

Brazilian SARS-CoV-2 Variants: Computational Approaches To Structural Insights Into RBD Mutations And Their Interactions With ACE2

Isabelle Pereira

Federal University of ABC (UFABC)

Yolanda Marcello

Federal University of ABC (UFABC)

Glauco Lima

Federal University of ABC (UFABC)

Ana Scott

analigiascott@gmail.com

Federal University of ABC (UFABC)

Research Article

Keywords: Elastic Network Models, Molecular Dynamics, Unsupervised Machine Learning, RBD, Spike, SARS-CoV-2

Posted Date: September 25th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10997331/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Mutations in the SARS-CoV-2 Spike Receptor-Binding Domain (RBD) may influence viral infectivity and transmissibility by modulating interactions with the human Angiotensin-Converting Enzyme 2 (ACE2) receptor. This study examined the impact of RBD point mutations by integrating Molecular Dynamics (MD) simulations, Normal Mode Analysis (NMA), and Machine Learning (ML). From the Global Initiative on Sharing All Influenza Data (GISAID), SARS-CoV-2 lineages circulating in Brazil between January 2023 and May 2024 were identified, and the prevalent RBD mutations, along with those reported in the literature, were selected for analysis. Mutant models were prepared using CHARMM-GUI and simulated for 50 ns with Nanoscale Molecular Dynamics (NAMD). Trajectories were analyzed in Visual Molecular Dynamics (VMD) through Root Mean Square Deviation, hydrogen bonds, residue-residue distances, and intermolecular contact persistence. NMA based on the Elastic Network Model, evaluated Root Mean Square Fluctuations and collective motions, while ML identified patterns across mutants. The analyses suggested mutation-specific differences in conformational dynamics and interactions. While some mutations increased structural flexibility, others maintained behaviors comparable to the wild type, with most preserving coordinated global motions. Observed differences were associated with compensatory structural adjustments, redistribution of contacts, and modulation of local flexibility. These findings suggest that, initially, RBD mutations may fine-tune the balance between rigidity and flexibility, enabling conformational adaptations that are potentially relevant to receptor engagement and membrane fusion. The results support the hypothesis that structural plasticity and transient conformational states are important for the Spike protein function while preserving global motion of the RBD-ACE2 complex.

1. INTRODUCTION

A novel respiratory disease emerged in Wuhan, China, in December 2019, and, by March 2020, the disease, then called Coronavirus Disease 2019 (COVID-19), was declared a pandemic (Flores-Vega et al. 2022; Haque et al. 2020; Hu et al. 2021). The disease is caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), a betacoronavirus from the Coronaviridae family, related to SARS-CoV (2002) and MERS-CoV (2012), both associated with severe respiratory infections (Jackson et al. 2022). It shares approximately 80% of its genome with SARS-CoV (Yan et al. 2020).

SARS-CoV-2 is a single-stranded RNA virus with a large genome that encodes two large polyproteins, four structural proteins, and nine accessory proteins (Hatmal et al. 2020; Jackson et al. 2022; Peng et al. 2021). The genome is organized in the 5'-3' order, starting with replicase (ORF1a/ORF1b), followed by the structural proteins Spike (S), Envelope (E), Membrane (M), and the Nucleocapsid (N), which is located inside the viral particle along with the RNA (Haque et al. 2020; Hatmal et al. 2020; Hu et al. 2021; Flores-Vega et al. 2022; Lomoio et al. 2023). Transmission occurs primarily through close contact with infected individuals via respiratory droplets and aerosols released during coughing, sneezing, or speaking, and transmission through contaminated surfaces may also occur (Balupuri et al. 2023).

The Spike protein is a homotrimeric glycoprotein anchored to the viral membrane by a transmembrane region, and its ectodomain extends outward and decorates the surface of the virus. The protein is glycosylated, which contributes to protein folding and stability, masks antigenic epitopes, facilitates immune

evasion, and mediates interactions with lectins and antibodies (Flores-Vega et al. 2022; Jackson et al. 2022). Each protomer contains 22 N-glycosylation sites and 2 O-glycosylation sites, and variations in the size and distribution of N-glycans can influence the dynamics and stability of receptor-binding domain (RBD) conformations (Jackson et al. 2022; Harbison et al. 2022; Zhao et al. 2021). The RBD mediates the interaction between the virus and the host cell receptor Angiotensin-Converting Enzyme 2 (ACE2) and represents the primary target of antibodies elicited by both vaccination and natural infection (Jackson et al. 2022). Structurally, it consists of a five-stranded antiparallel beta sheet connected to alpha helices and an extended loop known as the Receptor-Binding Motif (RBM) (Zhang et al. 2021). The Spike adopts two distinct conformations: an "up" position, to expose the RBD and RBM for receptor engagement; and a "down" position, where this domain remains closely packed, preventing receptor interaction (Jackson et al. 2022; Peng et al. 2020; Peng et al. 2021; Zhang et al. 2021).

Mutations within the RBD of the S protein have influenced its interaction with the ACE2 receptor, enhancing immune evasion and reducing the efficacy of monoclonal antibody (mAb) neutralization. These changes have contributed to increased infectivity and transmissibility and have supported the virus's adaptation to humans, with the potential to become a seasonal pathogen. This emphasizes the need for vaccines and therapies that remain effective across diverse Spike variants, with periodic updates as the virus continues to evolve (Balupuri et al. 2023; Chakraborty 2022; Fang et al. 2023; Magazine et al. 2022; Winger and Caspari 2021). The Spike comprises S1 and S2 subunits upon cleavage. Without this cleavage, S2 adopts a premature conformation, rendering the protein non-functional (Jackson et al. 2022). To counteract this instability and delay cleavage, the virus quickly acquired the D614G mutation, which became dominant in over 95% of sequenced SARS-CoV-2 variants by January 2021 (Jackson et al. 2022; Khan et al. 2021; Magazine et al. 2022). This mutation rapidly became dominant due to its association with enhanced viral transmission and entry efficiency. It has also been linked to an increased incidence of anosmia (loss of smell), which likely results from stronger interactions between the Spike protein and ACE2 receptors in the olfactory epithelium (Chen et al. 2022 ; Khan et al. 2021).

Substitutions at key residues in the Spike RBD, including N501, L452, N439, E484, T470, and Q498, have been reported to increase the virus binding affinity of the Spike-ACE2 complex, and mutations such as N501Y, N439K, E484K, L452R, and S477N have been reported to stabilize the open conformation of the RBD (Alaofi and Shahid 2021). In addition, co-occurring mutations such as K417N and E484K were shown to increase receptor binding affinity (Chakraborty 2022; Bhadane and Salo-Ahen 2022), with E484K increasing the number of interaction contacts and reducing HBs with ACE2 (Chakraborty 2022), while also being associated with reduced efficacy of neutralizing antibodies from natural infection, monoclonal therapies, and vaccination (Flores-Vega et al. 2022). Additionally, E484K and S477N, both located in flexible binding loops, contributed to reduced structural stability in the RBD (Alaofi and Shahid 2021), and alanine substitution at Y505 (Y505A) was reported to abolish ACE2 binding (Xu et al. 2021).

L452R and E484Q mutations were reported to disrupt electrostatic interactions with residue L31 of ACE2 and reduce intra- and intermolecular interactions (Antony and Vijayan 2021). L452R also showed increased interaction with surrounding water molecules and enhanced Spike-ACE2 binding (Flores-Vega et al. 2022; Antony and Vijayan 2021). Despite reduced intramolecular stability, L452R and N501Y were reported to form

strong HBs with ACE2 (Alaofi and Shahid 2021). Mutations G339D, S371L, S375F, N440K, S477N, and T478K were suggested to be the reason for the instability of the Omicron variant (Roy 2023), while N501Y enhanced the Spike-ACE2 binding at the same variant (Flores-Vega et al. 2022; Mittal et al. 2022). High-resolution Molecular Dynamics (MD) interaction analyses identified the N501Y substitution as a key determinant of Spike-ACE2 affinity, associated with the formation of novel hydrogen bonds (HBs) (Balupuri et al. 2023). Other studies reported that N501Y can reduce the total number of HBs and interfacial contacts, while introducing π - π stacking and cation- π interactions that strengthen the binding interface (Chakraborty 2022; Tian et al. 2021).

Computational approaches, such as MD simulations and Normal Mode Analysis (NMA), have become valuable tools in bioinformatics for exploring the structural behavior of viral proteins. These approaches enable the investigation of conformational transitions that macromolecules undergo between experimentally observable stable states, which are often critical for their functional and regulatory mechanisms (Bahar et al. 2010; Perahia and Mouawad 1995; Resende-Lara et al. 2020). This work studied the structural plasticity and dynamics of the RBD and ACE2 using Molecular Dynamics (all-atom), Normal Mode Analysis (NMA) using Elastic Network Models (ENM) (Gasparini et al. 2023), and Machine Learning techniques (ML) to compare the results with what is known in the literature.

2. MATERIALS AND METHODS

2.1 Database

The methodology is summarized in the Data Flow Diagram shown in Figure 1. Lineage frequency data were obtained from the Global Initiative on Sharing All Influenza Data - GISAID - (GISAID 2025) to identify the most prevalent SARS-CoV-2 variants circulating in Brazil during the one-year period from May 2023 to May 2024 (the maximum timeframe available for graphical visualization of the lineages on the platform). All selected lineages, as well as all the mutations studied in this work, correspond to subvariants of the Omicron variant (Arantes et al. 2025; Liu et al. 2024; Ma et al. 2024; Roy 2023).

Complete sequencing metadata that corresponds to these lineages were subsequently retrieved, and the dataset consisted of genomic metadata that was selected based on the following inclusion criteria: complete genomes (>29,000 nucleotides), virus designation “hCoV-19,” collection dates between January 1, 2023, and May 27, 2024, and geographic origin reported as Brazil. Additionally, only mutations occurring in the Spike protein were considered, specifically, those annotated with the prefix “Spike”; from those, only mutations observed at least twice were retained. Due to time constraints and study scope and considering localization within the RBD (residues 319-528), only the most frequent mutations in the dataset together with those most reported in experimental and in silico studies in literature (Bhadane and Salo-Ahen 2022; Chen et al. 2022; Cheng et al. 2022; Fang et al. 2023; Kodchakorn and Kongtawelert 2022; Magazine et al. 2022; Roy 2023; Rucker et al. 2023; Zhu et al. 2022), were included in the study: D405N, E484A, E484K, F486P, F490S, G339D, G339H, G446S, G496S, K417N, K417T, L452R, N440K, N460K, N501Y, Q493R, Q498R, R408S, S371F, S371L, S373P, S375F, S477N, T478K, T487K, V445P, and Y505H.

2.2 *In silico* models

PDB 6VW1 (resolution: 2.68 Å) corresponds to the RBD (chimeric RBD: SARS-CoV RBD and SARS-CoV-2 RBM) in complex with the human ACE2 receptor and was used for this study. An X-ray crystallographic structure that comprises four chains: A and B (ACE2; 597 residues each), and C and D (RBD, 217 residues each), derived from *Homo sapiens* and a chimeric Severe Acute Respiratory Syndrome Coronavirus/Severe Acute Respiratory Syndrome Coronavirus 2, respectively. The PDB contains no mutations. Structural preparation was performed using the PDB Reader and Manipulator module available in CHARMM-GUI (Jo et al. 2014; Kong et al. 2024; Park et al. 2023).

Chains A (ACE2) and C (RBD) were retained, missing residues were modeled, and the final sequence was aligned with the 2019 wild-type (WT) reference sequence retrieved from UniProt (ID: P0DTC2-SPIKE_SARS2) (UniProt 2025), to ensure consistency for subsequent analyses (Figure 2). We acknowledge that fully resolved SARS-CoV-2 RBD-ACE2 complexes were also available and may represent more direct structural models of the interaction; therefore, this choice could constitute a limitation of the present study. Nevertheless, we chose to use 6VW1, which was already available in CHARMM-GUI when the study started. Besides this, our analyses were intended to characterize general dynamical features of the mutants in the RBD-ACE2 complex rather than to compare alternative structural models, and the results should be interpreted within this context.

Through Solution Builder available in CHARMM-GUI, mutations were prepared and analysed individually to analyse their own characteristics, under physiological conditions: pH 7.0, 0.15 M NaCl, temperature of 310.15 K, and a rectangular water box with a 40.0 Å side length, under the CHARMM36 m force field (Huang and MacKerell 2013). Disulfide bonds were defined according to UniProt (UniProt 2025) annotations: for the RBD (chain C), between residues 336-361, 379-432, 391-525, and 480-488; and for ACE2 (chain A), between residues 133-141, 344-361, and 530-542. N-glycosylation was added at residues 33 and 331 for the RBD, and 53, 90, 103, 322, and 546 for ACE2.

2.3 Molecular Dynamics and ENM Simulations

All systems underwent a 10 ns equilibration phase under the NVT ensemble, followed by 50 ns of production MD simulations under the NPT ensemble. Simulations were performed using Nanoscale Molecular Dynamics (NAMD) to characterize local structural characteristics, evaluate the initial dynamic effects of point mutations on the Spike RBD-ACE2 complex, and generate representative conformations for subsequent ENM analyses (Phillips et al. 2020). Analyses were performed using VMD (Humphrey et al. 1996). Simulation energy, pressure, temperature, and volume were monitored using the NAMD Plot plugin. Intramolecular HBs were identified using a donor-acceptor distance cutoff of 3.5 Å and an angle cutoff of 20°. Hydrogen bonds of specific interacting residues were analyzed separately using the same criteria. All trajectories were structurally aligned prior to calculating the Root-Mean-Square Deviation (RMSD). The trajectories were then used as representative conformations for subsequent Elastic Network Model analyses, enabling the investigation of large-scale collective motions without the need for computationally expensive microsecond-scale simulations.

For the ENM analyses, a PDB structure was extracted from the final frame of each MD trajectory and analyzed using ProDy (Bakan et al. 2011). All ENM calculations were performed under vacuum-like

conditions, without explicit solvent molecules or dielectric screening, allowing a comparative assessment of the intrinsic flexibility predicted by the Anisotropic Network Model (ANM) through its collective motions. The extracted structures were used to construct the Hessian matrix and compute the normal modes using a Ca-based ENM. The coordinate set was then duplicated to define the initial and target conformations, and the first 20 ANM modes were calculated for subsequent analyses. Although ENMs are typically calculated using protein structures in vacuum, MD simulations were performed in explicit solvent to equilibrate each system and obtain representative conformations for subsequent ENM analyses. This preprocessing step provides structurally relaxed models that better reflect the solvated state before the vacuum-based ENM calculations.

2.4 Data used in the study for analyses

In the WT system, distances were measured between known interacting residue pairs through VMD: 417-30, 449-38, 449-42, 453-34, 455-41, 455-34, 456-30, 456-31, 475-19, 484-31, 486-330, 487-24, 489-31, 493-31, 493-35, 496-353, 498-353, 500-41, 501-41, 502-353, 505-37, 505-393, RBD and ACE2 residue respectively (Hatmal et al. 2020). The mutations are all from the subsequent lineages: XDR, XBB.2.3, XBB.1.9, XBB.1.5.86, XBB.1.5.70, XBB.1.5, XBB.1.18.1, XBB.1.16, XBB.1, JN.1.43.1, JN.1.4, JN.1.29, JN.1.23, JN.1.16.1, JN.1.1, JN.1, JD.1.2, JD.1.1.1, JD.1.1, JD.1, GK.1.1, GK.1, FL.1.5.1, FE.1.2, FE.1.1, FE.1, EG.5.1, that belong to Omicron variant (Arantes et al. 2025; Liu et al. 2024; Ma et al. 2024; Roy 2023). The pairs of residues studied, the known interactions between RBD and ACE2, are detailed in Table 1.

The selection was based on the closest non-hydrogen atoms identified by visual inspection, considering them in contact if at least one pair of non-hydrogen atoms was within 3.5 Å of each other (Table 1) (Zhu et al. 2022). In the mutant systems, the selection was replicated, and when the mutant changed the reference atoms, the nearest available non-hydrogen atoms within the mutated residues were selected for the measurements. The RMSD was also analyzed, and the HBs were calculated for the whole complex and between these highlighted residue pairs. The datasets from MD simulations were subsequently concatenated into a single dataframe to work with Machine Learning supervised and unsupervised models, with all variables aligned by trajectory frame. The class label was assigned in accordance with the type of amino acid substitution in each mutant, based on the physicochemical properties of the side chains (Table S1).

The final dataset comprised 14,000 rows and 51 columns. After processing and analysis, 38 of the 51 columns were retained for subsequent work (Table 2). For comparative analyses, the data were stored both in sequential order and in randomized form.

Table 1 Atom pairs for distance measurements in VMD. Distance measurements were performed in VMD by selecting atom pairs, enabling consistent comparisons between WT and mutant structures

Interactions (RBD-ACE2)	Interactions (RBD-ACE2)	Exceptions (RBD-ACE2)
417 (NZ) - 30 (OD2)	493 (OE1) - 31 (NZ)	E484A - 484 (CB) - 31 (CE)
449 (CE2) - 38 (OD2)	493 (NE2) - 34 (CB)	K417N - 417 (ND2) - 30 (CG)
449 (CZ) - 42 (NE2)	493 (NE2) - 35 (CG)	K417T - 417 (CB) - 30 (OD1)
453 (CE1) - 34 (CD2)	496 (O) - 353 (CE)	N501Y - 501 (CE2) - 41 (CD2)
455 (CD1) - 31 (N)	498 (OE1) - 353 (CG)	T487K - 487 (NZ) - 24 (OE1)
455 (CD2) - 34 (ND1)	500 (OG1) - 41 (CE2)	Y505H - 505 (CE1) - 37 (OE1)
456 (CD1) - 30 (OD1)	501 (ND2) - 41 (CZ)	Y505H - 505 (CE1) - 393 (CZ)
456 (CZ) - 31 (CB)	502 (C) - 353 (O)	-
475 (O) - 19 (O)	505 (CZ) - 37 (OE1)	-
484 (CD) - 31 (NZ)	505 (CZ) 393 (CZ)	-
489 (CD1) - 31 (CG)	-	-

To facilitate visualization of the Root-Mean-Square Fluctuation (RMSF) obtained from NMA, min-max normalization (feature scaling) was applied due to an extreme peak in the WT and some mutants. Values were rescaled to a range between 0 and 1. This transformation was performed exclusively for comparative visualization and not for statistical analysis, enabling the comparison of relative motion patterns between mutants and WT structures.

2.5 Exploratory and statistical analyses

The dataset described in Table 2 comprised 14,000 rows and 51 columns. To reduce dimensionality, an exploratory statistical analysis was conducted. Since the data did not follow a normal distribution, non-parametric approaches were considered appropriate. Feature redundancy was initially explored through Pearson correlation analysis to identify highly correlated variables and detect potentially redundant attributes. Subsequently, multicollinearity was quantitatively assessed through the Variance Inflation Factor (VIF) to allow the identification of variables whose information could be largely explained by combinations of other predictors. To further evaluate feature relevance, the Kruskal-Wallis test was used to assess whether individual features significantly differed across mutation groups. Features with statistically significant differences (low p-values <0,05) were considered informative for distinguishing mutation classes, whereas variables showing low discriminatory power or high redundancy were excluded. Following this feature selection process, 38 out of the original 51 columns were retained to use in machine learning analysis. Statistical differences were evaluated using the Mann-Whitney U test with Benjamini-Hochberg false discovery rate (FDR) correction for RMSD, contact, and HB analyses to adjust p-values since multiple comparisons were performed.

2.6 Machine Learning

Table 2 presents the dataset prepared for Machine Learning applications. All extracted features before the feature selection process are listed, whereas the features retained for Machine Learning analyses are highlighted.

Table 2 Complete set of features included in the dataset for Exploratory and Machine Learning analyses. Features highlighted are those retained after data processing and statistical evaluation for use in the clustering and Random Forest analyses

COLUMNS / ATTRIBUTES					
frame	hbond_500_41	hbond_456_31	dist_449_38	dist_500_41	dist_475_19
mutation	hbond_417_30	hbond_489_31	dist_502_353	dist_496_353	dist_501_41
hbond_453_34	hbond_502_353	hbond_484_31	dist_455_31	dist_493_34	dist_417_30
hbond_449_38	hbond_486_330	hbond_487_24	dist_456_31	dist_493_35	rmsd
hbond_493_31	hbond_455_31	hbond_505_393	dist_453_34	dist_455_34	hbond
hbond_501_41	hbond_498_353	hbond_493_34	dist_493_31	dist_486_330	substitution
hbond_496_353	hbond_449_42	hbond_455_34	dist_449_42	dist_505_37	-
hbond_493_35	hbond_505_37	dist_487_24	dist_489_31	dist_498_353	-
hbond_456_30	hbond_475_19	dist_456_30	dist_484_31	dist_505_393	-

K-means and DBSCAN were applied to both ordered and randomized datasets, with and without the inclusion of the frame index as an additional feature (Chadjipadelis T 2021), to enable a comparative assessment of clustering performance across different data configurations. To improve feature representation and reduce dimensionality, Principal Component Analysis (PCA) and Uniform Manifold Approximation and Projection (UMAP) were applied sequentially before clustering. Principal Component Analysis (PCA) was first applied as a linear dimensionality reduction step to reduce noise, minimize feature redundancy, and transform correlated variables into principal components while preserving most of the data variance (Ringnér 2008). A total of 31 principal components were retained, accounting for 95% of the cumulative explained variance. Subsequently, Uniform Manifold Approximation and Projection (UMAP) (McInnes et al. 2018) was employed as a non-linear dimensionality reduction method to capture the underlying manifold structure and generate a low-dimensional representation of the data (n_neighbors = 30, min_dist = 0.0). In this combined approach, PCA improved data compactness and computational efficiency by removing redundant information. At the same time, UMAP preserved complex non-linear relationships and local neighborhood structure that linear methods alone cannot capture (McInnes et al. 2018). Clustering performance was evaluated using the Davies-Bouldin, Calinski-Harabasz, and Silhouette indices (Davies and Bouldin 1979; Caliński and Harabasz 1974; Rousseeuw 1987).

Given that the dataset was not large enough to justify the use of more data-intensive supervised learning methods, a Random Forest classifier was applied for supervised learning analysis (Parmar et al. 2019) using the substitution class as the target variable and the remaining variables as input features. The frame index

was excluded from the feature set to avoid introducing explicit temporal ordering information into the model. The dataset was randomly divided into training (90%), test (5%), and validation (5%) subsets. Model performance was assessed through cross-validation (Kohavi 1995) and evaluated on the test and validation sets using the Confusion Matrix and the Receiver Operating Characteristic (ROC) curve (Fawcett 2006). Feature-importance scores derived from the model were analyzed to identify the attributes most relevant for class discrimination. All ML analyses were implemented using the Scikit-learn Python package (Pedregosa et al. 2012).

3. RESULTS AND DISCUSSION

It was observed that the residues substitutions revealed a diverse range of alterations in side chain properties, including changes in polarity (S371L and S373P), charge (D405N, N460K, Q493R, and others), and aromaticity (N501Y, S371F, and S375F). These molecular alterations provide the basis for the structural, dynamic, and interaction effects explored.

A comparative structural analysis of the WT RBD-ACE2 complex and each mutant complex was performed to identify differences in atomic distances, hydrogen bonds, and contact persistence throughout the simulations. Atomic distances were calculated as Euclidean distances between manually selected atom pairs using VMD. Interatomic contacts were represented as binary variables using a 3.5 Å distance cutoff: distances ≤ 3.5 Å were encoded as 1 (contact), whereas distances > 3.5 Å were encoded as 0 (no contact). Contact persistence was quantified by summing contact occurrences over consecutive 10-frame intervals, enabling the identification of temporal differences in intermolecular interactions throughout the MD trajectories.

The WT complex maintained one to two intermolecular contacts throughout nearly the entire simulation, with only a brief loss of contact near the end of the trajectory (Figure S1). Among the mutants that exhibited a balanced interaction profile, characterized by compensatory gains and losses of intermolecular contacts through interactions, D405N displayed a reduced number of contacts throughout the trajectory. E484A and G339D maintained a stable contact pattern, generally sustaining at least two contacts over extended periods with only minor fluctuations. G339H and N440K exhibited occasional contact losses, including brief intervals without detectable proximity. S371L maintained at least one contact throughout the trajectory with intermittent increases to two contacts lasting approximately 1-3ns. S477N showed additional contacts at the beginning of the trajectory and brief contact-free intervals near the end. Mutant V445P kept one to two contacts (Figures S2–S28).

Among the mutations that showed an overall tendency to losses through the different pairs, F490S and G446S exhibited losses and a lower number of contacts. G496S exhibited a more stable profile, maintaining at least 2 contacts throughout the trajectory. N501Y and Q498R lost contacts, particularly at the end of the simulations, whereas S373P, S375F, and Y505H maintained fewer contacts overall, frequently fluctuating between one and two contacts. In contrast, mutations that exhibited an overall tendency toward gains through the pairs showed increased contact persistence. F486P increased the residence time of multiple contacts and eliminated the contact-free gaps observed in the WT. E484K also exhibited compensatory effects through increased contact persistence, while L452R showed an increased number of contacts,

consistently maintaining two or more contacts throughout the trajectory. Other mutations, including K417N/T, N460K, R408S, S371F, T478K, and T487K showed a balanced behavior, alternating between contact loss and gain across different interaction pairs.

The residue pairs that exhibited the highest number of contacts across all frames were 501-41, 493-31, and 417-30 (RBD-ACE2), whereas the greatest number of HBs was observed in 502-353, 493-35, and 505-37 (RBD-ACE2) (Figure 3 and Figure S29). Despite the predominance of losses compared to the WT at interactions 501-41 and 417-30, the cumulative analysis along the trajectory revealed high occupancy of contacts. As for RMSD: E484A, F486P, G496S, K417T, L452R, N440K, Q493R, Q498R, R408S, and T487K exhibited higher median values compared to the WT, which suggested increased flexibility (Figure 4).

Considering that residues 0-596 correspond to ACE2 (19-615), while residues 597-813 correspond to the RBD (319-535), the WT exhibited a uniform RMSF, with minimal fluctuation, except for a pronounced peak at residue 322 of the RBD. This loose loop of the RBD displayed flexibility and was similarly highlighted in some mutants. A theory for this behavior is the absence of the portions of the Spike that were not included.

The RMSF, obtained through NMA, revealed that several mutants exhibited dynamic behavior similar to that of the WT: D405N, E484K, G339D, G339H, G446S, G496S, K417T, L452R, N460K, N501Y, Q493R, R408S, S371F, S375F, S477N, T487K, V445P and Y505H, which suggests limited impact on global dynamics. In addition, a subset of mutations that include F490S, N440K, Q498R, S371L, and S373P exhibited reduced fluctuations, which indicates increased local rigidity. Mutations such as E484A, F486P, K417N, and T478K showed increased fluctuations, which suggests enhanced local flexibility. Overall, RMSF results indicated that the mutations modulated protein flexibility through local motions, but most importantly, they maintained the global coordinated motion, as visually observed, leading to increased conformational plasticity while maintaining the overall integrity of the RBD-ACE2 complex (Figure S30 - Figure S43). These structural adaptations reveal compensatory mechanisms that may contribute to functional optimization of the Spike protein.

Feature-importance analysis using a Random Forest model allowed the identification of the variables that most contributed to the classification, with the distances between RBD and ACE2 residues (residue pairs 505-393 and 449-38) and RMSD being the most important. The model demonstrated good predictive performance, achieving a validation accuracy of 0.82 and a test accuracy of 0.86. The weighted F1-scores were 0.81 (validation) and 0.86 (test), while macro-average F1-scores reached 0.78 and 0.83, respectively, which indicates consistent classification performance across classes, although with some variability in recall (Figure S44). The slightly higher scores on the test set compared to the validation set suggest that the model generalizes well and is not overfitting. The relatively lower macro F1-score suggests that some classes remain more difficult to predict, likely reflecting the effects of class imbalance.

The 505-393 interaction exhibited more hydrogen bond losses than gains relative to the WT, whereas the 449-38 interaction showed substantial reductions in both residue contacts and hydrogen bonds. Notably, residues surrounding positions 38 and the ACE2 hotspot 353 have been reported to play key roles in determining SARS-CoV-2 binding patterns across different species (Chen et al. 2025). Moreover, for the

unsupervised analysis, eight different models were tested: K-means and DBSCAN with and without frame, with ordered and randomized data (Table 3).

Table 3 Performance metrics of the Unsupervised Machine Learning models. Comparison of clustering performance of K-means and DBSCAN under different conditions (ordered versus randomized data, with and without frame). Model performance was evaluated using Davies-Bouldin, Calinski–Harabasz, and Silhouette scores, along with the number of clusters identified

Algorithm	Frame	Data	Davies-Bouldin Score	Calinski-Harabasz Score	Silhouette Score	Nº clusters
K-means	yes	ordered	0.4136	45983.3125	0.6026	15
K-means	yes	random	0.3488	65984.7812	0.6348	18
K-means	no	ordered	0.4155	46140.7969	0.6077	15
K-means	no	random	0.4603	68459.6094	0.6479	20
DBSCAN	yes	ordered	0.2555	12165.8896	0.5694	12
DBSCAN	yes	random	0.2717	9988.3486	0.5964	11
DBSCAN	no	ordered	0.2469	18410.7988	0.6161	16
DBSCAN	no	random	0.2633	11410.9365	0.6000	15

The best performance among the evaluated approaches of unsupervised learning was DBSCAN when applied to the ordered dataset without frame. This method avoided the bias of clustering by frame, and it handles variable data more effectively, yet it yielded the lowest Davies-Bouldin index overall, as well as the highest Calinski-Harabasz and Silhouette scores among the DBSCAN analyses. Although the Silhouette score is not ideal for evaluating DBSCAN, as it tends to favor spherical clusters, it was still considered for comparison. Later, for data analysis, the frame was included to state some characteristics. Across all analyses, the mutants that were consistently highlighted were G496S, N501Y, and Y505H, with the last two at the same cluster in most parts of the analyses. In this analysis, N501Y and Y505H appeared in cluster 14, whereas G496S appeared in cluster 15. Cluster 14 presented an RMSD value between 4 and 6 Å, which may indicate moderate structural alterations (Figure 5). These values corresponded approximately to 30-40 ns for N501Y and 18-30 ns for Y505H. Cluster 15 exhibited an RMSD value between 5 and 6 Å, and it corresponds approximately to 33 ns until the end of the trajectory (Figure 6).

In both clusters (14 and 15), HBs were not very frequent and predominated at value zero, except for three pairs that showed slightly higher occurrences: 502-353, 493-35 and 505-37, which corresponded to the previously highlighted interactions. The residue distances followed the general behavior observed in other clusters (approximately 5 Å). However, some pairs exhibited significantly lower or higher values. All things considered, the clusters appeared to represent intermediate and more frequently visited states, in which neither proximity nor pronounced distance is observed.

The mutants highlighted in the clustering analyses showed behavior similar to the WT according to RMSF analysis obtained through NMA. In accordance with the literature, these three mutants (G496S, N501Y and Y505H) were shown to belong to a group of binding-energy hotspots, together with Q498R and other mutations (Verkhivker et al. 2022). They are suggested to contribute to allosteric communication between the RBD and ACE2, and to promote increased stability and allosteric coupling within the complex. These mutations have also been associated with immune evasion and antibody resistance (Pavia et al. 2024). In the present study, G496S, N501Y, and Y505H exhibited RMSF similar to the WT, which indicates preservation of collective motions. Although these mutants frequently presented losses in specific contacts and hbs, these reductions did not translate into marked global differences. With regard to N501Y specifically, the mutation has been reported to directly affect the ACE2-RBD interaction by forming HBs that stabilize the binding (Balupuri et al. 2023). However, it has also been observed that N501Y does not produce large changes in the overall flexibility of the RBD (Alaofi and Shahid 2021). In the present study, transient losses of HBs were observed for some interaction pairs during the simulations of the N501Y mutant. This behavior likely reflects local conformational rearrangements, which indicates that stabilization of the complex does not require persistence of the same HBs throughout the trajectory, but rather maintenance of the overall interaction network.

Taken together, these alterations suggest a pattern of compensatory adaptation. Instead of uniformly increasing or decreasing structural stability, these mutations appear to modulate the balance between rigidity and flexibility, thereby preserving overall structural integrity. This interpretation aligns with experimental findings that describe a mechanism in which the Spike protein undergoes a series of conformational transitions upon ACE2 binding, which begins with a twisting and opening motion and follows with increased flexibility, and ultimately reaches a conformation that facilitates viral entry into the host cell (Xu et al. 2021). A similar conclusion has been reported for another protein of the same virus, where a mutation was selected by the virus despite its destabilizing behavior (Saha et al. 2025), which reinforces the idea that selective advantage arises from the combined structural and functional effects of mutations rather than their stabilizing or destabilizing nature alone.

Some mutations exemplified this diversity of effects. L452R preserved or increased the number of contacts and hydrogen bonds while maintaining multiple contacts over extended periods. Although it exhibited a higher RMSD than the WT, its RMSF profile remained comparable, suggesting that the mutation altered the overall conformation without substantially affecting local residue flexibility. Similarly, F486P displayed increased contact persistence together with a balanced pattern of hydrogen bond gains and losses, indicating a redistribution rather than a net change in intermolecular interactions. This mutation also showed fluctuations in RMSF while maintaining coordinated global motions. In contrast, G496S led to local losses of interactions without significant changes in fluctuation patterns, as indicated by a similarity with the WT in the RMSF. Also, Q493R exhibited losses of both contacts and hydrogen bonds, even in interactions conserved in other mutants; however, its RMSF remained similar to the WT.

4. CONCLUSIONS

In conclusion, the results suggested that mutations primarily modulated local balances between rigidity and flexibility. Some exhibited increased local flexibility, whereas others maintained WT-like rigidity alongside coordinated global motions. Mutants highlighted in clustering analysis (G496S, N501Y, and Y505H) showed moderate structural rearrangements while preserving overall behavior. Although local alterations in contacts and hydrogen bonds were observed, the complex's global dynamic organization remained largely preserved. Mutation-associated functional differences were better explained by compensatory structural adjustments, redistribution of intermolecular contacts, and modulation of local flexibility. The results suggest that the mutations modulate the structural landscape by finely balancing rigidity and flexibility, thereby enabling conformational adjustments that are likely critical for receptor engagement and membrane fusion. Overall, structural plasticity emerges as a key determinant of receptor binding while preserving the global architecture of the complex. This finding underscores a delicate interplay between structural stability and conformational adaptability, both of which are essential for efficient viral attachment and entry. These results are consistent with our hypothesis that conformational changes, dynamics, and transient conformations are critically important for protein function.

Some limitations should be considered when interpreting the findings of this study. Nevertheless, these limitations do not diminish its scientific relevance; rather, they define the scope of the conclusions and provide directions for future investigations. The study focused on subvariants circulating in Brazil within a defined one-year period. Structural analyses were restricted to the RBD, using a chimeric structure, as modeling the full Spike protein - particularly in its trimeric form - poses a significant computational challenge. Consequently, potential contributions from other domains were not evaluated. The mutations were studied separately; the effects of combinations of mutations were not evaluated either. These findings provide a framework for understanding how emerging RBD mutations model the structure through dynamic compensation, offering insights that may support future surveillance and structure-guided therapeutic design.

Declarations

AUTHOR CONTRIBUTIONS

Isabelle Alves Pereira - Conceptualization, Data curation, Formal analysis, Methodology, Investigation, Visualization, Project administration, Writing - original draft

Yolanda M.B Marcello - Writing - review & editing

Glauco Lima - Writing - review & editing

Ana Lígia Scott - Financing, Project administration, Resources, Conceptualization, Methodology, Supervision, Writing - review & editing.

No potential conflict of interest was reported by the author(s).

Funding: This study was financed in part by the Brazilian Federal Agency for Support and Evaluation of Graduate Education (CAPES) - Finance Code 001.

References

1. Alaofi AL, Shahid M (2021) Mutations of SARS-CoV-2 RBD may alter its molecular structure to improve its infection efficiency. *Biomolecules* 11:1273. <https://doi.org/10.3390/biom11091273>
2. Antony P, Vijayan R (2021) Molecular dynamics simulation study of the interaction between human angiotensin converting enzyme 2 and spike protein receptor binding domain of the SARS-CoV-2 B.1.617 variant. *Biomolecules* 11:1244. <https://doi.org/10.3390/biom11081244>
3. Arantes I, Ito K, Gomes M, et al (2025) Rapid spread of the SARS-CoV-2 Omicron XDR lineage derived from recombination between XBB and BA.2.86 subvariants circulating in Brazil in late 2023. *Microbiol Spectr* 13:e0119324. <https://doi.org/10.1128/spectrum.01193-24>
4. Bahar I, Lezon TR, Yang L-W, Eyal E (2010) Global dynamics of proteins: bridging between structure and function. *Annu Rev Biophys* 39:23–42. <https://doi.org/10.1146/annurev.biophys.093008.131258>
5. Bakan A, Meireles LM, Bahar I (2011) ProDy: protein dynamics inferred from theory and experiments. *Bioinformatics* 27:1575–1577. <https://doi.org/10.1093/bioinformatics/btr168>
6. Balupuri A, Kim J-M, Choi K-E, et al (2023) Comparative computational analysis of spike protein structural stability in SARS-CoV-2 Omicron subvariants. *Int J Mol Sci* 24:16069. <https://doi.org/10.3390/ijms242216069>
7. Bhadane R, Salo-Ahen OMH (2022) High-throughput molecular dynamics-based alchemical free energy calculations for predicting the binding free energy change associated with the selected omicron mutations in the spike receptor-binding domain of SARS-CoV-2. *Biomedicines* 10:2779. <https://doi.org/10.3390/biomedicines10112779>
8. Caliński T, Harabasz J (1974) A dendrite method for cluster analysis. *Commun Stat* 3:1–27. <https://doi.org/10.1080/03610927408827101>
9. Chadjipadelis T, Lausen B, Markos A, et al (eds) (2021) *Data analysis and rationality in a complex world*, 1st edn. Springer Nature, Cham, Switzerland
10. Chakraborty S (2022) E484K and N501Y SARS-CoV 2 spike mutants Increase ACE2 recognition but reduce affinity for neutralizing antibody. *Int Immunopharmacol* 102:108424. <https://doi.org/10.1016/j.intimp.2021.108424>
11. Chen W, Lee J-Y, Kim J-S, et al (2025) Critical amino acid residues in human ACE2 for SARS-CoV-2 spike protein binding and virus entry. *Microbiol Spectr* 13:e0324424. <https://doi.org/10.1128/spectrum.03244-24>
12. Chen Y, Liu Q, Zhou L, et al (2022) Emerging SARS-CoV-2 variants: Why, how, and what's next? *Cell Insight* 1:100029. <https://doi.org/10.1016/j.cellin.2022.100029>
13. Cheng MH, Krieger JM, Banerjee A, et al (2022) Impact of new variants on SARS-CoV-2 infectivity and neutralization: A molecular assessment of the alterations in the spike-host protein interactions. *iScience* 25:103939. <https://doi.org/10.1016/j.isci.2022.103939>
14. Davies DL, Bouldin DW (1979) A cluster separation measure. *IEEE Trans Pattern Anal Mach Intell* 1:224–227. <https://doi.org/10.1109/tpami.1979.4766909>

15. Fang L, Xu J, Zhao Y, et al (2023) The effects of amino acid substitution of spike protein and genomic recombination on the evolution of SARS-CoV-2. *Front Microbiol* 14:1228128. <https://doi.org/10.3389/fmicb.2023.1228128>
16. Fawcett T (2006) An introduction to ROC analysis. *Pattern Recognit Lett* 27:861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>
17. Flores-Vega VR, Monroy-Molina JV, Jiménez-Hernández LE, et al (2022) SARS-CoV-2: Evolution and emergence of new viral variants. *Viruses* 14:653. <https://doi.org/10.3390/v14040653>
18. Gasparini P, Philot EA, Pantaleão SQ, et al (2023) Unveiling mutation effects on the structural dynamics of the main protease from SARS-CoV-2 with hybrid simulation methods. *J Mol Graph Model* 121:108443. <https://doi.org/10.1016/j.jmgm.2023.108443>
19. GISAID - gisaid.org. In: Gisaid.org. <https://gisaid.org/>. Accessed 30 Mar 2025
20. Haque SM, Ashwaq O, Sarief A, Azad John Mohamed AK (2020) A comprehensive review about SARS-CoV-2. *Future Virol* 15:625–648. <https://doi.org/10.2217/fvl-2020-0124>
21. Harbison AM, Fogarty CA, Phung TK, et al (2022) Fine-tuning the spike: role of the nature and topology of the glycan shield in the structure and dynamics of the SARS-CoV-2 S. *Chem Sci* 13:386–395. <https://doi.org/10.1039/d1sc04832e>
22. Hatmal MM, Alshaer W, Al-Hatamleh MAI, et al (2020) Comprehensive structural and molecular comparison of spike proteins of SARS-CoV-2, SARS-CoV and MERS-CoV, and their interactions with ACE2. *Cells* 9:2638. <https://doi.org/10.3390/cells9122638>
23. Hu B, Guo H, Zhou P, Shi Z-L (2021) Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev Microbiol* 19:141–154. <https://doi.org/10.1038/s41579-020-00459-7>
24. Huang J, MacKerell AD Jr (2013) CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J Comput Chem* 34:2135–2145. <https://doi.org/10.1002/jcc.23354>
25. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14:33–38. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5)
26. Jackson CB, Farzan M, Chen B, Choe H (2022) Mechanisms of SARS-CoV-2 entry into cells. *Nat Rev Mol Cell Biol* 23:3–20. <https://doi.org/10.1038/s41580-021-00418-x>
27. Jo S, Cheng X, Islam SM, et al (2014) CHARMM-GUI PDB manipulator for advanced modeling and simulations of proteins containing nonstandard residues. *Adv Protein Chem Struct Biol* 96:235–265. <https://doi.org/10.1016/bs.apcsb.2014.06.002>
28. Khan A, Khan T, Ali S, et al (2021) SARS-CoV-2 new variants: Characteristic features and impact on the efficacy of different vaccines. *Biomed Pharmacother* 143:112176. <https://doi.org/10.1016/j.biopha.2021.112176>
29. Kodchakorn K, Kongtawelert P (2022) Molecular dynamics study on the strengthening behavior of Delta and Omicron SARS-CoV-2 spike RBD improved receptor-binding affinity. *PLoS One* 17:e0277745. <https://doi.org/10.1371/journal.pone.0277745>
30. Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. *Int Jt Conf Artif Intell* 1137–1145

31. Kong L, Park S-J, Im W (2024) CHARMM-GUI PDB Reader and Manipulator: Covalent ligand modeling and simulation. *J Mol Biol* 436:168554. <https://doi.org/10.1016/j.jmb.2024.168554>
32. Liu Y, Zhao X, Shi J, et al (2024) Lineage-specific pathogenicity, immune evasion, and virological features of SARS-CoV-2 BA.2.86/JN.1 and EG.5.1/HK.3. *Nat Commun* 15:8728. <https://doi.org/10.1038/s41467-024-53033-7>
33. Lomoio U, Puccio B, Tradigo G, et al (2023) SARS-CoV-2 protein structure and sequence mutations: Evolutionary analysis and effects on virus variants. *PLoS One* 18:e0283400. <https://doi.org/10.1371/journal.pone.0283400>
34. Ma KC, Castro J, Lambrou AS, et al (2024) Genomic surveillance for SARS-CoV-2 variants: Circulation of Omicron XBB and JN.1 lineages - United States, May 2023-September 2024. *MMWR Morb Mortal Wkly Rep* 73:938–945. <https://doi.org/10.15585/mmwr.mm7342a1>
35. Magazine N, Zhang T, Wu Y, et al (2022) Mutations and evolution of the SARS-CoV-2 spike protein. *Viruses* 14:640. <https://doi.org/10.3390/v14030640>
36. McInnes L, Healy J, Melville J (2018) UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv [stat.ML]*
37. Mittal A, Khattri A, Verma V (2022) Structural and antigenic variations in the spike protein of emerging SARS-CoV-2 variants. *PLoS Pathog* 18:e1010260. <https://doi.org/10.1371/journal.ppat.1010260>
38. Park S-J, Kern N, Brown T, et al (2023) CHARMM-GUI PDB Manipulator: Various PDB structural modifications for biomolecular modeling and simulation. *J Mol Biol* 435:167995. <https://doi.org/10.1016/j.jmb.2023.167995>
39. Parmar A, Katariya R, Patel V (2019) A review on random forest: An ensemble classifier. In: *Lecture Notes on Data Engineering and Communications Technologies*. Springer International Publishing, Cham, pp 758–763
40. Pavia G, Quirino A, Marascio N, et al (2024) Persistence of SARS-CoV-2 infection and viral intra- and inter-host evolution in COVID-19 hospitalized patients. *J Med Virol* 96:e29708. <https://doi.org/10.1002/jmv.29708>
41. Pearson K (1901) LIII. On lines and planes of closest fit to systems of points in space. *Lond Edinb Dublin Philos Mag J Sci* 2:559–572. <https://doi.org/10.1080/14786440109462720>
42. Pedregosa F, Varoquaux G, Gramfort A, et al (2012) Scikit-learn: Machine Learning in Python. *arXiv [cs.LG]*
43. Peng C, Zhu Z, Shi Y, et al (2020) Computational insights into the conformational accessibility and binding strength of SARS-CoV-2 spike protein to human angiotensin-converting enzyme 2. *J Phys Chem Lett* 11:10482–10488. <https://doi.org/10.1021/acs.jpclett.0c02958>
44. Peng R, Wu L-A, Wang Q, et al (2021) Cell entry by SARS-CoV-2. *Trends Biochem Sci* 46:848–860. <https://doi.org/10.1016/j.tibs.2021.06.001>
45. Perahia D, Mouawad L (1995) Computation of low-frequency normal modes in macromolecules: improvements to the method of diagonalization in a mixed basis and application to hemoglobin. *Comput Chem* 19:241–246. [https://doi.org/10.1016/0097-8485\(95\)00011-g](https://doi.org/10.1016/0097-8485(95)00011-g)

46. Phillips JC, Hardy DJ, Maia JDC, et al (2020) Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J Chem Phys* 153:044130. <https://doi.org/10.1063/5.0014475>
47. Resende-Lara PT, Perahia D, Scott AL, Braz ASK (2020) Unveiling functional motions based on point mutations in biased signaling systems: A normal mode study on nerve growth factor bound to TrkA. *PLoS One* 15:e0231542. <https://doi.org/10.1371/journal.pone.0231542>
48. Ringnér M (2008) What is principal component analysis? *Nat Biotechnol* 26:303–304. <https://doi.org/10.1038/nbt0308-303>
49. Rousseeuw PJ (1987) Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math* 20:53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
50. Roy U (2023) Computational investigation of selected spike protein mutations in SARS-CoV-2: Delta, Omicron, and some circulating subvariants. *Pathogens* 13:10. <https://doi.org/10.3390/pathogens13010010>
51. Rucker G, Qin H, Zhang L (2023) Structure, dynamics and free energy studies on the effect of point mutations on SARS-CoV-2 spike protein binding with ACE2 receptor. *PLoS One* 18:e0289432. <https://doi.org/10.1371/journal.pone.0289432>
52. Saha A, Mukherjee D, Mukhopadhyay A (2025) Position-wise mutation analysis and temporal changes in SARS CoV-2 Envelope (E) protein variants. *Acad Mol Biol Genom* 2:. <https://doi.org/10.20935/acadmolbiogen8011>
53. Tian F, Tong B, Sun L, et al (2021) N501Y mutation of spike protein in SARS-CoV-2 strengthens its binding to receptor ACE2. *Elife* 10:e69091. <https://doi.org/10.7554/eLife.69091>
54. UniProt. In: UniProt. <https://www.uniprot.org/uniprotkb/P0DTC2/entry>. Accessed 30 Mar 2025
55. Verkhivker G, Agajanian S, Kassab R, Krishnan K (2022) Computer simulations and network-based profiling of binding and allosteric interactions of SARS-CoV-2 spike variant complexes and the host receptor: Dissecting the mechanistic effects of the Delta and Omicron mutations. *Int J Mol Sci* 23:4376. <https://doi.org/10.3390/ijms23084376>
56. Winger A, Caspari T (2021) The spike of concern-the novel variants of SARS-CoV-2. *Viruses* 13:1002. <https://doi.org/10.3390/v13061002>
57. Xu C, Wang Y, Liu C, et al (2021) Conformational dynamics of SARS-CoV-2 trimeric spike glycoprotein in complex with receptor ACE2 revealed by cryo-EM. *Sci Adv* 7:eabe5575. <https://doi.org/10.1126/sciadv.abe5575>
58. Yan R, Zhang Y, Li Y, et al (2020) Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* 367:1444–1448. <https://doi.org/10.1126/science.abb2762>
59. Zhang J, Xiao T, Cai Y, Chen B (2021) Structure of SARS-CoV-2 spike protein. *Curr Opin Virol* 50:173–182. <https://doi.org/10.1016/j.coviro.2021.08.010>
60. Zhao X, Chen H, Wang H (2021) Glycans of SARS-CoV-2 spike protein in virus infection and antibody production. *Front Mol Biosci* 8:629873. <https://doi.org/10.3389/fmolb.2021.629873>
61. Zhu R, Canena D, Sikora M, et al (2022) Force-tuned avidity of spike variant-ACE2 interactions viewed on the single-molecule level. *Nat Commun* 13:7926. <https://doi.org/10.1038/s41467-022-35641-3>

Figures

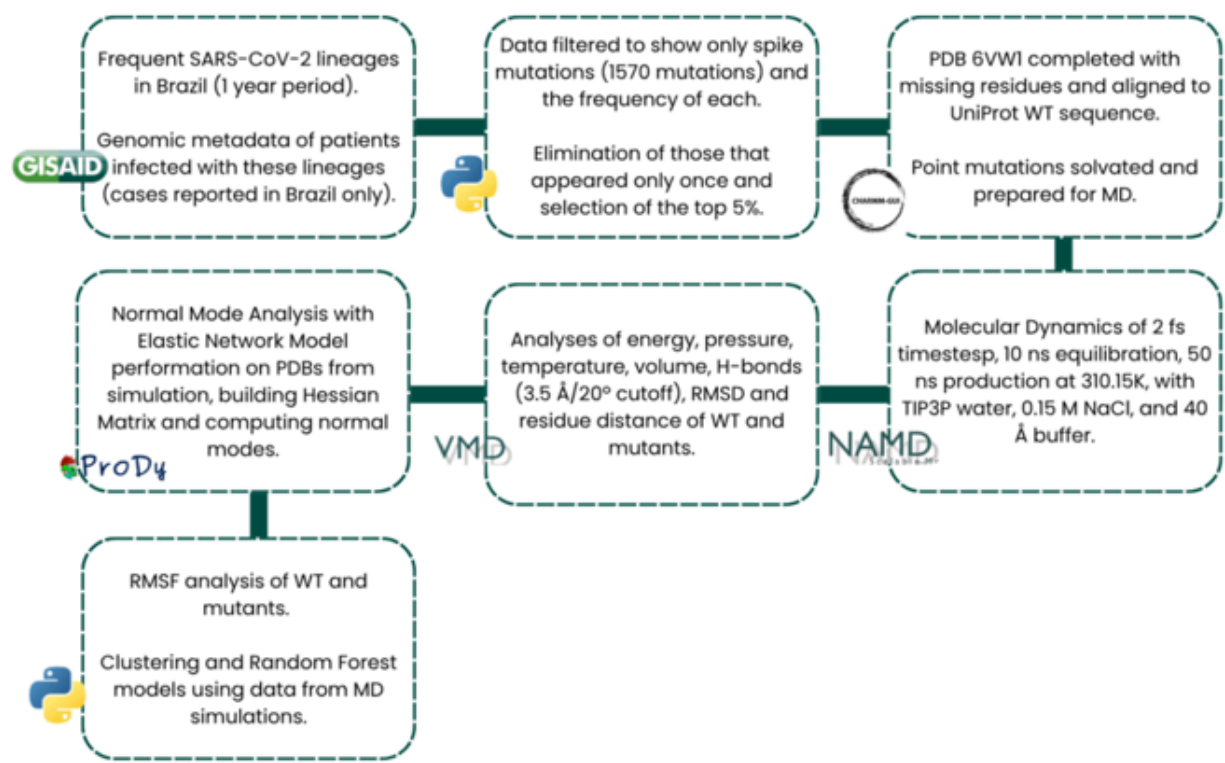


Figure 1

Data Flow Diagram of the methodology used in the study

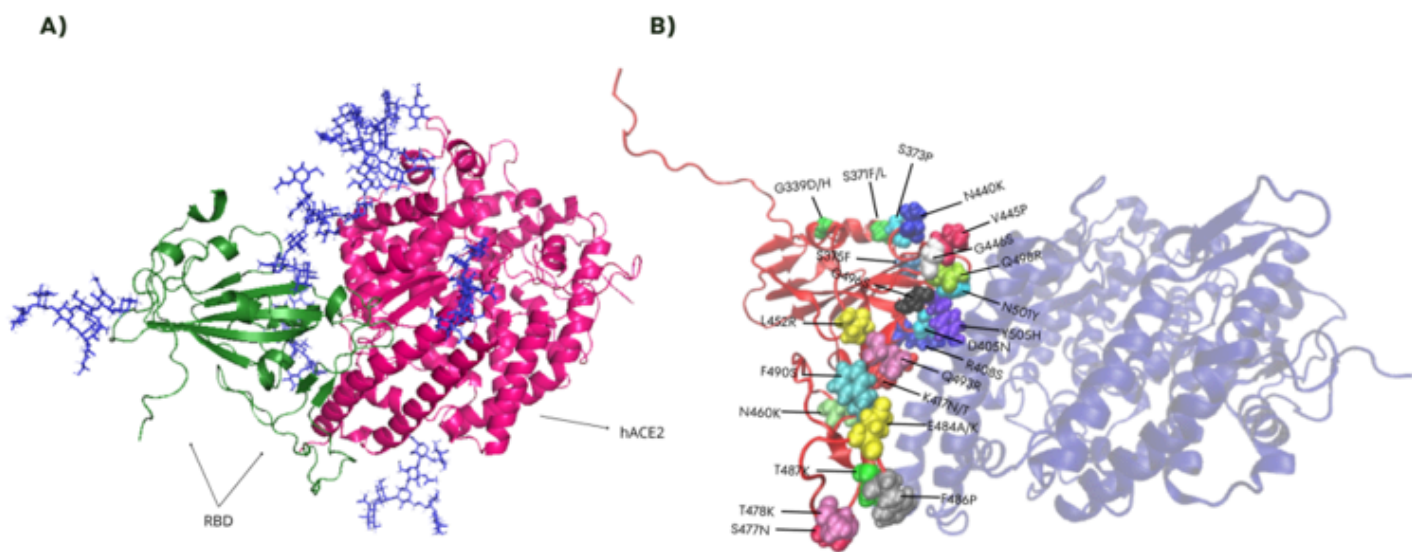


Figure 2

PDB 6VW1 and localization of the mutations in the RBD. A) PDB 6VW1 after reconstruction of the structure in CHARMM-GUI. In pink, ACE2 (PROA) and in green, the Spike RBD (PROB), with the glycans of both proteins in purple. B) WT protein after Molecular Dynamics simulation. ACE2 is in purple and RBD in red, with the localization of all the mutations used in this study in different colors

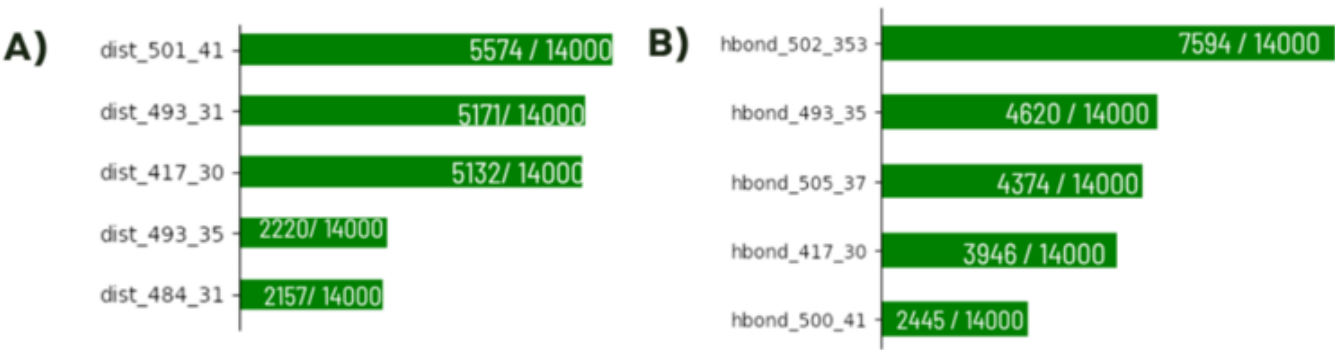


Figure 3

Top five frequencies of contacts and hydrogen bonds through the dataset A) Top five residue pairs ranked by contact frequency. All the distances were counted as 1 “contact” when ≤ 3.5 Å. Pairs 501-41, 493-31, and 417-30 showed the highest number of contacts. B) Top five residue pairs ranked by hydrogen-bond frequency. Pairs 502-353, 493-35, and 505-37 showed the highest number of HBs

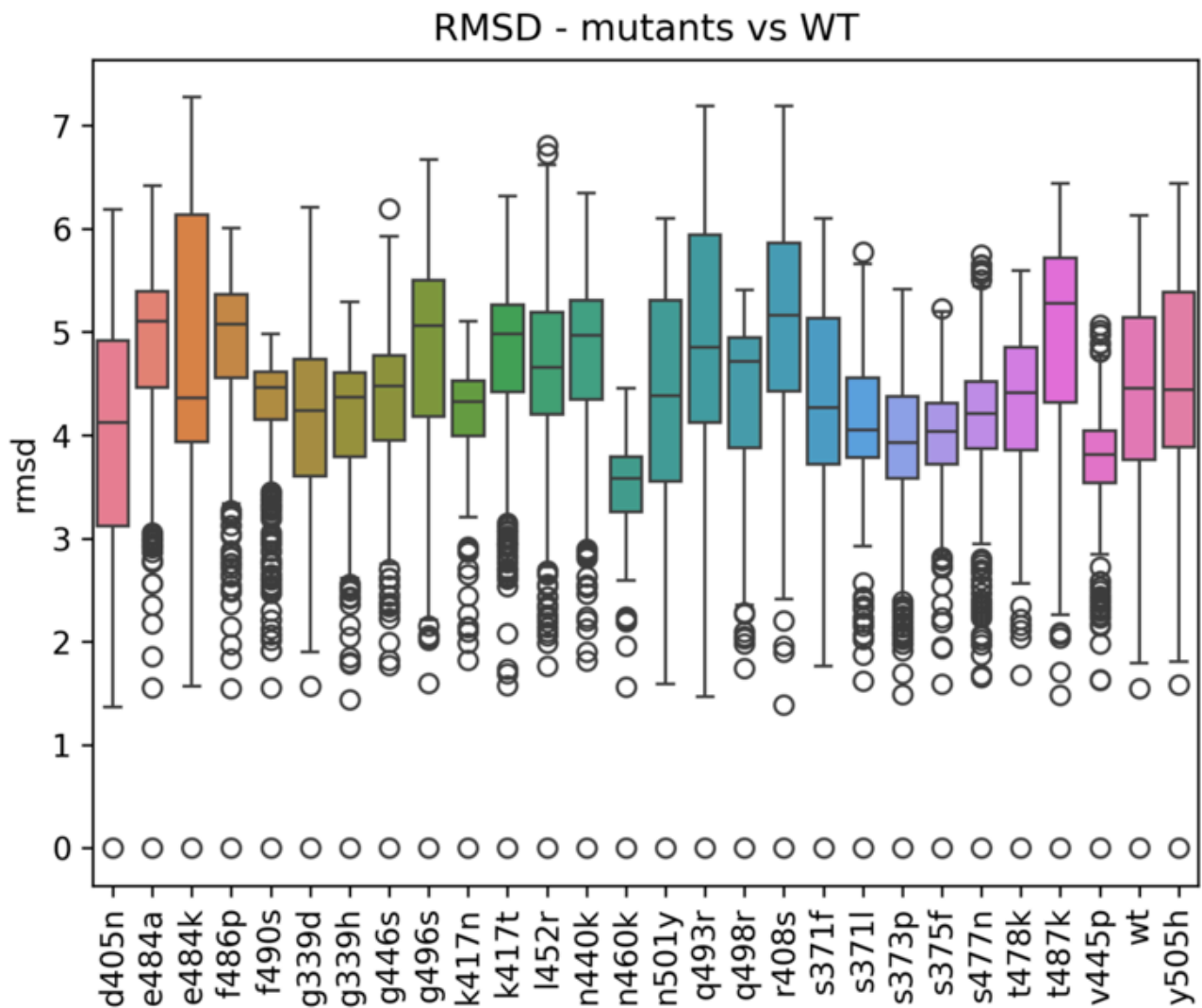


Figure 4

Boxplot representation of the RMSD of the mutants and WT. E484A, F486P, G496S, K417T, L452R, N440K, Q493R, Q498R, R408S, and T487K showed higher medians in comparison to the WT, suggesting higher flexibility

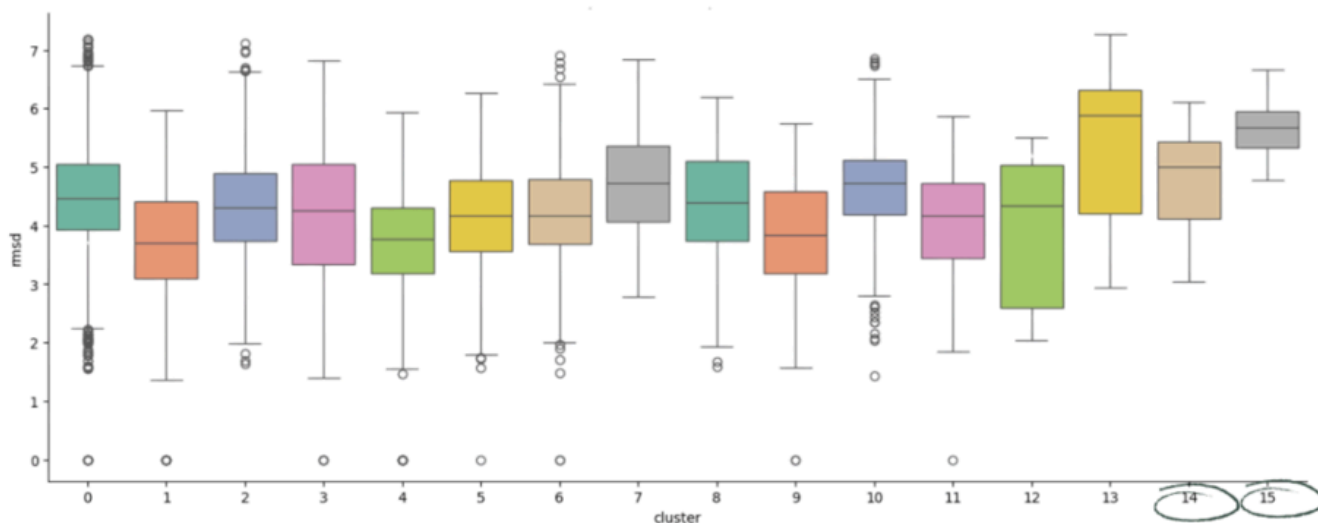


Figure 5

Clustering analysis. Distribution of RMSD values across clusters shown as boxplots. Clusters 14 and 15 showed higher median values than the rest of the clusters

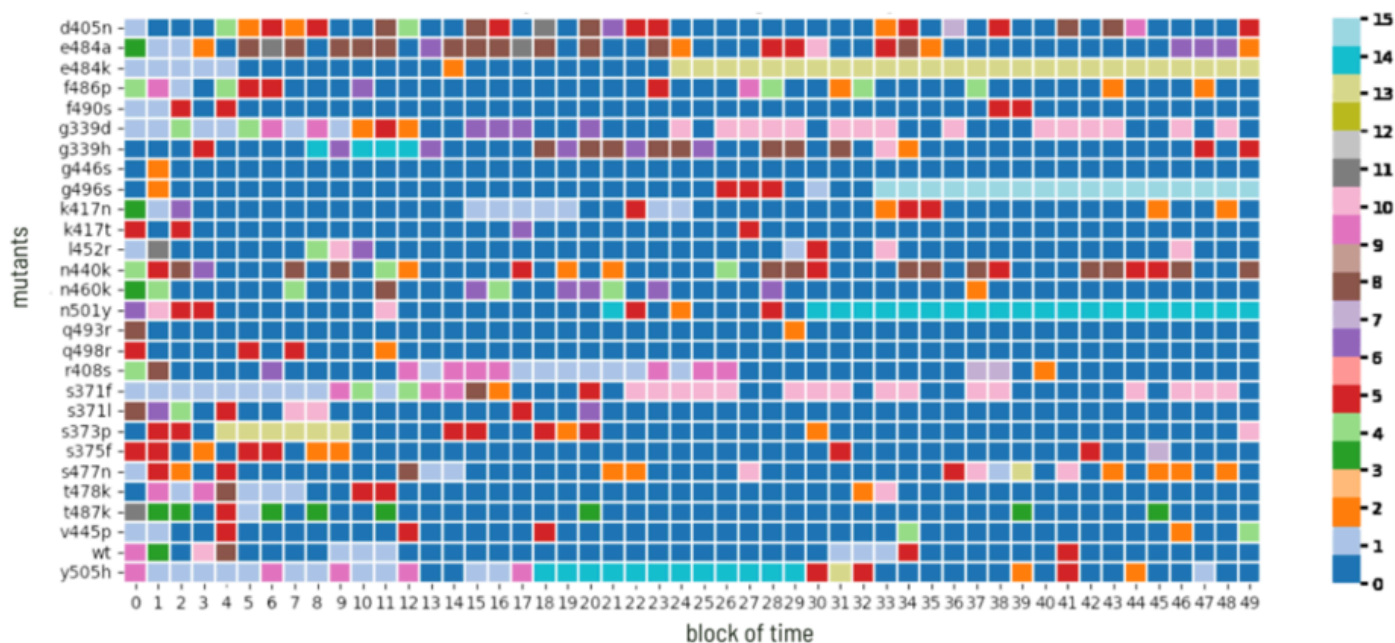


Figure 6

Clustering analysis. Time blocks of mutants and their clusters along the trajectory based on simulation frame index. Approximately 30-40 ns for N501Y and 18-30 ns for Y505H, and ~33 ns until the end of the

simulation for G496S, as representations of clusters 14 and 15, respectively

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementarymaterial.docx](#)