

## INVESTIGATING THE IMPACT OF SPIKE PROTEIN MUTATIONS ON SARS-COV-2 VIRULENCE IN BENIN USING NETWORK CENTRALITY AND MOLECULAR DOCKING APPROACHES

Quan Ke Thai<sup>1,✉</sup>, Phuoc Huynh<sup>2</sup>, Huyen Nguyen Thi Thuong<sup>3</sup>, Quoc-Dang Quan<sup>4</sup>

<sup>1</sup>Saigon University, 273 An Duong Vuong, Ward 3, District 5, Ho Chi Minh City, Vietnam

<sup>2</sup>University of Science, Vietnam National University Ho Chi Minh City, 227 Nguyen Van Cu, Ward 4, District 5, Ho Chi Minh City, Vietnam

<sup>3</sup>Department of Biology, HCMC University of Education, 280 An Duong Vuong, Ward 4, District 5, Ho Chi Minh City, Vietnam

<sup>4</sup>Thien Thanh Bio Co.,Ltd, 204/9 Truong Cong Dinh, Ward 14, Tan Binh District, Ho Chi Minh City, Vietnam

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

Received: 17.4.2023

Accepted: 27.6.2023

### SUMMARY

The COVID-19 pandemic is ongoing and spreading around the world, which means a continuous increase in the number of infections and death. SARS-CoV-2 constantly rapidly stored mutation in the Spike gene to adapt with the host cell. The Spike gene encoded spike protein directly interacts with hACE2 on the human cell surface. Herein, using the network centrality and molecular docking approaches, we detected key mutations that positively affect spike protein. Based on network centrality, we demonstrate that the A23403G (D614G) mutation in the Spike gene is the center of a network which means this mutation has a positive effect on the virus. In addition, analyzing the interaction of spike protein with hACE2, we highlighted that the mutation appeared in the RBD region by changing the electrostatic energy of the complex. Remarkably, mutations N440K, L452R, T478K, E484K, Q493R, and Q498R increased binding free energy of Spike-hACE2 complex due to the change of the side chain into a positive charge. The Eta, Delta, and Omicron variants existed in one or more of these mutations resulting in higher binding free energy and binding affinity than the Wuhan variant indicating sounder interaction with hACE2. In general, mutations appearing on the spike protein tended to cause the surface to become positively charged in order to interact easily with the negative surface of the hACE2 receptor.

**Keywords:** Benin, COVID-19, SARS-CoV-2, Spike gene, spike protein.

### INTRODUCTION

SARS-CoV-2 belongs to the *Betacoronavirus* genus and is the primary cause of the ongoing COVID-19 pandemic

from late 2019, causing new outbreaks worldwide. In the viral core, SARS-CoV-2 contains the +ssRNA genome and its stored four structure genes, which encoded for

Nucleocapsid protein (N), Spike (S) protein, Envelope (E), and Matrix protein (M) (Wu *et al.*, 2020; Zhou *et al.*, 2020). In the viral particle, the S protein is a homotrimeric protein located on the envelope of SARS-CoV-2, with each protomer consisting of two subunits, the S1 subunit and the S2 subunit. The S1 subunit includes the N-terminal domain (NTD) and the Receptor Binding Domain (RBD), which are responsible for recognizing and interacting with the hACE2 receptor. After binding to the receptor, the S2 subunit is activated by cleavage at the furin cleavage site (FCS) by host cell proteases like transmembrane serine protease 2 (TMPRSS2) on the host cell for membrane fusion (Lan *et al.*, 2020). Furthermore, the S gene is the longest and has a higher mutation rate than the genome (Berrio *et al.*, 2020; Chaw *et al.*, 2020). The S gene was positively selected in the evolution of SARS-CoV-2, while other structural genes were not (Berrio *et al.*, 2020). Therefore, during transmission, SARS-CoV-2 undergoes positive mutation in the S gene to adapt to the human host cells. In addition, the S protein is the main target for vaccine development as well as immunotherapy treatment (Martínez-Flores *et al.*, 2021; Tabynov *et al.*, 2022; Candido *et al.*, 2022). Consequently, mutations in this protein can enhance SARS-CoV-2 adaptation to the human immune system.

Understanding the importance of spike protein, WHO also classified the SARS-CoV-2 based on the mutation appeared in this protein, including Variants Under Monitor (VUM), Variants of Interesting (VOI), and Variants of Concern (VOC), the latter being associated with more phenotypes. In the past, the world recorded some VOCs have a high transmission rate, virulence, and the ability to evade or subvert the immune system. The VOC is the most

notable in Delta, the main cause for millions of deaths worldwide. Delta has an average transmission rate of 63% to 167% higher than Alpha – the previous VOC variant (Earnest, Uddin *et al.* 2022) and infected over 200 countries (Yang and Shaman). Therefore, Delta rapidly became the most dominant variant in the world. Delta conserves many mutations that have demonstrated an increasing affinity for interactions of spike protein with hACE2. Additionally, Delta carries deletion mutations ( $\Delta 69/70$ ,  $\Delta 144/145$ , and  $\Delta 156/157$ ) in the (NTD) and missense (K417N, L452R, and T478K) mutations in the (RBD) that play essential roles in escape antibodies and increase binding affinity (Han *et al.*, 2022; Di Giacomo *et al.*, 2021; Cherian *et al.*, 2021; Liu *et al.*, 2022). Furthermore, Delta contains two substitutions, D614G and P681R, in the Furin Cleavage Site (FCS) that directly increase the cleavage effectiveness (Kumar *et al.*, 2021; Fan *et al.*, 2021). A new VOC variant that emerged in December 2021, called Omicron, has more than 30 mutations in the spike protein (Viana *et al.*, 2022). Currently, Omicron has replaced Delta to become the most popular variant worldwide and proliferated numerous novel subvariants containing more dangerous phenotypes, such as BA.2, BA.2.75, or BQ.1. The numerous mutations in spike protein provide Omicron's with robust adaptability to the host. Much evidence demonstrates escape antibody and breakthrough of the immune system to reinfection events (Tan *et al.*, 2023; Cele *et al.*, 2022; Carreño *et al.*, 2022). Thus, the continuous analysis of the mutation that appears in spike protein will provide valuable information for tracking the emergence of novel variants with significant substitutions.

In the previous study, we examined the genetic diversity of the S gene of SARS-CoV-2 variants that appeared in the Benin Republic (data not shown). The preliminary results show that Benin has higher nucleotides and haplotype diversity levels. We indicated the parallel circling of B.1.318 and B.1.525 (Eta) variants in the early outbreak in March and April 2021. Then, Delta and Omicron will become the dominant variants in late 2021. We also exposed the quiet transmission in the Benin community along the time of B.1.318 until that was replaced by the Delta variant. In addition, we evaluated the sites of high-frequency nucleotide polymorphisms and the nucleotide variation in the S gene during the circulating period of SARS-CoV-2 in Benin. However, we have yet to investigate which nucleotide variation leads to successful transmission in Benin, nor have we evaluated the impact of mutations on spike protein. Therefore, in this study, we will examine the effectiveness of mutation in the S gene to protein spike through the network centrality and molecular docking.

## MATERIALS AND METHODS

### Identification node centrality of network

In our previous study, we divided 1230 the S gene sequences in Benin into 464 haplotypes (Hap) which were numbering from Hap\_1 to Hap\_464 with Hap\_1 as the Reference sequence. This study used these haplotypes to build the haplotype network using POPART software (Leigh, Bryant, 2015) based on the Minimum Spanning method. The POPART's outcomes were used to calculate the network centrality in Gephi software version 0.9.7. Then, the top 6 nodes with the highest betweenness

centrality were selected for downstream analysis.

### Homology modelling and protein-protein docking

After extracting nodes from the haplotype network, we translated these haplotypes into protein sequences and used SWISS-MODEL (<https://swissmodel.expasy.org/>) to build a three-dimensional (3D) protein structure. We used chain B of crystal spike protein – hACE2 (Spike-hACE2) complex in PDB (Protein Data Bank) (PDB ID: 7DF4) as the template structure (Xu *et al.*, 2021). The quality of homology structures was estimated by the Ramachandran plot in the Structure Assessment tool, PROCHECK in the PDBsum server, ProSA-web server (Wiederstein, Sippl 2007) and ERRAT (Colovos, Yeates, 1993) on the SAVES ver 6.0 package.

Six 3D homology spike protein structures will be docked with the hACE2 template structure by the HADDOCK (High Ambiguity Driven protein-protein DOCKing) server (Dominguez *et al.*, 2003). To enhance the reliability of the docking process, we used Optimize run for bioinformatics and the ambiguous restraints (AIRs) features of the HADDOCK server. The cross-validation docking structure was established in the Cluspro server (Kozakov *et al.*, 2017).

### Estimating binding free energy and binding affinity

The binding free energy ( $G_{bind}$ ) and binding affinity ( $\Delta G_{predicted}$ ) is the mutual indices for estimating the affinity of protein-protein complex. To calculate  $G_{bind}$  for the Spike-hACE2 complex, we used the

Hawkdock server with Molecular Mechanics Generalized Born Surface Area (MM/GBSA) approach, which is the best approach for predicting the binding free energy of protein-protein/RNA/DNA complex (Weng *et al.*, 2019; Chen *et al.*, 2016). To calculate MM/GBSA, the Hawkdock server minimized the complex in ff02 force fields for 5000 steps, and the maximum distance for van der Waals interactions is 12 Å. The total  $G_{bind}$  has indicated details in equation (1). Furthermore, to give more information about the complex, PRODIGY server was used to calculate binding affinity (Vangone, Bonvin 2015; Xue *et al.*, 2016).

$$G_{bind} = \Delta G_{Complex} - \Delta G_{Spike} - \Delta G_{hACE2} \quad (1)$$

### Protein-protein interface network bonding

After molecular docking, these complex structures were examined by the PDBsum server to identify the network bonding of the interface. The MM/GBSA approach in Hawkdock (as described above) was used to determine the hot spot residues of the spike directly interacting with hACE2 through the free energy per residue. We also use the free energy per residue of PyDockEneRes server as the cross-validation with Hawkdock (Romero-Durana *et al.*, 2020). Then, determination of the positive effect of mutation that appeared in spike protein, we used the output that were calculated in equation (2) which was suggested in the Lei Xu *et al.* publication (2015).  $\Delta\Delta E$  is below 0 which means the mutation enhances the binding

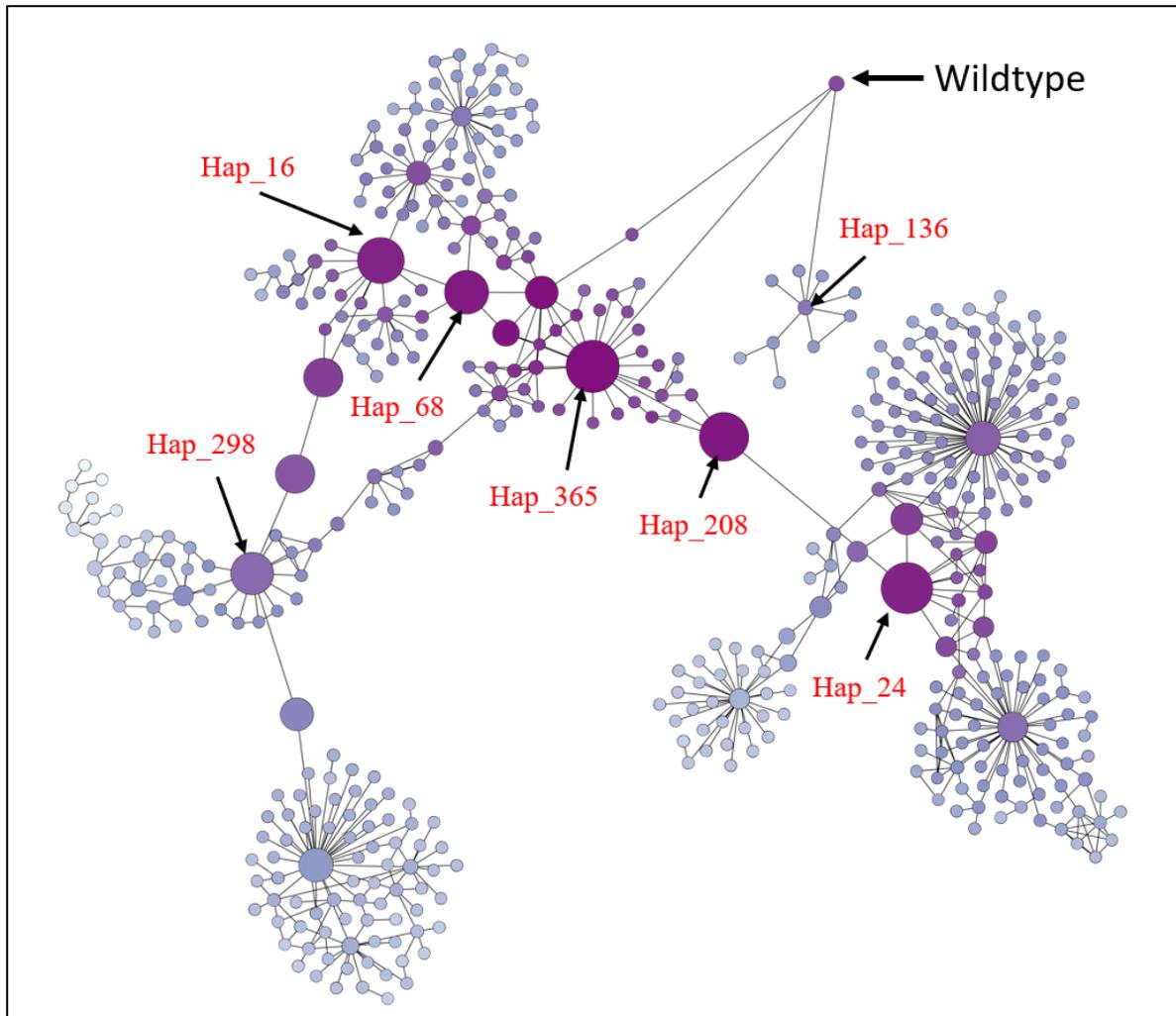
affinity of the spike with hACE and the opposite  $\Delta\Delta E$  is higher than 0.

$$\Delta\Delta E = \Delta E_{mutant} - \Delta E_{wildtype} \quad (2)$$

## RESULTS

### Identification network centrality

In this work, we used POPART software to construct a haplotype network and Gephi software to visualize and calculate the betweenness and closeness centrality of the network. In Gephi, the network was built by 464 nodes corresponding to 464 haplotypes with 610 edges (Figure 1). Analyzing the network centrality indicated that Hap\_365 exhibited the highest betweenness and closeness centrality, indicating that this haplotype is the essential linker of the network (Figure 1). Hap\_365 contains A23403G that changes Glutamic acid at residue 614 to Glycine in protein which is the beginning of the success of novel variants. Without A23403G (D614G) substitution, Hap\_136 contains C21614T (L18F), G22468T, T22917G (L452R), A23063T (N501Y), C23520T, C23525T (H655Y), G23948T, G25218T mutations but this mutant does not successfully proliferate create the descendants in the network (Figure 1). These findings emphasize the importance of the mutation in the S gene for haplotype with the highest betweenness centrality index. Therefore, based on the output of Gephi, we selected the top 6 haplotypes to translate their spike nucleotide sequences into protein sequences for subsequent analysis. The betweenness and closeness centrality indexes, along with the mutation profile of these six haplotypes, are listed in Table 1.



**Figure 1.** The haplotype network of Spike gene in Benin. In this network, the scale color displayed for the closeness centrality measures – more light is closer to 0, and darker is closer to 1. The size of the node presents the betweenness centrality index.

### Homology spike protein models

The spike gene sequence of six haplotypes was used to build the homology structure base on the SWISS-MODELLING server. The models was assessed in the Structure Assessment tool, ProSA-web server, PROCHECK of PDBsum server, and ERRAT in the SAVES ver 6.0 package. Analyzing our homology models showed that all models contained more than 93% of

the Ramachandran plot and more than 80 % of the overall quality factor in ERRAT. Specially, all models had G-Factor of PROCHECK about 0.31 to 0.35, being more than  $-0.5$ , which meant the structure was usually. To check the angle of the *C-alpha*, we superimposed the 3D homology models with reference structure. The RMSD (Root-mean-square deviation) of all models was less than  $0.03$  (Å), indicating that mutation in spike protein was set up successfully.

Generally, all models were passed, which could be used for downstream analysis. The detailed index of homology models is displayed in Table 2.

**Table 1.** The closeness, betweenness centrality, mutation profile of 6 haplotypes.

	<b>Closeness</b>	<b>Betweenness</b>	<b>Variant</b>	<b>Nucleotide mutation</b>	<b>Amino acid substitution</b>
<b>Hap_365</b>	0.164418	0.542285	B.1	A23403G	D614G
<b>Hap_298</b>	0.12951	0.406286	Omicron	C21762T, C21846T, G22578A, G22599A, T22673C, C22674T, T22679C, C22686T, G22813T, T22882G, G22898A, G22992A, C22995A, A23013C, A23040G, G23048A, A23055G, A23063T, T23075C, C23202A, A23403G, C23525T, T23599G, C23604A, C23854A, G23948T, C24130A, A24424T, T24469A, C24503T, C25000T	A67V, T95I, G339D, R346K, S371P, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y, Y505H, T547K, D614G, H655Y, N679K, P681H, N764K, D796Y, N856K, Q954H, N969K, L981F
<b>Hap_208</b>	0.160876	0.491819	B.1	G23012A, A23403G, G23593C, G23868T	E484K, D614G, Q677H
<b>Hap_68</b>	0.160319	0.425667	Delta	C21618G, G21987A, T22917G, C22995A, A23403G, C23604G, G24410A	T19R, G142D, L452R, T478K, D614G, P681R, D950N
<b>Hap_24</b>	0.157162	0.526946	Eta	Del21991-3TTA, G23012A, A23403G, G23593C	Del144Y, E484K, D614G
<b>Hap_16</b>	0.156525	0.459042	Delta	C21618G, G21987A, Del22029-34AGTTCA, T22917G, C22995A, A23403G, C23604G, G24410A	T19R, G142D, Del156/157, L452R, T478K, D614G, P681R, D950N

**Protein-protein docking**

After constructing the homology models, we utilized these spike mutants to form complexes with hACE2 using the HADDOCK server. The Spike-hACE2 complexes were successfully generated

with HADDOCK, and all structures, except for Hap\_298, exhibited stability with an RMSD (Root-mean-square deviation) ranging from 0.1 to 1.2 Å. The higher vibration observed in the RMSD of the Hap\_298 complex compared to other structures may be primarily attributed by

the presence of 29 amino acid substitutions in the spike protein. Consistent with the HADDOCK score, the Cluspro server also demonstrated similar energy profiles for these complexes. The Hap\_298 complex had the most mutations but only fluctuations and did not change in

energy. This indicated that mutations appearing on the protein did not seem to change much in terms of protein structure, as well as how spike protein interacted with hACE2. The HADDOCK scores, RMSD, and Cluspro energy are shown in Table 3.

**Table 2.** Quality of six homology models built by SWISS-MODELLING.

Models	Ramachandran (%)	ProSA-web (Z-scores)	PROCHECK (G-Factor)	ERRAT (%)	RMSD (A)
Hap_365	93.29	-12.29	0.34	90.41	0.201
Hap_298	93.55	-12.17	0.31	90.52	0.220
Hap_208	94.35	-12.25	0.32	87.70	0.179
Hap_68	93.64	-12.39	0.32	87.00	0.171
Hap_24	95.14	-12.19	0.35	81.00	0.073
Hap_16	93.98	-12.33	0.33	83.80	0.098

**Table 3.** The measures of Spike-hACE2 complex docked.

	HADDOCK scores	RMSD	Cluspro energy
WT	-141.5 ± 3.5	0.4 ± 0.3	-1425.7
Hap_16	-146.3 ± 1.9	0.7 ± 0.5	-1498.6
Hap_24	-137.1 ± 0.3	0.4 ± 0.2	-1425.6
Hap_68	-143.0 ± 2.5	0.4 ± 0.2	-1534
Hap_208	-139.6 ± 3.6	0.6 ± 0.4	-1474.9
Hap_298	-140.7 ± 1.6	1.3 ± 0.9	-1428.4
Hap_365	-142.9 ± 3.7	0.7 ± 0.4	-1431.5

### Estimate binding free energy and binding affinity of Spike-hACE2 complex

Based on the HADDOCK output, we submitted these complexes to the Hawkdock server to calculate the  $\Delta G_{bind}$  and PRODIGY server to estimate the binding affinity ( $\Delta G_{predicted}$ ). We recorded that Hap\_16 and Hap\_68 had the lowest  $\Delta G_{bind}$  and  $\Delta G_{predicted}$  (Table 4), indicating their strong binding affinity to

hACE2. These haplotypes were classified as Delta variants. Conversely, we also indicated that the Spike-hACE2 complexes of Hap\_208 and Hap\_365 had the  $\Delta G_{bind}$  and  $\Delta G_{predicted}$  that were not significantly different from the Wildtype (WT) structure. Hap\_208 and Hap\_365 belong to the B.1 variants; in particular, Hap\_365 only contains A23403G (D614G) mutation, which is present in the FCS region; therefore, this mutation did not

increase the binding affinity of spike with hACE2. Hap\_208 contained E484K in RBD but did not increase the  $\Delta G_{bind}$  and  $\Delta G_{predicted}$ . We also found that Hap\_24 only carried the E484K mutation in the RBD region, which increased the  $\Delta G_{bind}$  but did not change  $\Delta G_{predicted}$ . Thus, other modifications in Hap\_24 had a relative impact on the Spike-hACE2 complex. However, Hap\_24 and Hap\_208 have in common the enhancement of the complex's electrostatic energy ( $\Delta E_{electrostatics}$ ) that may be due to the E484K mutation. The  $\Delta E_{electrostatics}$  plays a vital role in the protein-protein interaction; therefore, this mutation may

positively affect the complex. In addition, Hap\_298 is classified into Omicron variants with the higher  $\Delta G_{bind}$  and  $\Delta E_{electrostatics}$  with WT, which means the spike protein of this haplotype binds stronger with hACE2. However, Hap\_298 had  $\Delta G_{predicted}$  lower than the WT; the main reason for this situation is this spike protein contains 29 amino acid mutations, which affects the interface of the Spike-hACE2 complex. Overall, the examination of the Spike-hACE2 complex indicated that haplotypes belonging to Delta and Omicron variants displayed higher  $\Delta G_{bind}$ , suggesting enhanced binding to the hACE receptor.

**Table 4.** Binding free energy and binding affinity of Spike-hACE2 complexes.

	WT	Hap_16	Hap_24	Hap_68	Hap_208	Hap_298	Hap_365
$\Delta E_{van\ der\ Waals}$	-123.37	-129.55	-125.77	-126.08	-133.15	-118.65	-125.09
$\Delta E_{electrostatics}$	-606.16	-	-1165.87	-	-1143.36	-2066.33	-689.67
		1304.52		1311.92			
$\Delta G_{Polar\ Solvation}$	662.26	1355.54	1215.76	1362.25	1211.2	2109.01	745.37
$\Delta G_{Nonpolar\ Solvation}$	-16.13	-17.26	-16.81	-16.53	-17.44	-15.48	-16.22
$\Delta G_{bind}$	-83.41	-95.79	-92.69	-92.28	-82.76	-91.45	-85.61
$\Delta G_{predicted}$	-14	-14.8	-13.8	-14.8	-14.1	-13.6	-14.6

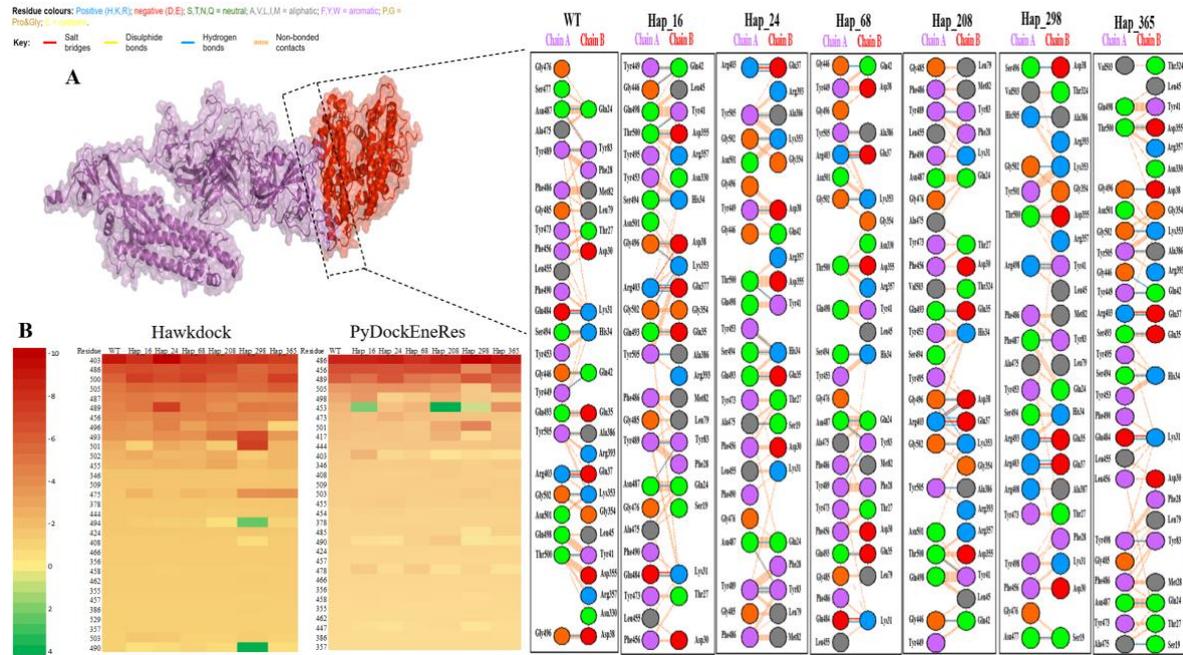
### Protein-protein network bonding and positive mutation

The amino acids in the interface were analyzed by the PDBsum server, and the free energy per residue of the complex output was obtained from Hawkdock and PyDockEneRes as shown in Figures 2A and 2B, respectively. Our results indicated that both the WT and variants Spike-hACE2 complexes shared a conservation 22 to 24 residues, which represented hotspot residues in the interface. The details

analyses of composition per residues of WT and mutant variants indicated that amino acid at positions 403, 408, 417, 424, 444, 447, 453, 454, 455, 456, 457, 458, 462, 466, 473, 475, 478, 485, 486, 487, 489, 490, 490, 493, 494, 496, 498, 500, 501, 502, 503, 505 and 509 were conserved in the Spike-hACE2 complex (Figure 2A). For exception, the amino acids 494, 490, and 453 have positive energy in the mutant, suggesting a potential reduction in binding affinity with hACE2 (Figure 2B). Nevertheless, the remaining mutants still

preserve amino acids in the interface. In general, the observed mutations in the spike protein did not appear to alter the structure

of the interface, but rather affected the interactions between amino acid side chains.



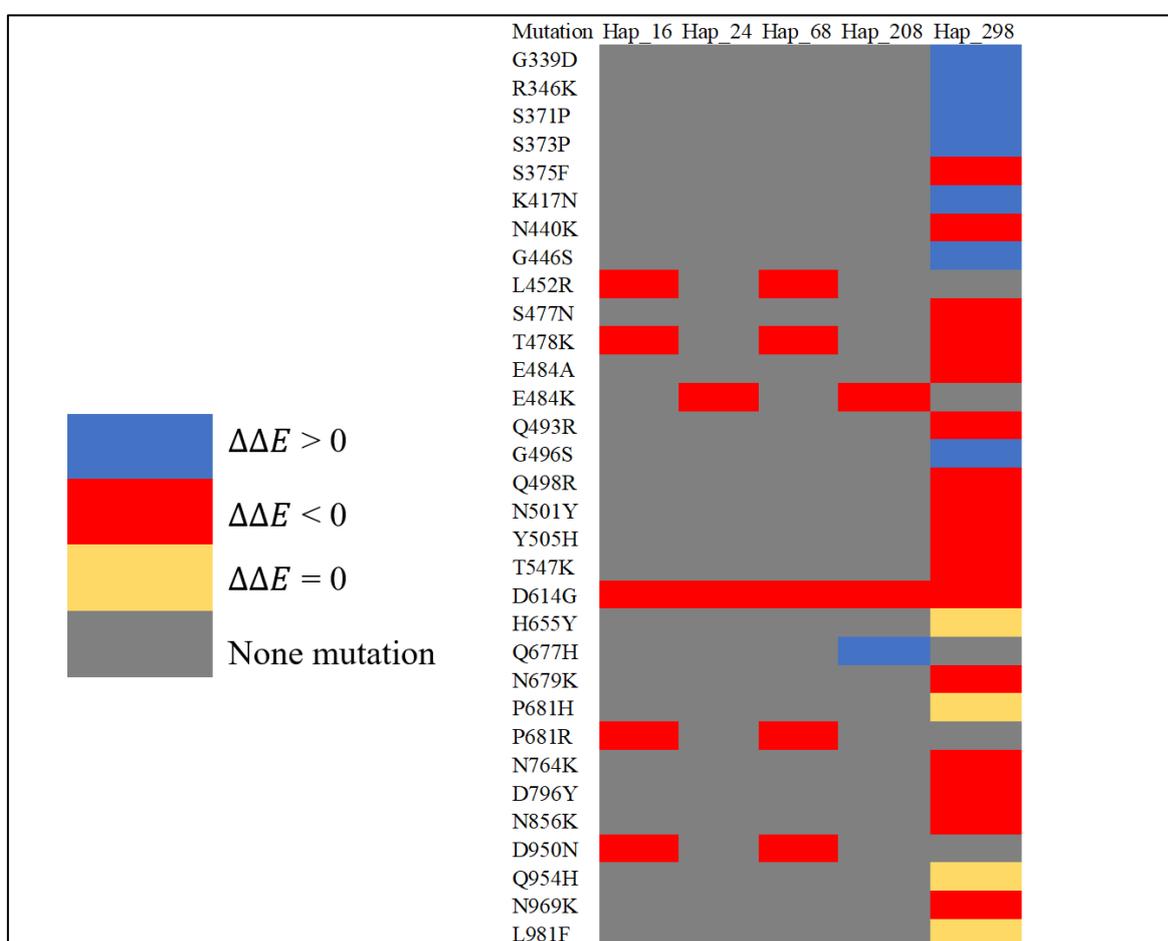
**Figure 2.** The network bonding and hotspot residues of interface. **A.** The spike and hACE2 proteins were colored purple and red, respectively. The amino acid bonding of Spike-hACE2 complex for each haplotype was displayed by PDBsum, in which the Salt bridges interaction, Disulphide bonds, Hydrogen bonds, and Non-bond contacts were displayed by red, yellow, blue line, and dashes-line, respectively. The color of residues presents for the characteristics of the amino acid. **B.** The hotspot residues were identified by Hawdock and PyDockEneRes through the free energy per residue.

To examine the effect of the mutation complex, we calculated the differences in energy values for the Spike-hACE2 mutant residues. In the RBD region, we identified the L452R, T478K mutations in Hap\_16 and Hap\_68, respectively, as positive mutations due to their  $\Delta\Delta E$  values being less than 0 (Figure 3). Although these mutations do not directly interact with hACE2, they contribute to an increase the  $\Delta E_{electrostatics}$  at the interface. The L452R mutation changes the hydrophobic side chain to a positive side chain, while the T478K mutation changes the polar

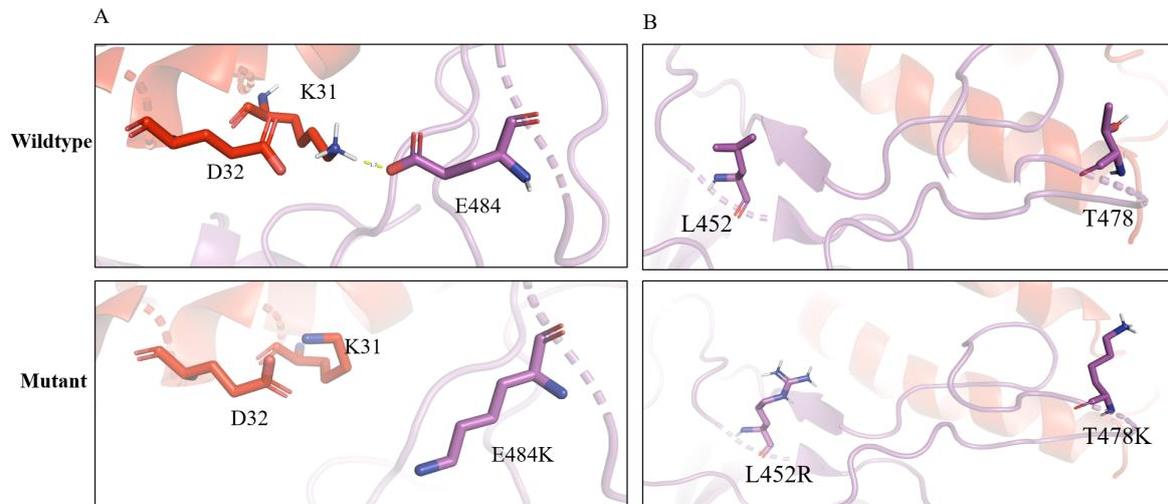
uncharged side chain to a positive side chain. Consequently, these mutations increase the electrostatic interactions with hACE2, considering that this receptor carries a negative charge (Xie, Guo *et al.* 2022). Similarly, the E484K mutation exhibits a negative  $\Delta\Delta E$  value, indicating a favorable effect (Figure 3). Although the E484K mutation abrogated the hydrogen bond with K31 of hACE2, it strongly increased the  $\Delta E_{electrostatics}$  of the complex (Figure 4). Furthermore, Hap\_298 carries eight novel positive mutations: S375F, N440K, S477N,

E484A, Q493R, Q498R, N501Y and Y505H (Figure 3). We found five of the seven positive mutations changed to positively charged side chain amino acids. Furthermore, at position E484, Hap\_298 mutates from Glutamic to Alanine which has an  $\Delta\Delta E$  less than 0, which means E484A support increasing binding affinity. Changing from Glutamic with a negatively charged side chain to Alanine with a hydrophobic side chain helps increase the

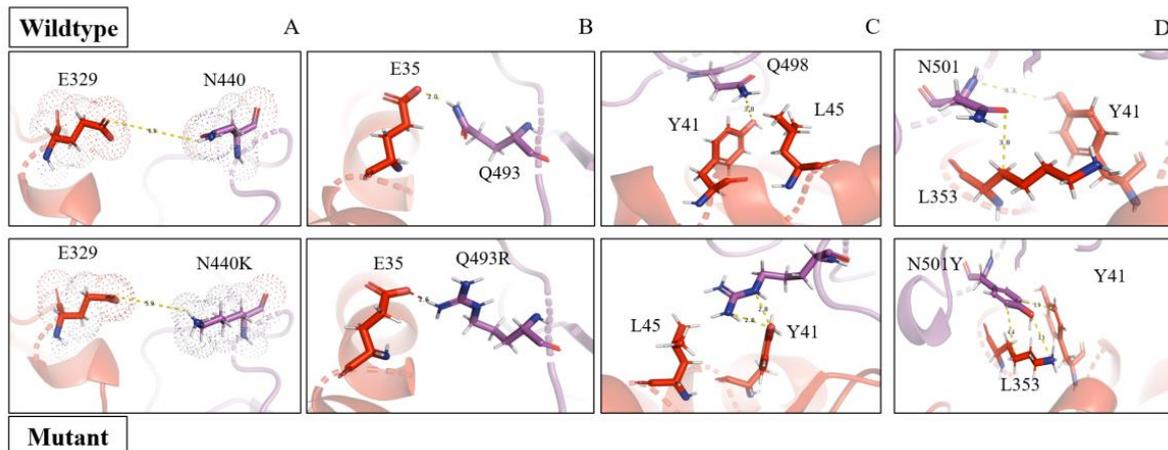
spike protein's electrostatic energy (Table 5). Therefore, Hap\_298 is consistent with the results of Table 4, the lowest  $\Delta E_{electrostatics}$  of -2066.33. Additionally, we evaluated mutations in the FCS region and indicated D614G, N679K, and P681R as enhancements to the free energy of the spike protein, which indirectly improves the affinity for hACE2. Details of the  $\Delta\Delta E$  changes of the mutants are shown in Figure 3.



**Figure 3.** Distribution of  $\Delta\Delta E$  of mutation in the interface. In this figure, we displayed the  $\Delta\Delta E$  of mutation in the interface. The  $\Delta\Delta E < 0$  was presented by blue, which exhibited positive transformation, which means increased binding affinity, and the opposite if  $\Delta\Delta E > 0$  in red. The grey and yellow color display for no mutation appear and  $\Delta\Delta E = 0$ , respectively.



**Figure 4.** The interaction of the E484K and T478K mutations. Spike and hACE2 proteins were displayed in purple and red, respectively. **A.** The yellow dashed line showed the interaction of E484 residues in the WT spike with hACE in the hydrogen bond. Below, the E484K mutation breaks the hydrogen bond with K31 of hACE2 because of the changes side chain to a positive charge. **B.** The position of L452R and T478K mutation in spike protein. Two mutations do not directly interact with hACE2, but changes in interface stored more positive charges.



**Figure 5.** The interaction of N440K, Q493R, Q498R and N501Y with hACE2. Spike and hACE2 proteins were displayed by purple and red color respectively. The hydrogen bond and salt bridges interaction were presented by yellow and red dashes line respectively.

Molecular interaction analysis revealed that the N440K mutation changes the side chain to a positive charge that attracts the E329 of hACE2 closer to forming an ionic interaction. The Q493R mutation enhances

interoperability by forming additional salt bridge interaction with E35 instead of creating a hydrogen bond as in the WT strain (Figure 2A and 5B). Likewise, the Q498R mutant enhances an additional hydrogen

bond as well as forms a hydrophobic interaction with L45 (Figure 5C). The common point of N440K, Q493R, and Q498R is that mutation into an amino acid with a positively charged side chain helps increase the complex's  $\Delta E_{electrostatics}$  (Table 5), assisting the spike protein to adhere better to hACE2. In contrast, the N501Y mutation did not directly affect the charge of the complex but contributed mainly to the formation of non-bonded contacts (Figure 2A). In the WT complex, N501 forms a weak hydrogen bond with Y41 and hydrophobic interaction with L353, whereas N501Y not only forms  $\pi$ - $\pi$  stacking

interaction between two aromatic benzene rings of Y501 and Y41 but also forms interaction hydrophobic and hydrogen bonded to L353 (Figure 5D). Therefore, N440K, Q493R, Q498R, and N501Y are critical mutations in the spike protein of Omicron. Based on the survey results, we found that from the VOC variant Eta to Delta and Omicron surveyed in this study, there is a tendency to accumulate mutations that enhance  $\Delta E_{electrostatics}$  through the accumulation of many amino acids that have a side chain that has a positive charge. The  $\Delta E_{electrostatics}$  of amino acids of each specific haplotype are shown in Table 5.

**Table 5.** The  $\Delta E_{electrostatics}$  of amino acids in variants Spike-hACE2 complex.

Amino acid position	WT	Hap_16	Hap_24	Hap_68	Hap_208	Hap_298
440	3.95	3.18	2.83	3.68	3.3	<b><u>-131.29</u></b>
452	-0.95	<b><u>-115.22</u></b>	-0.96	<b><u>-118.09</u></b>	-1.12	-0.81
478	-0.47	<b><u>-102.15</u></b>	-0.57	<b><u>-101.31</u></b>	-0.55	-95.15
484	100.39	100.96	<b><u>-111.09</u></b>	100.93	<b><u>-113.75</u></b>	<b><u>-1.06</u></b>
493	-1.28	-2.07	-1.45	-0.53	-6.87	<b><u>-141.2</u></b>
498	2.92	-7.78	1.89	2.39	4.34	<b><u>-154.46</u></b>

## DISCUSSION

In previous studies, we were able to identify A23403G (D614G) as an essential mutation in the evolution of SARS-CoV-2. In this study, based on network centrality, we could determine that although D614G was not directly related to the spike's ability to interact with hACE2, it played a significant role in the success of the phylogenetic process of SARS-CoV-2. The D614G substitution, located in the FCS region, was directly related to the ability of spike protein to cleave to activate the S2 subunit.

Experiment studies had shown that D614G accelerates the transmission of SARS-CoV-2 by enhancing the cleavage efficiency of the protein (Gobeil *et al.*, 2021; Zhang *et al.*, 2020; Korber *et al.*, 2020). In addition, our previous study demonstrated the predominance of the Eta variant in the early stages of the COVID-19 outbreak in Benin. In this study, we found out the Eta variant contains an E484K mutation, which located in RBD as a key mutation that helps the Eta variant infects more quickly. After that, the Delta variant appeared in July 2021 and gradually replaced the Eta variant in Benin.

Delta carries double and highly conserved mutations, L452R and T478K, in the RBD region and a new mutation in the FCS region, P681R. By *in silico* approach, we demonstrated the P681R mutation, which positively affected the Spike-hACE2 complex through increased the binding free energy of Spike-hACE2 complex. Moreover, in the previous experiment, the P681R mutation was one of the critical mutations to enhance Delta transmission fitness (Yang Liu *et al.*, 2022). Our initial results indicated molecular docking and network analysis could help find an effect of the new mutation in Spike-hACE2 complex.

Although Delta had spread vigorously in Benin and worldwide, this variant was replaced by Omicron, which was quickly classified into a VOC variant because of more than 30 amino acid substitutions on the spike protein. In this work, we showed that Omicron present in Benin has eight mutations that contribute to enhancing the free energy of the Spike-hACE2 complex. The number of mutations that increase the affinity of Omicron more than that of Delta can explain why the Omicron variant quickly prevailed in Benin and the world. Next, based on network bonding and hotspot residues of the interface, we found that the mutations do not change interface interaction of the Spike-hACE2 complex, and the amino acids in the interface were also highly conserved. Mutations in the interface help change the spike protein's electrostatic interaction with hACE2. In this study, we further noted that the S375F mutation also positively contributes to the exchange of spike protein with hACE2. Similarly, Hwang *et al.* (2022) showed that S375F enhanced the more substantial solvation energy of the Spike-hACE2 complex. Our study found out S375F mutation not only enhanced

$\Delta G_{Polar\ Solvation}$ , but also noted a change of  $\Delta E_{van\ der\ Waals}$  from -0.02 of WT going down to -0.09. The E484A mutation of Omicron variant in Benin has  $\Delta\Delta E < 0$ , which means an increased binding affinity of hACE2. Especially, Yao *et al.* (2023) showed that changes to the positively charged side chains of E484A and T478K enormously changed the redox of the disulfide bond, leading to a decrease in the stability of the RBD structure. Therefore, Spike-hACE2 complex structure of Omicron in Benin has the highest RMSD (Table 3).

In conclusion, we identified the key amino acids in the interface of Spike-hACE2 complex and explained the success of Eta, Delta, and Omicron variants that appeared in Benin. Furthermore, we have demonstrated that new mutations in spike protein positively affected the interaction with the hACE2 receptor. The high point of our study is the combination of network centrality and molecular docking methods to demonstrate the importance of novel variants appearing in the viral population. Therefore, a newly emerged variant with a high betweenness centrality means the variant has a high degree of influence in the viral population. This greatly assists in quickly tracking the movements in the spike protein of SARS-CoV-2 emerging variants in the world. From the initial results of this study, we anticipate that in the future, there will be a need for more profound concern because mutations tend to change the electrical side of spike protein, especially the substitution side chain, to positively charged or hydrophobic.

**Acknowledgement:** We would like to thank the Faculty of Natural Science Education, Saigon University, for providing us with the conditions to conduct this research.

REFERENCES

- Berrio A, Gartner V, Wray GA (2020) Positive selection within the genomes of SARS-CoV-2 and other Coronaviruses independent of impact on protein function. *PeerJ* 8: e10234.
- Candido KL, Eich CR, de Fariña LO, Kadowaki MK, da Conceição Silva JL, Maller A, Simão RCG (2022) Spike protein of SARS-CoV-2 variants: a brief review and practical implications. *Braz J Microbiol* 53(3): 1133-1157.
- Carreño JM, Alshammary H, Tcheou J, Singh G, Raskin AJ, Kawabata H, Sominsky LA, Clark JJ, Adelsberg DC, Bielak DA, Gonzalez-Reiche AS, Dambrauskas N, Vigdorovich V; PSP-PARIS Study Group; Srivastava K, Sather DN, Sordillo EM, Bajic G, van Bakel H, Simon V, Krammer F (2022) Activity of convalescent and vaccine serum against SARS-CoV-2 Omicron. *Nature* 602(7898): 682-688.
- Cele S, Jackson L, Khoury DS, Khan K, Moyo-Gwete T, Tegally H, San JE, Cromer D, Scheepers C, Amoako DG, Karim F, Bernstein M, Lustig G, Archary D, Smith M, Ganga Y, Jule Z, Reedoy K, Hwa SH, Giandhari J, Blackburn JM, Gosnell BI, Abdool Karim SS, Hanekom W; NGS-SA; COMMIT-KZN Team; von Gottberg A, Bhiman JN, Lessells RJ, Moosa MS, Davenport MP, de Oliveira T, Moore PL, Sigal A (2022) Omicron extensively but incompletely escapes Pfizer BNT162b2 neutralization. *Nature* 602(7898): 654-656.
- Chaw SM, Tai JH, Chen SL, Hsieh CH, Chang SY, Yeh SH, Yang WS, Chen PJ, Wang HY (2020) The origin and underlying driving forces of the SARS-CoV-2 outbreak. *J Biomed Sci* 27(1): 73.
- Chen F, Liu H, Sun H, Pan P, Li Y, Li D, Hou T. (2016) Assessing the performance of the MM/PBSA and MM/GBSA methods. 6. Capability to predict protein-protein binding free energies and re-rank binding poses generated by protein-protein docking. *Phys Chem Chem Phys* 18(32): 22129-22139.
- Cherian S, Potdar V, Jadhav S, Yadav P, Gupta N, Das M, Rakshit P, Singh S, Abraham P, Panda S, Team N (2021) SARS-CoV-2 Spike Mutations, L452R, T478K, E484Q and P681R, in the Second Wave of COVID-19 in Maharashtra, India. *Microorganisms* 9(7):1542.
- Colovos C, Yeates TO (1993) Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci* 2(9): 1511-1519.
- Di Giacomo S, Mercatelli D, Rakhimov A, Giorgi FM (2021) Preliminary report on severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) Spike mutation T478K. *J Med Virol* 93(9): 5638-5643.
- Dominguez C, Boelens R, Bonvin AM (2003) HADDOCK: A Protein-Protein Docking Approach Based on Biochemical or Biophysical Information. *J Am Chem Soc* 125(7): 1731-1737.
- Earnest R, Uddin R, Matluk N, Renzette N, Siddle KJ, Loreth C, Adams G, Tomkins-Tinch CH, Petrone ME, Rothman JE, Breban MI, Koch RT, Billig K, Fauver JR, Vogels CBF, Turbett S, Bilguvar K, De Kumar B, Landry ML, Peaper DR, Kelly K, Omerza G, Grieser H, Meak S, Martha J, Dewey HH, Kales S, Berenzy D, Carpenter-Azevedo K, King E, Huard RC, Smole SC, Brown CM, Fink T, Lang AS, Gallagher GR, Sabeti PC, Gabriel S, MacInnis BL; New England Variant Investigation Team; Tewhey R, Adams MD, Park DJ, Lemieux JE, Grubaugh ND (2022) Comparative transmissibility of SARS-CoV-2 variants Delta and Alpha in New England, USA. *Cell Rep Med* 3(4): 100583.
- Fan LQ, Hu XY, Chen YY, Peng XL, Fu YH, Zheng YP, Yu JM, He JS. Biological Significance of the Genomic Variation and Structural Dynamics of SARS-CoV-2 B.1.617 (2021) Biological Significance of the Genomic Variation and Structural Dynamics of SARS-CoV-2 B.1.617. *Front Microbiol* 12: 750725-750725.
- Gobeil SM, Janowska K, McDowell S, Mansouri K, Parks R, Manne K, Stalls V, Kopp MF, Henderson R, Edwards RJ, Haynes BF, Acharya P (2021) D614G Mutation Alters SARS-CoV-2

Spike Conformation and Enhances Protease Cleavage at the S1/S2 Junction. *Cell Reports* 34(2): 108630.

Han P, Li L, Liu S, Wang Q, Zhang D, Xu Z, Han P, Li X, Peng Q, Su C, Huang B, Li D, Zhang R, Tian M, Fu L, Gao Y, Zhao X, Liu K, Qi J, Gao GF, Wang P (2022) Receptor binding and complex structures of human ACE2 to spike RBD from omicron and delta SARS-CoV-2. *Cell* 185(4): 630-640.e610.

Hwang S, Baek SH, Park D (2022) Interaction Analysis of the Spike Protein of Delta and Omicron Variants of SARS-CoV-2 with hACE2 and Eight Monoclonal Antibodies Using the Fragment Molecular Orbital Method. *J Chem Inf Model* 62(7): 1771-1782.

Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, Hengartner N, Giorgi EE, Bhattacharya T, Foley B, Hastie KM, Parker MD, Partridge DG, Evans CM, Freeman TM, de Silva TI; Sheffield COVID-19 Genomics Group; McDanal C, Perez LG, Tang H, Moon-Walker A, Whelan SP, LaBranche CC, Saphire EO, Montefiori DC (2020) Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell* 182(4): 812-827.e819.

Kozakov D, Hall DR, Xia B, Porter KA, Padjhony D, Yueh C, Beglov D, Vajda S (2017) The ClusPro web server for protein-protein docking. *Nat Protoc* 12(2): 255-278.

Kumar V, Singh J, Hasnain SE, Sundar D (2021) Possible Link between Higher Transmissibility of Alpha, Kappa and Delta Variants of SARS-CoV-2 and Increased Structural Stability of Its Spike Protein and hACE2 Affinity. *Int J Mol Sci* 22(17): 9131.

Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, Zhang Q, Shi X, Wang Q, Zhang L, Wang X (2020) Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* 581(7807): 215-220.

Leigh JW, Bryant D (2015) Popart: full-feature

software for haplotype network construction. *Methods Ecol Evol* 6(9): 1110-1116.

Liu Y, Liu J, Johnson BA, Xia H, Ku Z, Schindewolf C, Widen SG, An Z, Weaver SC, Menachery VD, Xie X, Shi PY (2022) Delta spike P681R mutation enhances SARS-CoV-2 fitness over Alpha variant. *Cell Reports* 39(7): 110829.

Liu Y, Liu J, Plante KS, Plante JA, Xie X, Zhang X, Ku Z, An Z, Scharton D, Schindewolf C, Widen SG, Menachery VD, Shi PY, Weaver SC (2022) The N501Y spike substitution enhances SARS-CoV-2 infection and transmission. *Nature* 602(7896): 294-299.

Martínez-Flores D, Zepeda-Cervantes J, Cruz-Reséndiz A, Aguirre-Sampieri S, Sampieri A, Vaca L (2021) SARS-CoV-2 Vaccines Based on the Spike Glycoprotein and Implications of New Viral Variants. *Front Immunol* 12: 701501.

Romero-Durana M, Jiménez-García B, Fernández-Recio J (2020) pyDockEneRes: per-residue decomposition of protein-protein docking energy. *Bioinformatics* 36(7): 2284-2285.

Tabynov K, Turebekov N, Babayeva M, Fomin G, Yerubayev T, Yespolov T, Li L, Renukaradhya GJ, Petrovsky N, Tabynov K (2022) An adjuvanted subunit SARS-CoV-2 spike protein vaccine provides protection against Covid-19 infection and transmission. *NPJ Vaccines* 7(1): 24.

Tan ST, Kwan AT, Rodríguez-Barraquer I, Singer BJ, Park HJ, Lewnard JA, Sears D, Lo NC (2023) Infectiousness of SARS-CoV-2 breakthrough infections and reinfections during the Omicron wave. *Nat Med* 29(2):358-365.

Vangone A, Bonvin AM (2015) Contacts-based prediction of binding affinity in protein-protein complexes. *eLife* 4: e07454.

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 NY, 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(7902): 679-686.
- Weng G, Wang E, Wang Z, Liu H, Zhu F, Li D, Hou T (2019) HawkDock: a web server to predict and analyze the protein–protein complex based on computational docking and MM/GBSA. *Nucleic Acids Res* 47(W1): W322-W330.
- Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* 35(Web Server issue):W407-10.
- Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY, Yuan ML, Zhang YL, Dai FH, Liu Y, Wang QM, Zheng JJ, Xu L, Holmes EC, Zhang YZ (2020) A new coronavirus associated with human respiratory disease in China. *Nature* 579(7798): 265-269.
- Xie Y, Guo W, Lopez-Hernandez A, Teng S, Li L (2022) The pH Effects on SARS-CoV and SARS-CoV-2 Spike Proteins in the Process of Binding to hACE2. *Pathogens* 11(2):238.
- Xu C, Wang Y, Liu C, Zhang C, Han W, Hong X, Wang Y, Hong Q, Wang S, Zhao Q, Wang Y, Yang Y, Chen K, Zheng W, Kong L, Wang F, Zuo Q, Huang Z, Cong Y (2021) Conformational dynamics of SARS-CoV-2 trimeric spike glycoprotein in complex with receptor ACE2 revealed by cryo-EM. *Sci Adv* 7(1): eabe5575.
- 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(5): 3370-3382.
- Xue LC, Rodrigues JP, Kastritis PL, Bonvin AM, Vangone A (2016) PRODIGY: a web server for predicting the binding affinity of protein–protein complexes. *Bioinformatics* 32(23): 3676-3678.
- Yang W, Shaman J (2022) COVID-19 pandemic dynamics in India, the SARS-CoV-2 Delta variant and implications for vaccination. *J R Soc Interface* 19(191): 20210900.
- Yao Z, Geng B, Marcon E, Pu S, Tang H, Merluzza J, Bello A, Snider J, Lu P, Wood H, Stagljar I (2023) Omicron Spike Protein Is Vulnerable to Reduction. *J Mol Biol* 435(13):168128.
- Zhang L, Jackson CB, Mou H, Ojha A, Peng H, Quinlan BD, Rangarajan ES, Pan A, Vanderheiden A, Suthar MS, Li W, IZard T, Rader C, Farzan M, Choe H (2020) SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity. *Nat Commun* 11(1): 6013.
- Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD, Chen J, Luo Y, Guo H, Jiang RD, Liu MQ, Chen Y, Shen XR, Wang X, Zheng XS, Zhao K, Chen QJ, Deng F, Liu LL, Yan B, Zhan FX, Wang YY, Xiao GF, Shi ZL (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579(7798): 270-273.