

IMPACT OF MUTATIONS ONOMICRON'S SPIKE PROTEIN: INSIGHTS FROM VARIANTS FOUND IN VIETNAM

Thai Ke Quan^{1,✉}, Huynh Phuoc²

¹Faculty of Natural Science Education, Saigon University, 273 An Duong Vuong Road, Ward 3, District 5, Ho Chi Minh City, Vietnam

²VNU HCMC University of Science, 227 Nguyen Van Cu Road, Ward 4, District 5, Ho Chi Minh City, Vietnam

✉To whom correspondence should be addressed. E-mail: tkquan@sgu.edu.vn

Received: 29.11.2023

Accepted: 25.12.2023

SUMMARY

The SARS-CoV-2 virus is the primary agent responsible for the ongoing COVID-19 pandemic, leading to the loss of millions of lives worldwide. In 2023, the COVID-19 pandemic's impact is primarily driven by the Omicron variant, which features many unique mutations compared to previously identified SARS-CoV-2 variants. Through analysis of the network of Omicron variants collected in Vietnam, we determined variants expected to have high transmission efficiency and positive influence before studying the impact of mutations on the structural behavior of spike proteins and their affinity with hACE2. Our computational analysis indicates that these Omicron variants not only exhibit a significantly higher binding affinity than the Wildtype but also demonstrate increased binding free energy among variants. This enhanced binding affinity of Omicron variant is reflected in the enhanced stability of the hACE2 binding structure compared to the Wildtype. This study further elucidates the specific contributions of point mutations to the binding free energy. Notably, such mutations as W152R, F157L, G257S, S371F, S373P, S375F, T376A, D405N, N440K, S477N, T478K, E484A, Q493R, Q498R, N501Y, D614G, N679K, P681H, N764K, D796Y, and N969K were found increasing the binding free energy of Omicron's spike protein in complex with hACE2.

Keywords: Covid-19, SARS-CoV-2, Omicron, Spike gene, Spike protein

INTRODUCTION

Coronavirus Disease 2019 (COVID-19) stands as a global pandemic precipitated by the SARS-CoV-2 virus, initially identified in late 2019 in Wuhan, China. The genome of SARS-CoV-2 comprises a positive single-strand RNA containing four structural genes,

each encoding structural protein: the Nucleocapsid protein (N), Spike protein (S), Envelope protein (E), and Matrix protein (M). Many studies reported high mutation rate inherent in the genomic makeup of SARS-CoV-2, giving rise to numerous variations across the globe (Li *et al.*, 2021; V'kovski *et al.*, 2021). Based on the

profound genetic diversity, the World Health Organization (WHO) has classified SARS-CoV-2 variants into three groups: Variants Under Monitoring (VUM), Variants of Interest (VOI), and Variants of Concern (VOC). Among these, VOCs are the most consequential regarding public health impact. The world has documented several SARS-CoV-2 variants that have been classified into the VOCs, including Alpha, Beta, Gamma, and Delta (Shahhosseini *et al.*, 2021). However, in late 2021, a new SARS-CoV-2 variant, Omicron, was quickly classified as a VOC with over 30 amino acid mutations in the spike protein (Elliott *et al.*, 2022). Notably, Omicron has outcompeted other VOC variants, ascending to dominance on a global scale (Viana *et al.*, 2022).

The success of VOC variants has been attributed to numerous mutations within the S gene (Harvey *et al.*, 2021; Fang *et al.*, 2023). This gene encodes the S protein, comprised of two major subunits: S1 and S2. The S1 subunit, consisting of two prominent domains: the N-terminal Domain (NTD) and the Receptor Binding Domain (RBD), plays an important role in the direct interaction with hACE2 (Yan *et al.*, 2020). The S2 subunit encompasses a fusion peptide (FP), two heptad-repeat domains (HR1/HR2), a transmembrane domain (TD), and a cytoplasmic domain (CD), all of which play a crucial roles in virus-host cell fusion (Huang *et al.*, 2020). Unlike other VOCs, the Omicron variant has accumulated over 30 amino acid substitutions in the S protein and approximately 10 substitutions in the RBD. Previous studies have revealed that the Omicron RBD exhibits a higher binding affinity to hACE2 than previous VOC variants, which reasonably explains its global spread (Duong *et al.*, 2022; Poudel *et*

al., 2022). Understanding the underlying mechanisms ruling the spike protein's binding affinity to ACE2 is of the most significance for variant detection, epidemic management, and inhibitor development.

Hence, continuous examination of the diversity and tracking mutations within the S gene of Omicron variants is crucial for assessing the dynamic effectiveness of these viruses on a global scale. The analyzed of 6348 spike gene sequences in our previous studies explored the high genetic diversity of Omicron and its subvariants collected in Vietnam (data not shown). Additionally, we identified the Omicron subvariant, BA.2*, as a key motivation for Omicron transmission in Vietnam. Therefore, in this study, we utilized computational methods to conduct molecular dynamics simulations based on the variants predicted to be highly transmissible in Vietnam.

MATERIALS AND METHODS

Building homology spike protein

Through analysis of the network of Omicron variants collected in Vietnam, we determined three variants expected to have high transmission efficiency, namely EBC1, EBC2, and EBC3. The S gene sequences were translated into protein sequences by MEGA11 software (Tamura *et al.*, 2021), which were then used to build homology structures using the SWISS-MODEL server (Waterhouse *et al.*, 2018) based on template protein (PDB ID: 8DM5). Structure Assessment and Structure Comparison with template structure in SWISS server revealed the quality of homology protein models (Studer *et al.*, 2020).

Protein-protein docking

The S protein of EBC1, EBC2, and EBC3

were made in complex with hACE2 using the HADDOCK 2.4 server. Amino acids involved in the interface interaction between the S protein and hACE2 were estimated using the PRODIGY server (Xue *et al.*, 2016): residue 449, 453, 455, 456, 475, 476, 477, 486, 487, 489, 493, 496, 498, 500, 501, 502, and 505 of S protein and residue 19, 24, 27, 28, 31, 34, 35, 38, 41, 42, 79, 82, 83, 353, 354, 355, and 357 of hACE2. The docking parameters of HADDOCK were set in defaults, and surface contact and ambiguous restraints (AIRs) features were used to grid the interface of the complex. The best models, determined by HADDOCK scores and Z-scores, were collected, and these complexes were refined in HADDOCK Refinement (Neijenhuis *et al.*, 2022) in order to increase the models' quality. EBC1-hACE2, EBC2-hACE2, and EBC3-hACE2 complexes quality were checked by Structure Comparison with template structure in SWISS-MODEL server.

Molecular dynamics simulation

20ns molecular dynamics simulation of the Wildtype, EBC1-hACE2, EBC2-hACE2, and EBC3-hACE2 complex was carried out by GROMACS version 2023.1 in the CHARMM36m force field (Huang *et al.*, 2017). All complexes were put in a cubic box with TIP3P water models, neutralized with Na⁺ and Cl⁻ ions, and the distance between any protein atom and the box edge was more significant than 2.5 nm. The simulation systems were first relaxed in the Energy minimization phase in 50,000 steps

to determine the lowest energy structure. Then, the systems were heated into 310°K and stabilized at 200 ps for the NVT (Number of particles, Volume, and Temperature) phase and 1000 ps for the NPT ensemble (Number of particles, Pressure, and Temperature) phase to reach the temperature and 1atm pressure equilibrium. In all simulations, the parameters, and conditions of the production MD runs were: the LINCS algorithms were used to constrain the hydrogen bond lengths with the integration time step of 2 fs; the long-range electrostatic interactions were calculated by PME (Particle Mesh Ewald); van der Waals interaction was modeled by Verlet scheme with a cut-off distance of 12.0 Å. The Root means square deviation (RMSD), Root means square fluctuations (RMSF), and Radius of gyration (R_g) were calculated by GROMACS to evaluate the stability, flexibility, and folding of the complexes. Pymol, VMD software were used to visualize and calculate hydrogen bond of complexes, respectively.

Estimating binding free energy

To estimate the binding free energy (BFE) of complexes after simulation, the MM/PBSA and MM/GBSA (Molecular Mechanics with Poisson-Boltzmann (Generalised Born) and Surface Area solvation) approach was used. The gmx_MMPBSA tool (Valdés-Tresanco *et al.*, 2021) was used to calculate BFE for all systems. The MM/PBSA and MM/GBSA equation is explained in below:

$$\Delta G_{GBSA} = \Delta G_{Complex} - (\Delta G_{Spike\ protein} + \Delta G_{hACE2})$$

Each component of the total free energy was estimated using the following equation:

$$\Delta G = \Delta G_{vdW} + \Delta G_{ele} + \Delta G_{pol} + \Delta G_{npol}$$

where ΔG_{ele} , and ΔG_{vdW} displayed electrostatic, and van der Waals energies, respectively. ΔG_{pol} and ΔG_{npol} are polar and nonpolar solvated energies.

The effect of mutations on BFE ($\Delta\Delta E$) was estimated in the below equation (Xu *et al.*, 2015). The negative $\Delta\Delta E$ inferred the increase and the positive inferred the decrease of BFE.

$$\Delta\Delta E = \Delta E_{mutant} - \Delta E_{wildtype}$$

RESULTS AND DISCUSSION

Homology protein and protein-protein docking

Amino acid sequences of EBC1, EBC2, and EBC3 generated from the nucleotide sequences were aligned with the reference sequence for identification of the amino acid substitutions profile (Figure 1). Quality models analysis showed that all homology

proteins exhibited high quality, with more than 90% Ramachandran favored regions and QMEANDisCo (Model Quality) scores exceeding 0.7 (Table 1). Upon comparing these homology models to the template structure, we observed low RMSD values of 0.093 Å, 0.083 Å, and 0.089 Å for EBC1, EBC2, and EBC3, respectively (Figure 1). These values indicate the suitability of these models for further analyses.

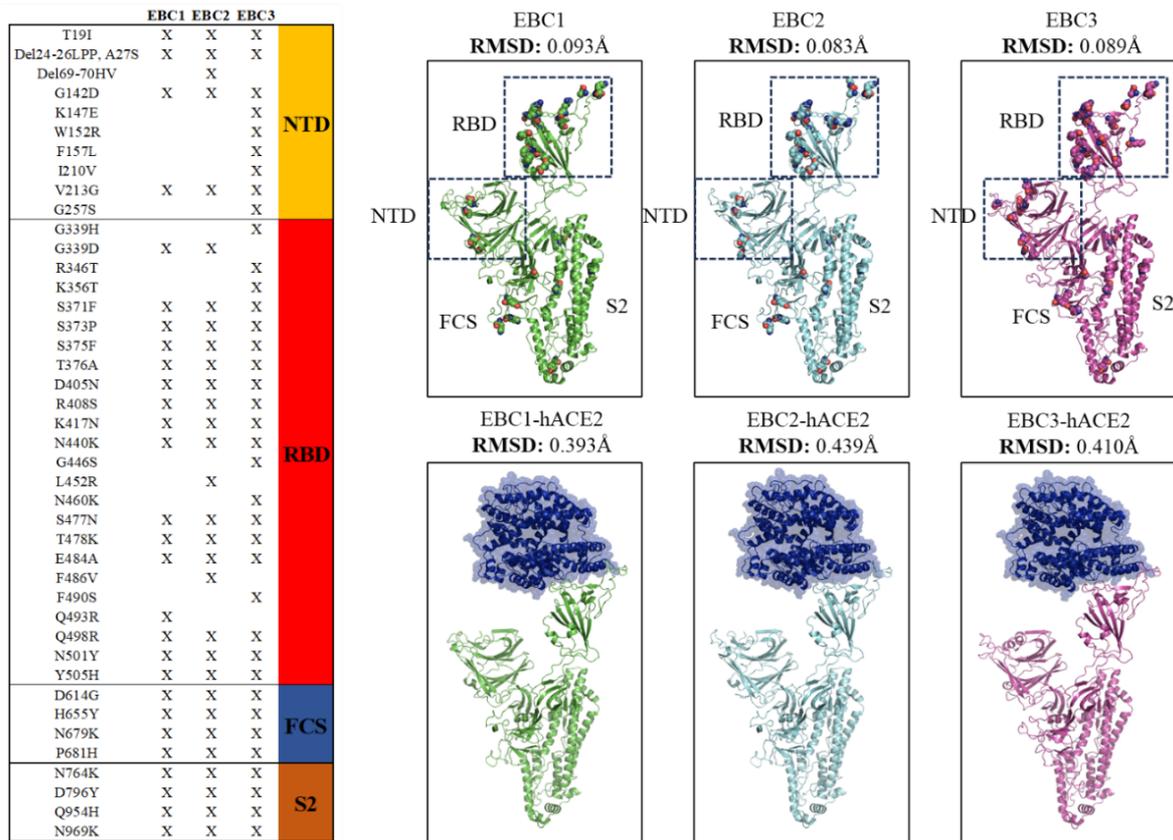


Figure 1. The mutation profile and structural of homology S protein. Mutations in the S protein of each Omicron variant (EBC1, EBC2, and EBC3) are marked as 'X.' The superimposing method in Pymol with the template structure to estimate the RMSD values of homology S proteins and the EBC1-hACE2, EBC2-hACE2, and EBC3-hACE2 complexes.

EBC1, EBC2, and EBC3 spike variants were then docked with hACE2 using HADDOCK to form EBC1-hACE2, EBC2-hACE2, and EBC3-hACE2 complexes. The docking models generally converge, as shown by the high Cluster size and the low RMSD values (Table 2). Consequently, the EBC1-hACE2, EBC2-hACE2, and EBC3-hACE2 complexes were refined in HADDOCK Refinement, showing high quality (Table 3). Accordingly, all models had Ramachandran favour over 90% and QMEAndisCo scores of around 0.75. Comparing complex models with template structure, the QS-score, DockQ, and DockQ-wave indices of EBC1-hACE2, EBC2-hACE2, and EBC3-hACE2 are all above 0.8 (Table 3). The superimpositions of complex models and templates showed low RMSD (Figure 1). This could be considered that models are suitable for further analysed through molecular dynamics simulations.

Table 1. Spike protein homology models quality.

Models	Ramachandran Favoured	IDDT	Model Quality
EBC1	94.96%	0.87	0.73 ± 0.05
EBC2	95.53%	0.86	0.74 ± 0.05
EBC3	94.96%	0.84	0.72 ± 0.05

Table 2. The HADDOCK predicted docking score for complexes is given in table.

	EBC1	EBC2	EBC3
HADDOCK score	-145.7 ± 3.9	-132.5 ± 0.9	-144.0 ± 1.5
Cluster size	199	200	200
RMSD	0.3 ± 0.2	0.9 ± 0.9	0.8 ± 0.7
Van der Waals energy	-78.3 ± 5.1	-83.9 ± 2.2	-79.3 ± 4.1
Electrostatic energy	-215.9 ± 22.4	-148.6 ± 5.5	-176.1 ± 11.3
Desolvation energy	-26.2 ± 5.1	-23.8 ± 3.6	-32.1 ± 4.3
Restraints violation energy	19.5 ± 8.3	49.8 ± 35.9	26.8 ± 18.1
Buried Surface Area	2122.7 ± 88.8	2135.3 ± 73.9	2110.8 ± 107.3
Z-Score	0.0	0.0	0.0

Table 3. Spike protein in complex hACE2 models quality.

Complexes	Ramachandran favoured	IDDT	QS score	DockQ	DockQ wave	Model Quality
EBC1-hACE2	92.74%	0.87	0.89	0.81	0.81	0.76 ± 0.05
EBC2-hACE2	91.93%	0.86	0.91	0.87	0.87	0.75 ± 0.05
EBC3-hACE2	92.43%	0.85	0.91	0.84	0.84	0.75 ± 0.05

Protein-protein dynamics simulation

The structural impact of the mutation was studied through molecular simulation of the S protein in complex with hACE2 of Wildtype, EBC1-hACE2, EBC2-hACE2, and EBC3-hACE2 models that were assessed through measurements: RMSD, R_g , and RMSF. Figures 2a and 2b illustrate the conformational dynamics of each complex. The RMSD evolution of the molecular systems during the simulations is shown in Figure 2a. RMSD values revealed that all simulations reached equilibrium between 15 to 20 ns. The EBC3-hACE2 complex

exhibited a notable instability at the 10 ns time point but gradually stabilized after 15 ns. The average RMSD values for each variant were determined to be 1.3 ± 0.4 nm for Wildtype, 1.0 ± 0.3 nm for EBC1-hACE2, 1.1 ± 0.3 nm for EBC2-hACE2, and 1.4 ± 0.5 nm for EBC3-hACE2. These collected RMSD values are consistent with earlier observations made in S protein simulations (Kwarteng *et al.*, 2021; Abdalla *et al.*, 2022). Overall, RMSD values exhibited insignificant differences between the simulations, with the EBC3-hACE2 complex displaying the highest value, signifying its relatively higher flexibility.

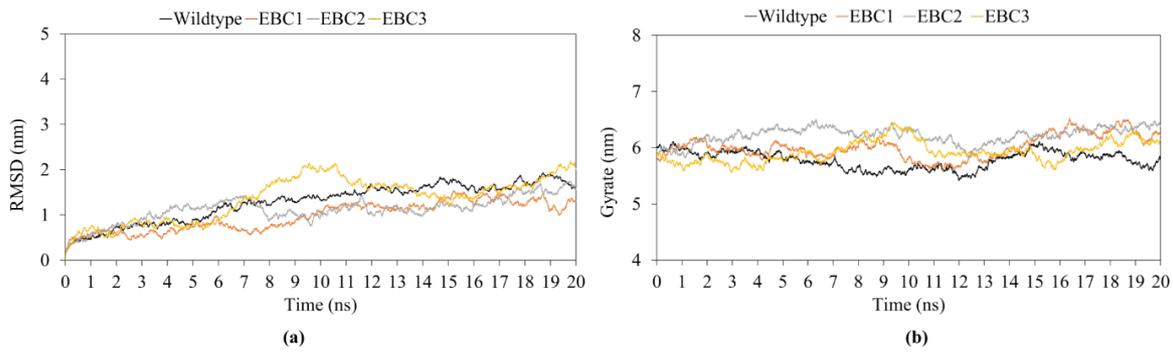


Figure 2. Structural and dynamic stability analysis of Wildtype, EBC1-hACE2, EBC2-hACE2, and EBC3-hACE2 complexes by RMSD and R_g values. (a) RMSD values; (b) R_g values of the Wildtype, EBC1-hACE2, EBC2-hACE2, and EBC3-hACE2 complexes, with each complex depicted in black, orange, gray, and yellow, respectively.

All the systems exhibit fluctuations in R_g values during the simulation (Khan *et al.*, 2021). These periodic increases and decreases in R_g values at different intervals can be attributed to the dynamic binding and unbinding events within the complexes (Khan *et al.*, 2021). Our analysis of R_g values for the simulations revealed minor fluctuations from the early stages up to the 10 ns time point of the simulation. Subsequently, R_g values consistently after 15 ns simulation at 5.8 ± 0.2 nm, 6.0 ± 0.2 nm, 6.2 ± 0.1 nm, and 5.9 ± 0.2 nm for Wildtype,

EBC1-hACE2, EBC2-hACE2, and EBC3-hACE2, respectively. As highlighted in the previous study, the higher R_g value observed in Omicron variants is acknowledged to result from the persistent formation and disruption of chemical bonds in the RBD-hACE2 complex (Abeywardhana *et al.*, 2023).

To compare the structural flexibility of the Omicron's S protein variants with the Wildtype, we performed an analysis of the RMSF values of the C_α atoms. Higher or

lower RMSF value indicates flexible or stable regions, respectively. As illustrated in Figure 3a, the RMSF values exhibit the most significant differences in the NTD region. In most cases, EBC1-hACE2, EBC2-hACE2, and EBC3-hACE2 had lower RMSF values, which display the less flexible NTD regions compared to the Wildtype, as observed across all NTD residues (Figure 3a). The RBD region shows minimal deviation between each

simulation. However, in the S2 region, we noticed higher fluctuations in the EBC3-hACE2 complex than in EBC1-hACE2 and EBC2-hACE2. This suggests that the S2 subunit may exhibit increased structural dynamics in EBC3-hACE2 complexes. Therefore, EBC3-hACE2 had the highest RMSD value (Figure 2a). Our analysis of RMSD, RMSF, and R_g values of the Omicron S protein shows its adaptability in interacting with the hACE2 receptor.

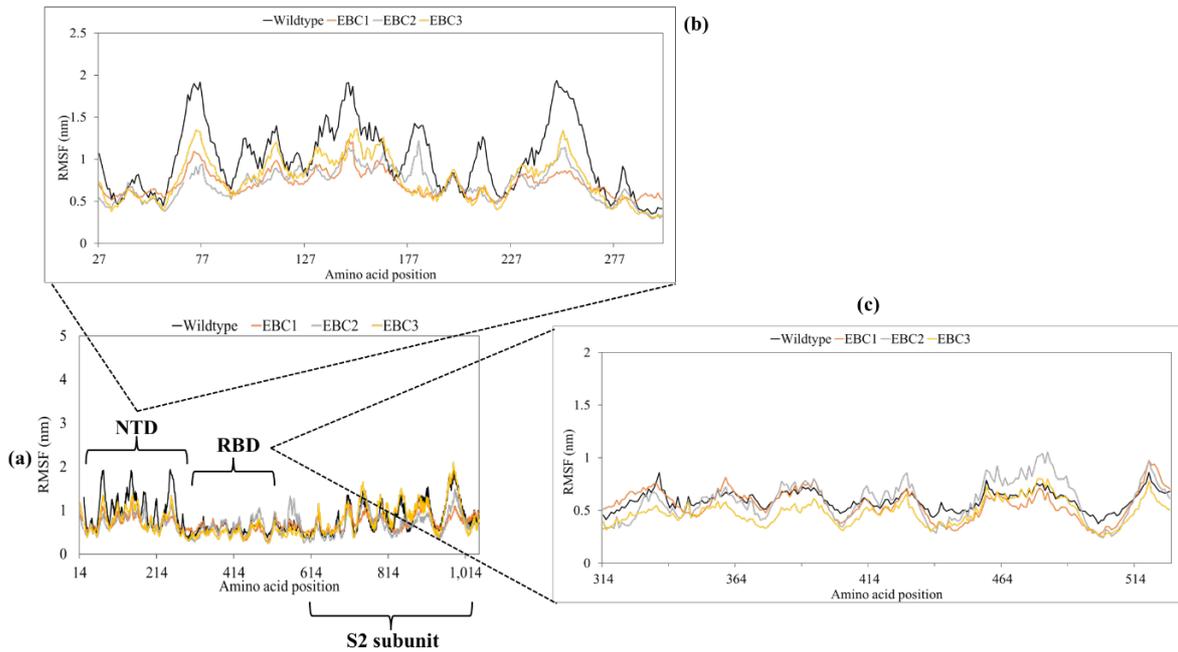


Figure 3. The residual flexibility index of the Wildtype, EBC1-hACE2, EBC2-hACE2, and EBC3-hACE2 complexes. In each panel, EBC1-hACE2, EBC2-hACE2, and EBC3-hACE2 are respectively represented as black, orange, gray, and yellow. (a) RMSF values per residue of the S protein for each variant; (b) RMSF values per residue in the NTD region for each variant; (c) RMSF values per residue in the RBD region for each variant.

The distinction in structural flexibility within each region is visually depicted in the structural analysis presented in Figure 4. Specifically, the NTD region of the Wildtype tends to approach the central area, potentially reverting to the native unbound state of the S protein. In contrast, the NTD region of the

EBC1-hACE2, EBC2-hACE2, and EBC3-hACE2 complexes demonstrates relative stability throughout the simulation. EBC1-hACE2, EBC2-hACE2, and EBC3-hACE2 complexes are the preservation of deletion and point mutations in the NTD that may reduce the flexibility of this region.

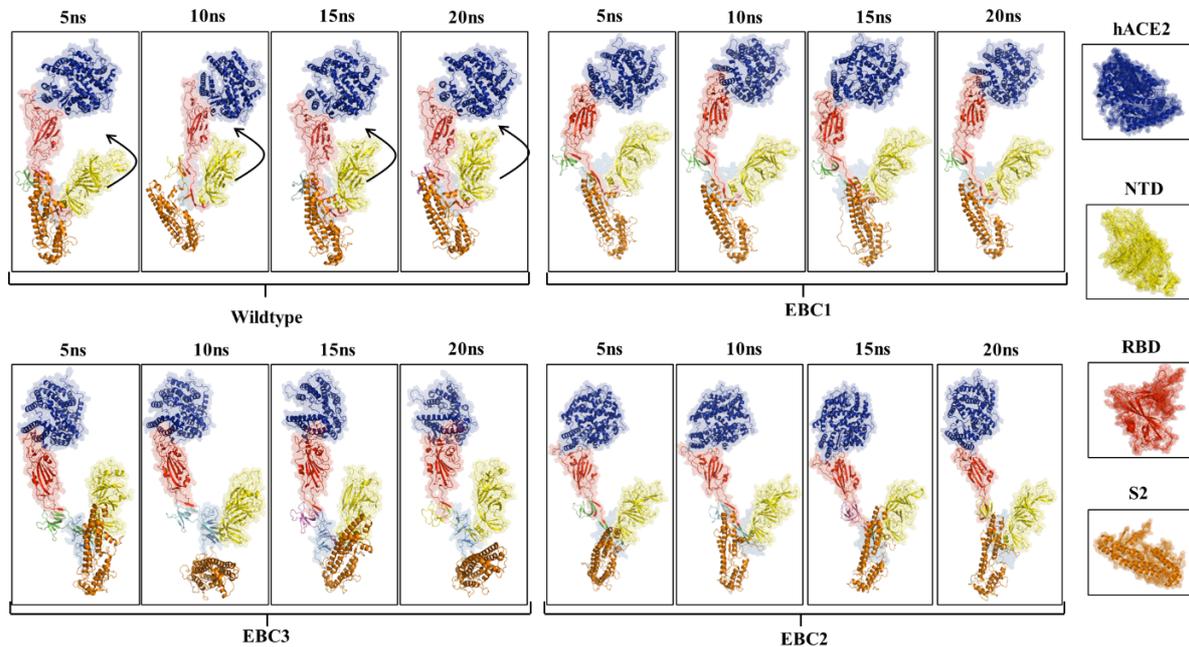


Figure 4. Snapshots of the structural conformations of Wildtype, EBC1-hACE2, EBC2-hACE2, and EBC3-hACE2 complexes at different time points in 20 ns simulation. Different domains are highlighted: NTD in yellow, RBD in red, S2 in orange; the hACE2 receptor is highlighted in blue.

The interaction and movement patterns of hACE2 within the RBD region show similarity across the simulations. Notably, in the EBC2-hACE2 complex, the RBD region tends to move away from the NTD region, explaining the higher RMSF observed in the RBD region of EBC2-hACE2 (Figure 3b). Regarding the S2 region, simulations of both the Wildtype and EBC3-hACE2 complexes indicate that this region tends to separate from S1, consistent with the inherent properties of the S protein when S1-S2 protomer, which was formed by non-covalent interactions (Walls *et al.*, 2020). In the simulations of EBC1-hACE2 and EBC2-hACE2, the dissociation tendency of S2 is less apparent. Therefore, a more extended simulation may clarify the structural changes within this subunit.

The hydrogen bonds analysis formed between spike and hACE2 in each

simulation by VMD software showed that EBC1-hACE2, EBC2-hACE2, and EBC3-hACE2 variants exhibited more hydrogen bonds than the Wildtype from 10 ns to 20 ns simulation (Figure 5). The number of hydrogen bonds in the Wildtype exhibited significant instabilities during the early stages of the simulation. Subsequently, it gradually decreased, stabilizing at 5.9 ± 2.4 bonds (Figure 5a and 5e) after 15 ns. This observed hydrogen bonds of Wildtype aligns with previous research findings (Lupala *et al.*, 2022), thus validating the accuracy of our simulation. On the other hand, simulations of EBC1-hACE2 and EBC2-hACE2 reached a relatively steady state after 15 ns and displayed a similar pattern, maintaining at 6.3 ± 2.3 bonds (Figure 5b, 5c, and 5e). EBC3-hACE2, however, stood out by exhibiting the highest number of hydrogen bonds after 15 ns simulation at 7.9 ± 2.5

bonds (Figure 5d and 5e). At the interface, we noted that the Omicron variants, EBC1-hACE2, EBC2-hACE2, and EBC3-hACE2, carry S477N and Q498R mutations that elucidate why these variants exhibit more hydrogen bonds than the Wildtype. Moreover, the frequency and the number of

amino acids involved in the interface differed among the variants, with EBC3-hACE2 displaying the highest frequency and the most significant number of amino acids engaged in the interface region, which correlates with the higher number of hydrogen bonds observed in this variant.

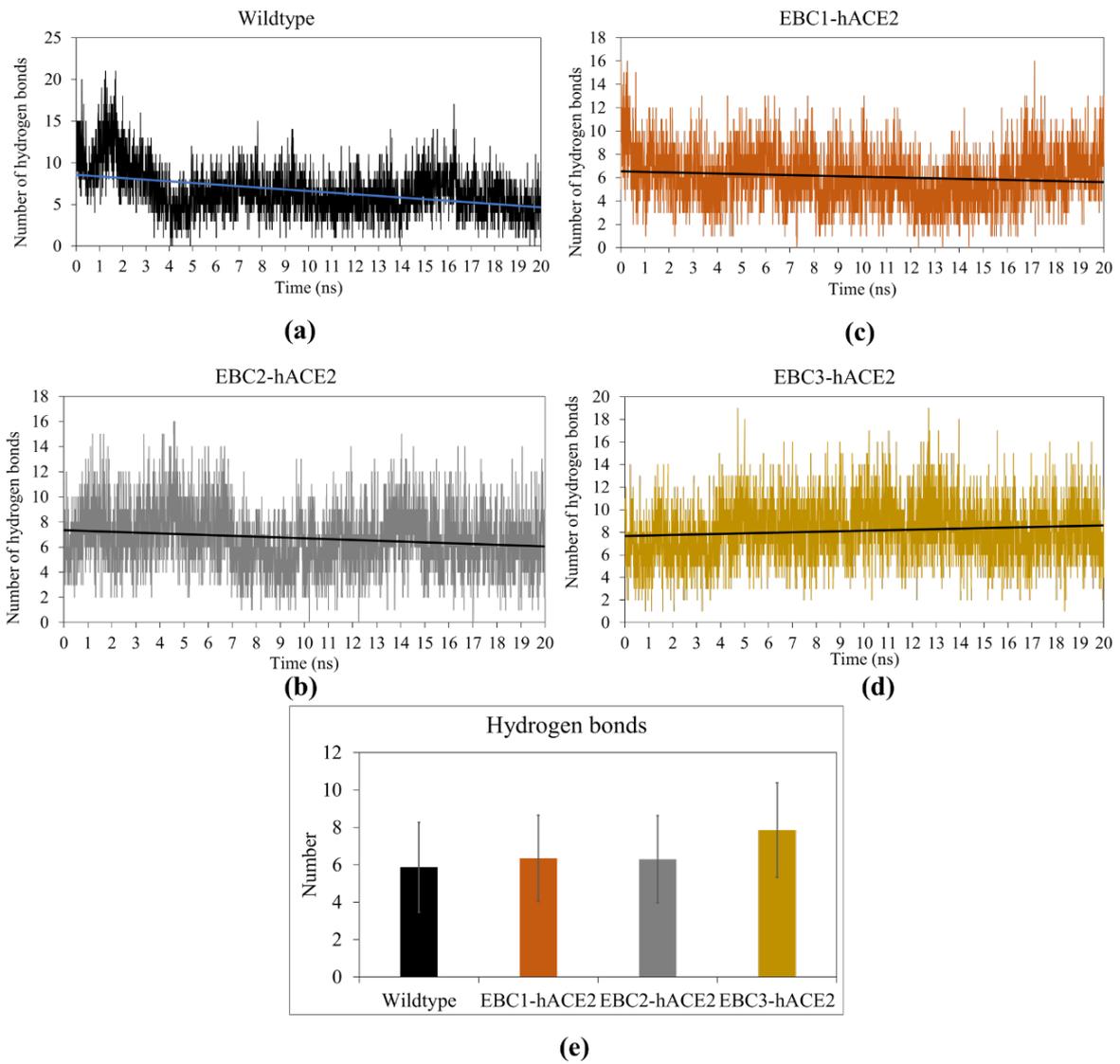


Figure 5. The hydrogen bonds analysis of Wildtype (a), EBC1-hACE2 (b), EBC2-hACE2 (c), EBC3-hACE2 (d) complexes displayed with the trendline, and their average hydrogen bonds (e).

We conducted a binding free energy analysis to assess the binding affinity in each simulation. In this study, using the MM/GBSA and MM/PBSA methods, revealed that the EBC1, EBC2, and EBC3 variants exhibited a higher binding affinity to hACE2 in comparison to the Wildtype. The BFE values were calculated by MM/GBSA as follows: -51.27 ± 4.45 kcal.mol⁻¹ for EBC1, -51.85 ± 6.14 kcal.mol⁻¹ for EBC2, -60.05 ± 6.17 kcal.mol⁻¹ for EBC3, and -41.33 ± 5.67 kcal.mol⁻¹ for Wildtype (Table 4). By MM/PBSA methods, BFE values were shown as follows: -58.57 ± 6.21 kcal.mol⁻¹ for EBC1, -63.89 ± 7.36 kcal.mol⁻¹ for EBC2, -73.92 ± 6.73 kcal.mol⁻¹ for EBC3, and -51.68 ± 7.23 kcal.mol⁻¹ for Wildtype (Table 5). The increased BFE observed in the variants is primarily attributable to a significant enhancement in

electrostatic energy (Table 4 and Table 5). Furthermore, when comparing the BFE of EBC1 and EBC2, we observed indices that lightly increased BFE in EBC3. EBC3, on the other hand, emerged as the variant with the highest BFE, aligning with the prior results of this study. Interestingly, EBC3, which emerged in September 2022, arrived later than EBC2 and EBC1, identified in July 2022 and late January 2022, respectively. EBC3 stands out as the Omicron variant with the highest number of mutations compared to EBC2 and EBC1. Our BFE analysis suggests that the assembly of more mutations of Omicron variants is considered related to stronger binding efficiency to hACE2. We conducted a detailed assessment to substantiate further our findings, including evaluating the energy changes associated with mutation.

Table 4. BFE (kcal.mol⁻¹) and their components, calculated using MM/GBSA methods, for each simulation.

	Wildtype	EBC1	EBC2	EBC3
ΔG_{vdW}	-80.38 ± 1.44	-80.39 ± 1.45	-78.74 ± 1.53	-80.21 ± 1.57
ΔG_{ele}	-892.04 ± 4.17	-1845.32 ± 3.46	-1808.40 ± 4.28	-1650.62 ± 4.72
ΔG_{pol}	943.39 ± 3.42	1885.48 ± 2.13	1846.31 ± 3.99	1682.89 ± 3.49
ΔG_{npol}	-12.3 ± 0.11	-11.04 ± 0.09	-11.02 ± 0.10	-12.11 ± 0.11
ΔG_{GBSA}	-41.33 ± 5.67	-51.27 ± 4.45	-51.85 ± 6.14	-60.05 ± 6.17

Table 5. BFE (kcal.mol⁻¹) and their components, calculated using MM/PBSA methods, for each simulation.

	Wildtype	EBC1	EBC2	EBC3
ΔG_{vdW}	-81.15 ± 4.82	-80.57 ± 4.00	-79.82 ± 4.58	-78.18 ± 5.07
ΔG_{ele}	-907.04 ± 26.53	-1846.41 ± 49.91	-1798.49 ± 45.88	-1634.69 ± 41.68
ΔG_{pol}	947.01 ± 28.12	1877.71 ± 48.53	1824.22 ± 40.93	1648.45 ± 40.71
ΔG_{npol}	-10.51 ± 0.28	-9.30 ± 0.25	-9.80 ± 0.30	-9.51 ± 0.27
ΔG_{PBSA}	-51.68 ± 7.23	-58.57 ± 6.21	-63.89 ± 7.36	-73.92 ± 6.73

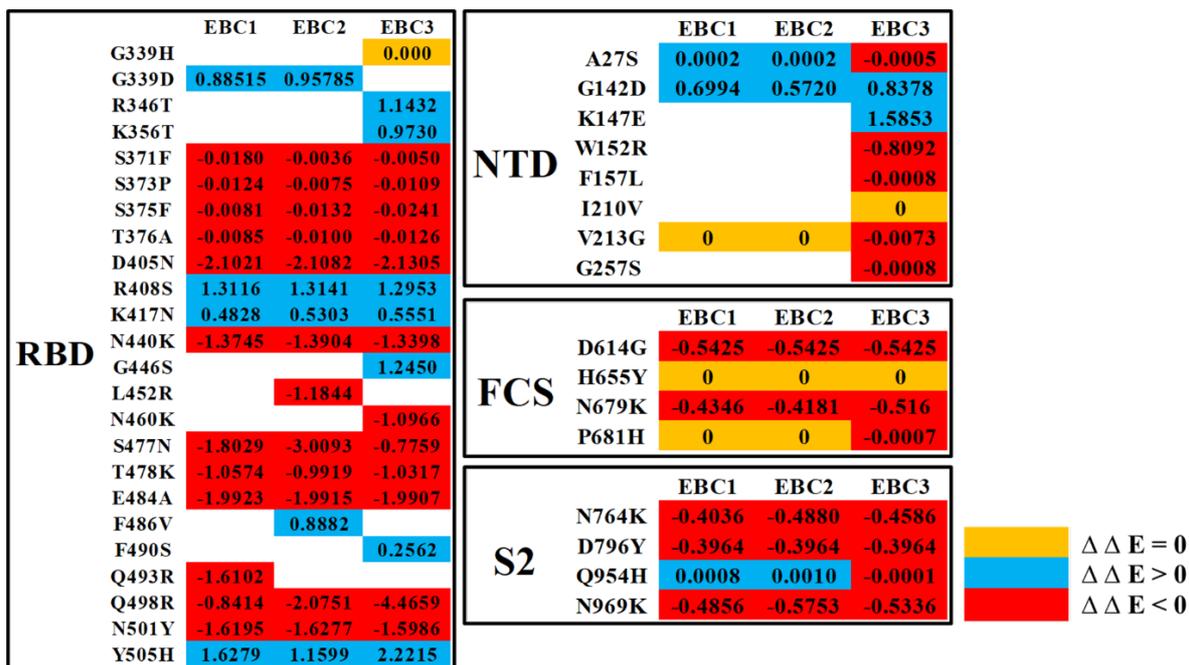


Figure 6. The energy change - $\Delta\Delta E$ (kcal.mol⁻¹) analysis in point mutation of EBC1, EBC2 and EBC3. Yellow (corresponding to $\Delta\Delta E = 0$) indicates that the mutation does not alter the BFE; Blue (related to $\Delta\Delta E > 0$) represents mutations reducing BFE; Red indicates $\Delta\Delta E < 0$, signifying an increase in BFE due to point mutation. Standard deviation is not shown.

The assessment of per-residue contributions of each point mutation to the BFE value was conducted and compared between each simulation. In the NTD, we observed that mutations in EBC1 and EBC2 resulted in $\Delta\Delta E > 0$, signifying a reduction in the BFE of the complex. Conversely, most mutations in the NTD region of EBC3 had $\Delta\Delta E < 0$, contributing to increased BFE (Figure 6). The specific mutations in NTD within the EBC3 variants stood out for their impact on BFE enhancement, including W152R, F157L, and G257S.

The conserved mutations of EBC1, EBC2, and EBC3 variants in the RBD region, include S371F, S373P, S375F, T376A, D405N, N440K, S477N, T478K, E484A, Q498R, and N501Y, exhibit $\Delta\Delta E < 0$. These mutations are associated with an increased affinity for hACE2 (Figure 6).

However, the remaining mutations in the RBD, G339D, R408S, K417N, and Y505H, had $\Delta\Delta E > 0$, reducing BFE (Figure 6). EBC3 introduces two mutations in RBD, R346T and K356T, replacing positively charged side chains with polar uncharged ones. Therefore, R346T and K356T reduce the electrostatic energy of EBC3 when compared to EBC1 and EBC2 (Figure 6) (Table 4). Other mutations that occur in the other S protein region and increase the affinity for hACE2 encompass D614G, N679K, P681H, N764K, D796Y, and N969K (Figure 6). In general, mutations that enhance BFE involve the substitution of polar uncharged and negatively charged side chains with positively charged ones, thereby elevating the electrostatic energy within the interaction complex (Table 4 and Table 5).

In previous research, we built a haplotype network and predicted the impact of each haplotype in the network based on graph theory analysis (data not shown). That motivated us to determine the effect of mutations on the transmission efficiency of Omicron variants. In this study, we evaluated the impact of the mutation on the structural behavior of the Omicron S protein, as well as determined the binding affinity with hACE2 using *in silico* methods.

In our structural analysis, dynamic simulations revealed that Omicron variants exhibit a notable reduction in the flexibility of the NTD. In order to bind with hACE2, the NTD region is required to extend away from the central trimeric structure to maintain the 'up' conformation of the RBD (Benton *et al.*, 2020). In contrast, the Wildtype typically exhibits the NTD region moving towards the center of the trimeric structure, impeding the interaction between RBD and hACE2. Notably, the NTD regions of the Omicron variants in our simulations remained in an 'up' state, consequently not impeding the interaction between RBD and hACE2. Our study lends support to and elucidates the proposal that the substantial number of mutations in the Omicron variant enhances transmission efficiency by sustaining a higher prevalence of 'up' states in the trimeric structures (Hirabara *et al.*, 2021; Kwarteng *et al.*, 2021; Zhang *et al.*, 2022). Furthermore, our dynamics structural analysis, based on RMSD values, indicates that the EBC3 variant displays the most significant structural variability. According to the principles of conformational selection theory (Csermely *et al.*, 2010; Weikl, Paul, 2014; Sang *et al.*, 2022), greater structure flexibility of protein is a crucial prerequisite for improved recognition and binding of ligands. This elucidates why, among the

simulated Omicron variants, EBC3 exhibits the highest affinity for hACE2.

In our analysis of BFE, we observed an electrostatic energy contribution in the interaction with hACE2. This finding aligns with previous studies on the RBD-hACE2 interaction (Khan *et al.*, 2022; Zhang *et al.*, 2022). However, our study goes outside the RBD region. We've discovered that the NTD featuring W152R and mutations in other structural areas, including D614G, N679K, P681H, N764K, D796Y, and N969K, also increases the BFE. These mutations alter the protein's electrostatic surface, emphasizing the importance of examining the entire protein structure to understand the role of S protein mutations. A noteworthy case involves the two mutations, R346T and K356T, in the EBC3 variant, which lead to a decrease in BFE. This is attributed to transforming positively charged side chains into polar side chains. Nonetheless, it is essential to consider that resistance to antibodies and immune evasion are also critical factors in determining the transmissibility of SARS-CoV-2. These mutations have been shown to reduce antibodies' effectiveness (Guo *et al.*, 2023; Planas *et al.*, 2023) which may indirectly explain EBC3's high transmission efficiency in haplotype network analysis. Furthermore, compared to EBC1 and EBC2, EBC3 exhibits higher polarization energy, a factor believed to contribute to its enhanced affinity for hACE2 (Khan *et al.*, 2022; Sang *et al.*, 2022). New mutations in the S protein in the future may alter the structural behavior and improve the transmissibility for new emergence variants by increasing their binding affinity to the hACE2 receptor.

However, it is crucial to evaluate the limitations of our study. Firstly, our predictions regarding the structural behavior

of proteins are grounded *in silico* modeling, lacking experimental validation. Another limitation is the absence of extended, replicated simulations, which would offer a more comprehensive understanding of result reproducibility and a robust foundation for accurate interactions. Hence, the necessity for further in-depth *in silico* investigations and subsequent *in vitro* assessments becomes evident.

CONCLUSION

In conclusion, our study of mutations in the S protein reveals substantial changes in structural behavior and binding affinity. The enhanced interaction efficiency of Omicron variants is attributed to their capacity to maintain a more stable 'up' state structure than the Wildtype. Omicron variants, characterized by numerous mutations, exhibit stronger binding energy with hACE2 in comparison to the Wildtype. Notable mutations that enhance affinity include W152R, F157L, G257S in the NTD; S371F, S373P, S375F, T376A, D405N, N440K, S477N, T478K, E484A, Q498R, N501Y in the RBD; and D614G, N679K, P681H, N764K, D796Y, N969K in other structural regions. The presence of point mutations enhances electrostatic energy on the protein surface by transforming positively charged side chains.

REFERENCES

Abdalla M, Eltayb WA, El-Arabey AA, Singh K, Jiang X (2022) Molecular dynamic study of SARS-CoV-2 with various S protein mutations and their effect on thermodynamic properties. *Comput Biol Med* 141: 105025.

Abeywardhana S, Premathilaka M, Bandaranayake U, Perera D, Peiris LDC (2023) In silico study of SARS-CoV-2 spike protein RBD and human ACE-2 affinity dynamics

across variants and Omicron subvariants. *J Med Virol* 95: e28406.

Benton DJ, Wrobel AG, Xu P, Roustan C, Martin SR, Rosenthal PB, Skehel JJ, Gamblin SJ (2020) Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion. *Nature* 588: 327-330.

Csermely P, Palotai R, Nussinov R (2010) Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends Biochem Sci* 35: 539-546.

Duong BV, Larpruenrudee P, Fang T, Hossain SI, Saha SC, Gu Y, Islam MS (2022) Is the SARS CoV-2 Omicron Variant Deadlier and More Transmissible Than Delta Variant? *Int J Environ Res Public Health* 19: 4586.

Elliott P, Eales O, Bodinier B, Tang D, Wang H, Jonnerby J, Haw D, Elliott J, Whitaker M, Walters CE, Atchison C, Diggle PJ, Page AJ, Trotter AJ, Ashby D, Barclay W, Taylor G, Ward H, Darzi A, Cooke GS, Chadeau-Hyam M, Donnelly CA (2022) Dynamics of a national Omicron SARS-CoV-2 epidemic during January 2022 in England. *Nat Commun* 13: 4500.

Fang L, Xu J, Zhao Y, Fan J, Shen J, Liu W, Cao G (2023) The effects of amino acid substitution of spike protein and genomic recombination on the evolution of SARS-CoV-2. *Front Microbiol* 14: 1228128.

Guo H, Jiang J, Shen S, Ge X, Fan Q, Zhou B, Cheng L, Ju B, Zhang Z (2023) Additional mutations based on Omicron BA.2.75 mediate its further evasion from broadly neutralizing antibodies. *iScience* 26: 106283.

Harvey WT, Carabelli AM, Jackson B, Gupta RK, Thomson EC, Harrison EM, Ludden C, Reeve R, Rambaut A, Peacock SJ, Robertson DL, Consortium C-GU (2021) SARS-CoV-2 variants, spike mutations and immune escape. *Nat Rev Microbiol* 19: 409-424.

Hirabara SM, Serdan TDA, Gorjao R, Masi LN, Pithon-Curi TC, Covas DT, Curi R, Durigon EL (2021) SARS-COV-2 Variants: Differences and

- Potential of Immune Evasion. *Front Cell Infect Microbiol* 11: 781429.
- Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, Grubmüller H, MacKerell AD, Jr. (2017) CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods* 14: 71-73.
- Huang Y, Yang C, Xu XF, Xu W, Liu SW (2020) Structural and functional properties of SARS-CoV-2 spike protein: potential antiviral drug development for COVID-19. *Acta Pharmacologica Sinica* 41: 1141-1149.
- Khan A, Waris H, Rafique M, Suleman M, Mohammad A, Ali SS, Khan T, Waheed Y, Liao C, Wei DQ (2022) The Omicron (B.1.1.529) variant of SARS-CoV-2 binds to the hACE2 receptor more strongly and escapes the antibody response: Insights from structural and simulation data. *Int J Biol Macromol* 200: 438-448.
- Khan A, Zia T, Suleman M, Khan T, Ali SS, Abbasi AA, Mohammad A, Wei DQ (2021) Higher infectivity of the SARS-CoV-2 new variants is associated with K417N/T, E484K, and N501Y mutants: An insight from structural data. *J Cell Physiol* 236: 7045-7057.
- Kwarteng A, Asiedu E, Sylverken AA, Larbi A, Sakyi SA, Asiedu SO (2021) Molecular characterization of interactions between the D614G variant of SARS-CoV-2 S-protein and neutralizing antibodies: A computational approach. *Infect Genet Evol* 91: 104815.
- Li J, Lai S, Gao GF, Shi W (2021) The emergence, genomic diversity and global spread of SARS-CoV-2. *Nature* 600: 408-418.
- Lupala CS, Ye Y, Chen H, Su X-D, Liu H (2022) Mutations on RBD of SARS-CoV-2 Omicron variant result in stronger binding to human ACE2 receptor. *Biochem Biophys Res Commun* 590: 34-41.
- Neijenhuis T, van Keulen SC, Bonvin AMJJ (2022) Interface refinement of low- to medium-resolution Cryo-EM complexes using HADDOCK2.4. *Structure* 30: 476-484.e473.
- Planas D, Bruel T, Staropoli I, Guivel-Benhassine F, Porrot F, Maes P, Grzelak L, Prot M, Mougari S, Planchais C, Puech J, Saliba M, Sahraoui R, Fémy F, Morel N, Dufloo J, Sanjuán R, Mouquet H, André E, Hocqueloux L, Simon-Loriere E, Veyer D, Prazuck T, Péré H, Schwartz O (2023) Resistance of Omicron subvariants BA.2.75.2, BA.4.6, and BQ.1.1 to neutralizing antibodies. *Nat Commun* 14: 824.
- Poudel S, Ishak A, Perez-Fernandez J, Garcia E, León-Figueroa DA, Romaní L, Bonilla-Aldana DK, Rodriguez-Morales AJ (2022) Highly mutated SARS-CoV-2 Omicron variant sparks significant concern among global experts - What is known so far? *Travel Med Infect Dis* 45: 102234.
- Sang P, Chen YQ, Liu MT, Wang YT, Yue T, Li Y, Yin YR, Yang LQ (2022) Electrostatic Interactions Are the Primary Determinant of the Binding Affinity of SARS-CoV-2 Spike RBD to ACE2: A Computational Case Study of Omicron Variants. *Int J Mol Sci* 23: 14796.
- Shahhosseini N, Babuadze GG, Wong G, Kobinger GP (2021) Mutation Signatures and In Silico Docking of Novel SARS-CoV-2 Variants of Concern. *Microorganisms* 9: 926.
- Studer G, Rempfer C, Waterhouse AM, Gumienny R, Haas J, Schwede T (2020) QMEANDisCo—distance constraints applied on model quality estimation. *Bioinformatics* 36: 1765-1771.
- Tamura K, Stecher G, Kumar S (2021) MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol Biol Evol* 38: 3022-3027.
- V'kovski P, Kratzel A, Steiner S, Stalder H, Thiel V (2021) Coronavirus biology and replication: implications for SARS-CoV-2. *Nat Rev Microbiol* 19: 155-170.
- Valdés-Tresanco MS, Valdés-Tresanco ME, Valiente PA, Moreno E (2021) gmx_MMPBSA: A New Tool to Perform End-State Free Energy Calculations with GROMACS. *J Chem Theory Comput* 17: 6281-6291.

- Viana R, Moyo S, Amoako DG, Tegally H, Scheepers C, Althaus CL, Anyaneji UJ, Bester PA, Boni MF, Chand M, Choga WT, Colquhoun R, Davids M, Deforche K, Doolabh D, du Plessis L, Engelbrecht S, Everatt J, Giandhari J, Giovanetti M, Hardie D, Hill V, Hsiao N-Y, Iranzadeh A, Ismail A, Joseph C, Joseph R, Koopile L, Kosakovsky Pond SL, Kraemer MUG, Kuate-Lere L, Laguda-Akingba O, Lesetedi-Mafoko O, Lessells RJ, Lockman S, Lucaci AG, Maharaj A, Mahlangu B, Maponga T, Mahlakwane K, Makatini Z, Marais G, Maruapula D, Masupu K, Matshaba M, Mayaphi S, Mbhele N, Mbulawa MB, Mendes A, Mlisana K, Mnguni A, Mohale T, Moir M, Moruisi K, Mosepele M, Motsatsi G, Motswaledi MS, Mphoyakgosi T, Msomi N, Mwangi PN, Naidoo Y, Ntuli N, Nyaga M, Olubayo L, Pillay S, Radibe B, Ramphal Y, Ramphal U, San JE, Scott L, Shapiro R, Singh L, Smith-Lawrence P, Stevens W, Strydom A, Subramoney K, Tebeila N, Tshiabuila D, Tsui J, van Wyk S, Weaver S, Wibmer CK, Wilkinson E, Wolter N, Zarebski AE, Zuze B, Goedhals D, Preiser W, Treurnicht F, Venter M, Williamson C, Pybus OG, Bhiman J, Glass A, Martin DP, Rambaut A, Gaseitsiwe S, von Gottberg A, de Oliveira T (2022) Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. *Nature* 603: 679-686.
- Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D (2020) Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* 181: 281-292.e286.
- Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T (2018) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* 46: W296-w303.
- Weickl TR, Paul F (2014) Conformational selection in protein binding and function. *Protein Sci* 23: 1508-1518.
- Xu L, Li Y, Li D, Xu P, Tian S, Sun H, Liu H, Hou T (2015) Exploring the binding mechanisms of MIF to CXCR2 using theoretical approaches. *Phys Chem Chem Phys* 17: 3370-3382.
- Xue LC, Rodrigues JP, Kastiris PL, Bonvin AM, Vangone A (2016) PRODIGY: a web server for predicting the binding affinity of protein-protein complexes. *Bioinformatics* 32: 3676-3678.
- Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou QJS (2020) Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* 367: 1444-1448.
- Zhang ZB, Xia YL, Shen JX, Du WW, Fu YX, Liu SQ (2022) Mechanistic Origin of Different Binding Affinities of SARS-CoV and SARS-CoV-2 Spike RBDs to Human ACE2. *Cells* 11: 1274.