free energy decomposition analysis allowed us to identify the residues that mostly contributed to ΔG_{bind} (Table 1). Table 2 indicates that the number of residues that contributed to ΔG_{bind} for wild-type and mutant S-protein$_{WT}$-ACE2 complexes was higher for mutants. This analysis indicates that K417, L455, F456, A475, G476, Y489, Q493, Q498, N501, G502, and Y505 (Table 2) contributed the most to ΔG_{bind} in the S-protein$_{WT}$-ACE2 complex. Among these key residues, A475, G476, Q493, G502, and Y505 were observed to form H-bonds with residues of ACE2 (Fig. 4A). For S-protein$_{H49Y}$-ACE2 complex, it was found that L455, Y473, Y489, Q493, Y495, G496, Q498, N501, G502, and Y505 are the main source of ΔG_{bind} (Table 1). Of these residues, Y489, Q493, G496, Q498, Y495, G502, and Y505 form H-bonds with the ACE2 receptor (Fig. 4B). In the case of S-protein$_{D614G}$-ACE2, the residues L455, F456, A475, Y489, Q493, Q498, T500, N501, G502, and Y505 are the major source of ΔG_{bind} (Table 1). Of these residues, Y489, Q493, T500, and G502 formed interactions with ACE2 (Fig. 4C). This analysis allowed us to identify that L455, F456, Y489, Q493, Q498, N501, G502, and Y505 of wild-type and mutant S-protein are key residues in molecular recognition. Analysis of the key residues for the three systems pointed out a higher degree of similarity between S-protein$_{WT}$-ACE2 and S-protein$_{H49Y}$-ACE2 complexes with respect to S-protein$_{D614G}$-ACE2, indicating a more significant impact of D614G substitution in the stabilization of the protein-protein interface.
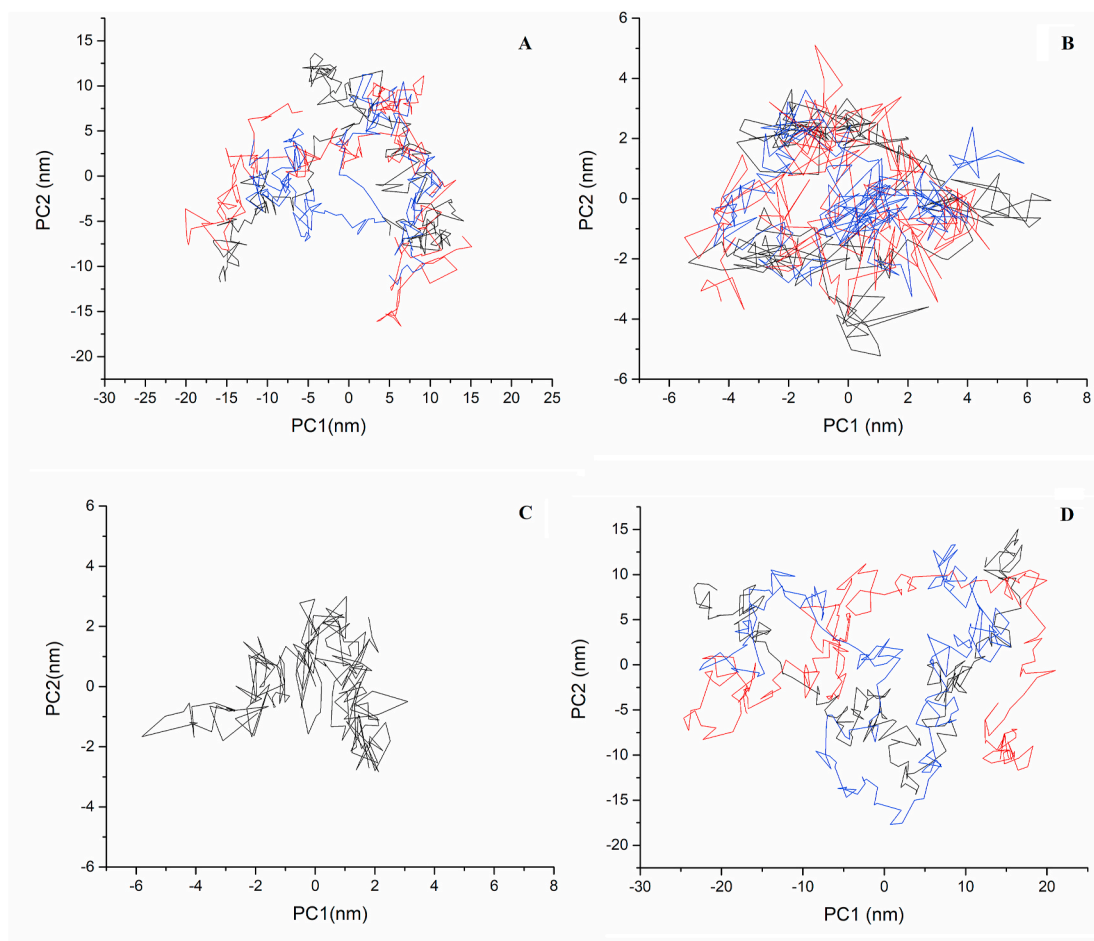
3.7. Principal component analysis of the free and bound SARS-CoV-2 S-protein-ACE2 systems

The total flexibility of free and bound S-protein-ACE2 systems was evaluated through PCA. 2D projection of the systems onto eigenvectors 1 (PC1) and 2 (PC2) allows observation into the essential space. Fig. 5A illustrates that the fluctuations of S-protein$_{WT}$ (black line), S-protein$_{H49Y}$ (red line), and S-protein$_{D614G}$ (blue line) in complex with ACE2 are confined within the two eigenvectors. However, S-protein$_{WT}$ and S-protein$_{H49Y}$ cover a larger region of phase space compared to S-protein$_{D614G}$, suggesting a higher degree of heterogeneity for S-protein$_{WT}$, and S-protein$_{H49Y}$ forming complex with ACE2. For ACE2 forming complexes with S-protein$_{WT}$ (black line), S-protein$_{H49Y}$ (red line), and S-protein$_{D614G}$ (blue line) (Fig. 5B), it was appreciated that they were also confined along PC1 and PC2, but it was found that ACE2 forming a complex with S-protein$_{WT}$ showed a slightly higher degree of heterogeneity with respect to S-protein$_{H49Y}$ and S-protein$_{D614G}$, indicating greater flexibility for S-protein$_{WT}$. For free ACE2, it was found to be confined more along PC2 than PC1 (Fig. 5C). Comparison of the results of free and bound ACE2 with S-protein, S-protein$_{H49Y}$, and S-protein$_{D614G}$ indicates that ACE2 experiences an increase in heterogeneity in the bound state, suggesting an increase in the conformational mobility upon complex conformation for the three systems.

Evaluation of the free species shows that free S-protein$_{WT}$ (black line), S-protein$_{H49Y}$ (red line), and S-protein$_{D614G}$ (blue line) are placed along the two eigenvectors (Fig. 5D), where S-protein$_{WT}$ and S-protein$_{D614G}$ exhibited similar heterogeneity, whereas S-protein$_{H49Y}$ showed a higher degree of heterogeneity, suggesting greater conformational flexibility with respect to the other two systems. Comparative analysis of the free and bound S-protein$_{WT}$, S-protein$_{H49Y}$, and S-protein$_{D614G}$ systems indicates an overall decreased flexibility of bound with regarding to the Free States. In contrast, bound ACE2 experiences an increase in mobility concerning the Free State, which could be linked to entropic contributions that may impact the binding affinity.

4. Conclusions

To date, the major mutations identified worldwide for SARS-CoV-2 have been localized on the viral S-protein, while in the Mexican population S-protein$_{H49Y}$ and S-protein$_{D614G}$ substitutions have been

area (MM/GBSA) approaches. Table 1 indicates that all the systems using both approaches were energetically favorable and indicated a similar affinity. The affinity was mainly guided through favorable non-polar interactions formed by van der Waals energy (ΔE_{vdw}) and non-polar solvation free energy ($\Delta G_{npol,sol}$), as well as electrostatic contributions (ΔE_{ele}). The S-protein$_{H49Y}$-ACE2 complex was the one with the most favorable ΔG_{bind} value, followed by S-protein$_{D614G}$-ACE2 and SARS-CoV-2 S-protein-ACE2. This result indicates that both mutations cause an increase in the affinity of S-protein by ACE2, but this affinity is stronger for S-protein$_{H49Y}$ than S-protein$_{D614G}$. This result also suggests improved accessibility of these two mutants to ACE2 with respect to wild type, which is consistent with greater cell entry of the viruses that carry these mutations [37,38]. However, considering the standard deviations of ΔG_{bind} values of S-protein$_{D614G}$-ACE2 and S-protein$_{WT}$-ACE2 indicates an overlap of both binding free energy values, indicating that D614G mutation does not impact importantly the binding free energy with respect to S-protein$_{WT}$-ACE2 system.

Fig. 5. Projection of wild-type and mutant S-protein-ACE2 complexes systems in phase space. A) 2D projection in the phase space for S-protein$_{WT}$ (black line), S-protein$_{H49Y}$ (blue) and S-protein$_{D614G}$ (red) in complex with ACE2. B) Projection for ACE2 in complex with S-protein$_{WT}$ (black line), S-protein$_{H49Y}$ (blue) or S-protein$_{D614G}$ (red). C) Projection for free ACE2. D) Projection for free S-protein$_{WT}$ (black line), S-protein$_{H49Y}$ (blue) and S-protein$_{D614G}$ (red) systems. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

experimentally identified and linked with an increase in infectivity for wild-type. However, structural insight into how these mutations impact the protein-protein interface in S-protein-ACE2 complexes is still missing. In this contribution, we combined MD simulations with the MMGB(PB)SA approach starting from structural data to explore the structural and energetic basis for the molecular recognition of the free and bound S-protein$_{WT}$-ACE2, S-protein$_{H49Y}$-ACE2, and S-protein$_{D614G}$-ACE2 systems. RMSF analysis allowed us to identify that for S-protein$_{WT}$-ACE2, S-protein$_{H49Y}$-ACE2, and S-protein$_{D614G}$-ACE2 systems, the more flexible regions in S-protein are far from the RBD, whereas, in the case of ACE2, the more dynamic regions are close to regions involved in forming the protein-protein interface of the wild-type and mutant S-protein-ACE2 complexes. RMSF analysis of free S-protein$_{WT}$, S-protein$_{H49Y}$ and S-protein$_{D614G}$ systems showed that only S-protein$_{WT}$ exhibited high flexibility in residues involved in forming the RBD, suggesting that these mutations contributed to decreased flexibility at RBD involved in molecular recognition with ACE2, whereas a decrease in the conformational flexibility was present in residues involved in the molecular interaction with RBD of wild-type and mutant S-protein systems. The per-residue free energy study allowed us to identify that L455, F456, Y489, Q493, Q498, N501, G502, and Y505 of wild-type and mutant S-protein are key residues for binding affinity. The MM/GBSA results revealed that the S-protein$_{H49Y}$-ACE2 complex has the highest affinity, followed by S-protein$_{D614G}$-ACE2 and SARS-CoV-2 S-protein-ACE2, suggesting improved accessibility of these mutants to ACE2, which is in line with experimental findings showing that these two mutations

favored cell entry of the virus. PCA indicated that ACE2 experiences an increase in heterogeneity in the bound state that is associated with an increase in conformational mobility upon complex formation and which could be linked to favorable entropic contributions that unfavorably impact the binding affinity. In contrast, the free and bound S-protein$_{WT}$, S-protein$_{H49Y}$, and S-protein$_{D614G}$ systems indicated decreased heterogeneity that could be linked to unfavorable entropic and favorably enthalpic contributions that favorably impact the binding affinity determined with the MMGB(PB)SA approach.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

References

[1] C. Huang, et al., Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China, Lancet 395 (10223) (Feb. 2020) 497–506, https://doi.org/10.1016/S0140-6736(20)30183-5.

[2] W. Guan, et al., Clinical characteristics of coronavirus disease 2019 in China, N. Engl. J. Med. 382 (18) (Apr. 2020) 1708–1720, https://doi.org/10.1056/nejmoa2002032.

[3] J. Liu, et al., Overlapping and discrete aspects of the pathology and pathogenesis of the emerging human pathogenic coronaviruses SARS-CoV, MERS-CoV, and 2019-nCoV, J. Med. Virol. 92 (5) (May 2020) 491–494, https://doi.org/10.1002/jmv.25709.

[4] S. Ludwig, A. Zarbock, Coronaviruses and SARS-CoV-2: A Brief Overview, Anesth. Analg., 2020, pp. 93–96, https://doi.org/10.1213/ANE.0000000000004845.

[5] Coronavirus update (live): 63,199,555 cases and 1,467,511 deaths from COVID-19 virus pandemic - worldometer (accessed Nov. 30, 2020), https://www.worldometers.info/coronavirus/.

[6] Z. Li, et al., Caution on Kidney Dysfunctions of COVID-19 Patients, " *SSRN Electron. J.*, 2020, https://doi.org/10.1101/2020.02.08.20021212, 02.08.20021212, Mar. 2020.

[7] J.L. Nguyen, W. Yang, K. Ito, T.D. Matte, J. Shaman, P.L. Kinney, Seasonal influenza infections and cardiovascular disease mortality, JAMA Cardiol. 1 (3) (Jun. 2016) 274–281, https://doi.org/10.1001/jamacardio.2016.0433.

[8] K. Jc, S. Kl, C. Ma, Acute myocardial infarction after laboratory-confirmed influenza infection, N. Engl. J. Med. 378 (26) (Jun. 2018) 2538–2541, https://doi.org/10.1056/nejmc1805679.

[9] S. Ullrich, C. Nitsche, The SARS-CoV-2 main protease as drug target, Bioorg. Med. Chem. Lett 30 (17) (Sep. 2020), https://doi.org/10.1016/j.bmcl.2020.127377, 127377.

[10] D.E. Gordon, et al., A SARS-CoV-2 protein interaction map reveals targets for drug repurposing, Nature 583 (7816) (Jul. 2020) 459–468, https://doi.org/10.1038/s41586-020-2286-9.

[11] Y. Chen, Q. Liu, D. Guo, Emerging coronaviruses: genome structure, replication, and pathogenesis, J. Med. Virol. 92 (4) (Apr. 2020) 418–423, https://doi.org/10.1002/jmv.25681.

[12] J. Ziebuhr, E.J. Snijder, A.E. Gorbalenya, Virus-encoded proteinases and proteolytic processing in the Nidovirales, J. Gen. Virol. 81 (4) (2000) 853–879, https://doi.org/10.1099/0022-1317-81-4-853.

[13] R.A. Khailany, M. Safdar, M. Ozaslan, Genomic characterization of a novel SARS-CoV-2, Gene Reports 19 (Jun) (2020), https://doi.org/10.1016/j.genrep.2020.100682.

[14] Mutations in spike protein of SARS-CoV-2 - cusabio (accessed Nov. 25, 2020), https://www.cusabio.com/mutations-in-spike-protein-of-SARS-CoV-2?fbclid=IwAR3Dizd-PUNRwJVcRSvlXxgcfIg8bfizcqAfRCgeWu7Miilx-EfwaaC2Hx0.

[15] R. Lu, X. Zhao, J. Li, P. Niu, B. Yang, H. Wu, W. Wang, H. Song, B. Huang, N. Zhu, et al., Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding, Lancet 395 (2020) 565–574.

[16] J. Shang, et al., Structural basis of receptor recognition by SARS-CoV-2, Nature 581 (7807) (May 2020) 221–224, https://doi.org/10.1038/s41586-020-2179-y.

[17] M. Wang, et al., International expansion of a novel SARS-CoV-2 mutant, J. Virol. 94 (12) (Apr 2020), https://doi.org/10.1128/jvi.00567-20.

[18] S. Mohammad, et al., Infection , Genetics and Evolution Exploring the genomic and proteomic variations of SARS-CoV-2 spike glycoprotein : a computational biology approach, Genet. Evol. 84 (2020), 104389.

[19] M. Becerra-Flores, T. Cardozo, SARS-CoV-2 viral spike G614 mutation exhibits higher case fatality rate, J. Clin. Pract. (2020) 4–7, https://doi.org/10.1111/ijcp.13525.

[20] L. Zhang, et al., The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity, bioRxiv (2020), https://doi.org/10.1101/2020.06.12.148726, 06.12.148726.

[21] S. Laha, et al., Characterizations of SARS-CoV-2 mutational prole, spike protein stability and viral transmission, bioRxiv 85 (2020), 2020.05.03.066266.

[22] B. Taboada, et al., Genomic analysis of early SARS-CoV-2 variants introduced in Mexico, Virology (2020), https://doi.org/10.1128/JVI.01056-20.

[23] Yudibeth Sixto-López, Martiniano Bello, José Correa-Basurto, et al., Structural insights into spike protein and its natural variants of SARS-CoV-2 found on Mexican population. https://doi.org/10.21203/rs.3.rs-57627/v1, 12 August 2020.

[24] W.L. DeLano, The PyMOL molecular graphics system, DeLanoScientific, PaloAlto (2002).

[25] D.A. Case, T.E. Cheatham, T. Darden, H. Gohlke, R. Luo, K.M. Merz, A. Onufriev, C. Simmerling, B. Wang, R.J. Woods, The Amber biomolecular simulation programs, J. Comput. Chem. 26 (16) (2005) 1668–1688.

[26] Y. Duan, C. Wu, S. Chowdhury, M.C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, P. Kollman, A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations, J. Comput. Chem. 24 (16) (2003) 1999–2012.

[27] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, T. Darden, D. York, L. Pedersen, J. Chem. Phys. 79 (1983). J. Chem. Phys. 98 (1993) 10089e10092.

[28] T. Darden, D. York, L. Pedersen, Particle mesh Ewald-an N.Log(N) method for Ewald sums in large systems, J. Chem. Phys. 98 (1993) 10089–10092.

[29] W.F. Van Gunsteren, H.J.C. Berendsen, Mol. Phys. 34 (1977), 1311e1327.

[30] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola, J.R. Haak, B. R. Miller, T.D. McGee, J.M. Swails, N. Homeyer, H. Gohlke, A.E. Roitberg, J. Chem. Phys. 81 (1984). J. Chem. Theory Comput. 8 (2012) 3314e3321.

[31] M. Feig, J. Karanicolas, C.L. Brooks III, MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology, J. Mol. Graph. Model. 22 (5) (2004) 377–395.

[32] H. Gohlke, D.A. Case, Converging free energy estimates: MM-PB (GB) SA studies on the protein–protein complex Ras–Raf, J. Comput. Chem. 25 (2) (2004) 238–250.

[33] B.R. Miller, T.D. McGee, J.M. Swails, N. Homeyer, H. Gohlke, A.E. Roitberg, J. Chem. Theor. Comput. 8 (2012), 3314e3321.

[34] A. Onufriev, D. Bashford, D.A. Case, Exploring protein native states and large-scale conformational changes with a modified generalized born model, Proteins: Structure, Function, and Bioinformatics 55 (2) (2004) 383–394.

[35] M. Bello, E. García-Hernández, Ligand entry into the calyx of β-lactoglobulin, Biopolymers 101 (7) (2014) 744–757.

[36] S.M.C. Gobeil, K. Janowska, S. McDowell, K. Mansouri, R. Parks, K. Manne, P. Acharya, D614G mutation alters SARS-CoV-2 spike conformation and enhances protease cleavage at the S1/S2 junction, Cell Rep. 34 (2) (2021), 108630.

[37] S. Ozono, et al., Naturally mutated spike proteins of SARS-CoV-2 variants show differential levels of cell entry, bioRxiv (2020), https://doi.org/10.1101/2020.06.15.151779, 06.15.151779.

[38] B. Korber, et al., Spike Mutation Pipeline Reveals the Emergence of a More Transmissible Form of SARS-CoV-2, 2020, https://doi.org/10.1101/2020.04.29.069054 bioRxiv 04.29.069054.