Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Computers in Biology and Medicine 141 (2022) 105170

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Comparative mutational analysis of SARS-CoV-2 isolates from Pakistan and structural-functional implications using computational modelling and simulation approaches

Abdullah Shah [a,1], Saira Rehmat [b], Iqra Aslam [c], Muhmmad Suleman [d,1], Farah Batool [e], Abdul Aziz [f], Farooq Rashid [g,h], Midrarullah [a], Muhmmad Asif Nawaz [a], Syed Shujait Ali [d], Muhammad Junaid [i], Abbas Khan [i,*], Dong-Qing Wei [i,j,k,**]

[a] Department of Biotechnology, Shaheed Benazir Bhutto University Sheringal, Dir (U), Pakistan
[b] Sharif Medical and Dental College, Lahore, Pakistan
[c] Nawaz Shareef Medical College, Gujrat, Pakistan
[d] Centre for Biotechnology and Microbiology, University of Swat, Khyber Pakhtunkhwa, Pakistan
[e] Institute of Pharmacy and Allied Health Sciences, Lahore College for Women University, Jail Road, Lahore, Pakistan
[f] Molecular Biology Research Center, School of Life Sciences, Central South University, Changsha, China
[g] Dermatology Hospital, Southern Medical University, Guangzhou, China
[h] Guangdong Provincial Key Laboratory of Tropical Disease Research, School of Public Health, Southern Medical University, Guangzhou, China
[i] Department of Bioinformatics and Biological Statistics, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, 200240, PR China
[j] State Key Laboratory of Microbial Metabolism, Shanghai-Islamabad-Belgrade Joint Innovation Center on Antibacterial Resistances, Joint Laboratory of International Laboratory of Metabolic and Developmental Sciences, Ministry of Education and School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, 200030, PR China
[k] Peng Cheng Laboratory, Vanke Cloud City Phase I Building 8, Xili Street, Nashan District, Shenzhen, Guangdong, 518055, PR China

ARTICLE INFO

Keywords:
SARS-CoV-2
COVID-19
ACE2
Mutation
Phosphorylation serine/threonine kinases

ABSTRACT

SARS-CoV-2, an RNA virus, has been prone to high mutations since its first emergence in Wuhan, China, and throughout its spread. Its genome has been sequenced continuously by many countries, including Pakistan, but the results vary. Understanding its genomic patterns and connecting them with phenotypic features will help in devising therapeutic strategies. Thus, in this study, we explored the mutation landscape of 250 Pakistani isolates of SARS-CoV-2 genomes to check the genome diversity and examine the impact of these mutations on protein stability and viral pathogenesis in comparison with a reference sequence (Wuhan NC 045512.2). Our results revealed that structural proteins mainly exhibit more mutations than others in the Pakistani isolates; in particular, the nucleocapsid protein is highly mutated. In comparison, the spike protein is the most mutated protein globally. Furthermore, nsp12 was found to be the most mutated NSP in the Pakistani isolates and worldwide. Regarding accessory proteins, ORF3A is the most mutated in the Pakistani isolates, whereas ORF8 is highly mutated in world isolates. These mutations decrease the structural stability of their proteins and alter different biological pathways. Molecular docking, the dissociation constant (K_D), and MM/GBSA analysis showed that mutations in the S protein alter its binding with ACE2. The spike protein mutations D614G-S943T-V622F (−75.17 kcal/mol), D614G-Q677H (−75.78 kcal/mol), and N74K-D614G (−73.84 kcal/mol) exhibit stronger binding energy than the wild type (−66.34 kcal/mol), thus increasing infectivity. Furthermore, the simulation results strongly corroborated the predicted protein servers. Our analysis findings also showed that E, M, ORF6, ORF7A, ORF7B, and ORF10 are the most stable coding genes; they may be suitable targets for vaccine and drug development.

* Corresponding author.
** Corresponding author. Department of Bioinformatics and Biological Statistics, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, 200240, PR China.
E-mail addresses: Tamimrai18@gmail.com (S. Rehmat), Habibrai36@gmail.com (I. Aslam), fblcwu@gmail.com (F. Batool), abbaskhan@sjtu.edu.cn (A. Khan), dqwei@sjtu.edu.cn (D.-Q. Wei).
[1] Contributed Equally.

https://doi.org/10.1016/j.compbiomed.2021.105170
Received 18 September 2021; Received in revised form 21 December 2021; Accepted 21 December 2021
Available online 25 December 2021

1. Introduction

In December 2019, the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus appeared in Wuhan, China, instigated a global coronavirus pandemic (COVID-19) [1,2]. Till October 2021, 241,199,505 cases have been reported with about 4,910,487 deaths recorded (WHO, 2020, (https://covid19.who.int). The transmission of SARS-CoV-2 is very high, with common symptoms attributed to this disease includes, high fever, breathing complexities, dry cough, and acute respiratory distress [3–8].

SARS-CoV-2 is a single-stranded, positive-sense RNA virus with a genome size of ~29.8 kb which encodes 16 non-structural proteins (nsp1–16) to form the replicase machinery of the virus [5,9]. It also codes for four structural proteins, namely Spike (S), envelope (E), membrane (M), and nucleocapsid (N). Besides structural and non-structural protein, SARS_CoV-2 codes for seven accessory proteins (ORF3a, ORF3b, ORF6, ORF7a, ORF7b, ORF8, and ORF10), which modulate host responses against viral infections to assist effective infection and pathogenesis [10–14].

Genome sequencing insights have shown nucleotide substitution rate as $\sim 1 \times 10^{-3}$ per year for SARS-CoV-2 [15,16]. Variations in different proteins of SARS-CoV-2, particularly the spike glycoproteinlead to a drift in the antigenicity of vaccines and other therapeutics [17]. The mutation rate of RNA viruses, including, SARS-CoV-2 is exceptionally high, which in turn, increase the virulence and transmission of these viruses [16,18–20]. These mutations occur in almost all the proteins of SARS-CoV-2, thereby rendering this virus to escape the host immune system more efficiently and consequently increases the pathogenicity and transmissibility [21–28]. These amino acid substitutions also alter the protein structures and consequently viral replication. The depletion of 9 amino acids in Orf6 caused the shifting of transmembrane localization that results in IFN resistance during antiviral therapy [29]. Mutations in the proteins affect the structure, function, and binding with the other proteins necessary for disease progression. Other mutations result in a new RNA polymerase variant that is reliant on RNA [30]. Some of the reported mutations also expedite the transmission of the virus i.e., D614G mutation in the spike protein has been reported to be associated with unusual increased in the transmission fitness of SARS-CoV-2 [23,31–34]. The genome of the virus is unceasingly mutating since its outbreak in late 2019 and many variants are reported until now. The B.1.617 lineage was identified in India in October 2020 [24,35]. It then became a dominant strain in India and spread to other countries of the world as well. Furthermore, the B1.617.1, B.1.617.2, and B.1.617.3 subtypes of this lineage also harbours different mutations in the *N*-terminal domain (NTD) and the receptor-binding domain (RBD) of the spike protein [27,36].

Since the virus mutates often, it is hard to develop effective therapeutics which can hinder all the variants. To come up with a successful treatment approach or vaccination, it is necessary to have a greater understanding of both the virus and host genome [37]. An idea of the efficacy of the vaccines being researched may be extrapolated from sequencing the entire genome sequences of isolates collected from diverse locations across the world [38]. The race for a final magic drug still continues and according to the updated data provided by the COVID-19 vaccine and therapeutics tracker, 408 clinical trials of 138 vaccine candidates are under consideration and are in different clinical phases (https://covid19.trackvaccines.org/). Globally 6.65 billion doses of COVID-19 vaccines are administered among which 2.83 billion are fully vaccinated which counts for only 36% of the whole population. Understanding the mutational landscape of SARS-CoV-2 would provide a better view of the therapeutic interventions. For instance, *Abbas* et al., used computational approaches to study the impact of different mutations on the binding and infectivity [18,20,39,40]. Moreover, using these approaches many small molecules drugs are tested against different proteins of SARS-CoV-2 [41–46]. Therefore, herein we used computational approaches to study the reported 250 genomes of SARS-CoV-2 isolated from Pakistan. In Pakistan, the first case of COVID-19 was reported on February 26, 2020. As of October 2021, about 1.27 million cases were reported, with total mortality of 28,280 were recorded in Pakistan. Hence, we used GSAID to extract the mutational spectrum in different proteins and their impact on the structure and function was estimated. Furthermore, protein-protein docking and molecular dynamics simulation approaches were used to grasp the impact of the most prevalent mutations on the binding and stability of the spike protein particularly. Consequently, the current study provides distinguishable features for the discovery of novel therapeutics and Pakistan-specific *anti*-COVID-19 strategies.

2. Materials and methods

2.1. Collection of genome data for SARS-CoV-2

The genome sequence of 250 Pakistani isolates OF SARS-CoV-2 was downloaded from a global initiative on sharing all influenza data (GISAID) (https://www.gisaid.org/) [47] and NCBI [48]. Only coding sequences were analyzed for mutational analysis at the protein level in the present study.

2.2. Multiple sequence alignment

DNASTAR, Lasergene, and Clustal W tools align the 250 Pakistani SARS-CoV-2 isolates [49,50]. The previously described isolate of SARS-CoV-2 (Wuhan NC 045512.2) was utilized as a reference sequence for mutational analysis at the amino acid level.

2.3. Sequence-based mutational effects prediction

The I-mutant online tool was utilized to assess the structural stability change [51]. Mutpred2 was also utilized to examine the molecular consequences and functional impacts of mutations found [52].

2.4. Wild type structure retrieval and variants modelling

The wild-type structure of SARS-CoV-2 spike protein (6VSB) was obtained from the Protein Data Bank, and Chimera v 1.15 software was used for SARS-CoV-2 spike protein variant modeling [53,54]. The wild type structure 6VSB was used as a template that shares 97% identity with the sequence of the variant. The.FASTA format was used as input and BLAST was carried out to select the template based on sequence identity and coverage.

2.5. Protein-protein docking

HDOCK (http://hdock.phys.hust.edu.cn/) online server algorithm was used for protein-protein interaction of both wild type-ACE2 and variants-ACE2, to check the respective binding efficiencies [55]. HDOCK uses a hybrid algorithm that combines both template-based and free approaches to estimate the docking conformation and affinity.

2.6. Dissociation constant (K_D) determination

Estimation of the strength of biomolecular association is an important parameter in biological processes. Herein, to estimate the strength of the wild type and mutants RBD with ACE2 PRODIGY (PROtein binDIng enerGY prediction)(https://wenmr.science.uu.nl/prodigy/) webserver was used [56].

2.7. Molecular dynamic simulation

MD simulation of the wild type and mutant complexes was completed by using the AMBER20 simulation package [57,58]. The default parameters used by Abbas et al., were employed to complete the

200ns simulation for each complex [59,60]. Post-simulation analyses such as RMSD and RMSF were calculated using CPPTRAJ and PTRAJ modules [61].

2.8. Binding free energy calculations

The binding energy between WT and mutant spike protein with the human receptor ACE2 was calculated using the MM/GBSA approach [62]. The spike mutants include N74K, D614G, V622F, Q677H, S943T, D115G with the human receptor ACE2 using the HAWKDOCK server [63].

The energy is calculated with the following equation:

$$\Delta G(\text{bind}) = \Delta G(\text{complex}) - [\Delta G(\text{receptor}) + \Delta G(\text{ligand})]\frac{1}{2}$$

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

$$G = G_{\text{bond}} + \quad G_{\text{ele}} + G_{\text{vdW}} + G_{\text{pol}} + G_{\text{npol}}$$

3. Results and discussion

3.1. Retrieved genome of SARS-CoV-2

A total of 250 SARS-CoV-2 isolates from Pakistan were obtained via the Global Initiative on GISAIDand the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/). Only coding sequences were examined for mutational analysis at the protein level (Table 1).

3.2. Worldwide mutational distribution of SARS-CoV-2 proteins

SARS-CoV-2 protein mutations were examined globally as part of a global initiative program for sharing all influenza data (GISAID). We observed mutations in most of the proteins of the SARS-CoV-2 virus. However, some proteins like envelop, ORF6, ORF7A, and ORF7B showed resistance to mutations. Additionally, specific proteins were more prone to mutations than others. Among the Structural proteins, spike protein was mutated the most, followed by nucleocapsid. Within spike, D614G was found to be mutated in 171 countries. Mutation R203K in nucleocapsid protein was observed in 151 countries (Fig. 1A). Among the NSPs, the mutation P4715L in NSP12 was observed in 172 countries globally, followed by L3606F in NSP3, which occurred in 134

Table 1
Retrieved genomes of SARS-CoV-2 from Pakistan and their mutational landscape.

S. No	Isolates	ORF1AB	ORF1A	Spike	N	M	ORF3A	ORF8	ORF10
1	MW447643.1	Q63L, L3293F, P4075, SA4489V, P4716L	Q63L P4075S	D614G			Q57H		
2	MW447622.1	G82S, L2564F, P4716L, A4921V	G82S L2564F	D614G					
3	MW447639.1	G82S, P4716L	G82S	D614G			Q57H		
4	MT240479.1	R207C, V378I, P2965L, L3606F	R207C V378I P2965L L3606F						
5	MW447634.1	K261 N, P4716L, T5451I	K261 N	D614G			Q57H		
6	MW447614.1	F275L, V2133A, P4716L		D614G	S194L		Q57H		
7	MW242667.1	K292E, V627F, T2018K, L3606F	K292E V627F T2016K L3606F		P13L				V30L
8	MT500122.1	D448 N, N449, deletion	D448 N N449delete						
9	MW447612.1	G934C, P4716L		D614G	R209I		Q57H		
10	MW031799.1	S944L, T1246I, K1305 N, G3278S, P4716L	S944LT1246I K1305 N G3278S	D614G	R203K				
11	MW031800.1	S944L, T1246I, K1305 N, P4716L	S944L T1246I K1305 N G3278S	D614G	R203K				
12	MW031801.1	S944L, T1246I, K1305 N, P4716L	S944L T1246I G3278S	D614G	R203K				
13	MW031802.1	S944L, T1246I, K1305 N, P4716L	S944L T1246I K1305 N G3278S	N74K D614G	R203K				
14	MW447609.1	L1175I, P4716L		D614G	R203K		Q57H		
15	MW447644.1	T1754I, T2846I, P4716L	T1754I T2846I	D614G			K21 N		
16	MW447645.1	T2408I, P4716L, L6082F	V2133A T2408I	D614G	S194LT379I		Q57H		
17	MW447642.1	A2618V, P4716L	A2618V	D614G	R209I		Q57H T223I	W45L	
18	MW447611.1	Q2702H, K3353R P4716L, V4979L	Q2702H K3353R	D614G	S327L		Q57H		
19	MT879619.1	Q2702H, P4716L, A5561T	Q2702H L3606F	D614G	S327L		Q57H		V30L
20	MT262993.1	(T3153Y, WFF3157, deleted) (L6418, YLDAY6424 N deleted)	(T3153YWFF3157 deleted)						
21	MW447636.1	P3447S, P4716L	P3447S	D614G	R209I		Q57H		
22	MW031803.1	P4716L		D614G					
23	MW400961.1	P4716L		D614G V622F S943T			Q57H		
24	MW447638.1	P4716L, V6600F		D614G, Q677H	S194L		Q57H		
25	MW447640.1	P4716L, V6600F	A2956V	D614G	S194L	H25Y			
26	MW447641.1	P4716L		D614G	R209I		Q57H		
27	MW447644.1	P4716L			R209I				
28	MW403500.1			D614G	S194L		Q57H		
29	MW411960.1			D614G	R209I				
30	MW447610.1			D614G			Q57H		
31	MW447619.1			D614G	S194L		Q57H		
32	MW447620.1			D614G	R209I		Q57H		
34	MW447618.1				S202 N			L84S	
35	MW447621.1			D614G D1153G	R209I		Q57H		

countries. The L1174I mutation in NSP3 was observed only in 2 countries (Fig. 1B). Among the accessory proteins, ORF8 was found to be the most mutated protein. Several mutations have been reported in the ORF8 protein, but L84S is a highly recurred mutation observed in 100 countries (Fig. 1A).

3.3. Mutational spectrum of the structural proteins of Pakistani SARS-CoV-2 isolates

Mutational analysis of the structural proteins of 250 Pakistani isolates of SARS-COV2 harbors six different amino acid substitutions in spike protein, one in membrane protein, nine amino acids substitutions, and a large deletion in nucleocapsid phosphoprotein. At the same time, no mutation was observed in envelope protein. The spike protein of 29 out 50 samples of SARS-COV2 Pakistani isolates harbors a signature D614G mutation indicative of the widespread mutation in Pakistani isolates. The D614G mutation in spike protein has been reported to strengthen the folding stability of the spike protein. This mutation was first reported in China and is the second-largest mutation in the United States, and its ratio is increasing as time goes on. The D614G mutation of the spike protein where Aspartic acid (D) with a polar negative side charged amino acid is substituted with Glycine (G) with a nonpolar side chain amino acid. This D614G mutation of spike protein has reported a critical mutation that makes the SARS-CoV-2 more contagious and enhances its infectivity [64]. The other less common mutation that co-occurred with D614G was N74K, V622F, Q677H, S943T, and D1153G (Table 1). Surprisingly, in the spike protein, the D614G substitution was associated with two consecutive series of two substitutions R203K, and G204R of the nucleocapsid phosphoprotein in five isolates (MW031799.1, MW031800.1, MW031801.1, MW0831802.1, and MW031803.1). These two mutations in nucleocapsid phosphoprotein were previously reported in the United States [64]. The amino acid arginine (R) and Lysine (L) are both positively charged, and its substitution may not affect the function of the nucleocapsid phosphoprotein. In contrast, the Glycine (a nonpolar) substitution by arginine (R) may affect the hydrophilicity of the nucleocapsid phosphoprotein of SARS-CoV2. Furthermore, a more significant deletion was observed in the nucleocapsid phosphoprotein of isolate MT240479.1 (Table 1). While the envelop protein was 100% identical in all isolates compared with the reference sequence (Wuhan NC_045512.2).

3.4. Mutational spectrum of the accessories proteins of Pakistani SARS-CoV-2 isolates

Among accessories protein of SARS-CoV-2 of Pakistani isolates, three proteins (ORF3A, ORF8, and ORF10) showed different amino acid substitution while no deletion was observed in these proteins of the selected isolates (Table 1). ORF3a protein (274 amino acids) of SARS-CoV-2 is expressed abundantly in infected cells and plays an essential role in pathogenesis. In these SARS-COV2 isolates, a signature Q57H substitution in ORF3 protein was observed in 23 out of 50 samples which is the second frequent mutation in these Pakistani isolates. This substitution of Q57H was first reported in Singapore and occurred in 70% of US Covid-19 cases [64]. The prevalence of Q57H substitution in ORF3 protein was reported relatively high (62.7%) in South Africa [65]. The Q57H substitution in ORF3 co-occurs with two other amino acids mutations K21 N and T223I, in isolates MW447644.1, MW447642.1 respectively. The SARS-COV-2, ORF8 has been reported to reduce the translocation of IRF3 to the nucleus and consequently reduce expression of the interferon production and to down-regulate MHC-1 in cells [66, 67]. Interestingly, in Pakistani SARS-COV-2 isolates, a new mutation in ORF8, W45L was observed in the present study, while 35 out of 50 isolates have the ORF8-L genotype. The substitution of Tryptophan (W) by lysine (L) may affect the hydrophilicity of the ORF8 of SARS-CoV-2 protein. In ORF10, single substitution V30L has been observed in two isolates MT 879619.1 and MW242667.1. While the other accessories protein ORF6, ORF7A, and ORF7B were found 100% identical to the reference sequence Wuhan NC_045512.2).

3.5. Mutational spectrum of non-structural proteins of Pakistani SARS-CoV-2 isolates

Mutational analysis of non-structural proteins of 50 Pakistani SARS-CoV-2 isolates harbored 75 substitution mutations and deletions compared to the reference sequence (NC_045512.2). Among substitution, two in nsp1 (Q63L, G82S), seven in nsp2 (R197C, K261 N, F275L, K292E, V378I, D448 N, V627F) and one deletion at N449, Nine amino acids substitutions in nsp3 (L1174I, T1246I, K1308 N, T1754I, T2016K, T2408F, L2564F, A2619V, Q2702H), two substitutions (T2848I, P2926L and five amino acid deletion 3153–3157) in nsp4, four in nsp5 (G3278S, L3393F, K3353R, P3447S), Two in nsp6 (L3606F, A3656V), one in nsp8 (P4075S), four in nsp12 (A4489V, P4715L, A4921V, V4979L), two in nsp13 (T5451I, A5561T), one substitution (L6082F), seven amino acid deletion (6418–6424) in nsp14, one in nsp15 (V6600F) while no mutation was observed in nsp7, nsp9, nsp10, and nsp16. Among the substation, two substations of nsp3 T1246I and K1308 N co-occurred with substitution G3278S and P4715L of nsp5 and nsp12, respectively in isolates MW031799.1, MW031800.1, MW031801.1, MW031802.1. In isolate MT500122 the substation D448 N has preceded a deletion of N449 amino acid in nsp2. Furthermore, two consecutive deletions of five amino acids FYWFF (3153–3157) and seven amino acids LYLDAYN (6418–6424) in nsp4 have been observed in the isolate MT262993 as compared to the reference sequence (Table 1).

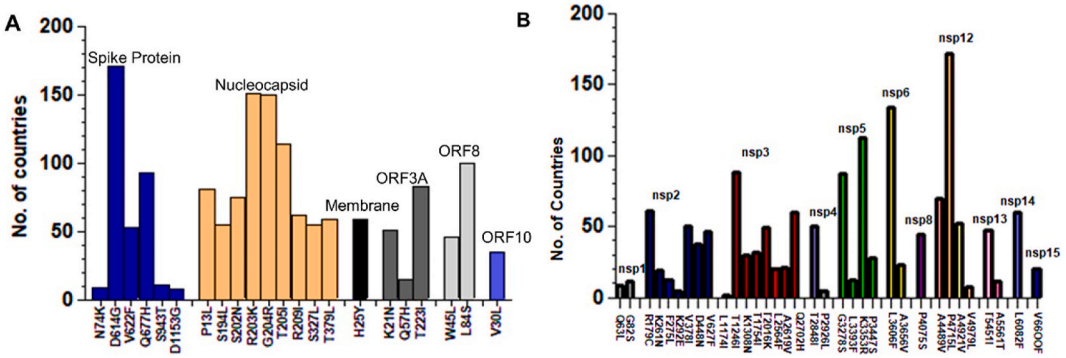

Fig. 1. Worldwide mutational distribution of SARS-CoV-2 proteins. (A) Among the Structural proteins, D614G mutation in spike protein was found the most reoccurring mutation in 171Countries while L84S in ORF8 was observed in 100 countries. **(B)** Among the Non-Structural proteins, the mutation P4715L in NSP12 was observed in 172 countries of the world, followed by L3606F in NSP3 which occurred in 134 countries while the L1174I mutation in NSP3 was observed only in 2 countries.

3.6. Identification of hyper-variable genomic hotspot in SARS-CoV-2 genomes of Pakistani isolates

Interestingly, among all recurrent mutations in structural protein, the spike protein has one D614G (82%), nucleocapsid had three S194L (20%), R203K (14%), G204R (14%) R209I (25%) mutation hotspot (Fig. 2). The accessory protein the ORF3A has one mutation hotspot Q57H (63%). Among the non-structural proteins, nsp3 have four mutation hotspots S944L (11%), T1246I (11%), K1305 N (8%), NSP12 have one P4715L (57%), NSP15 have one V6600F (5%) mutation hotspots in Pakistani isolates as compared with the reference (Wuhan NC_045512.2) (Fig. 2).

3.7. Mutational effects on structure stability

It has been determined that 27 mutations decrease structural stability, whereas six mutations enhance structural stability due to I-mutant server analysis of 96 missense mutations. These proteins' structural stability was expected to be lowered by most mutations in surface glycoprotein and nucleocapsid phosphoprotein, whereas specific changes in ORF1AB and ORF1A were projected to improve it (Table 2).

3.8. Molecular consequences of reported mutations in different pathways

Additionally, three mutations occurring in ORFIA were predicted to alter the molecular consequences, including Loss of Disulfide linkage, Loss of Catalytic site, Loss of Allosteric site, Gain of Proteolytic cleavage, and Gain of Sulfation. Similarly, one mutation each in Membrane Glycoprotein and ORF8 were predicted to alter transmembrane protein, altered DNA binding, as well as altered disordered interface respectively Table 3.

3.9. Modelling of the wild type and spike mutant structures

Continued mutations in SARS-CoV-2 Genome cause potential differences in its transmission, virulence, and pathogenesis compared to their wild counterparts. Recently reported mutations in the spike proteins of SARS-CoV-2 from South Africa (Lys417Asn, Glu484Lys, Asn501Tyr) (501Y.V2Variant) and Brazil (Lys417Asn, Glu484Lys, Asn501Tyr) had led these strains to evade the vaccines successfully [18]. Therefore, it was necessary to explore the mutations in the spike protein of Pakistani isolates. SARS-CoV-2 virus Spike proteins have an RBD domain, which aids in binding to ACE2 receptors on the host cell. This binding sets off a chain reaction of events that lead to fusing the host cell and viral membranes for cell penetration. SARS-spike CoV-2's RBD was submitted to comparative binding and biophysical study upon interaction with ACE2 because of its relevance in virulence and pathogenesis. RCSB was used to get the structure of the Spike and ACE2 proteins (PDB ID:6M0J). Docking the spike protein with the human ACE2 receptor, HDOCK online server was used. To verify whether the binding affinity of the RBD domain for the human ACE2 can be affected by the six mutations present in the spike protein of Pakistani isolates, we made a double mutant, which consists of the N74K-D614G, N74K-Q677H, D614G-V622F, D614G-S943F, D614G-V622F–S943F, and D614G-Q677H. Structural organization, domain separation, and binding interface of spike protein are shown in Fig. 3a. The mutants were generated by Chimera and are given in Fig. 3b–e.

Table 2
Prediction of the mutational effects on the structural stability. A negative value indicates destabilizing effect.

Index	Protein	Amino Acid Changes	SVM2 Prediction Effect	DDG Value (kcal/mol)
1	Surface Glycoprotein	N74K	Decrease	−0.49
2	Surface Glycoprotein	D614G	Decrease	−0.93
3	Surface Glycoprotein	V622F	Decrease	−1.01
4	Surface Glycoprotein	Q677H	Decrease	−0.67
5	Surface Glycoprotein	S943T	Decrease	−0.24
6	Surface Glycoprotein	D1153	Decrease	−1.08
7	Nucleocapsid phosphoprotein	P13L	Decrease	−0.48
8	Nucleocapsid phosphoprotein	R203K	Decrease	−0.93
9	Nucleocapsid phosphoprotein	G204R	Decrease	−0.52
10	Nucleocapsid phosphoprotein	T205I	Decrease	−0.53
11	ORF1AB Polyprotein	R207C	Decrease	−0.60
12	ORF1AB Polyprotein	K292E	Decrease	−0.42
13	ORF1AB Polyprotein	V378I	Decrease	−0.23
14	ORF1AB Polyprotein	S944L	Decrease	0.63
15	ORF1AB Polyprotein	T1246I	Increase	0.14
16	ORF1AB Polyprotein	K1305 N	Increase	−0.10
17	ORF1AB Polyprotein	Q2702H	Decrease	−0.68
18	ORF1AB Polyprotein	G3278S	Decrease	−0.89
19	ORF1AB Polyprotein	L3606F	Decrease	−1.00
20	ORF1AB Polyprotein	A4489L	Decrease	−0.67
21	ORF1AB Polyprotein	P4715L	Decrease	−0.83
22	ORF1A Polyprotein	R207C	Decrease	−0.60
23	ORF1A Polyprotein	K292E	Decrease	−0.42
24	ORF1A Polyprotein	S318I	Increase	0.36
25	ORF1A Polyprotein	D448 N	Decrease	−0.98
26	ORF1A Polyprotein	V627F	Decrease	−1.19
27	ORF1A Polyprotein	S944L	Increase	0.63
28	ORF1A Polyprotein	T1246I	Increase	0.14
29	ORF1A Polyprotein	K1305 N	Increase	−0.10
30	ORF1A Polyprotein	P2965L	Decrease	−0.67
31	ORF1A Polyprotein	G3278S	Decrease	−0.89
32	ORF1A Polyprotein	L3606F	Decrease	−1.00
33	0RF3A Polyprotein	Q57H	Decrease	−0.90

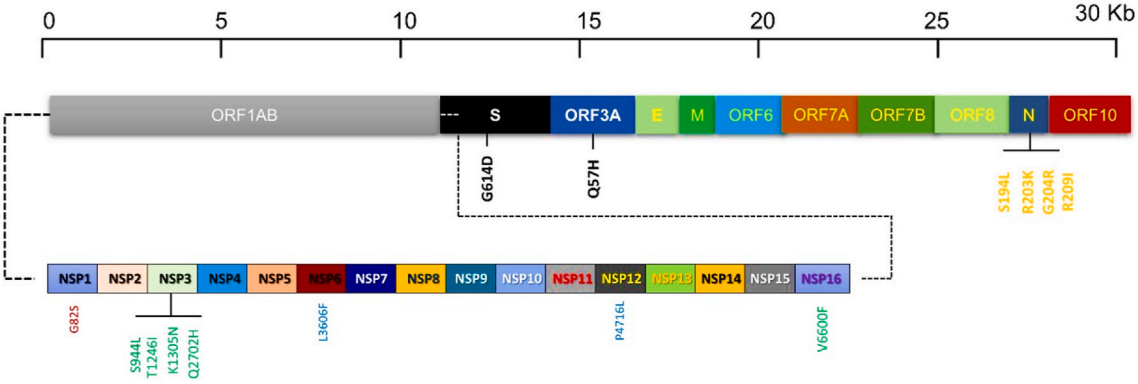

Fig. 2. Schematic representation illustrating the distribution of recurrent mutations in structural (S, E, M, N), Accessories (ORF3A, ORF6, ORF7A, ORF7B, ORF8, ORF10), and nonstructural (nsp1 to nsp 16) proteins along with the SARS-CoV-2 genome. Recurrent mutations are represented by vertical lines.

Table 3
Prediction of effects of the mutation on different biological pathways.

Protein Name	Mutation	Effects
Membrane Glycoprotein	H125Y	Altered Transmembrane protein Altered Ordered interface
ORF8	W45L	Altered Ordered interface Altered Disordered interface Altered DNA binding Altered Transmembrane protein
ORF1A	T2408I	Altered Transmembrane protein Gain of strand Loss of Disulfide linkage at C2404 Gain of *N*-linked glycosylation at N2405 Loss of Catalytic site at C2404 Loss of GPI-anchor amidation at N2405
	P2965L	Altered Transmembrane protein Altered Ordered interface Loss of Relative solvent accessibility Altered Metal binding
	P3447S	Altered Transmembrane protein Altered Ordered interface Loss of Allosteric Site at Y3445 Gain of Strand Loss of Relative solvent accessibility Gain of Proteolytic cleavage at D3450 Loss of Pyrrolidone carboxylic acid at Q3452 Gain of Sulfation at Y3445
ORF1AB	L3606	Loss of Helix Gain of Loop Altered Ordered interface Gain of Sulfation at Y3607 Gain of Strand

3.10. Docking of the wild type and mutant spike structures

Keeping in mind the involvement of spike protein in a particular process, including attachment and pathogenesis, we investigated the critical analysis of both wild and mutant versions of the spike protein (Spike RBD) with ACE2 receptor. Structural study of protein interactions and their binding energy is an important initial step in biological process knowledge. Binding affinity, which defines whether a complex is formed under specific conditions, is essential to regulate molecular interactions, create new treatments (i.e. directing rational drug design) and anticipate the impact variation on the protein interfaces [68]. Prior to docking of Spike and ACE2 superimposition of the mutants was performed to check the structural variations caused by the mutations. The generated mutants were superimposed on wild-type spike protein, and the RMSD values were recorded (Fig. 4a, b, c).

Therefore, we used HDOCK to perform the docking of ACE2 with the wild-type and double mutant spike protein (D614G-V622F–S943F D614G-Q677H and N74K-D614G) to explain how these mutations led to the higher infectivity of SARS-CoV-2 variants. HDOCK resulted in a docking score of −12.40 kcal/mol for the ACE2-wild type spike complex. PDBsum interaction study revealed that the interface is made up of 37 residues. 18 residues are supplied by ACE2 and 15 by spike-RBD, which are responsible for interface formation. One salt bridge, 11 hydrogen bonds, and 125 non-bonded interactions were found at the interface. The hydrogen bonds formed by the ACE2-spike (RBD) wild include Gln493-Glu35, Gln493-Lys31, Lys417-Asp30, Gly476-Glu23, Tyr505-Glu37, Gly502-Lys353, Gly496- Lys353, Gln498- Lys353, and double hydrogen bond between Thr500-Tyr41(Fig. 4d). One salt bridge was reported between Asp30 (ACE2) and Lys417 (spike RBD).

The HDOCK docking score for D614G-V622F–S943F (ACE2-spike RBD) was reported as −13.70 kcal/mol. This distinction between wild-

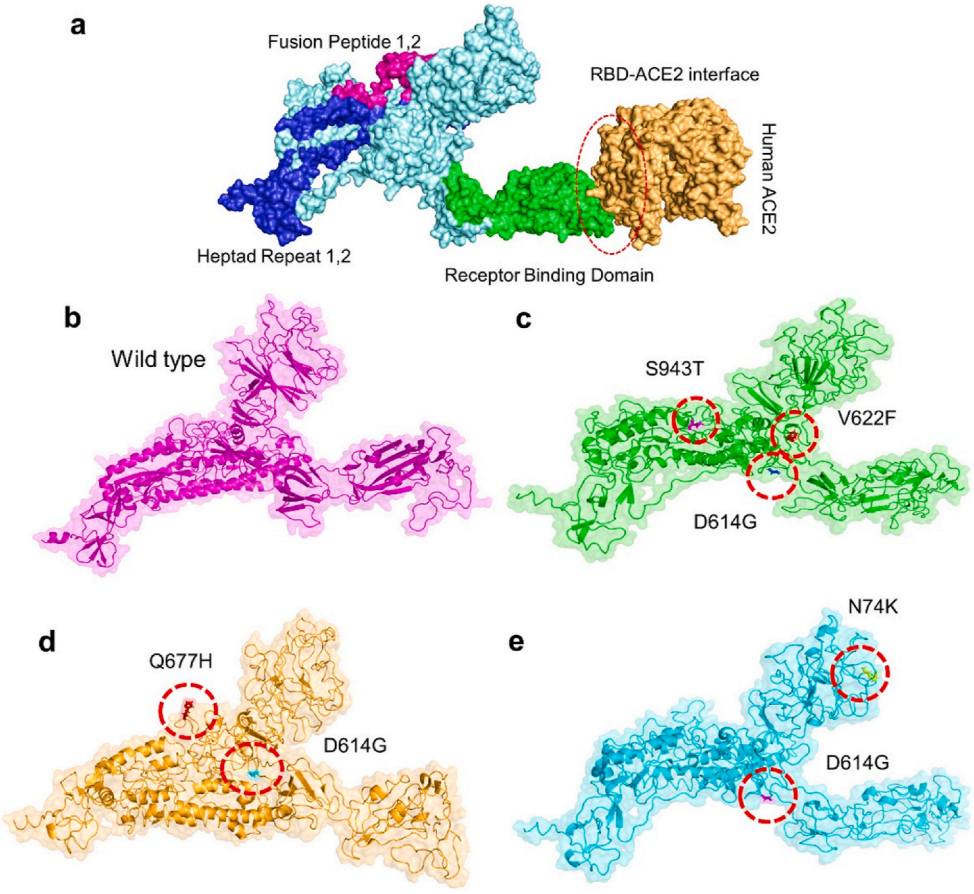

Fig. 3. Representation of spike glycoprotein structure (PDB ID:6M0J**) and SARS-CoV-2 binding domain** (**a**) Different domains of spike protein with different colors. The green color shows the RBD domain while the orange color shows the ACE2 receptor. The right-side figure shows the binding interface along with its stick representation of the key hydrogen interactions of ACE2 and the spike RBD. (**b**) wild type spike protein, (**c**) D614G, V622F, S943T (**d**) D614G, Q677H (**e**) N74K, D614G. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

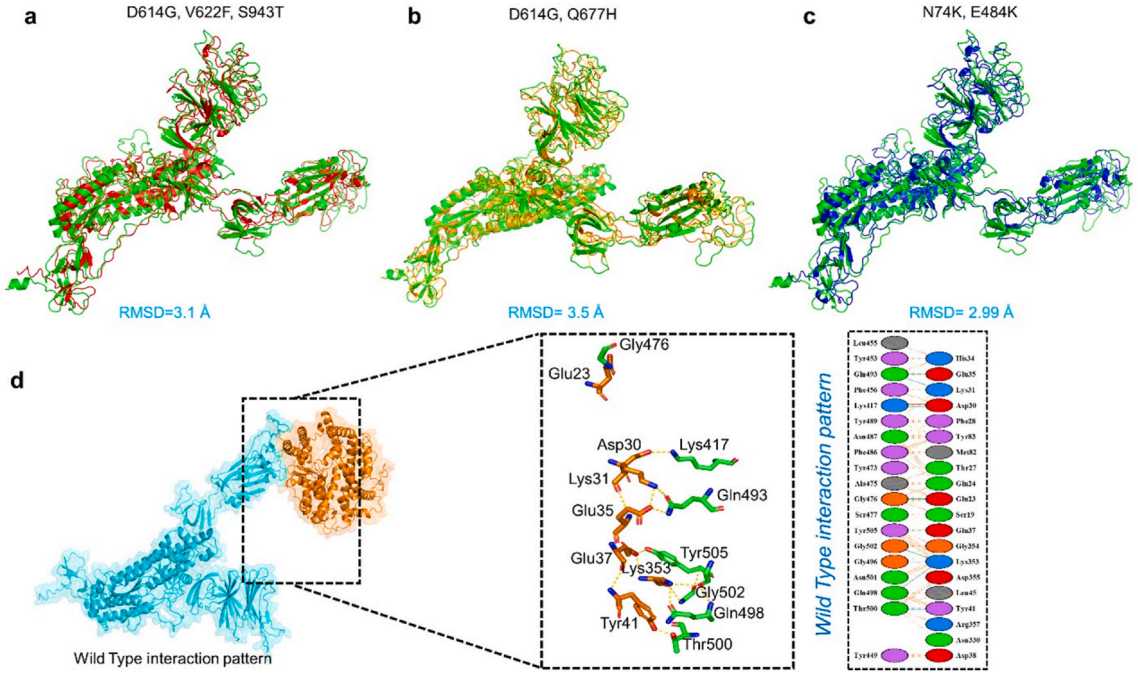

Fig. 4. (a, b, c) Superimposed structure of wild-type spike protein (green) with D614G-V622F–S943F (red), and D614G-Q677H (orange) and. N74K-D614 (blue). The RMSD of each superimposition was reported to be 3.1 Å, 3.5 Å, and 2.99 Å. **(d)** Docking complex of wild-type spike protein with ACE2. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

type and mutant was investigated further by examining the molecular interactions. The interaction analysis showed two salt bridges and an increased number of hydrogen bonds (12 hydrogen bonds) in mutant form. The enhanced interaction is due to substituted residue D614G, V622F, S943F in the spike protein complex (Fig. 5a). The key residues Lys31-Glu484 and Asp30-Lys417 established the salt bridges. Among the hydrogen bonds Thr747-Asp609, Lys417-His195, Glu406-His195, Thr415-Arg195, Gln755-His493, Asn501-Gln86, Gln498-Gln86, Tyr489-Ser77, Tyr489-Lys74, and Arg457-Ser109 residues are involved. As previously stated, the D614G-V622F–S943F mutation is responsible for enhancing the dissemination and infectivity of the SARS-CoV-2 variant and increasing the binding affinity and infectivity. The HDOCK docking score for D614G, Q677H (ACE2-spike RBD), was −13.42 kcal/mol. The difference between the HDOCK score of wild-type and mutant is explained in terms of molecular interactions. Among the hydrogen bonds Gly476-Ser19, Gly476-Glu23, Ser477-Gln24, Tyr489-Tyr83, Tyr449-Asp38, Gln493-Glu35, Thr500-Tyr41, Gln489-Lys353, Gly502-Lys353, and Tyr505-Glu37 residues are involved (Fig. 5b). As a result, the D614G, Q677H double mutant enhanced binding efficiency, and this mutant, as well as other interface residues, are important hot-spots for therapeutic development against SARS-CoV-2 variants.

The predicted score of HDOCK for the ACE2-N74K, D614G mutant spike protein was −12.50 which is comparable with the wild-type complex. According to a PDBsum study of the complex, the interface

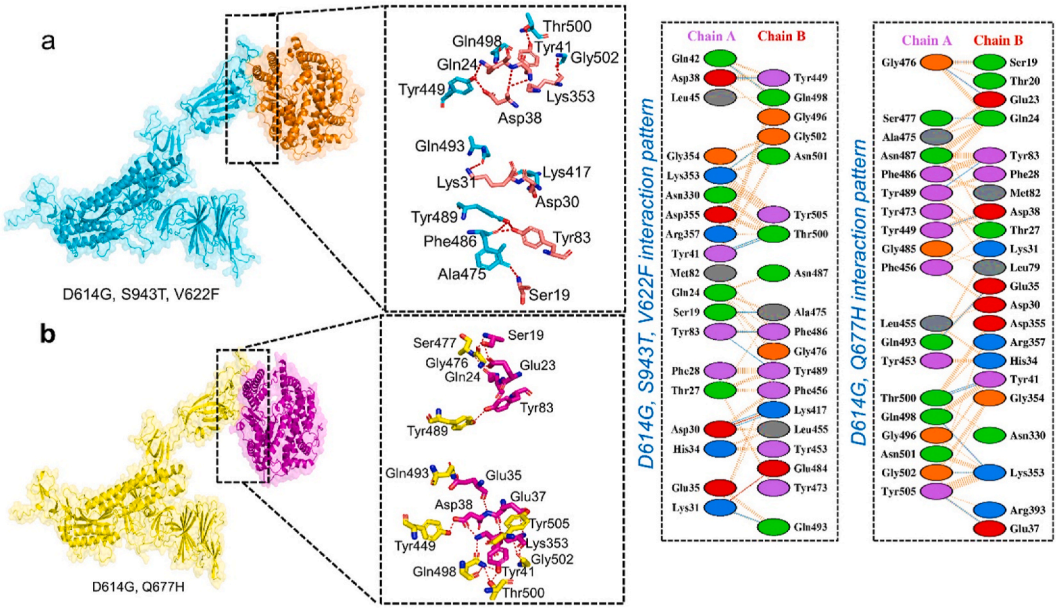

Fig. 5. Docking complexes of D614G-V622F–S943F spike protein with ACE2. **(b)** Docking complexes of D614G-Q677H spike protein with ACE2.

is formed by 37 residues, 18 of which are supplied by ACE2 and 15 by the spike-RBD. According to the interaction study, both structures create 1 salt bridge, 11 hydrogen bonds, and 125 non-bonded interactions. The hydrogen bonds formed by the ACE2- N74K, D614G spike (RBD) include Gln493-Glu35, Gln493-Lys31, Tyr453-His34, Lys417-Asp30, Thr500-Tyr41, Tyr449-Asp38, Tyr505-Glu37, Gly502-Lys353, Tyr449- Lys353, Gln498-Lys353, and double hydrogen bond between Thr500-Tyr41 (Fig. 6). In this complex, most of the hydrogen binding pattern is similar to the wild-type complex.

3.11. Estimation of K_D

Estimation of the K_D revealed that D614G-S943T-V622F (2.0 E^{-10}) variant binds stronger than all. Moreover, the strength of the wild type and mutants was ranked as D614G-Q677H variant (1.0 E^{-09}), wild type (1.9 E^{-09}) while for the N74K-D614G variant the predicted K_D value was 2.2 E^{-09}. Conclusively this shows that the two mutants possess stronger affinity than the wild type which might be responsible for the increased cases in Pakistan. The docking score and the predicted K_D values are given in Table 4.

3.12. Dynamic stability and flexibility analysis

To understand the dynamic behavior of the wild type and these variants, MD simulation of each complex was performed. The stability of each complex was assessed as root mean square deviation (RMSD) and residual flexibility as root mean square fluctuation. The RMSD graphs shown in Fig. 7A-C demonstrate that the wild type exhibit more stable dynamics than the variants. From Fig. 7A, it can be seen that the wild type complex attained the equilibrium at 25ns and reached stability at 2.0 Å. On the other hand, the D614G-S943T-V622F variant complex demonstrated unstable dynamics at different times interval thus corroborated with the servers' predicted results. Unlike the D614G-S943T-V622F variant, the other two variants i.e., D614G-Q677H and N74K-D614G reported a significantly destabilizing effect throughout the simulation. From Fig. 7B and C, it can be seen that these two complexes demonstrated significant structural perturbation and hence reports uniform results with the aforementioned results predicted by different servers.

On the other hand, the RMSF results demonstrated a differential flexibility index for each complex thus showing the impact of these mutations on the structural-dynamic features. The RMSF results of each complex are shown in Fig. 8.

Table 4
The predicted docking scores and K_D values for each complex.

Complex	ΔG (kcal mol^{-1})	K_D (M) at 38.0 °C
Wild type	−12.40	1.9E^{-09}
D614G-S943T-V622F Variant	−13.42	2.0E^{-10}
D614G-Q677H Variant	−13.70	1.0E^{-09}
N74K-D614G Variant	−12.50	2.2E^{-09}

3.13. Binding free energy calculations

To connect protein conformational changes with the binding, induced by mutations in the proteins, we used the MM/GBSA technique to calculate the binding energy of spike RBD to hACE2. The binding comparison revealed significant variation in the spike RBD binding between the wild type, D614G-V622F–S943F, and D614G-Q677H double mutants. As given in Table 5, the total binding energy for the D614G-Q677H variant was −75.78 kcal/mol followed by the D614G-S943T-V622F variant −75.17 kcal/mol and −73.84 kcal/mol for the N74K-D614G variant. The total binding energy for the wild type was −66.34 kcal/mol. The vdW results reported a comparable result however the electrostatic energies are reported to be higher in the case of the variants which strongly corroborate with the previous papers which reported an increment in the electrostatic energy by variants. The notion of electrostatic energy has been a key factor in the enhanced binding [20,59,60,69]. In sum, the polar solvation energy supplies a non-favorable contribution by comparing with non-polar which reported a good contribution. The dynamic's environment's charge density seems to engage with itself infrequently, explaining the non-significant effect in binding. In brief, the solvation energy of variants adhering to ACE2 is unfavorable, whilst gas-phase energy is critical for the variants' strong binding to ACE2. Decisively this shows that the re-ranking of the complex using MM/GBSA validated that the mutants possess a stronger affinity towards ACE2 than the wild type thus enhancing the transmission and infectivity. The MM/GBSA results including vdW (van der Waal), electrostatic (ELE), GB (generalized Born), SA (Surface Area), and the total binding energy are given in Table 5.

4. Conclusions

In summary, this study used genome-based analysis and molecular modeling approaches to determine a basis for the increased infection and transmission of SARS-CoV-2 in Pakistan. The data revealed that the

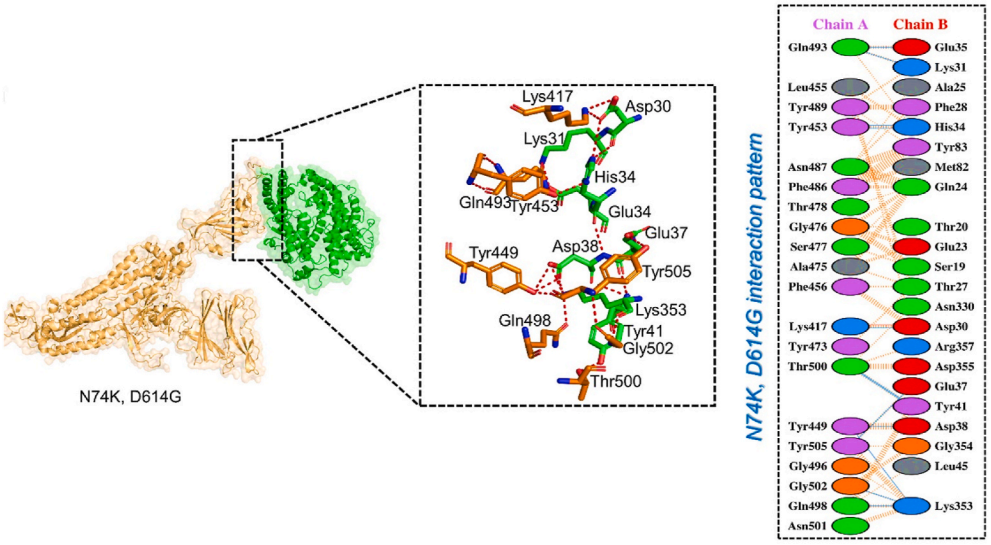

Fig. 6. Docking complexes of ACE2-N74K, D614G spike protein with ACE2. The right panel shows the 2D interaction pattern generated with PDBsum.

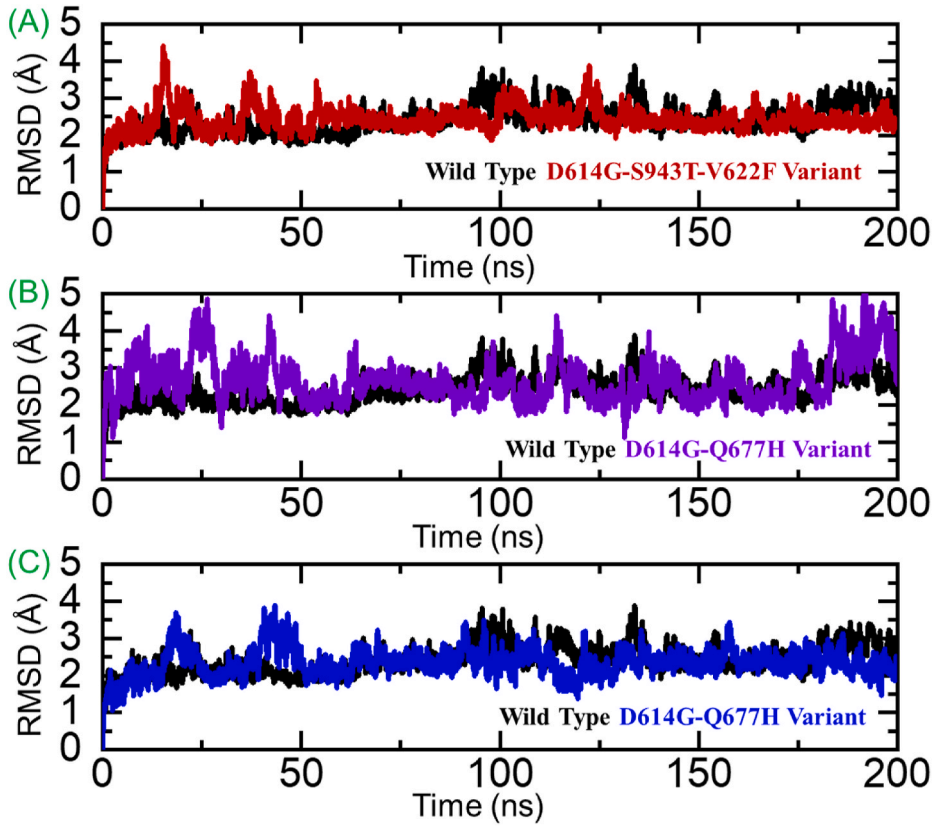

Fig. 7. Dynamic stability analysis of the wild type and mutant complexes. (A) RMSD of the wild type. **(B)** RMSD of D614G-S943T-V622F, **(C)** RMSD of D614G-Q677H and **(D)** RMSD N74K-D614G variant complex.

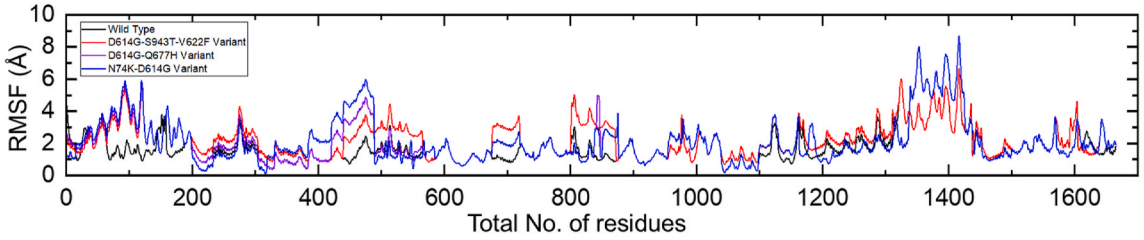

Fig. 8. Residual flexibility of the wild type, D614G-S943T-V622F, D614G-Q677H, and RMSD N74K-D614G variant complexes.

Table 5
MM/GBSA analysis of the wild type and mutant complexes. All the results are calculated in kcal/mol.

Complex	VDW	ELE	GB	SA	TOTAL
Wild Type	−108.39	−659.94	716.45	−14.46	**−66.34**
D614G-S943T-V622F Variant	−114.64	−697.17	751.1	−14.46	**−75.17**
D614G-Q677H Variant	−109.08	−700.02	747.87	−14.55	**−75.78**
N74K-D614G Variant	−109.81	−687.13	737.56	−14.46	**−73.84**

D614G mutation particularly contributes to the enhanced infectivity. A significant bonding reprogramming was identified, and the structural dynamic features strongly corroborated the predicted results. However, such polymorphic sites should be further studied to explore the evolutionary paradigm and any emerging variants. This work provides a strong impetus for the development of novel drugs against the strains circulating in Pakistan and around the world.

Data availability

All the data is available on RCSB, UniProt, and any simulation data would be provided on reasonable demand. The accession numbers to access this data are given in the manuscript.

Ethical approval

Not Applicable.

Declaration of competing interest

Declared None.

Acknowledgments

Dong-Qing Wei is supported by grants from the Key Research Area Grant 2016YFA0501703 of the Ministry of Science and Technology of China, the National Science Foundation of China (Grant No. 32070662, 61832019, 32030063), the Science and Technology Commission of

Shanghai Municipality (Grant No.: 19430750600), as well as SJTU JiRLMDS Joint Research Fund and Joint Research Funds for Medical and Engineering and Scientific Research at Shanghai Jiao Tong University (YG2021ZD02). The computations were partially performed at the PengCheng Lab. and the Center for High-Performance Computing, Shanghai Jiao Tong University.

References

[1] D. Benvenuto, M. Giovanetti, A. Ciccozzi, S. Spoto, S. Angeletti, M. Ciccozzi, The 2019-new coronavirus epidemic: evidence for virus evolution, J. Med. Virol. 92 (2020) 455–459.
[2] F. Wu, S. Zhao, B. Yu, Y.M. Chen, W. Wang, Z.G. Song, Y. Hu, Z.W. Tao, J.H. Tian, Y.Y. Pei, M.L. Yuan, Y.L. Zhang, F.H. Dai, Y. Liu, Q.M. Wang, J.J. Zheng, L. Xu, E. C. Holmes, Y.Z. Zhang, A new coronavirus associated with human respiratory disease in China, Nature 579 (2020) 265–269.
[3] V. Stadnytskyi, C.E. Bax, A. Bax, P. Anfinrud, The airborne lifetime of small speech droplets and their potential importance in SARS-CoV-2 transmission, Proc. Natl. Acad. Sci. U.S.A. 117 (2020) 11875–11877.
[4] M. Klompas, M.A. Baker, C. Rhee, Airborne transmission of SARS-CoV-2: theoretical considerations and available evidence, JAMA 324 (2020) 441–442.
[5] H. Hilal El Idrissi, COVID-19: what you need to know, Gene. Rep. 20 (2020) 100756.
[6] N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, P. Niu, F. Zhan, X. Ma, D. Wang, W. Xu, G. Wu, G.F. Gao, W. Tan, A novel coronavirus from patients with pneumonia in China, 2019, N. Engl. J. Med. 382 (2020) 727–733.
[7] H.S. Rahman, M.S. Aziz, R.H. Hussein, H.H. Othman, S.H. Salih Omer, E.S. Khalid, N.A. Abdulrahman, K. Amin, R. Abdullah, The transmission modes and sources of COVID-19: a systematic review, Intern. J. Surg. Open 26 (2020) 125–136.
[8] W.J. Guan, Z.Y. Ni, Y. Hu, W.H. Liang, C.Q. Ou, J.X. He, L. Liu, H. Shan, C.L. Lei, D. S.C. Hui, B. Du, L.J. Li, G. Zeng, K.Y. Yuen, R.C. Chen, C.L. Tang, T. Wang, P. Y. Chen, J. Xiang, S.Y. Li, J.L. Wang, Z.J. Liang, Y.X. Peng, L. Wei, Y. Liu, Y.H. Hu, P. Peng, J.M. Wang, J.Y. Liu, Z. Chen, G. Li, Z.J. Zheng, S.Q. Qiu, J. Luo, C.J. Ye, S. Y. Zhu, N.S. Zhong, Clinical characteristics of coronavirus disease 2019 in China, N. Engl. J. Med. 382 (2020) 1708–1720.
[9] D. Tang, P. Comish, R. Kang, The hallmarks of COVID-19 disease, PLoS Pathog. 16 (2020), e1008536.
[10] S.M. Haque, O. Ashwaq, A. Sarief, A.K. Azad John Mohamed, A comprehensive review about SARS-CoV-2, Future Virol. 15 (2020) 625–648.
[11] T. Day, S. Gandon, S. Lion, S.P. Otto, On the evolutionary epidemiology of SARS-CoV-2, Curr. Biol. 30 (2020) R849–R857.
[12] R. Yan, Y. Zhang, Y. Li, L. Xia, Y. Guo, Q. Zhou, Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2, Science 367 (2020) 1444–1448.
[13] J. Shang, G. Ye, K. Shi, Y. Wan, C. Luo, H. Aihara, Q. Geng, A. Auerbach, F. Li, Structural basis of receptor recognition by SARS-CoV-2, Nature 581 (2020) 221–224.
[14] D.C. Dinesh, D. Chalupska, J. Silhan, V. Veverka, E. Boura, Structural Basis of RNA Recognition by the SARS-CoV-2 Nucleocapsid Phosphoprotein, bioRxiv, 2020.
[15] S. Duchene, L. Featherstone, M. Haritopoulou-Sinanidou, A. Rambaut, P. Lemey, G. Baele, Temporal signal and the phylodynamic threshold of SARS-CoV-2, Virus evol. 6 (2020) veaa061.
[16] I. Hussain, N. Pervaiz, A. Khan, S. Saleem, H. Shireen, D.-Q. Wei, V. Labrie, Y. Bao, A.A. Abbasi, Evolutionary and structural analysis of SARS-CoV-2 specific evasion of host immunity, Gene Immun. (2020) 1–11.
[17] T.P. Peacock, D.H. Goldhill, J. Zhou, L. Baillon, R. Frise, O.C. Swann, R. Kugathasan, R. Penn, J.C. Brown, R.Y. Sanchez-David, The furin cleavage site in the SARS-CoV-2 spike protein is required for transmission in ferrets, Nat. Microbiol. (2021) 1–11.
[18] A. Khan, T. Zia, M. Suleman, T. Khan, S.S. Ali, A.A. Abbasi, A. Mohammad, D.-Q. Wei, 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. (2021) n/a.
[19] A. Khan, M. Khan, S. Saleem, Z. Babar, A. Ali, A.A. Khan, Z. Sardar, F. Hamayun, S. S. Ali, D.-Q. Wei, Phylogenetic analysis and structural perspectives of RNA-dependent RNA-polymerase inhibition from SARs-CoV-2 with natural products, Interdiscipl. Sci. Comput. Life Sci. 12 (2020) 335–348.
[20] A. Khan, D.-Q. Wei, K. Kousar, J. Abubaker, S. Ahmad, J. Ali, F. Al-Mulla, S.S. Ali, N. Nizam-Uddin, A.M. Sayaf, Preliminary Structural Data Revealed that the SARS-CoV-2 B. 1.617 Variant's RBD Binds to ACE2 Receptor Stronger than the Wild Type to Enhance the Infectivity, ChemBioChem.
[21] W.F. Garcia-Beltran, E.C. Lam, K.S. Denis, A.D. Nitido, Z.H. Garcia, B.M. Hauser, J. Feldman, M.N. Pavlovic, D.J. Gregory, M.C. Poznansky, A. Sigal, A.G. Schmidt, A.J. Iafrate, V. Naranbhai, A.B. Balazs, Multiple SARS-CoV-2 Variants Escape Neutralization by Vaccine-Induced Humoral Immunity, medRxiv, 2021, 2021.2002.2014.21251704.
[22] K. Nagano, C. Tani-Sassa, Y. Iwasaki, Y. Takatsuki, S. Yuasa, Y. Takahashi, J. Nakajima, K. Sonobe, N. Ichimura, Y. Nukui, SARS-CoV-2 R. 1 Lineage Variants Prevailed in Tokyo in March 2021, medRxiv, 2021.
[23] B. Zhou, T.T.N. Thao, D. Hoffmann, A. Taddeo, N. Ebert, F. Labroussaa, A. Pohlmann, J. King, J. Portmann, N.J. Halwe, SARS-CoV-2 Spike D614G Variant Confers Enhanced Replication and Transmissibility, bioRxiv, 2020.
[24] M. Hoffmann, H. Hofmann-Winkler, N. Krüger, A. Kempf, I. Nehlmeier, L. Graichen, A. Sidarovich, A.-S. Moldenhauer, M.S. Winkler, S. Schulz, H.-M. Jäck, M.V. Stankov, G.M.N. Behrens, S. Pöhlmann, SARS-CoV-2 Variant B.1.617 Is Resistant to Bamlanivimab and Evades Antibodies Induced by Infection and Vaccination, bioRxiv, 2021, 2021.2005.2004.442663.
[25] M. Hoffmann, P. Arora, R. Groß, A. Seidel, B.F. Hörnich, A.S. Hahn, N. Krüger, L. Graichen, H. Hofmann-Winkler, A. Kempf, SARS-CoV-2 variants B. 1.351 and P. 1 escape from neutralizing antibodies, Cell 184 (2021) 2384–2393, e2312.
[26] J. Singh, S.A. Rahman, N.Z. Ehtesham, S. Hira, S.E. Hasnain, SARS-CoV-2 variants of concern are emerging in India, Nat. Med. (2021) 1–3.
[27] D. Focosi, M. Tuccori, A. Baj, F. Maggi, SARS-CoV-2 Variants: A Synopsis of in Vitro Efficacy Data of Convalescent Plasma, Currently Marketed Vaccines, and Monoclonal Antibodies, Multidisciplinary Digital Publishing Institute, 2021.
[28] T. Koyama, D. Platt, L. Parida, Variant analysis of SARS-CoV-2 genomes, Bull. World Health Organ. 98 (2020) 495.
[29] Á. Nagy, S. Pongor, B. Győrffy, Different mutations in SARS-CoV-2 associate with severe and mild outcome, Int. J. Antimicrob. Agents 57 (2021) 106272.
[30] S. Wu, C. Tian, P. Liu, D. Guo, W. Zheng, X. Huang, Y. Zhang, L. Liu, Effects of SARS-CoV-2 mutations on protein structures and intraviral protein–protein interactions, J. Med. Virol. 93 (2021) 2132–2140.
[31] L. Zhang, C.B. Jackson, H. Mou, A. Ojha, E.S. Rangarajan, T. Izard, M. Farzan, H. Choe, The D614G Mutation in the SARS-CoV-2 Spike Protein Reduces S1 Shedding and Increases Infectivity, BioRxiv, 2020.
[32] D.C. Groves, S.L. Rowland-Jones, A. Angyal, The D614G mutations in the SARS-CoV-2 spike protein: implications for viral infectivity, disease severity and vaccine design, Biochem. Biophys. Res. Commun. 538 (2021) 104–107.
[33] A. Mohammad, E. Alshawaf, S.K. Marafie, M. Abu-Farha, J. Abubaker, F. Al-Mulla, Higher binding affinity of furin for SARS-CoV-2 spike (S) protein D614G mutant could be associated with higher SARS-CoV-2 infectivity, Int. J. Infect. Dis. 103 (2021) 611–616.
[34] L. Yurkovetskiy, X. Wang, K.E. Pascal, C. Tomkins-Tinch, T.P. Nyalile, Y. Wang, A. Baum, W.E. Diehl, A. Dauphin, C. Carbone, Structural and functional analysis of the D614G SARS-CoV-2 spike protein variant, Cell 183 (2020) 739–751, e738.
[35] J.L. Bernal, N. Andrews, C. Gower, E. Gallagher, R. Simmons, S. Thelwall, J. Stowe, E. Tessier, N. Groves, G. Dabrera, R. Myers, C. Campbell, G. Amirthalingam, M. Edmunds, M. Zambon, K. Brown, S. Hopkins, M. Chand, M. Ramsay, Effectiveness of COVID-19 Vaccines against the B.1.617.2 Variant, medRxiv, 2021, p. 2021, 2005.2022.21257658.
[36] T. Tada, H. Zhou, B.M. Dcosta, M.I. Samanovic, M.J. Mulligan, N.R. Landau, The Spike Proteins of SARS-CoV-2 B. 1.617 and B. 1.618 Variants Identified in India Provide Partial Resistance to Vaccine-Elicited and Therapeutic Monoclonal Antibodies, BioRxiv, 2021.
[37] Y. Ruan, C.L. Wei, A.E. Ling, V.B. Vega, H. Thoreau, S.Y.S. Thoe, J.-M. Chia, P. Ng, K.P. Chiu, L. Lim, Comparative full-length genome sequence analysis of 14 SARS coronavirus isolates and common mutations associated with putative origins of infection, Lancet 361 (2003) 1779–1785.
[38] A. Ali, R. Vijayan, Dynamics of the ACE2–SARS-CoV-2/SARS-CoV spike protein interface reveal unique mechanisms, Sci. Rep. 10 (2020) 1–12.
[39] M.T. Khan, A. Khan, A.U. Rehman, Y. Wang, K. Akhtar, S.I. Malik, D.-Q. Wei, Structural and free energy landscape of novel mutations in ribosomal protein S1 (rpsA) associated with pyrazinamide resistance, Sci. Rep. 9 (2019) 1–12.
[40] A. Khan, M.T. Khan, S. Saleem, M. Junaid, A. Ali, S.S. Ali, M. Khan, D.-Q. Wei, Structural Insights into the mechanism of RNA recognition by the N-terminal RNA-binding domain of the SARS-CoV-2 nucleocapsid phosphoprotein, Comput. Struct. Biotechnol. J. (2020).
[41] V.K. Bhardwaj, R. Singh, P. Das, R. Purohit, Evaluation of acridinedione analogs as potential SARS-CoV-2 main protease inhibitors and their comparison with repurposed anti-viral drugs, Comput. Biol. Med. 128 (2021) 104117.
[42] R. Singh, V.K. Bhardwaj, J. Sharma, R. Purohit, S. Kumar, In-silico evaluation of bioactive compounds from tea as potential SARS-CoV-2 nonstructural protein 16 inhibitors, J. tradition. complemen. med. (2021).
[43] J. Sharma, V. Kumar Bhardwaj, R. Singh, V. Rajendran, R. Purohit, S. Kumar, An in-silico evaluation of different bioactive molecules of tea for their inhibition potency against non structural protein-15 of SARS-CoV-2, Food Chem. 346 (2021) 128933.
[44] R. Singh, V.K. Bhardwaj, J. Sharma, D. Kumar, R. Purohit, Identification of potential plant bioactive as SARS-CoV-2 Spike protein and human ACE2 fusion inhibitors, Comput. Biol. Med. 136 (2021) 104631.
[45] S.K. Miryala, S. Basu, A. Naha, R. Debroy, S. Ramaiah, A. Anbarasu, S. Natarajan, Identification of bioactive natural compounds as efficient inhibitors against Mycobacterium tuberculosis protein-targets: a molecular docking and molecular dynamics simulation study, J. Mol. Liq. 341 (2021) 117340.
[46] S. Basu, B. Veeraraghavan, S. Ramaiah, A. Anbarasu, Novel cyclohexanone compound as a potential ligand against SARS-CoV-2 main-protease, Microb. Pathog. 149 (2020) 104546.
[47] Y. Shu, J. McCauley, GISAID: global initiative on sharing all influenza data–from vision to reality, Euro Surveill. 22 (2017) 30494.
[48] K.D. Pruitt, T. Tatusova, D.R. Maglott, NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins, Nucleic Acids Res. 33 (2005) D501–D504.
[49] M. Li, T.-f. Yu, The application of DNAStar soft in research of animal virus, J. Sci. Teach. Coll. Univ. 30 (2010) 61–63.
[50] J.D. Thompson, T.J. Gibson, D.G. Higgins, Multiple sequence alignment using ClustalW and ClustalX, Curr. protoc. bioinform. (2003) 2.3. 1–2.3. 22.

[51] L. Wang, J. Zheng, Y. Luo, T. Xu, Q. Zhang, L. Zhang, M. Xu, J. Wan, M.B. Wang, C. Zhang, Construction of a genomewide RNA i mutant library in rice, Plant biotechnol. j. 11 (2013) 997–1005.
[52] V. Pejaver, J. Urresti, J. Lugo-Martinez, K.A. Pagel, G.N. Lin, H.-J. Nam, M. Mort, D.N. Cooper, J. Sebat, L.M. Iakoucheva, MutPred2: Inferring the Molecular and Phenotypic Impact of Amino Acid Variants, BioRxiv, 2017, p. 134981.
[53] P.W. Rose, A. Prlić, A. Altunkaya, C. Bi, A.R. Bradley, C.H. Christie, L.D. Costanzo, J.M. Duarte, S. Dutta, Z. Feng, The RCSB Protein Data Bank: Integrative View of Protein, Gene and 3D Structural Information, Nucleic acids research, 2016, p. gkw1000.
[54] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera—a visualization system for exploratory research and analysis, J. Comput. Chem. 25 (2004) 1605–1612.
[55] Y. Yan, H. Tao, J. He, S.-Y. Huang, The HDOCK server for integrated protein–protein docking, Nat. Protoc. 15 (2020) 1829–1852.
[56] L.C. Xue, J.P. Rodrigues, P.L. Kastritis, A.M. Bonvin, A. Vangone, PRODIGY: a web server for predicting the binding affinity of protein–protein complexes, Bioinformatics 32 (2016) 3676–3678.
[57] D.A. Case, T.E. Cheatham III, T. Darden, H. Gohlke, R. Luo, K.M. Merz Jr., A. Onufriev, C. Simmerling, B. Wang, R.J. Woods, The Amber biomolecular simulation programs, J. Comput. Chem. 26 (2005) 1668–1688.
[58] D.A. Pearlman, D.A. Case, J.W. Caldwell, W.S. Ross, T.E. Cheatham III, S. DeBolt, D. Ferguson, G. Seibel, P. Kollman, AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules, Comput. Phys. Commun. 91 (1995) 1–41.
[59] A. Khan, T. Zia, M. Suleman, T. Khan, S.S. Ali, A.A. Abbasi, A. Mohammad, D. Q. Wei, 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 (2021) 7045–7057.
[60] A. Khan, J. Gui, W. Ahmad, I. Haq, M. Shahid, A.A. Khan, A. Shah, A. Khan, L. Ali, Z. Anwar, The SARS-CoV-2 B. 1.618 variant slightly alters the spike RBD–ACE2 binding affinity and is an antibody escaping variant: a computational structural perspective, RSC Adv. 11 (2021) 30132–30147.
[61] D.R. Roe, T.E. Cheatham III, PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data, J. Chem. Theor. Comput. 9 (2013) 3084–3095.
[62] H. Sun, Y. Li, S. Tian, L. Xu, T. Hou, Assessing the performance of MM/PBSA and MM/GBSA methods. 4. Accuracies of MM/PBSA and MM/GBSA methodologies evaluated by various simulation protocols using PDBbind data set, Phys. Chem. Chem. Phys. 16 (2014) 16719–16729.
[63] G. Weng, E. Wang, Z. Wang, H. Liu, F. Zhu, D. Li, T. Hou, HawkDock: a web server to predict and analyze the protein–protein complex based on computational docking and MM/GBSA, Nucleic Acids Res. 47 (2019) W322–W330.
[64] R. Wang, J. Chen, K. Gao, Y. Hozumi, C. Yin, G.-W. Wei, Analysis of SARS-CoV-2 mutations in the United States suggests presence of four substrains and novel variants, Commun. bio. 4 (2021) 1–14.
[65] M. Laamarti, T. Alouane, S. Kartti, M. Chemao-Elfihri, M. Hakmi, A. Essabbar, M. Laamarti, H. Hlali, H. Bendani, N. Boumajdi, Large scale genomic analysis of 3067 SARS-CoV-2 genomes reveals a clonal geo-distribution and a rich genetic variations of hotspots mutations, PLoS One 15 (2020), e0240345.
[66] F. Rashid, E.E. Dzakah, H. Wang, S. Tang, The ORF8 protein of SARS-CoV-2 induced endoplasmic reticulum stress and mediated immune evasion by antagonizing production of interferon beta, Virus Res. 296 (2021) 198350.
[67] Y. Zhang, J. Zhang, Y. Chen, B. Luo, Y. Yuan, F. Huang, T. Yang, F. Yu, J. Liu, B. Liu, The ORF8 Protein of SARS-CoV-2 Mediates Immune Evasion through Potently Downregulating MHC-I, BioRxiv, 2020.
[68] G.R. Smith, M.J. Sternberg, Prediction of protein-protein interactions by docking methods, Curr. Opin. Struct. Biol. 12 (2002) 28–35.
[69] A. Khan, T. Khan, S. Ali, S. Aftab, Y. Wang, W. Qiankun, M. Khan, M. Suleman, S. Ali, W. Heng, SARS-CoV-2 New Variants: Characteristic Features and Impact on the Efficacy of Different Vaccines, Biomedicine & Pharmacotherapy, 2021, p. 112176.