


SARS-CoV-2 spike evolutionary behaviors; simulation of N501Y mutation outcomes in terms of immunogenicity and structural characteristic

Neda Rostami¹ | Edris Choupani² | Yaeren Hernandez³ | Seyed S. Arab⁴ |
Seyed M. Jazayeri⁵ | Mohammad M. Gomari² 

¹Department of Chemical Engineering, Faculty of Engineering, Arak University, Arak, Iran

²Department of Medical Biotechnology, Faculty of Allied Medicine, Iran University of Medical Sciences, Tehran, Iran

³Department of Cellular and Molecular Medicine, University of Arizona, Tucson, Arizona, USA

⁴Department of Biophysics, School of Biological Sciences, Tarbiat Modares University, Tehran, Iran

⁵Department of Virology, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran

Correspondence

Mohammad M. Gomari, Department of Medical Biotechnology, Faculty of Allied Medicine, Iran University of Medical Sciences, Tehran 14535, Iran.

Email: mmahmoudig@razi.tums.ac.ir

Abstract

Since the emergence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a large number of mutations in its genome have been reported. Some of the mutations occur in noncoding regions without affecting the pathobiology of the virus, while mutations in coding regions are significant. One of the regions where a mutation can occur, affecting the function of the virus is at the receptor-binding domain (RBD) of the spike protein. RBD interacts with angiotensin-converting enzyme 2 (ACE2) and facilitates the entry of the virus into the host cells. There is a lot of focus on RBD mutations, especially the displacement of N501Y which is observed in the UK/Kent, South Africa, and Brazilian lineages of SARS-CoV-2. Our group utilizes computational biology approaches such as immunoinformatics, protein–protein interaction analysis, molecular dynamics, free energy computation, and tertiary structure analysis to disclose the consequences of N501Y mutation at the molecular level. Surprisingly, we discovered that this mutation reduces the immunogenicity of the spike protein; also, displacement of Asn with Tyr reduces protein compactness and significantly increases the stability of the spike protein and its affinity to ACE2. Moreover, following the N501Y mutation secondary structure and folding of the spike protein changed dramatically.

KEYWORDS

ACE2, computational biology, COVID-19, N501Y, protein folding, SARS-CoV-2, spike

1 | INTRODUCTION

The emergence of COVID-19 can be considered as one of the main human health challenges in the present century. COVID-19 is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which is a positive-strand RNA virus.¹ The genome length of this virus is 29 903 nucleotides (Wuhan strain) and capable of encoding 12 structural and 16 nonstructural proteins.² In addition to the high rate of

transmission, mutations in genome-encoding regions have made it difficult to control the SARS-CoV-2 pandemic.³ Mutation in all SARS-CoV-2 proteins can be significant; however, amino acid displacement in spike protein is important due to its role in binding and the virus entering into the host cells through interaction with the angiotensin-converting enzyme 2 (ACE2).⁴ Although not much time has passed since the onset of COVID-19, many mutations and polymorphisms have been reported in the gene encoding the

spike protein. Mutations and polymorphisms as an unfortunate consequence of evolutionary pressures,^{5–8} based on their position in the spike sequence they are divided into two categories; inside the ACE2 binding interface and outside this region.⁹ Both types of mutations can affect the stability and affinity of the spike to its receptor.¹⁰ An important mutation in the ACE2-binding domain is N501Y, which has been observed in the UK/Kent (B.1.1.7; N501Y.V1), South Africa (B.1.351; N501Y.V2), and Brazilian (B.1.1.28; P1; N501Y.V3) spike protein variants, generating the explanation of variants of concern and suggestion for lineage-specific surveillance.¹¹

So far, many studies have evaluated the biological consequences of the N501Y mutation in SARS-CoV-2 behavior. In an investigation which 496 patients were enrolled to evaluate the difference between clinical signs and genetic characteristics of B.1.1.7 versus non-B.1.1.7 variants, it was found that viral load was higher in B.1.1.7. These samples have a lower cycle threshold in a quantitative polymerase chain reaction assay, and the read depth was higher in B.1.1.7 carrying samples.¹² Examination of other variants carrying the N501Y mutation, such as B.1.351 and B.1.1.28, have shown that the ability to escape from neutralizing antibodies responses increases in these lineages.^{13,14} In addition to characterized lineages, the prevalence of unknown mutants of SARS-CoV-2 carrying the N501Y has been reported and raised concerns. For example, a clinical study that was conducted to investigate the prevalence of COVID-19 in Switzerland identified many patients infected by incognito mutants of spike that did not belong to any lineage.¹⁵ SARS-CoV-2 lineages that introduced K417N/E484K/N501Y spike mutations show a greater transmission and infectivity potency compared to the native form. The ability of the Pfizer vaccine to immunize against variants containing these three mutations is reduced by about 6–8-fold.¹⁶ Liu et al.¹⁷ in an investigation that evaluated the transmission capability of the B.1.1.7 variant, they realized that N501Y substitution is the main factor in increasing the rate of transmission and spike affinity to its receptor. Also, spike^{N501Y} exhibits a high ability to neutralize the therapeutic effects of antibodies composed of V-region IGHV3-53. Regarding the importance of the N501Y mutation in the spike protein biological behavior, our group studied the structural characteristics of this variant using computational biology and structural bioinformatics approaches, such as immunoinformatics, docking, molecular dynamics (MD), and free energy computations.¹⁸ Since limited numbers of studies were conducted to analyze the three-dimensional structure of spike^{N501Y}, the results of this study can provide an atomic-level information for biology researchers and other related disciplines.

2 | MATERIALS AND METHODS

2.1 | Immunogenicity analysis

The N501Y mutation is located in the interaction interface of spike protein with its receptor. Clinical studies demonstrate that this replacement can affect the neutralization resistance potency of SARS-CoV-2 and its viral load, which may indicate a change in immune responses against this variant. Immunogenicity of the N501Y variant was compared with native spike using the immune epitope database (IEDB) server which is funded by the National Institute of Allergy and Infectious Diseases.^{19,20} This server is a comprehensive immunoinformatics database for examining T-cells and B-cells antigens.²¹ The sequences of native and N501Y spike variants were retrieved from UniProt in FASTA format.²² Then, for identification and scoring of T-cells epitopes, major histocompatibility complex Class I (MHC-I) and MHC-II binding predictions of IEDB were implemented. For this purpose, IEDB recommended 2.22 (MHC-I) and NetMHC-Pan 4.1 EL (MHC-II) methods were chosen.²³ As well Linear and discontinuous B-cells epitopes were predicted by Antigen Sequence Properties (Bepipred Linear Epitope Prediction 2.0) and DiscoTope tools of IEDB, respectively.^{24,25}

2.2 | Docking and residue scanning methodology

Molecular docking is a versatile and high throughput approach for molecular interactions modeling. Docking and related items such as virtual screening have revolutionized the world of drug design and biomolecule studies.²⁶ Due to the importance of the subject, many algorithms and docking tools have been developed. The first category is global algorithms (also called blind docking) which docking are performed without accurate information about the binding pockets, like the implemented algorithm in the MDockPeP server.²⁷ Another category is called local docking algorithms and requires sufficient information from the binding pocket, like the algorithm employed on the HADDOCK server.²⁸ Some tools also provide a hybrid environment that allows performing both types of docking such as Rosetta, Schrödinger, ClusPro, and AutoDock.²⁹ In the present study, since the residues in the spike–ACE2 interface was previously identified, local docking was performed by Rosetta v3.12.³⁰ Throughout docking with Rosetta, parameters such as rotation, translation, and the number of outputs (-nstruct) were set at 8°, 3 Å, and 1000, respectively. In the present study, desired structures of native (Protein Data Bank [PDB] ID: 7DF4) and mutant (PDB ID: 7MJN) spike–ACE2 complexes^{31,32} were retrieved from RCSB (<http://www.rcsb.org>). All input structures before docking conducted to the

relaxation and repack process for backbone and sidechains clashes fixation, respectively. Furthermore, a local refinement protocol was done for local docking outputs to improve predicted quality poses. In parallel with docking, residual scanning analysis was performed using DynaMut server to investigate the effect of the mutation at position 501 on the stability and affinity of the spike for its receptor.³³ The highest score and best conformations of the spike-ACE2 complex obtained in docking were conducted to other analysis steps.

2.3 | MD simulation

Various laboratory techniques such as X-ray, nuclear magnetic resonance, and cryo-electron microscopy examine changes in the secondary and tertiary structure of proteins.^{25,34} Despite the high accuracy of these approaches, some disadvantages, such as complexity and high cost limit their use. In parallel with these techniques, computational biology has advanced significantly in recent years. Affordability, high speed, and flexibility are the main advantages of computational and in-silico methods.³⁵ One of the most popular techniques in the computational biology of biomolecules is MD simulation (also called computational molecular microscope) which enables investigators to determine changes at the atomic level.²⁵ In MD, the behavior of molecules over time is investigated using Newtonian equations.³⁶ Our group utilizes MD to investigate the effect of N501Y displacement on the structural characteristics of the spike. All MD simulation steps were performed using

GROMACS v2021 (CUDA compiled) on Linux operating system (Ubuntu v20.04 LTS).³⁷ CHARMM36 all-atom force field was implemented to parameterize the structures and generation of topologies.³⁸ To equilibrate systems, the Canonical ensemble (constant number, volume, temperature) was performed for temperature coupling at 310 K by V-rescale algorithm for 150 ps, as well as isothermal-isobaric ensemble (constant number, pressure, temperature) by Parrinello-Rahman algorithm for 500 ps was accomplished for pressure coupling at 1 bar.^{39,40} During equilibration, constraints were applied for h-bonds; furthermore, rculomb and rvdw were set to 1.2 nm. The TIP3P water model was used to solvation of systems and Na⁺ and Cl ions at a concentration of 0.15 M were added for neutralization. In the MD production step, all systems were simulated for 30 ns. Finally, structural terms such as root-mean-square deviation (RMSD), root-mean-square fluctuation (RMSF), solvent accessible surface area (SASA), the radius of gyration (Rg), free energy landscape (FEL), principal component analysis (PCA), secondary structure, hydrogen bonds, minimum distance, and the number of contacts were extracted from the output trajectory files.

TABLE 1 IEDB server outputs, as seen native spike exhibited a higher immunogenicity potency for immune cells (higher score indicates higher affinity)

Position	DiscoTope	ASP	MHC-I	MHC-II
498 ^W	1.715	0.574	N	0.484 N 69
498 ^M	0.196	0.561		
500 ^W	3.006	0.584		
500 ^M	1.737	0.558		
501 ^W	-0.797	0.579		
501 ^M	-1.935	0.550	M	0.483 M 69
502 ^W	1.204	0.585		
502 ^M	0.107	0.560		
505 ^W	-2.396	0.563		
505 ^M	-3.248	0.535		

Abbreviations: ASP, antigen sequence properties; IEDB, immune epitope database; M, mutant; MHC-I, major histocompatibility complex Class I; N, native.

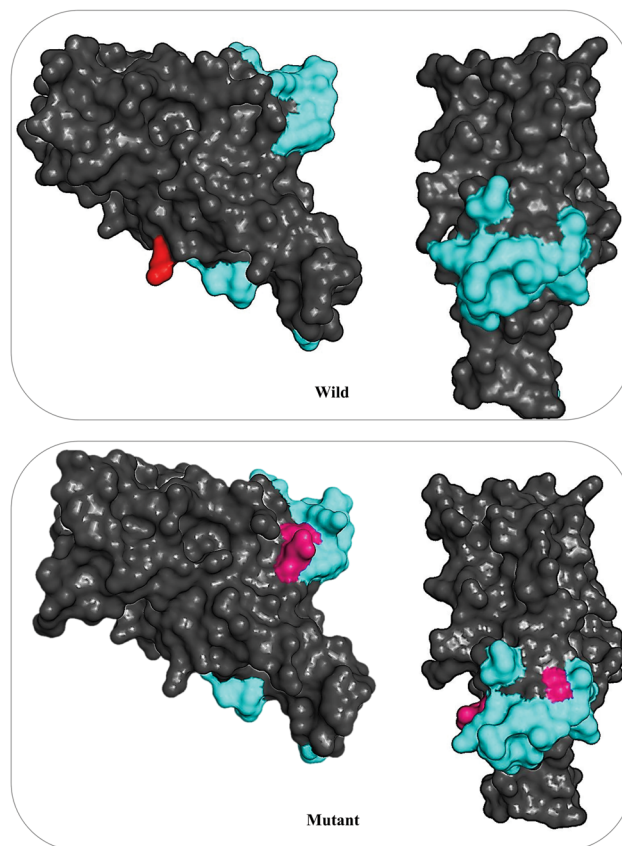


FIGURE 1 Structural epitopes examined using DiscoTope (same and different epitopes are shown in cyan and hot pink, respectively)

TABLE 2 Some energy terms reported by Rosetta (the unit of all energy terms is REU)

Spike	fa_atr	fa_elec	hbond_sr_bb	hbond_lr_bb	hbond_bb_sc	hbond_sc	Omega	Total
Native	-4950.7	-1242.1	-307.867	-87.296	-138.802	-104.010	82.678	-1945.57
Mutant	-4949.6	-1242.2	-307.503	-88.195	-141.318	-101.883	83.503	-1948.51

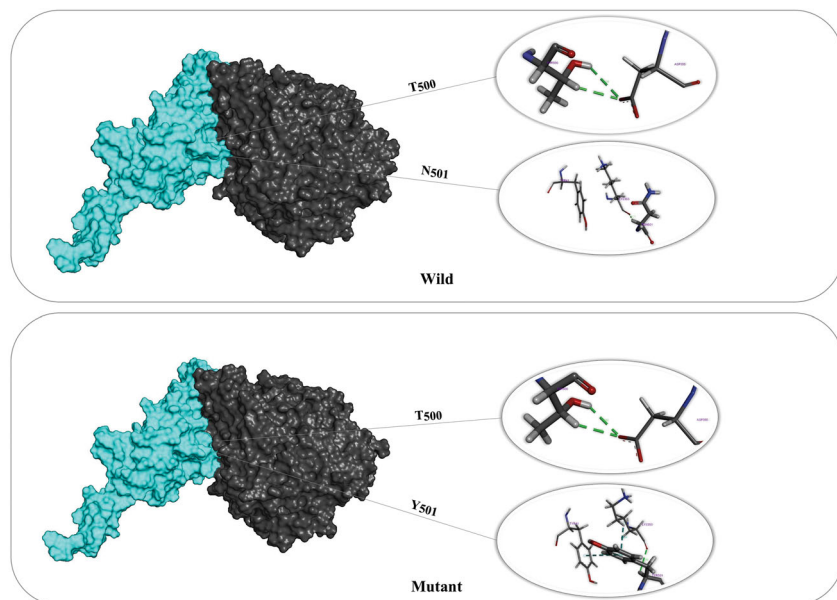


FIGURE 2 Rosetta docking outputs for native and N501Y spike variants

TABLE 3 As can be seen, all variables coupled to desired conditions

System	Temperature (K)	Pressure (bar)	Density (kg/m ³)
Native	309.99	1.2	1044.4
Mutant	310	0.99	1044

2.4 | Free energy computation

Investigating the affinity and stability, following mutations is a fundamental step in the study and design of proteins.⁶ One of the main ways to evaluate these biophysical parameters is free energy calculation. Although numerous algorithms including free energy perturbation and thermodynamic integration were developed for this purpose, low speed limits their use to small molecules studies. Another available method for calculating free energy is molecular mechanics Poisson–Boltzmann surface area (MM-PBSA), which has high speed and accuracy.⁴¹ This algorithm decomposes free energy terms into potential energy and solvation energy. The potential energy consists of both bonded and nonbonded interactions energies and is computed using MM functions. In addition, the solvation energy term is divided into polar solvation energy (electrostatic term) and nonpolar solvation

energy (nonelectrostatic term). To calculate the binding free energy of the native and N501Y spike–ACE2 complexes, *g_mmpbsa* which is equipped with the MM-PBSA algorithm was implemented.⁴² During free energy calculation, the SASA method was chosen for nonpolar solvation energy computation and *-incl_14* set at yes. Besides, we employed Rosetta (scoring protocol), FoldX,⁴³ and OSPREY (v3.2.2) for the structure extracted from the last frame of the simulation trajectory to computation structural energy of native and mutant spike in unbounded mode. OSPREY and FoldX are versatile tools for tracking mutation in proteins.^{44,45}

2.5 | Tertiary structure topology analysis

To rigorously investigate the impact of the N501Y mutation on the folding characteristics of spike protein, structural alignment along with three functions including the probability density function (PDF),⁴⁶ contact map,⁴⁷ and tunnel analysis⁴⁸ were applied for the residues around the mutated position. PDF was examined based on dihedral angle and triangle area using *G_Measure* (v.0.8); in addition, contact maps for carbon alpha and tunnel analysis were analyzed by utilizing *Pymod* (v3.0) and *Caver* (3.0.3), respectively.^{49,50}

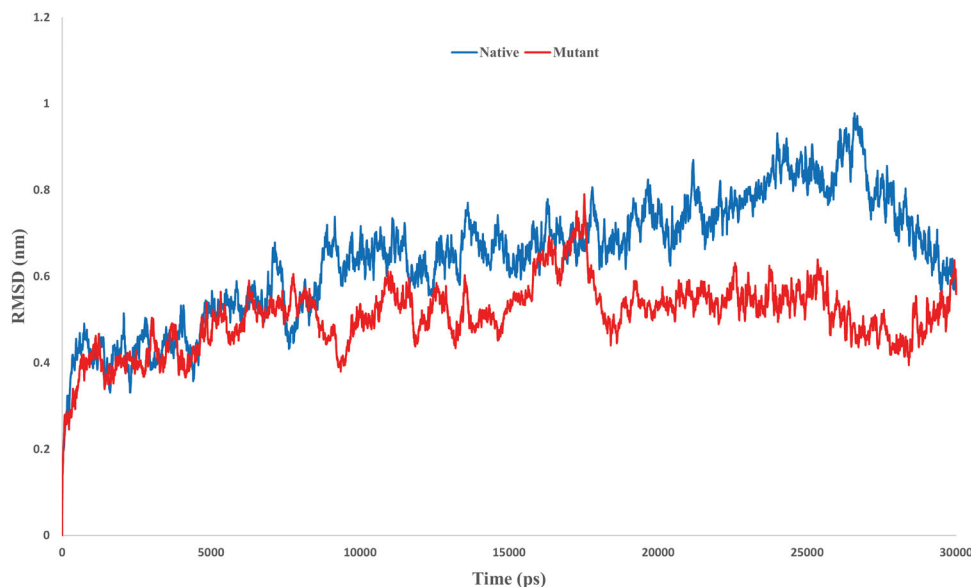


FIGURE 3 Root-mean-square deviation (RMSD) curve for the native and mutant spike, after 18 ns, the amount of conformational shift in both systems has been minimized

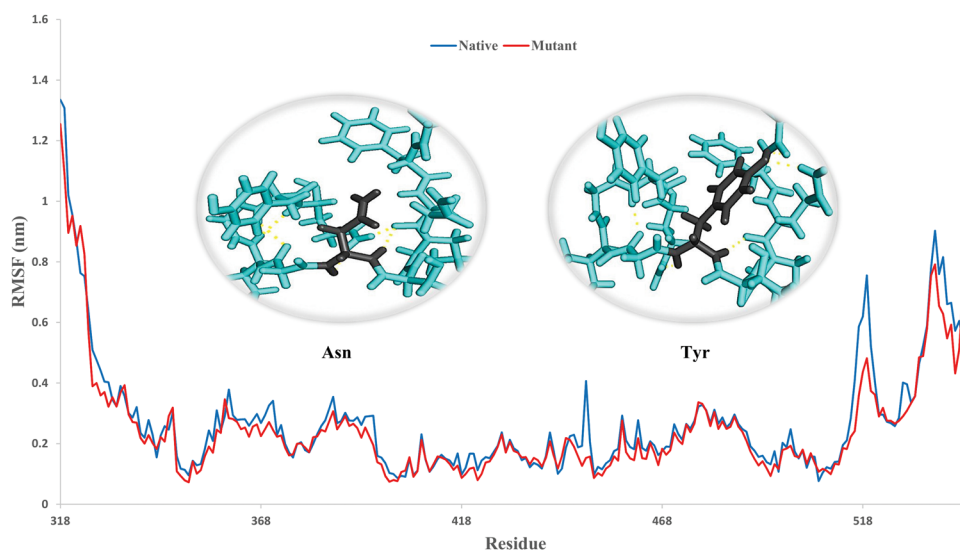


FIGURE 4 Root-mean-square fluctuation (RMSF) diagram, the higher number of polar interactions in the mutant compared to the native spike (4 vs. 2) reduces the structural fluctuations rate following the N501Y mutation

3 | RESULTS AND DISCUSSION

3.1 | Immunogenicity analysis

Insight into immunological characteristics of native and N501Y variants of spike comprehensively was obtained in two steps. First, MHC-I and MHC-II epitopes were predicted for the human leukocyte antigen allele reference set. It was observed that the N501Y spike has the less binding ability for MHC-1 compared to the native type. Then, linear and discontinuous epitopes for B-cells were anticipated. The

mutation at position 501 significantly reduced the immunogenicity of adjacent residues which forms the main epitope for B-cells (Table 1).

In addition to the higher binding affinity, the number of epitope regions in the native spike was higher than the N501Y variant (Figure 1). In the native form of a spike, 23 residues were identified as epitopes, while the value for the mutant variant was 21 residues. In previous studies,^{12,16,31} confirmation of lineages carrying the N501Y have higher viral load and resistance to antibody neutralization, the

observed decrease in immunogenicity can be attributed to the occurrence of this phenomenon.

3.2 | Docking and residue scanning

The pattern of residues interactions for spike-ACE2 complex in both native and N501Y variants was tracked using Rosetta docking protocol as a state-of-the-art, free, and open-source suite. Rosetta docking energy calculates using physics-statistics functions and is reported in the form of a Rosetta energy unit. For native spike, Asn at position 501 forms a hydrogen bond (2.71 Å) with Lys₃₅₃. Furthermore, Thr₅₀₀ interacts with residue Asp₃₅₅ with

two hydrogen bonds (2.61 and 3.06 Å) versus in the N501Y variant, Tyr at position 501 is able to form one hydrogen bond with Lys₃₅₃ (2.93 Å), while its aromatic ring forms two hydrophobic bonds with Tyr₄₁ (5.13 Å) and Lys₃₅₃ (3.94 Å) ACE2. In addition, Thr₅₀₀, like the native spike, forms two hydrogen bonds with Asp₃₅₅, but with a shorter distance (both are 2.63 Å). Rosetta energies for native and N501Y variants were -1945.63 and -1948.516, respectively. Indeed, following the N501Y mutation affinity of spike protein increases for its receptor (Table 2).

In addition to the distance and number of interactions between the residues in the spike interface with its receptor, another factor that increases the binding

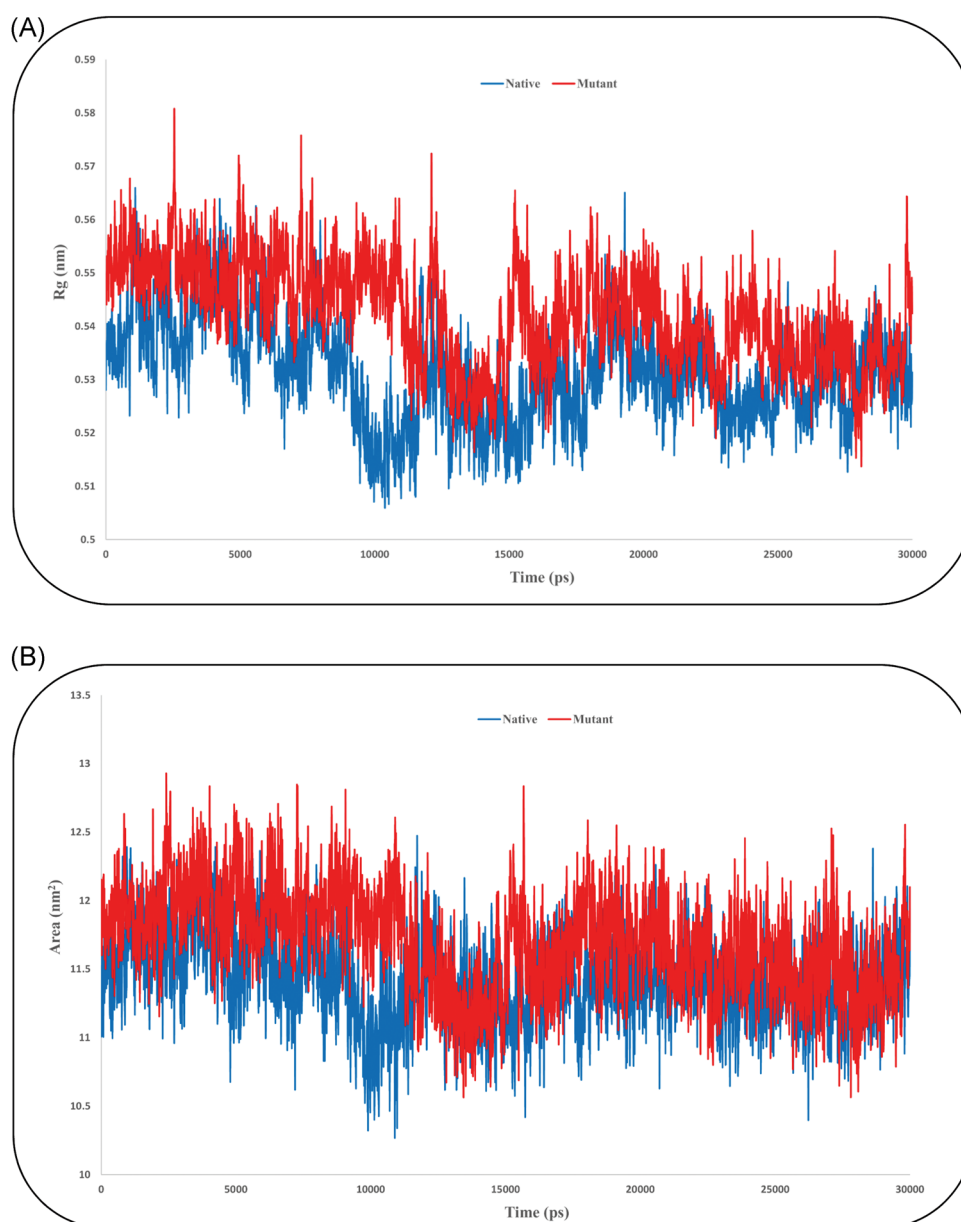


FIGURE 5 (A) Radius of gyration (Rg) and (B) solvent accessible surface area diagram for both native and mutant spike

potency in the N501Y is the shape of the Tyr side chain. Unlike Asn₅₀₁ which is adjacent to Lys₃₅₃, the side chain and aromatic ring of the Tyr₅₀₁ are surrounded by Y₄₁ and Lys₃₅₃ (Figure 2). This characteristic leads to the better formation of the binding pocket of the N501Y SARS-CoV-2 lineage. Residue scanning calculations of DynaMut also confirm the docking results and show that the stability of the spike-ACE2 complex improves following the N501Y mutation (-0.522 kcal/mol).

3.3 | MD investigations

Investigation of system equilibration in terms of thermodynamic parameters is an essential step before analyzing the MD outputs. For this purpose, some thermodynamics parameters such as pressure, temperature, and density were extracted for the studied systems. Examination of the mentioned parameters indicated that the target systems were equilibrated (Table 3).

To accurately assess the equilibration and to make sure that MD time was sufficient, RMSD curves were extracted for the target systems. We found that during the simulation, this structural parameter also reached equilibrium (Figure 3).

To investigate the role of amino acids in the rate of structural changes, the RMSF factor that indicates residue fluctuations was assigned. As can be seen in Figure 4, the

number of structural fluctuations following the mutation at position 501 decreases, this is consistent with the results of RMSD output. Replacing Asn with Tyr reduces deviations and variations in the RBD domain of spike. The occurrence of this phenomenon can be related to the interaction potency of Tyr. Since Tyr compared to Asn has a higher ability to interact with adjacent residues in the RBD domain; thus, it applies less fluctuation to the spike structure and can improve the stability of this fragment of the spike. The mean fluctuations in the native and mutant structures were 0.28 and 0.26 nm, respectively.

The effect of the N501Y mutation on spike folding was tracked by evaluating the Rg and SASA functions. At first, the Rg that referred to the protein compactness was measured in both native and mutant species. Rg represents the average distance of the atoms that make up a molecule from its center. The average Rg was 0.53 nm in the native and 0.54 nm in the mutant. This finding indicates a reduction in the compactness of the spike protein following mutation. In addition, the average of SASA for native and mutant spike were 11.4 and 11.7 nm², respectively (Figure 5). The results of the Rg and SASA investigations were highly consistent. This signifies that placement Tyr at position 501 has a great impact on the folding of adjacent areas. The increase in accessibility and interaction interface observed in mutant variants can be one of the main factors in promoting spike affinity to ACE2 receptors.

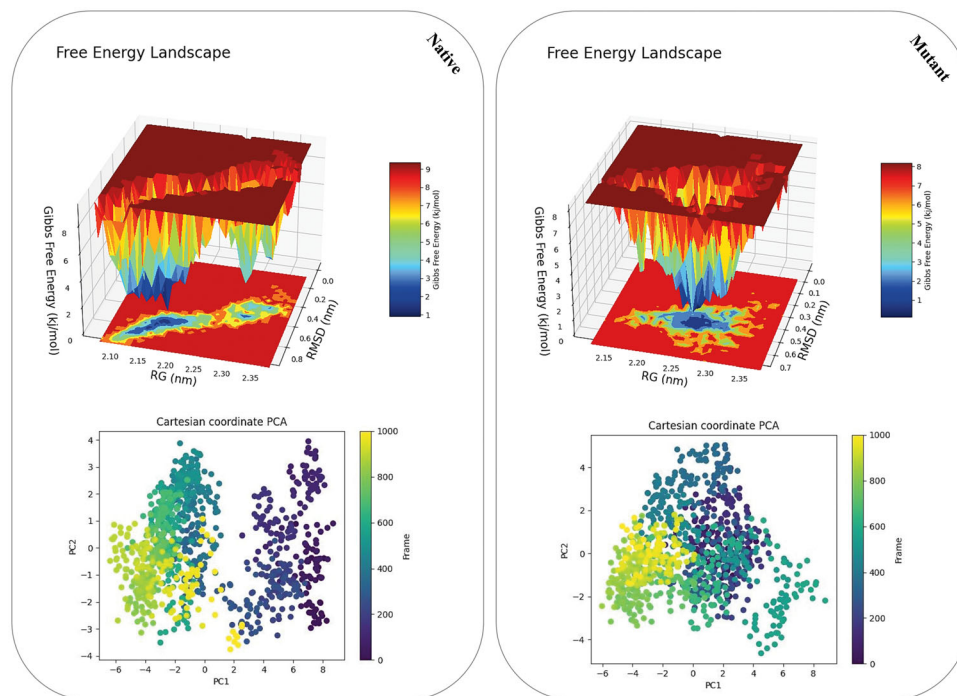


FIGURE 6 Free energy landscape and principal component analysis (PCA) diagram. Rg, radius of gyration; RMSD, root-mean-square deviation

In the next step, FEL and PCA functions were extracted to track the effect of the N501Y mutation on the folding and the conformational energy patterns of the target structures (Figure 6). Following the mutation, spike folding significantly affects the conformational energy patterns in the mutant species, where it was more concentrated compared to the native form. These results confirm the RMSD, RMSF,

and Rg outputs and indicate that the flexibility and structural variation decrease following N501Y mutation.

Analysis of secondary structure elements using the defined secondary structure of proteins module of GROMACS (residues 490–510) was another examined parameter. We observed that the N501Y displacement could affect the composition of the secondary structural

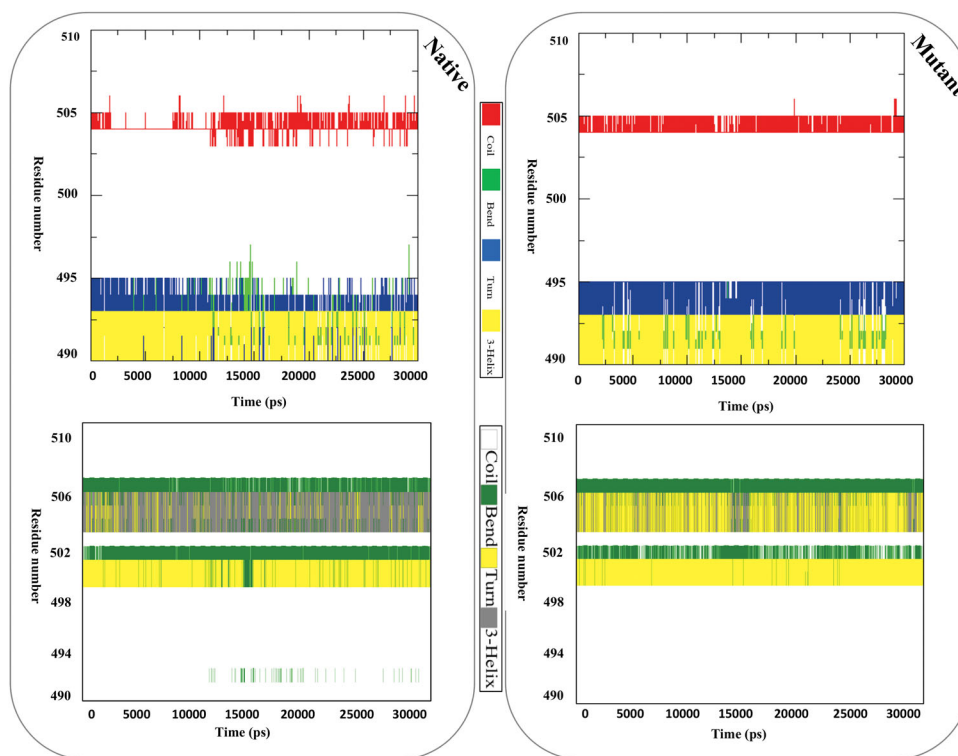


FIGURE 7 Schematic representation of secondary structure for the native and mutant spike

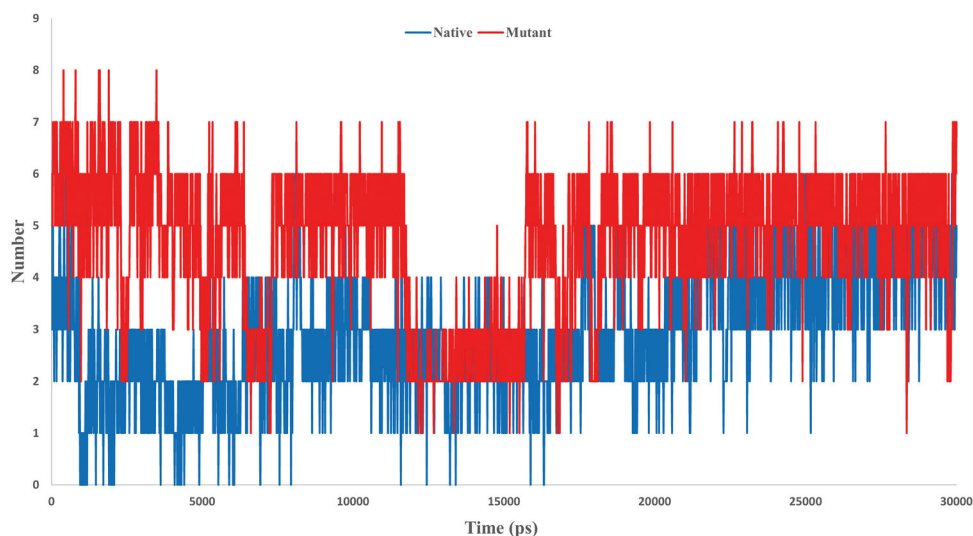


FIGURE 8 Hydrogen bonds state throughout molecular dynamics simulation

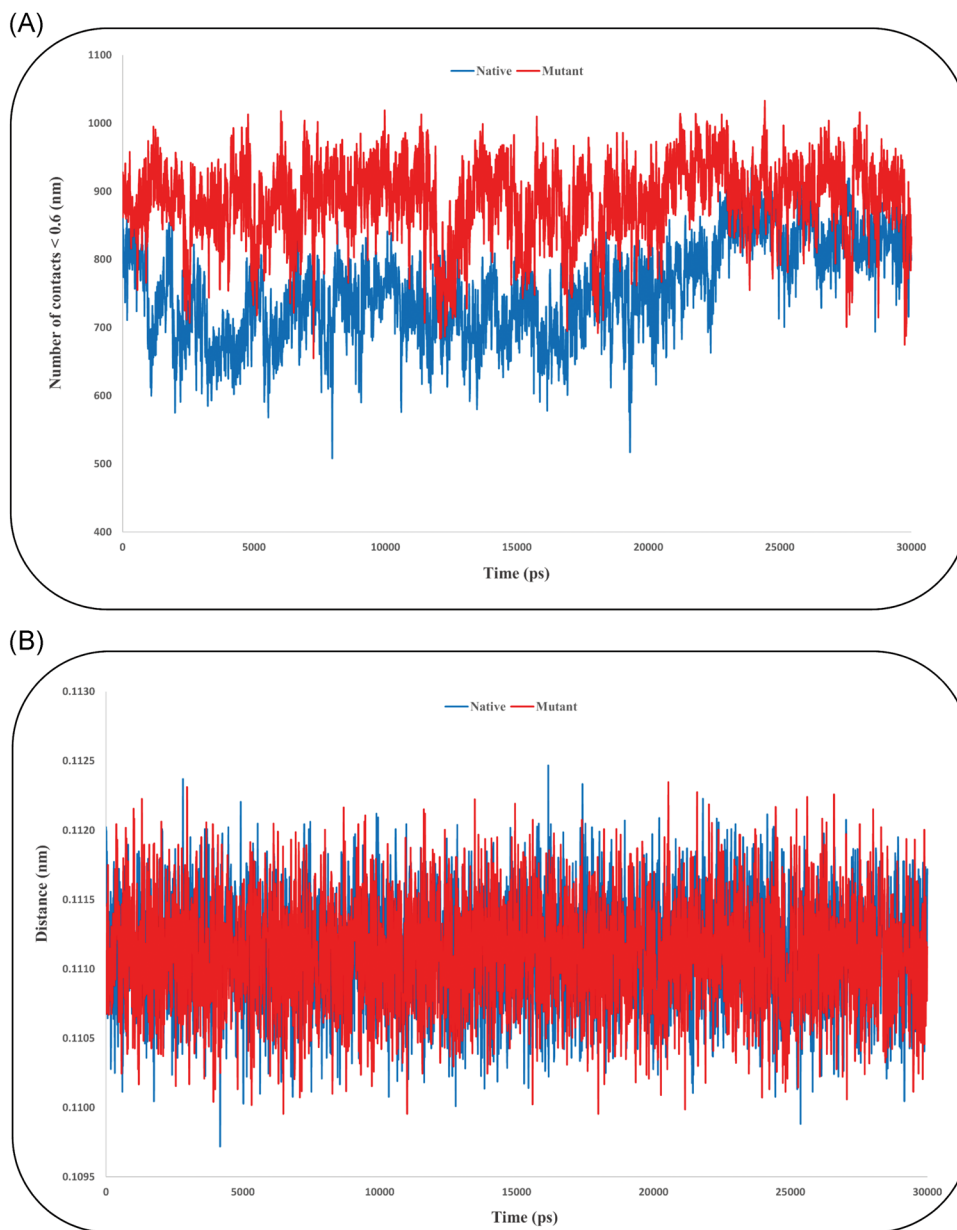


FIGURE 9 (A) Number of contacts and (B) minimum distance for interface residues in spike–angiotensin-converting enzyme 2 complexes

components of its adjacent amino acids. Compared to the native structure, in the mutated spike amount of bend and 3-helix were decreased, while the coil elements frequency had increased (Figure 7).

In the final step, MD trajectory files were examined to track the effect of N501Y mutation on the affinity of spike for its receptor using functions such as a number of hydrogen bonds, the distance between interface residues, and the number of contacts at distance less than 0.6 nm for the spike–ACE2 complexes. Hydrogen bond analysis revealed that minimum, maximum, and the average of hydrogen bonds for complex of native spike with ACE2 were 0, 6, and

2.95, respectively. However, these values for mutant spike were 1, 8, and 4.7 (Figure 8).

Evaluation of distance between Asn and Tyr at position 501 with interface residues of ACE2 showed that the mean distance in mutant species (0.111 nm) was less than the native (0.112 nm), which could be evidence for greater ability of spike to ACE2 cell surface receptor. Finally, contacts at distance less than 0.6 nm for native and mutant complexes were investigated and the results were significantly consistent with other observations. We found that the number of contacts for mutant complex (886) is higher than native (755) lineage (Figure 9). These results along with the output of hydrogen bonds investigation confirm the higher

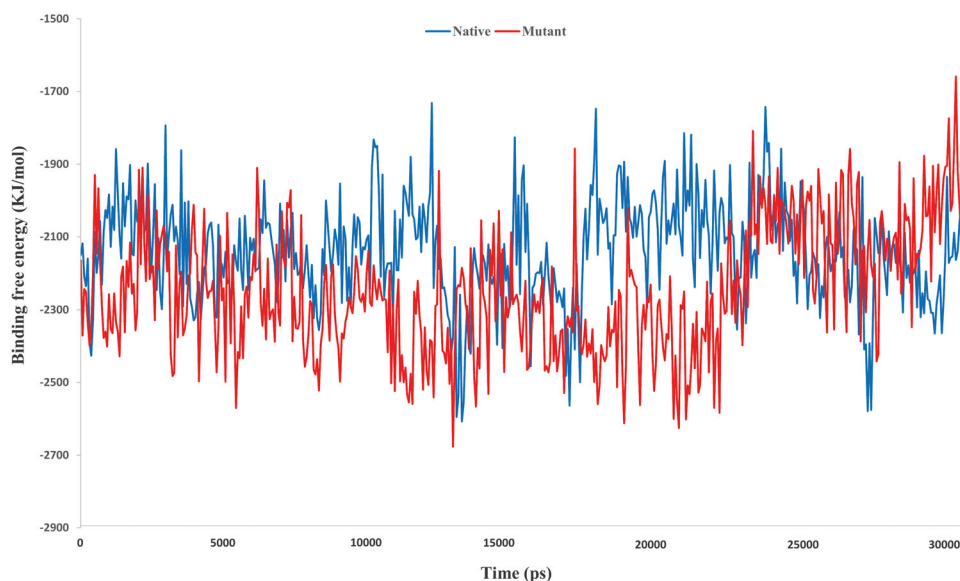


FIGURE 10 Schematic representation of the binding energy for native and N501Y carrying complexes

Complex	VdW energy	ES energy	PS energy	SASA energy	Binding energy
Native	-283.415	-2353.393	533.526	-37.414	-2140.696
Mutant	-313.698	-2487.918	593.290	-39.733	-2248.058

TABLE 4 Binding free energy decomposition for spike-ACE2 complexes

Abbreviations: ACE2, angiotensin-converting enzyme 2; ES, electrostatic energy; PS, polar solvation energy; SASA, solvent accessible surface area; VdW, Vander walls energy.

binding affinity in mutant spike for ACE2 compared to native form.

Regarding the results observed during MD simulation, it can be claimed that the N501Y mutation has a significant effect on the RBD domain of SARS-CoV-2 spike protein and can increase the spike stability and its affinity to ACE2.

3.4 | Free energy analysis

Energy calculation was implemented for two aims. First, the affinity of native and mutant spike for ACE2 was investigated using binding free energy approximation. The findings of this step confirmed the results of docking and MD. We observed that the binding affinity for the ACE2 increases following the replacement of Asn with Tyr (Figure 10).

Further examination was done by decomposition of binding free energy. We found that, with the exception of polar solvent energy, the other constituent functions of the binding energy were more optimal for the complex containing the N501Y mutation (Table 4).

Furthermore, per residue binding free energy decomposition was carried out for residues 41, 353, and 355 ACE2 and residues 500–502 spike. As expected, compared to Asn₅₀₁, placement of Tyr imposes much higher binding energy to the spike for binding to ACE2 (Figure 11).

After studying the binding free energy, to accurately examine the effect of the mutation at position 501 on the stability of the spike protein, we investigated the folding energy of this protein in the unbounded form. For this purpose, FoldX, Rosetta, and OSPREY were employed. Examination of the results of FoldX, Rosetta, and OSPREY showed that the N501Y mutation increases the stability of the spike protein (Table 5). The outputs of the structural energy investigation agreed with the results of other analysis steps and indicate the positive effect of the N501Y on the stability and affinity of the spike to its receptor.

3.5 | Tertiary structure topology analysis

During structural alignment, we observed a structural shift (2.2 nm) in the structure of the mutant spike

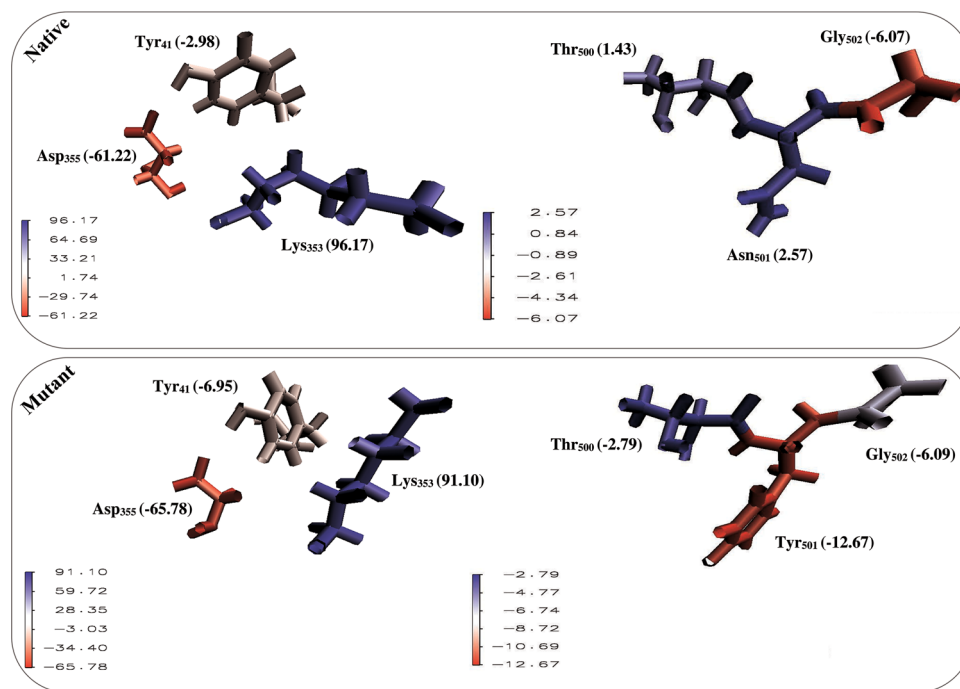


FIGURE 11 Binding free energy (kJ/mol) of position 501 and surrounding amino acids for spike and their interface residues in angiotensin-converting enzyme 2

TABLE 5 Stability assessment of spike in native and mutant forms in terms of folding energy

Lineage	Tools		
	FoldX _(kJ/mol)	Rosetta _(REU)	OSPREY _(kJ/mol)
Native	-117.78	-382.720	-24.76
Mutant	-141.9	-402.669	-38.56

Abbreviation: REU, Rosetta energy unit.

compared to the native structure, which is a reason for the change in the binding affinity of the spike to interaction with ACE2 (Figure 12).

PDF analysis consistent with results of MD and structural alignments revealed an extensive change in the folding of the spike. We found that the pattern of dihedral angles and triangles area changes following the N501Y displacement. Furthermore, analysis of the contact map for carbon alpha at position 501 showed that the arrangement of residues around this position undergoes extensive changes, in the mutant structure Tyr₅₀₁ interacts with residues 497–506, while in the native spike, Asn₅₀₁ interacts with residues 498–506 (excluding position 504). More stable interaction of position 501 with surrounding residues in mutant spike maybe reduce the structural fluctuations and increase its stability. Besides, the study of the tunnels around the mutated position revealed there is more space in the mutant spike that can facilitate the interaction of spike with ACE2

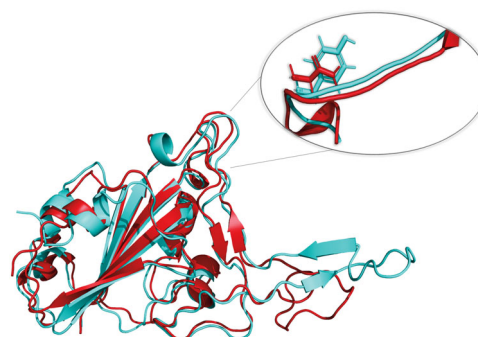


FIGURE 12 Structural alignment of native (red) and mutant (cyan) spike

(Figure 13). The Caver results showed that the length and radius of the available tunnel around the Tyr₅₀₁ (2 and 1.2 Å) are greater than the Asn₅₀₁ (1.9 and 1.1 Å).

4 | CONCLUSION

One of the main challenges in detecting and controlling SARS-CoV-2 is mutation occurring in functional regions of the virus genome, such as the coding sequence of the spike protein. So far, many mutations (E484K, Q493N, and N501T) have been identified in the RBD domain of this protein. Given that this protein is the main target in

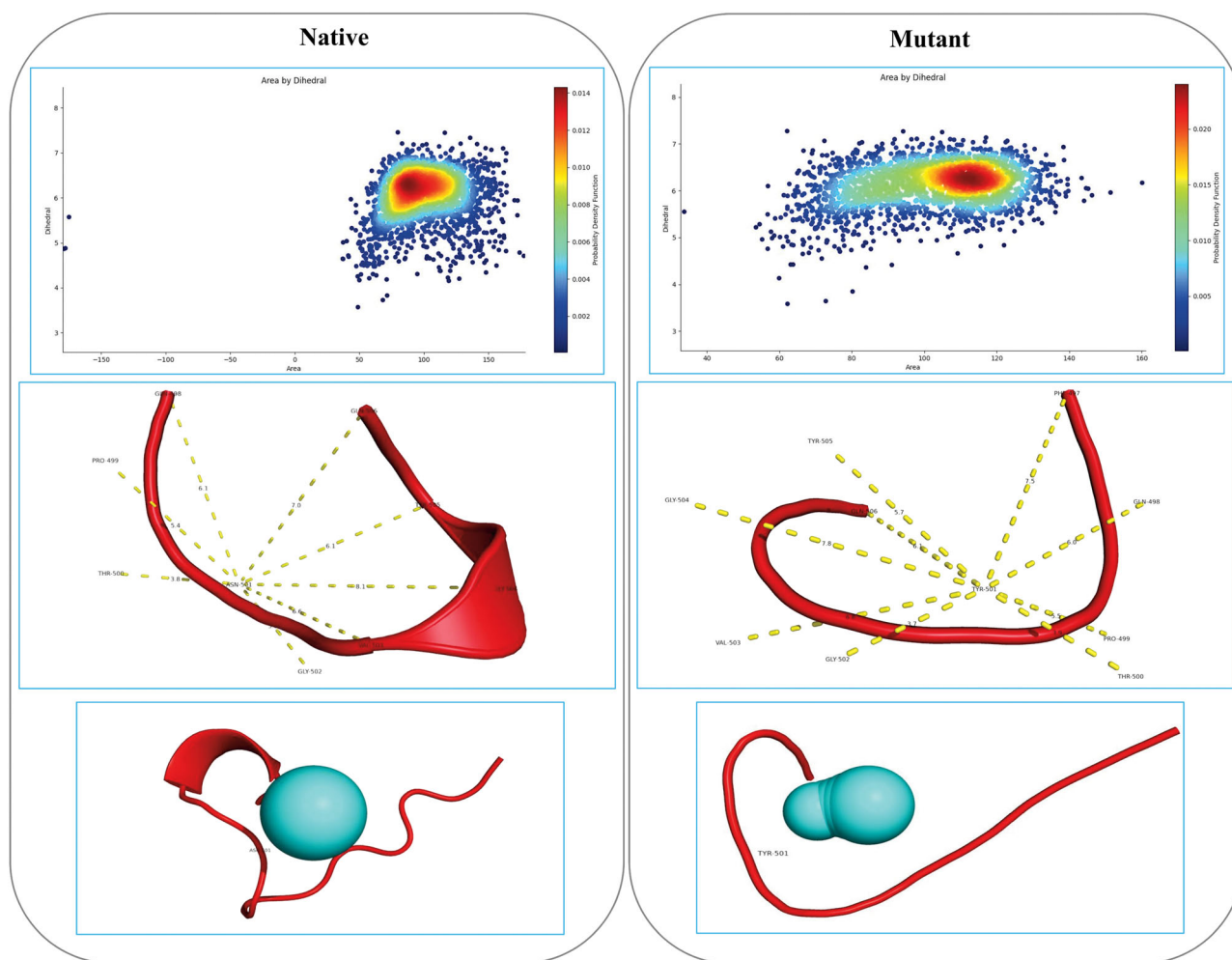


FIGURE 13 Representation of a probability density function, contact map, and tunnel analysis outputs for desired structures

vaccination against COVID-19, tracking the effect of these mutations, and changing the function of the spike is very valuable. Our team in this current study uses various computational biology approaches examining the significant effects on the function of the spike molecular behavior due to the N501Y mutation. We discovered that this mutation reduces immunogenicity, increases stability, and affinity for the spike receptor.

Since this mutation can facilitate the potency of the virus entering the cell and exacerbate its clinical symptoms. Due to the N501Y mutation effect on the immunogenicity of the virus induces immune responses, we recommend that more attention be paid to the identification of SARS-CoV-2 lineage during the admittance of COVID-19 patients. Having efficient diagnostic and therapeutic protocols implemented for patients carrying this mutation compared to the reference lineage (Wuhan strain).

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

AUTHOR CONTRIBUTIONS

Methodology and software: Neda Rostami. *Analyzed data and co-wrote the paper:* Edris Choupani. *Software and co-wrote the paper:* Yaeren Hernandez. *Performed bioinformatic analyses:* Seyed S. Arab. *Conceptualization:* Seyed M. Jazayeri. *Designed experiment, data analysis, and conceptualization:* Mohammad M. Gomari.

ORCID

Mohammad M. Gomari  <http://orcid.org/0000-0003-4143-2208>

REFERENCES

- Pagliari F, Marafioti MG, Genard G, et al. ssRNA virus and host lipid rearrangements: is there a role for lipid droplets in SARS-CoV-2 infection? *Front Mol Biosci.* 2020;7:578964.
- Sedova M, Jaroszewski L, Alisoltani A, Godzik A. Coronavirus3D: 3D structural visualization of COVID-19 genomic divergence. *Bioinformatics.* 2020;36(15):4360-4362. doi:10.1093/bioinformatics/btaa550

3. Hosseini SA, Zahedipour F, Mirzaei H, Oskuee RK. Potential SARS-CoV-2 vaccines: concept, progress, and challenges. *Int Immunopharmacol*. 2021;97:107622.
4. Verkhivker GM, Di Paola L. Dynamic network modeling of allosteric interactions and communication pathways in the SARS-CoV-2 spike trimer mutants: differential modulation of conformational landscapes and signal transmission via cascades of regulatory switches. *J Phys Chem B*. 2021;125(3):850-873. doi:10.1021/acs.jpcc.0c10637
5. Gomari MM, Farsimadan M, Rostami N, et al. CD44 polymorphisms and its variants, as an inconsistent marker in cancer investigations. *Mutation Res/Rev Mutation Res*. 2021;787:108374. doi:10.1016/j.mrrev.2021.108374
6. Harris H. Polymorphism and protein evolution. The neutral mutation-random drift hypothesis. *J Med Genet*. 1971;8(4):444-452. doi:10.1136/jmg.8.4.444
7. Siahpoosh Z, Farsimadan M, Pazhohan M, Vaziri H, Mahmoudi Gomari M. KISS1R polymorphism rs58777844 (Tyr313His) is linked to female infertility. *Br J Biomed Sci*. 2021;78(2):98-100. doi:10.1080/09674845.2020.1856496
8. Fang J. A critical review of five machine learning-based algorithms for predicting protein stability changes upon mutation. *Brief Bioinform*. 2020;21(4):1285-1292. doi:10.1093/bib/bbz071
9. Starr TN, Greaney AJ, Hilton SK, et al. Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell*. 2020;182(5):1295-1310.e1220. doi:10.1016/j.cell.2020.08.012
10. Mahmoudi Gomari M, Rostami N, Omidi-Ardali H, Arab SS. Insight into molecular characteristics of SARS-CoV-2 spike protein following D614G point mutation, a molecular dynamics study. *J Biomol Struct Dyn*. 2021;39:1-9. doi:10.1080/07391102.2021.1872418
11. Gobeil SM, Janowska K, McDowell S, et al. Effect of natural mutations of SARS-CoV-2 on spike structure, conformation and antigenicity. *bioRxiv*. 2021. doi:10.1101/2021.03.11.435037
12. Frampton D, Rampling T, Cross A, et al. Genomic characteristics and clinical effect of the emergent SARS-CoV-2 B.1.1.7 lineage in London, UK: a whole-genome sequencing and hospital-based cohort study. *Lancet Infect Dis*. 2021;21:1246-1256. doi:10.1016/s1473-3099(21)00170-5
13. Hoffmann M, Arora P, Groß R, et al. SARS-CoV-2 variants B.1.351 and P.1 escape from neutralizing antibodies. *Cell*. 2021;184:2384-2393. doi:10.1016/j.cell.2021.03.036
14. Garcia-Beltran WF, Lam EC, St. Denis K, et al. Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. *Cell*. 2021;184:2372-2383. doi:10.1016/j.cell.2021.03.013
15. Goncalves Cabecinhas AR, Roloff T, Stange M, et al. SARS-CoV-2 N501Y introductions and transmissions in Switzerland from beginning of October 2020 to February 2021—implementation of Swiss-wide diagnostic screening and whole genome sequencing. *Microorganisms*. 2021;9(4):677. doi:10.3390/microorganisms9040677
16. Kuzmina A, Khalaila Y, Voloshin O, et al. SARS-CoV-2 spike variants exhibit differential infectivity and neutralization resistance to convalescent or post-vaccination sera. *Cell Host Microbe*. 2021;29(4):522-528. doi:10.1016/j.chom.2021.03.008
17. Liu Y, Liu J, Plante KS, et al. The N501Y spike substitution enhances SARS-CoV-2 transmission. *bioRxiv*. 2021. doi:10.1101/2021.03.08.434499
18. Mahmoudi Gomari M, Rostami N, Ghodrati A, et al. Implementation of docking, molecular dynamics and free energy to investigate drug potency of novel BCR-ABL315I inhibitors as an alternative to ponatinib. *Comput Toxicol*. 2021;20:100180. doi:10.1016/j.comtox.2021.100180
19. Sefid F, Baghban R, Payandeh Z, Khalesi B, Mahmoudi Gomari M. Structure evaluation of iron for designing a vaccine against *Escherichia Coli*, an in silico approach. *J Med Microbiol Infect Dis*. 2019;7(4):93-106. doi:10.29252/JoMMID.7.4.93
20. Dhanda SK, Mahajan S, Paul S, et al. IEDB-AR: immune epitope database-analysis resource in 2019. *Nucleic Acids Res*. 2019;47(W1):W502-w506. doi:10.1093/nar/gkz452
21. Goumari MM, Farhani I, Nezafat N, Mahmoodi S. Multi-epitope vaccines (MEVs), as a novel strategy against infectious diseases. *Curr Proteomics*. 2020;17(5):354-364.
22. UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res*. 2021;49(D1):D480-D489. doi:10.1093/nar/gkaa1100
23. Trolle T, Metushi IG, Greenbaum JA, et al. Automated benchmarking of peptide-MHC class I binding predictions. *Bioinformatics*. 2015;31(13):2174-2181. doi:10.1093/bioinformatics/btv123
24. Habibi M, Bakhshi PK, Aghdam R. LRC: a new algorithm for prediction of conformational B-cell epitopes using statistical approach and clustering method. *J Immunol Methods*. 2015;427:51-57. doi:10.1016/j.jim.2015.09.006
25. Jafari D, Malih S, Gomari MM, Safari M, Jafari R, Farajollahi MM. Designing a chimeric subunit vaccine for influenza virus, based on HA2, M2e and CTxB: a bioinformatics study. *BMC Mol Cell Biol*. 2020;21(1):89. doi:10.1186/s12860-020-00334-6
26. Jiao X, Jin X, Ma Y, et al. A comprehensive application: molecular docking and network pharmacology for the prediction of bioactive constituents and elucidation of mechanisms of action in component-based Chinese medicine. *Comput Biol Chem*. 2021;90:107402. doi:10.1016/j.compbiolchem.2020.107402
27. Xu X, Zou X. MDockPeP: A web server for blind prediction of protein-peptide complex structures. *Methods Mol Biol*. 2020;2165:259-272. doi:10.1007/978-1-0716-0708-4_15
28. Koukos PI, Roel-Touris J, Ambrosetti F, et al. An overview of data-driven HADDOCK strategies in CAPRI rounds 38-45. *Proteins*. 2020;88(8):1029-1036. doi:10.1002/prot.25869
29. Mavromoustakos T, Kellici TF. *Rational Drug Design*. Springer; 2018.
30. Leman JK, Weitzner BD, Lewis SM, et al. Macromolecular modeling and design in Rosetta: recent methods and frameworks. *Nat Methods*. 2020;17(7):665-680. doi:10.1038/s41592-020-0848-2
31. Supasa P, Zhou D, Dejnirattisai W, et al. Reduced neutralization of SARS-CoV-2 B.1.1.7 variant by convalescent and vaccine sera. *Cell*. 2021;184(8):2201-2211. doi:10.1016/j.cell.2021.02.033
32. Zhu, X, Mannar D, Srivastava SS, et al. Cryo-electron microscopy structures of the N501Y SARS-CoV-2 spike protein in complex with ACE2 and 2 potent neutralizing

- antibodies. *PLoS Biol.* 2021;19(4):e3001237. doi:10.1371/journal.pbio.3001237
33. Rodrigues CH, Pires DE, Ascher DB. DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. *Nucleic Acids Res.* 2018;46(W1):W350-W355. doi:10.1093/nar/gky300
34. Rutsdottir G, Härmark J, Weide Y, et al. Structural model of dodecameric heat-shock protein Hsp21: flexible N-terminal arms interact with client proteins while C-terminal tails maintain the dodecamer and chaperone activity. *J Biol Chem.* 2017;292(19):8103-8121. doi:10.1074/jbc.M116.766816
35. Müller J, Siemann-Herzberg M, Takors R. Modeling cell-free protein synthesis systems—approaches and applications. *Front Bioeng Biotechnol.* 2020;8:584178. doi:10.3389/fbioe.2020.584178
36. Konermann L, Metwally H, McAllister RG, Popa V. How to run molecular dynamics simulations on electrospray droplets and gas phase proteins: basic guidelines and selected applications. *Methods.* 2018;144:104-112. doi:10.1016/j.ymeth.2018.04.010
37. Nayeem SM, Sohail EM, Sudhir GP, Reddy MS. Computational and theoretical exploration for clinical suitability of Remdesivir drug to SARS-CoV-2. *Eur J Pharmacol.* 2021;890:173642. doi:10.1016/j.ejphar.2020.173642
38. Huang J, MacKerell AD Jr. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J Comput Chem.* 2013;34(25):2135-2145. doi:10.1002/jcc.23354
39. Karwasra R, Fatihi S, Raza K, et al. Filgrastim loading in PLGA and SLN nanoparticulate system: a bioinformatics approach. *Drug Dev Ind Pharm.* 2020;46(8):1354-1361. doi:10.1080/03639045.2020.1788071
40. Martonák R, Laio A, Parrinello M. Predicting crystal structures: the Parrinello-Rahman method revisited. *Phys Rev Lett.* 2003;90(7):075503. doi:10.1103/PhysRevLett.90.075503
41. Botello-Smith WM, Luo R. Applications of MMPBSA to membrane proteins I: efficient numerical solutions of periodic Poisson-Boltzmann equation. *J Chem Inf Model.* 2015;55(10):2187-2199. doi:10.1021/acs.jcim.5b00341
42. Kumari R, Kumar R, Lynn A. g_mmpbsa—a GROMACS tool for high-throughput MM-PBSA calculations. *J Chem Inf Model.* 2014;54(7):1951-1962. doi:10.1021/ci500020m
43. Buß O, Rudat J, Ochsenreither K. FoldX as protein engineering tool: better than random based approaches? *Comput Struct Biotechnol J.* 2018;16:25-33. doi:10.1016/j.csbj.2018.01.002
44. Hallen MA, Martin JW, Ojewole A, et al. OSPREY 3.0: open-source protein redesign for you, with powerful new features. *J Comput Chem.* 2018;39(30):2494-2507. doi:10.1002/jcc.25522
45. Ojewole A, Lowegard A, Gainza P, et al. OSPREY predicts resistance mutations using positive and negative computational protein design. *Methods Mol Biol.* 2017;1529:291-306. doi:10.1007/978-1-4939-6637-0_15
46. Oya M, Taniguchi Y, Fujimura N, Miyamoto K, Oya M. Kinetic analysis of hemoglobin detergency by probability density functional method. *PLoS One.* 2020;15(8):e0237255. doi:10.1371/journal.pone.0237255
47. Xie J, Ding W, Chen L, Guo Q, Zhang W. Advances in protein contact map prediction based on machine learning. *Med Chem.* 2015;11(3):265-270. doi:10.2174/1573406411666141230095427
48. Brezovsky J, Kozlikova B, Damborsky J. Computational analysis of protein tunnels and channels. *Methods Mol Biol.* 2018;1685:25-42. doi:10.1007/978-1-4939-7366-8_3
49. Janson G, Paiardini A. PyMod 3: a complete suite for structural bioinformatics in PyMOL. *Bioinformatics.* 2021;37(10):1471-1472. doi:10.1093/bioinformatics/btaa849
50. Chovancova E, Pavelka A, Benes P, et al. CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. *PLoS Comput Biol.* 2012;8(10):e1002708. doi:10.1371/journal.pcbi.1002708

How to cite this article: Rostami N, Choupani E, Hernandez Y, Arab SS, Jazayeri SM, Gomari MM. SARS-CoV-2 spike evolutionary behaviors; simulation of N501Y mutation outcomes in terms of immunogenicity and structural characteristic. *J Cell Biochem.* 2022;123:417-430. doi:10.1002/jcb.30181