Comprehensive Machine Learning Analysis of Key Residues in Variants and Polymorphisms for Ace2-Spike Interaction in SARS-CoV-2

Ana Luísa Rodrigues de Ávila, Ana Carolina
Damasceno Sanches, Levy Bueno Alves,
Silvana Giuliatti

Departament of Genetics. Ribeirão Preto Medical School,
University of São Paulo, USP
Ribeirão Preto, Brazil
e-mail: ana.avila@usp.br, anacarolina2001@usp.br,
levybuenoalves@usp.br, silvana@fmrp.usp.br

Arthur Scorsolini Fares

Departament of Engineering, Computing and Exact,
University of Ribeirão Preto, UNAERP
Ribeirão Preto, Brazil
e-mail: arthur.fares@sou.unaerp.edu.br

Abstract— **The invasion of SARS-CoV-2 into host cells depends on the interaction of the Spike protein with the human angiotensin-converting enzyme 2 (Ace2). Specific Ace2 polymorphisms have been associated with increased susceptibility to SARS-CoV-2, potentially affecting the risk of infection and the severity of COVID-19. Furthermore, SARS-CoV-2 has a high probability of mutating and adapting to the environment. However, the effect of these genetic variations on the stability and affinity of the Spike-Ace2 interaction is not well understood. For a deeper understanding of this interaction, molecular dynamics simulations are used. Despite generating extensive data, these simulations do not easily facilitate the identification of essential residues that influence protein interaction. To address this challenge, we combined molecular dynamics simulations and supervised machine learning techniques to identify the residues that are subtly important in the interaction and dynamics of the complexes. The molecular dynamics simulations revealed subtle trajectory variations, emphasizing key residues and loop regions residues. While complexes show stable behavior with slight differences, machine learning techniques offer deep insights into how genetic variations in both the virus and host receptor influence the interaction region of these proteins.**

Keywords-COVID-19; Bioinformatics; Virus-host interaction; Polymorphism; Variants.

I. INTRODUCTION

On March 11, 2020, the World Health Organization characterized COVID-19 as a pandemic, identifying it as an infectious disease caused by the Severe Acute Respiratory Syndrome of Coronavirus-2 (SARS-CoV-2) [1] [2]. To date, August 2023, more than 691 million cases have been confirmed, with the global death toll surpassing 6.9 million [3]. In Brazil, listed as one of the most impacted countries, records exceed 37 million confirmed cases and almost 704 thousand deaths [4]. COVID-19 is a respiratory disease primarily transmitted through virus-containing particles that are expelled when an infected person coughs, sneezes, or talks. The severity of the disease can vary from mild cases to severe cases that can lead to Acute Respiratory Distress Syndrome and, in more serious situations, organ failure [5]. People with pre-existing comorbidities and/or who are experiencing some degree of immunosuppression are generally more susceptible to developing severe forms of the disease. Although some people may experience the severe form of COVID-19, others remain asymptomatic [5][6].

The entry of the virus into the host cell is one of the most important processes in viral infection. The virus establishes interactions with specific receptors present on the cell surface, followed by a fusion or endocytosis process, which enables the release of its genetic material into the cell cytoplasm. This viral entry step is a critical target in the development of vaccines and antiviral drugs, as inhibiting or blocking this process can effectively prevent or limit viral replication and spread. The invasion of SARS-CoV-2 into host cells depends on the interaction of the Spike protein with the human angiotensin-converting enzyme 2 (Ace2), which is present in the cell membrane. Certain polymorphisms of the Ace2 protein have been associated with increased susceptibility to SARS-CoV-2 [5][6]. These genetic variations in Ace2 can influence how effectively the virus attaches and enters into host cells, potentially affecting the risk of infection and the severity of the resulting COVID-19 disease.

Furthermore, SARS-CoV-2 has a high probability of mutating and adapting to the environment [7]. The virus has multiple Variants of Concern (VOC) during the course of the pandemic, each with specific mutations that have raised global health concerns. Notable VOCs include Omicron (B.1.1.529 – several countries), Alpha (B.1.1.7 - United States), Beta (B.1.351 – South Africa), Gamma (P.1 - Brazil) and Delta (B.1.617.2 - India) [2]. In addition, there were region-specific variants of interest, such as the P2 (or Zeta variant) (B.1.1.28.2), which was detected in the Rio de Janeiro city, Brazil, in October 2020. The mutations observed in SARS-CoV-2 variants, in conjunction with the Ace2 polymorphisms, raise questions about whether genetic variability of both the virus and the host could explain the different degrees of severity observed in infection cases.

Understanding the complex interaction between viral mutations and host genetic variations is crucial to unraveling the factors that influence disease outcomes. One of the areas of investigation is how these variations impact the stability and affinity of the Spike-Ace2 complexes, which are critical for viral entry into host cells. Certain mutations in the Spike

can increase the ability of the virus to interact more tightly with the Ace2 receptor, potentially leading to increased viral replication and infectivity. Unraveling the mechanisms by which these mutations influence viral entry and replication could open new avenues for therapeutic interventions.

Molecular dynamics (MD) simulations offer valuable information for exploring the effects of mutations and evaluating the stability and affinities between complex structures. However, the trajectories resulting from these simulations generate large amounts of data from thousands of atoms at each time interval. The analysis of complex trajectories can be performed through various approaches, including temporal trajectory analysis, evaluation of thermodynamic properties, and investigation of the bonds and interactions present.

Despite these analytical approaches, the highly dimensional nature and noisy output for the simulations present significant challenges in extracting crucial features from the trajectories. Consequently, it becomes difficult to gain a deeper understanding of molecular processes, such as regions or residues that may subtly contribute to protein interactions. Interpreting and extracting significant information from these trajectories requires robust analysis and is not a simple task.

Machine learning techniques are utilized to analyze extensive datasets, helping to identify crucial distinctions between trajectories obtained in MD simulations, even when these differences are subtle. Fleetwood et al. (2020) demonstrated the utility and potential of machine learning techniques in understanding biomolecular processes [8]. Their work involved the successful application of both supervised and unsupervised methods to investigate three distinct biological systems. In the field of viral interactions with human hosts, Pavdola et al. (2021) employed MD simulations and machine learning techniques to investigate the differences in how SARS-CoV and SARS-CoV-2 interact with the human Ace2 receptor [9]. Inspired by these studies, our research aimed to further explore the interaction between SARS-CoV-2 and Ace2, aiming to fill the knowledge gap about the interaction between viral mutations and genetic variations of the host.

In our previous study, we investigated the effects of genetic variability in SARS-CoV-2 on the interaction with wild type Ace2 [1]. Extensive MD simulations were performed to evaluate the stability of the formed protein complexes and, subsequently, supervised machine learning methods were used considering the trajectories obtained in the simulations as input data. In this study, we expand our analyzis to address the interaction of Spike variants with wild type Ace2, in addition to including new investigations into the effects of Ace2 polymorphisms. The combining of MD simulations and machine learning methods has allowed us to gain deeper insights into how genetic variations in both the virus and the host receptor can impact the region of interaction between these essential proteins.

The structure of this work is outlined as follows: Section II elucidates the methods employed at each stage of this study. Moving forward, Section III elaborates on the results and subsequent discussions, while Section IV summarizes the conclusion and describes the next steps of this research. The article concludes with acknowledgements.

II. MATERIAL AND METHODS

In this section, we will outline the methods employed to perform molecular dynamics simulations and implement machine learning architectures.

A. Molecular Dynamics

The tertiary structure of the complex Spike receptor-binding domain (RBD) and Ace2 (PDB ID: 6LZG) was obtained from the Protein Data Bank [10] and the Modeller software v9.23 [11] was used to fill the missing atoms and residues. The mutant complexes for Spike variants and Ace2 polymorphisms were generated by using the UCFS CHIMERA software version 1.14 [12]. Seven complexes Ace2_Spike-RBD were analyzed: Ace2_Spike-RBD (Wild complex), Ace2_Spike-RBD variants (Omicron, Delta and P2) and Ace2_Spike-RBD complexes with the Ace2 polymorphisms (K26R, R219C, K341R). The Ace2 polymorphisms were selected using the Genome Aggregation Database (GnomAD – https://gnomad.broad.institute.org) and Brazilian Online Mutations Archive (ABraOM – http://abraom.ib.usp.br). The selection of non-synonymous Ace2 mutations was carried out using the following criteria: amino acid residues located in the region of the peptidase domain of Ace2 (19-614 residues) and identified, according to the literature, as critical residues in the interaction between Ace2 and Spike-RBD, polymorphisms found in samples of the Brazilian population deposited in the ABraOM database and which are in high frequency in the population according to GnomAD data.

The systems were solvated in a cubic box with a minimum distance of 1.25 nm from the solute to the edge of the box. GROMACS package version 2020.5 [13] was used in the MD simulations of complexes. The force field used was CHARMM36 [14]. The molecules were solvated with TIP3P water molecules and neutralized by adding the appropriate number of Na+Cl ions considering the ionic concentration of 0.15 M. The energy minimization was performed using the steepest descent method with a maximum force of 1000 KJ.mol^{-1}.nm^{-1}. After minimization, the systems were equilibrated in two stages: a canonical NVT set followed by an isothermal-isobaric NPT set. The NVT equilibrium was performed with a constant temperature of 300 K for 500 ps. The NPT equilibrium was performed with a constant pressure of 1 bar and a constant temperature of 300 K for 500 ps. The v-rescale and Parrinello-Rahman algorithms were utilized to keep constant temperature and pressure. The production step was conducted at 300 K for 100 ns and the trajectories were saved every 10 ps.

The Root Mean Square Deviation (RMSD) and Root Mean Square Fluctuation (RMSF) were calculated as metrics to evaluate the structural stability and dynamic fluctuations of the systems. While RMSD measures the average distances between matching atoms in two structures, usually comparing frames obtained during MD simulations with the initial frame (t = 0 ns), RMSF calculates the average squared

fluctuations of atom positions in relation to their average positions throughout a simulation.

To estimate the binding energy and determine the energetic contributions of residues in protein-protein interactions, we utilized the MM/PBSA method [15]. For the MM/PBSA calculations, we included all frames from the final 10 ns of production for each complex. We selected the last 10 ns of the simulations, as during an extended simulation, the systems attain a state of dynamic equilibrium where relevant properties for calculating binding free energy become stabilized. This time window is appropriate for sampling the conformation and properties of the system, not requiring high computational power. Additionally, a shorter time interval helps minimize the effects of initial system fluctuations, allowing for a more accurate estimation of the binding energy.

B. Machine Learning

Based on Fleetwood et al. [8], correlation matrices of filtered contact maps from MD trajectories were used as inputs for supervised ML techniques. In order to reduce the influence of a single model and enhance the stability of our results, we utilized two different classification strategies: Multilayer Perceptron (MLP) and Random Forest (RF). Both methods were used to identify residues that contribute to the difference in dynamic behavior between the complexes (Fig. 1).
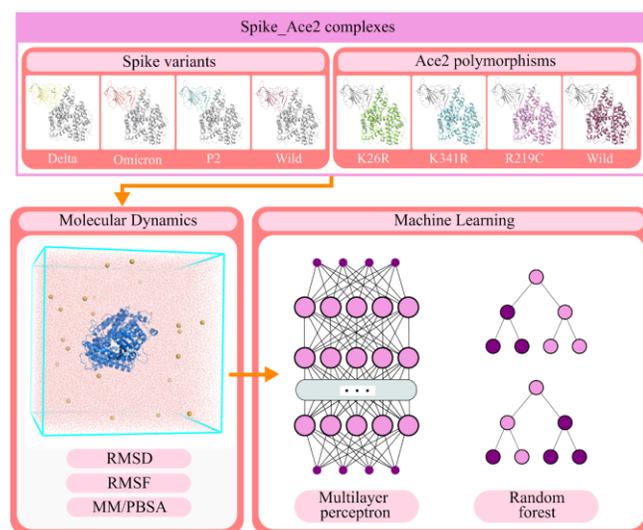
Figure 1. Workflow of the Machine Learning methods used in this study.

The MLP is an artificial neural network with multiple layers between the input and output layers. It is particularly suited for capturing complex, non-linear relationships in data. On the other hand, RF is an ensemble learning technique that builds multiple decision trees and combines their outputs through majority voting. The performance and capability of RF to manage noisy and incomplete data make it a valuable resource in various scenarios. By utilizing both RF and MLP, we ensure a comprehensive exploration of intrinsic relationships and distinctive patterns within the dataset, which ultimately results in more accurate predictions and valuable insights in our analysis.

What constitutes the input features for the MLP are correlation matrices obtained from contact distances between Ace2 and Spike-RBD residues. The distances were calculated as the minimum distance between the heavy atoms of residues in the interaction region and then filtered, leaving only the distances less than 15 Å, in order to establish a predetermined range of analysis for the studied regions. The values were then inverted and normalized to be used for the calculation of the correlation matrix, which was also filtered. Correlations over 0.9 were discarded, as the objective was to identify residues that are not easily recognized as significant contributors to the interaction.

In the MLP, four additional profiles were generated for each complex using bootstrapping, aiming to enhance the classifiers performance. As a result, five profiles were obtained for each complex.

The MLP was implemented using the open-source ML library Scikit-learn in Python [16]. We also used the data analysis and manipulation library Pandas [17], and the numerical computing library NumPy [18]. For the structure of the MLP, 8 hidden layers were used, with 100, 75, 50, 40, 30, 20, 10, and 5 neurons respectively, each with the ReLU function as activation. The labels (Ace2_Spike-RBD complex) were one-hot encoded to represent categorical data numerically and the training process used the Adam optimizer [19] to adjust node weights. A train-test split was applied, with 80% of the data in the training set and 20% in the test set.

This network was trained with each of the profiles, resulting in 5 total MLPs for the complexes with Spike-RBD variants and Wild type (WT) and 5 for the complexes with the Ace2 polymorphisms and WT.

Since the classification task itself does not directly indicate which were the important features that influenced the prediction, an explanation algorithm was applied for the model. The one selected was the Layer-Wise Relevance Propagation (LRP) [20] with the LRP-0 rule. This algorithm indicates which inputs had the most impact on a specific prediction made by the model, obtaining this through the allocation of a normalized relevance score to each individual feature. Therefore, making the decision-making of the neural network more transparent.

The RF classifier, also implemented with Scikit-learn, receives as inputs the distances matrix, since it uses an internal bootstrapping process to produce consistent profiles, and the number of decision trees was set to 100. Our model utilized the Gini impurity coefficient, ranging from zero to one. Zero indicating a pure split and one indicating maximum impurity. We aimed to select splits that would lower Gini impurity, resulting in more homogeneous distribution of classes within the leaves of the tree.

To calculate the importance of a specific state in the RF model, the one-versus-the-rest approach was employed. This strategic method decomposes the problem into multiple binary classification instances and endeavors to discriminate each individual case.

III. RESULTS AND DISCUSSION

The outcomes achieved at each step of our work will be detailed in the subsequent subsections.

A. Analysis of trajectory stability

The simulation data was used to compute RMSD for Ace2_Spike-RBD complexes, considering two distinct approaches: the first one considering the interaction between Ace2 WT and Spike-RBD variants; and the second one analyzing the interaction between Ace2 polymorphisms and Spike-RBD WT. Fig. 2 shows the RMSD values for the Ace2 and Spike-RBD proteins.
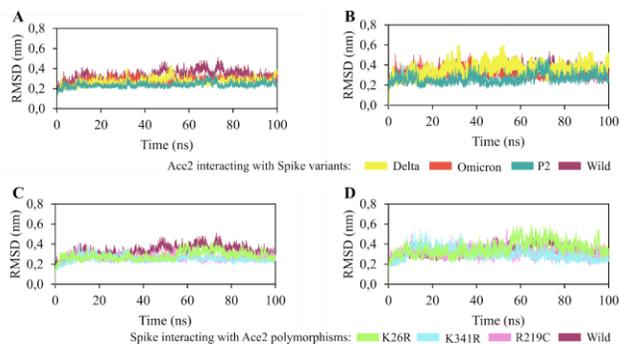
Figure 2. Analysis of the RMSD trajectories obtained in the MD simulations. (A) RMSD of Ace2 WT interacting with Spike-RBD variants; (B) RMSD of Spike-RBD variants interacting with Ace2 WT; (C) RMSD of Ace2 polymorphisms interacting with Spike-RBD WT; (D) RMSD of Spike-RBD WT interacting with Ace2 polymorphisms.

In the Ace2 chain trajectory (Figs. 2A and 2C) involving the interaction with Spike-RBD, the RMSD showed similar values (between 0.2 nm and 0.4 nm) and low standard deviation, implying their stability. The trajectory of the Ace2 WT had a subtle higher RMSD value compared to the other complexes, revealing greater structural variation over the analyzed time. The Spike-RBD trajectories (Figures 2B and 2D) of all complexes remained in equilibrium, with RMSD values between 0.2 and 0.6 nm. All analyzed trajectories of the Spike-RBD complex interacting with Ace2 polymorphisms similarly exhibited stability throughout the simulation.

Stable trajectories indicate that the simulation is converging to an equilibrium state, where the properties of the systems stop showing significant variations. This stability enhances the reliability and precision of the simulation data, providing valuable insights into the behavior and interactions of the studied complexes.

B. Analysis of the Atomic Position Variation

The RMSF represents the degree of variation in the position of a given atom during the course of time. Higher values of RMSF per residue characterize greater flexibility, and vice versa [21]. The RMSF results of the interaction between Ace2 WT and Spike-RBD variants, along with the interaction between Ace2 polymorphisms and Spike-RBD WT, are presented and discussed in sections 1 and 2, respectively.

1) Mobility Analysis of Ace2 WT Associated with Spike-RBD Variants.: The RMSF analysis revealed the residues that exhibited the most significant fluctuations in the trajectories of the Ace2 protein (Fig. 3A) and Spike-RBD (Fig. 3B). On the Ace2 trajectories, the most pronounced fluctuations occur predominantly in loop regions, specifically in residues near Pro138, Gln287, Asn290, Gln340, and Phe428 (Fig. 3C). Notably, Gln340 exhibited the highest peak in the trajectory of the Delta variant (0.67 nm), with a difference of 0.36 nm from the other trajectories. Among these residues, Gln340 is the only one located in a loop relatively close to the Spike-RBD.
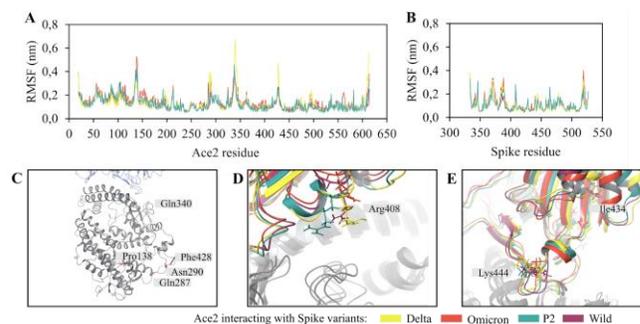
Figure 3. Analysis of residual RMSF from Ace2 protein interacting with Spike-RBD variants. (A) Residual RMSF of Ace2 WT in the Ace2_Spike-RBD complex; (B) Residual RMSF of Spike-RBD variantes in the Ace2_Spike-RBD complex; (C) Fluctuations in Ace2 WT loop regions; (D) Fluctuations in Arg408 residue of Spike-RBD variants; (E) Fluctuations in residues Ile434 and Lys444 of Spike-RBD variants.

Regarding the Spike-RBD trajectories (Fig. 3B), notable fluctuations were observed for the Arg408, Ile434 and Lys444 residues. The RMSF values showed that Arg408 residue of the Delta variant obtained a slightly lower RMSF value (0.13 nm) compared to the WT (0.21 nm), Omicron (0.25 nm) and P2 (0.28 nm) variant complexes. This residue is situated in an alpha-helix, near the interaction interface of Spike-Ace2 (Fig. 3D).

Arg408 is adjacent to residue 417, which has mutated into the Omicron variant, resulting in an amino acid switch from lysine to asparagine (K417N). In the Spike WT, Lys417 forms a very stable salt bridge with the aspartate at residue 30 of the Ace2 receptor. The replacement of lysine for asparagine or threonine largely disrupts binding at this position, as it induces a loss of the salt bridge at this position [22].

As for the Omicron variant complex, the Ile434 residue showed a slightly higher fluctuation (0.12 nm) compared to the Spike-RBD WT (0.7 nm), in addition to Delta (0.7 nm) and P2 (0.6 nm) variants. Ile434 is located in a beta-sheet, but it is not close to the Spike-RBD (Fig. 3E). Isoleucine, being a non-polar amino acid, plays an important role in the structural stabilization of proteins due to hydrophobic interactions within its interior. Furthermore, Ile434 is close to the S375F and N440K mutations in the Omicron variant.

Such mutations decrease the protein stability, which may explain the greater fluctuation observed in this variant [22].

The Lys444 residue of the Delta variant had the highest fluctuation peak of 0.20 nm, followed by the Omicron variant (0.16 nm), P2 variant (0.16 nm), and Spike-RBD WT (0.14 nm). Lys444 is located in a loop close to Gly446, Tyr449, Gln498, Thr500 and Asn501 in Spike WT, which are involved in polar interactions with Ace2 (Fig. 3E) [23]. The L452R mutation of the Delta variant, which is relatively close to the Lys444 residue, reduces protein stability [24].

2) Mobility Analysis of Spike WT Associated with Ace2 Polymorphisms: The RMSF results for Ace2 polymorphisms and Spike-RBD are shown in Figs. 4A and 4B, respectively. The most prominent RMSF fluctuations are observed in loop regions, which were previously highlighted in the previous section and are represented in Fig. 3C. However, a few subtle fluctuations have been identified near the Ace2 polymorphisms regions, such as those observed in Gln325, Trp328, and Arg582.

The Gln325 residue is situated in an alpha-helix near the Spike-RBD and exhibited a slightly reduced fluctuation in the trajectory of the complex with the K341R polymorphism (0.18 nm), followed by WT (0.26 nm), R219C (0.27 nm), and the largest, K26R (0.30 nm). Similarly, Trp328 also showed a smaller fluctuation in the trajectory of the K341R polymorphism, with a difference of 0.16 nm compared to the other complexes. The tryptophan residue is located in an alpha-helix next to the Spike-RBD. Both Gln325 and Trp328 are positioned near the K341R polymorphism (Fig. 4C).

The residue Arg582 of the Ace2 WT obtained a slightly lower RMSF value (0.07 nm) compared to the complexes with the polymorphisms K26R (0.25 nm), R219C (0.20 nm) and K341R (0.19 nm). Residue Arg582 is located in an alpha helix close to the R219C polymorphism (Fig. 4D).
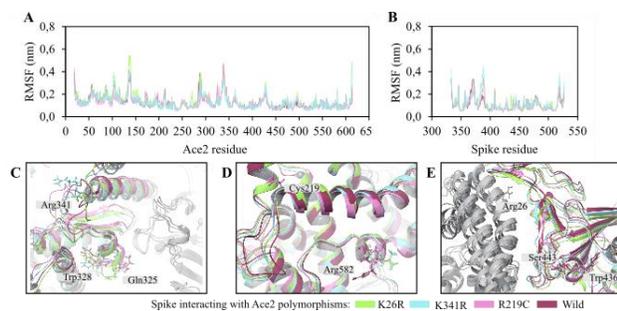
Figure 4. Analysis of Residual RMSF from Spike WT interacting with Ace2 polymorphisms. (A) Residual RMSF of Ace2 polymorphisms in the Ace2_Spike-RBD complex; (B) Residual RMSF of Spike-RBD in the Ace2_Spike-RBD complex; (C) Fluctuations in residues Gln325 and Trp328 of Ace2 polymorphisms; (D) Fluctuations in Arg582 residue of Ace2 polymorphisms; (E) Fluctuations in Trp436 and Ser443 residues of Spike-RBD.

In the RMSF plot of the Spike-RBD chain (Fig. 4B), the Ace2_Spike-RBD complex featuring the K26R polymorphism demonstrated a slightly elevated RMSF value at residue Trp436 within the Spike protein (with a 0.1 nm

difference) compared to complexes involving other polymorphisms. The Arg26 polymorphism represents a polar residue located within an alpha helix, near to the Spike-RBD, while Trp436 is an aromatic amino acid with non-polar characteristics, positioned within a beta sheet structure (Fig. 4E). Additionally, the complex with the G211R polymorphism exhibited marginally higher RMSF values in the Ser443 residue, showcasing a difference of 0.27 nm in relation to the other polymorphisms. Situated adjacent to Spike-RBD within an alpha helix, the residue Ser443 emerges as an uncharged polar element.

C. MM/PBSA Binding Free Energy Analysis

The MM/PBSA calculation was performed to estimate the binding energies between the Ace2 and Spike proteins, along with to comprehend the factors contributing to the stability or instability of the interaction. The binding energy values are summarized in Table 1, revealing that the Omicron variant demonstrates a more affinity with Ace2, showing a binding energy of -2572.23 ± 144.92 KJ.mol^{-1}, followed by the Delta and P2 variants, which record values of -1875.71 ± 132.97 and -1837.21 KJ.mol^{-1}, respectively. The analysis of MM/PBSA residual energy decomposition has revealed that it is the electrostatic interactions that predominantly influence the stability of the Spike-RBD variants in the protein-protein interaction. On the other hand, the values obtained for the Van der Waals energy components, solvation polar energy and SASA did not show significant variations. Regarding the interaction of Spike-RBD WT with Ace2 polymorphisms, the estimated values for binding energy did not exhibit significant variations. This can be attributed to the point-wise changes in Ace2 polymorphisms, which were not sufficient to differentiate the binding interaction energy.

TABLE I.   RESIDUAL DECOMPOSITION AND BINDING ENERGY (GIVEN IN KJ.MOL^{-1}) IN THE PROTEIN-PROTEIN INTERACTION.

ACE2	SPIKE	ΔEvdW	ΔEelec	Polar sovatation energy	SASA energy	Binding energy
Wild	Wild	-321.60 ± 22.54	-1429.29 ± 59.43	785.31 ± 107.76	-40.91 ± 3.42	-1006.49 ± 115.09
Wild	Delta	-315.86 ± 19.37	-2046.18 ± 73.33	527.73 ± 131.85	-41.40 ± 3.56	-1875.71 ± 132.97
Wild	Omicron	-328.23 ± 20.67	-2866.46 ± 96.49	664.30 ± 135.75	-41.83 ± 3.54	-2572.23 ± 144.92
Wild	P2	-322.53 ± 24.99	-2190.35 ± 71.59	719.82 ± 115.05	-44.16 ± 3.58	-1837.21 ± 119.87
K26R	Wild	-323.33 ± 26.13	-1243.89 ± 76.66	612.92 ± 157.23	-41.41 ± 3.32	-995.71 ± 140.33
K341R	Wild	-333.92 ± 21.91	-1394.38±78.38	801.89 ± 159.18	-42.52 ± 3.73	-968.94 ± 171.39
R219C	Wild	-337.37 ± 20.15	-1433.61 ± 65.65	669.35 ± 174.18	-41.37 ±4.13	-1142.99 ± 161.39

Fig. 5 illustrates the per-residue energy decomposition through MM/PBSA of the analyzed complexes. Although Ace2 residues showed subtle variations in interaction with the Spike variants (Fig. 5A), variations within Spike residues themselves were more pronounced (Fig. 5B). Specifically, the residues at positions 408, 417, 440, 452, 478, 484, 493, and 498 exhibited significant energetic variations.

In the Spike-RBD chain (Fig. 5B), the Omicron variant shows a slightly decreased binding energy at the Arg408 residue in comparison to WT and other variants. Furthermore, Lys417 residue in the Omicron variant shows a weaker binding energy than in the other complexes (-4.72 KJ.mol^{-1}), with a discrepancy greater than 200 KJ.mol^{-1}. As previously mentioned, the substitution of lysine with asparagine, as observed in the Omicron variant, disrupts a

highly stable salt bridge interaction between Lys417 and Ace2 residue 30, consequently destabilizing the complex. In addition, the N440K mutation present in the Omicron variant records a value of -225,52 KJ.mol^{-1}, differing by more than 200 KJ.mol^{-1} from the other complexes.
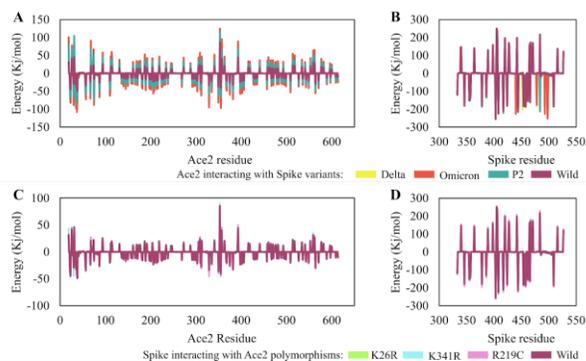
Figure 5. MM/PBSA per-residue energy decomposition for the trajectories obtained in the MD simulations. (A) Contribution energy of Ace2 WT residues in interaction with Spike-RBD variants; (B) Contribution energy of Spike-RBD variants residues in interaction with Ace2 WT; (C) Contribution energy of Ace2 polymorphism residues in interaction with Spike-RBD WT; (D) Contribution energy of Spike-RBD WT residues in interaction with Ace2 polymorphisms.

In the Delta variant complex, the L452R mutation showed a binding energy of -200.19 KJ.mol^{-1}, while in the other complexes the residue reached values around -1 KJ.mol^{-1}. The residue L452R is located in the hydrophobic region of the Spike protein and does not interact with the Ace2 receptor. However, there is a possibility that the mutation induces structural changes by promoting its interaction with the Ace2 receptor [25].

The T478K mutation, identified in the Delta and Omicron variants, resulted in binding energies of -186.69 and -175.78 KJ.mol^{-1}, respectively. In contrast, the WT and P2 complexes showed binding energies around -1 KJ.mol^{-1}. Replacement of the polar and uncharged threonine residue with the positively charged basic amino acid lysine increases the electrostatic potential contribution within the Spike-RBD, promoting stronger affinity with the Ace2 receptor. Furthermore, the elongated lysine side chain could increase the steric effects of the Delta variant, potentially elucidating the increased interaction between Spike-RBD and the Ace2 receptor [26].

The Glu484 residue is mutated in the Omicron variant, involving the replacement of glutamine by alanine, and likewise in the P2 variant, but with the replacement of glutamine by lysine. Notably, the E484K mutation in the P2 variant produced a higher binding energy (-211.87 KJ.mol^{-1}), contrasting with the E484A mutation observed in the Omicron variant (-1.76 KJ.mol^{-1}). The E484A mutation within the Omicron variant abolishes the weak binding of Glu484 to Ace2 in the WT, while mitigating the destabilization arising from conceivable electrostatic repulsion between Glu484 from WT and Glu35 from Ace2 after transition to Ala484. Consequently, the E484A

mutation has no influence on the binding free energy, in contrast to the E484K mutation observed in P2, which increases the interaction.

Q493R and Q498R are mutations present in the Omicron variant that showed a reduction in binding energy, with values of -226 and -252 KJ.mol^{-1}, respectively, with a difference of more than 200 KJ.mol^{-1} from the Delta and WT. The combination of the Q498R and N501Y mutations significantly increases Ace2 binding capacity due to the formation of two new strong salt bridges between Arg493 and Arg498 of Omicron, and Glu35 and Asp38 in Ace2 [27].

Regarding the complexes between Ace2 polymorphisms and Spike-RBD WT, no significant values were observed in the MM/PBSA results for the residues in question (Figs. 5C and 5D). However, the residues Val343 and His345, located in the Ace2 chain of the complex containing the K341R polymorphism, showed slightly weaker binding energy compared to the other complexes (Fig. 5C). Val343 and His345 are situated close to the K341R polymorphism, all within the loop region and relatively near the Spike-RBD.

D. Discriminatory Residues of Spike Variants through Machine Learning

Table 2 shows the five most significant residue pairs for each complex. In the MLP, the importance values of each residue pair were derived by calculating the average of the associated LRP-0 attributes. On the other hand, in the RF, the importance was evaluated based on the reduction of Gini impurity. Key residues responsible for variations in binding between Spike-RBD variants and Ace2 WT were identified, some of them already reported in previous studies.

TABLE II.   IMPORTANT RESIDUES OBTAINED FROM MLP AND RF FOR SPIKE VARIANTS INTERACTING WITH ACE2 WT. THE PAIRS HIGHLIGHTED IN BOLD ARE SUPPORTED BY THE LITERATURE AND ARE DISCUSSED IN THIS STUDY.

Spike variants	MLP			Spike variants	RF		
	ACE2	SPIKE	Importance		ACE2	SPIKE	Importane
Wild	SER106	GLY485	1.00	Wild	SER19	VAL483	1.00
	VAL107	PHE486	0.99		SER19	CYS488	0.95
	GLN89	SER477	0.98		SER19	CYS480	0.60
	SER19	**PRO479**	**0.89**		SER44	TYR505	0.52
	ALA71	GLU484	0.82		SER19	GLN474	0.52
Delta	ASP30	GLU484	1.00	Delta	ALA36	ASN501	1.00
	GLN24	LYS417	0.71		GLY66	ASN501	0.98
	GLY352	ARG408	0.68		ALA342	THR500	0.93
	ALA65	SER443	0.62		ASN103	TYR505	0.71
	ASN33	GLN498	0.60		LYS68	ASN501	0.64
Omicron	GLU329	SER438	1.00	Omicron	**ALA25**	**ASN417**	**1.00**
	GLN42	SER349	0.96		**GLN24**	**ASN417**	**0.98**
	TYR381	GLY502	0.92		**ILE21**	**ASN417**	**0.97**
	GLY352	ASN448	0.91		**LYS353**	**ARG498**	**0.94**
	GLY354	GLY504	0.90		**THR27**	**ASN417**	**0.88**
P2	PRO321	ARG403	1.00	P2	**SER106**	**LYS484**	**1.00**
	SER19	**ASN477**	**0.92**		SER19	CYS480	0.90
	SER19	**PRO479**	**0.87**		SER105	ASN487	0.87
	GLN325	SER371	0.84		GLY104	ASN487	0.80
	GLU37	THR415	0.81		**SER105**	**LYS484**	**0.79**

Fig. 6 illustrates the importance values of the pairs obtained through MLP (Fig. 6.A) and RF (Fig. 6B), highlighting the most relevant residues that distinguish the variants from the WT (Figs. 6C-E).
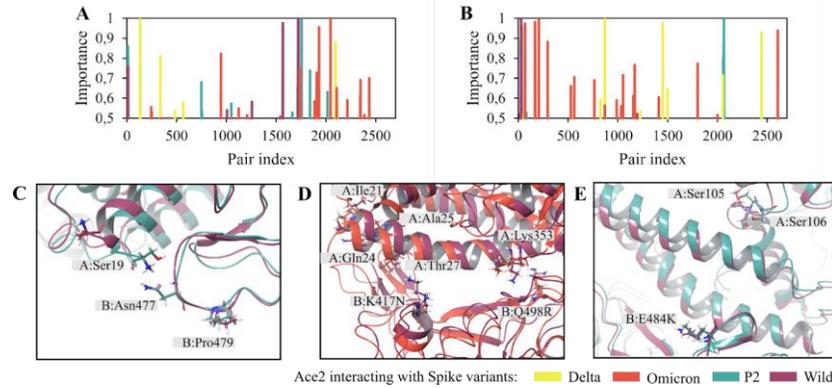
Figure 6. Pairs of residues determined as most important for distinguishing the binding between Ace2 WT and Spike-RBD variants. (A) Pairs identified by MLP; (B) Pairs identified by RF; (C) Pairs of important residues found for Spike-RBD WT and P2; (D) Pairs of important residues found for Spike-RBD WT and Omicron; (E) Pairs of important residues found for Spike-RBD WT and P2. Important residues are composed of highly distinct pairs of distances between Ace2 WT and the studied Spike-RBD variants.

The residue pairs identified by the RF model differed from those identified by the MLP model. However, some residues were identified by both methods. The analysis of the results emphasizes the importance of the main residues of Ace2 in the interaction with Spike-RBD, as mentioned by Ali and Vijayan (2020), which include Gln24, Thr27, Asp30, Glu37, Gln42, and Lys353 [28]. Furthermore, the Ser19 residue of Ace2 protein, which was commonly seen among pairs, is also important. Ser19 participates in a network of hydrogen interactions, particularly interacting with Pro462 [29].

The results obtained by the MLP suggest that Ser19 can subtly interact with residues near Pro462, forming pairs with Pro479 in the Spike-RBD WT, and Asn477 and Pro479 in the P2 variant. The S477N mutation, present in the P2 variant, may favor a greater interaction with Ser19 of Ace2 protein. In addition, a mutation S19P increased the interaction between Ace2 and Spike-RBD [30]. This suggests that the Ser19 residue plays a critical role in modulating the interaction between Ace2 and Spike, potentially influencing the infection capability and viral transmissibility of the P2 variant. The mentioned pairs may be of importance in distinguishing between the binding of Spike-RBD WT and P2 to Ace2 (Fig. 6C).

In the RF model, the mutations K417N and Q498R of the Omicron variant formed interesting residue pairs. The K417N mutation may have significant effects with residues Ile21, Gln24, Ala25, and The27, while the Q498R mutation may affect Lys353 of Ace2 protein (Fig. 6D). Studies indicate that the K417N mutation leads to a reduction in the binding affinity of the Spike to Ace2; however, it is an immune escape mutation, which helps SARS-CoV-2 escape the natural immune defenses of the host, contributing to increased viral transmissibility. On the other hand, Q498R is associated with an increase in viral infection. The Lys353-Arg498 pair had an importance of 0.94. The study by Zhang et al. (2022) suggests that Q498R is structurally incompatible with Lys353 in Ace2, but is structurally adapted to Asp38 [31].

The E484K mutation, found in the P2 variant, is associated with a reduction in neutralizing antibodies. This mutation results in a tighter binding interface between Spike-RBD and Ace2 protein, contributing to an increase in binding affinity [32]. Our results indicate the presence of two pairs related to this residue, namely Ser106-Lys484 with a significance of 1.00, and Ser105-Lys484 with a significance of 0.79 (Fig. 6E).

E. Discriminatory Residues of Ace2 polymorphisms through Machine Learning

Table 3 shows the five pairs responsible for variations in the binding between Ace2 polymorphisms and Spike-RBD. Both the RF and MLP models identified identical residue pairs, such as Phe356-Tyr495 in K26R, and Gly104-Phe486 and Ser105-Tyr489 in K341R. For ACE2 WT and R219C polymorphism, the residue pairs identified by the RF model differed from those identified by the MLP model. However, the residue Phe486 (in R219C) and Gln325 (in WT) were identified in both methods. Notably, the Ser19 residue in the Ace2 protein received high scores for all the polymorphisms, including K26R, K341R, and R219C identified by MLP, in addition to the WT identified by RF.

TABLE III.   IMPORTANT RESIDUES OBTAINED FROM MLP AND RF FOR ACE2 POLYMORPHISMS INTERACTING WITH SPIKE-RBD WT. THE PAIRS HIGHLIGHTED IN BOLD AND MARKED WITH ASTERISKS APPEARED IN BOTH MLP AND RF ANALYSES. THE REMAINING BOLD PAIRS ARE ALSO DISCUSSED IN THIS STUDY.

MLP				RF			
ACE2 polymorphisms	ACE2	SPIKE	Importance	ACE2 polymorphisms	ACE2	SPIKE	Importance
K26R	SER19	SER477	1.00	K26R	**PHE356***	**TYR495***	**1.00***
	SER19	ASP467	0.91		LYS353	TYR495	0.94
	PHE356*	**TYR495***	**0.80***		GLY352	TYR495	0.94
	LEU351	TYR495	0.41		ASP355	ASN439	0.79
	ASN194	PHE486	0.39		ASP350	TYR495	0.66
K341R	**GLY104***	**PHE486***	**1.00***	K341R	**GLY104***	**PHE486***	**1.00***
	SER19	PRO479	1.00		GLN102	ASN487	0.79
	ALA193	PHE486	0.86		**SER105***	**TYR489***	**0.78***
	SER105*	**TYR489***	**0.80***		LYS26	GLY504	0.73
	ASN103	TYR505	0.77		GLN102	TYR489	0.65
R219C	SER19	PRO479	1.00	R219C	SER105	TYR489	1.00
	ASP38	LYS444	0.54		GLY104	**PHE486**	0.98
	PHE32	GLY446	0.43		GLN102	GLY485	0.91
	PHE40	TYR495	0.43		ASN103	GLY485	0.86
	ALA193	**PHE486**	0.41		ASN103	TYR489	0.84
Wild	ASN103	TYR505	1.00	Wild	SER19	ALA475	1.00
	ASN103	TYR489	0.85		**GLN325**	THR500	0.96
	ASN103	GLY485	0.80		SER19	GLN474	0.87
	HIS401	TYR505	0.74		**GLN325**	GLN498	0.81
	GLN325	TYR449	0.54		**GLN325**	ASN448	0.76
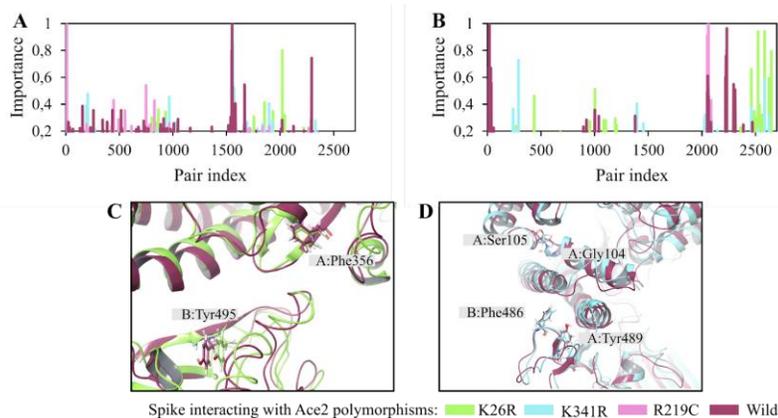
Figure 7. Pairs of residues determined as most important for distinguishing the binding between Spike-RBD WT and Ace2 polymorphisms. (A) Pairs identified by MLP; (B) Pairs identified by RF; (C) Pairs of important residues found for Ace2 WT and K26R; (D) Pairs of important residues found for Ace2 WT and K341R. Important residues are composed of highly distinct pairs of distances between Ace2 polymorphisms and the Spike-RBD.

Fig. 7 shows the significance of the pair values obtained using MLP (Fig. 7A) and RF (Fig. 7B), emphasizing the key residues that differentiate the variants from the WT, as demonstrated in Figs. 7C and 7D.

Ace2 K26R increases susceptibility to SARS-CoV-2 due to a higher binding affinity with the Spike protein [30]. In this polymorphism, Phe326 forms a pair with Tyr495. Phe356 is located near residues Tyr41, Gln42, Lys353, and Arg357, which interact with the Spike-RBD, while Tyr495 participates in a hydrogen bonding network with Ace2. However, there is currently no literature available on the specific importance of Phe356 in the K26R polymorphism when interacting with the Spike-RBD.

The K341R mutation, which replaces lysine with arginine, results in a larger mutant residue, which can cause protrusions [33]. Our research findings highlight significant variations in the pairs Gly104-Phe486 and Ser105-Tyr489. The roles of some of these residues have been documented by Ali et al. (2020) [28]. Specifically, Phe486 participates in important polar interactions with Tyr83 and hydrophobic interactions with Leu79 of Ace2. Tyr489 is involved in polar interactions with Thr27 and Lys31, as well as hydrophobic interactions with Phe28, Tyr83, Thr27, Phe32, and Phe72 from Ace2.

IV. CONCLUSIONS AND FUTURE WORK

The interaction between the Spike and Ace2 proteins plays a crucial role in determining the replication rate of SARS-CoV-2 and has implications for the disease progression in infected individuals. The virus exhibits a pronounced propensity for mutations, as evidenced by the emergence of various variants in recent years. Ace2 genetic polymorphisms have the potential to influence susceptibility to the disease, along with its subsequent intensity and clinical outcome. However, a comprehensive understanding of how mutations and polymorphisms impact the stability and interaction dynamics within the SARS-CoV-2-Ace2 complex remains an ongoing effort.

In our research, we combined MD simulations and machine learning techniques to explore the interaction between different SARS-CoV-2 variants and human Ace2 polymorphisms. Through these simulations, we obtained valuable information about the protein-protein interaction. Concurrently, employing machine learning techniques allowed us to pinpoint critical amino acid residues within the binding region that subtly contribute to this interaction.

The MD simulations revealed similar stability patterns among the studied complexes. Furthermore, the resulting trajectories indicated a convergence of the simulations into an equilibrium state. The stability of the Ace2 protein complex with Spike-RBD WT was slightly diminished, as evidenced by the RMSD values, in contrast to the SARS-CoV-2 variant complexes. This observation is consistent with the anticipated effect of mutations in the Spike interaction region leading to increased stability.

Although the most significant fluctuations were observed in loop regions, some residues near the interaction interface exhibited notable fluctuations. Arg408 and Lys444 of the Spike-RBD showed slightly higher RMSF values in the Delta variant. Gln325 and Trp328 residues of the Ace2 protein showed lower fluctuations in the trajectories of the K341R polymorphism, whereas Trp436 and Ser443 exhibited higher fluctuations for WT and G211R, respectively. Significant changes in RMSF in these regions may suggest important conformational alterations for the biological activity in the Ace2-Spike interaction.

The Omicron variant demonstrates a stronger affinity with Ace2, as evidenced by the MM/PBSA values, where the Q493R and Q498R mutations contributed more significantly to the binding energy. Regarding the complexes formed between Ace2 polymorphisms and the Spike-RBD, no significant differences were identified in the MM/PBSA results for the considered residues. These point polymorphisms were not sufficient to generate detectable notable variations using the MM/PBSA method.

Regarding the Machine Learning methods, we achieved a precision score of 1 and loss values below 0.005 for both

approaches using the test dataset. The high precision and low loss on the test data suggest that the model is performing well, but they do not ensure the absence of overfitting. A more comprehensive evaluation, utilizing other data sources such as cross-validation, is necessary to determine the presence of overfitting.

The ML and RF approaches successfully identified key residues from both proteins responsible for differences in binding region, some of which have been previously reported in the literature. This demonstrates that our method was able to identify residues that significantly contribute to the distinction between virus and host interaction due to Spike variants and Ace2 polymorphisms, extending even to those pairs of residues that have been not previously documented in the existing literature.

Our study shows that machine learning can simplify the complexity of virus-host interactions by reducing dimensionality and identifying crucial residues. Our findings indicate that there may be additional important residues beyond those previously considered above that may impact the interaction between Spike and Ace2 proteins. These residues may account for differences in stability and affinity, leading to varying levels of susceptibility to SARS-CoV-2 and resulting in varying degrees of disease severity. In our work, we aim to gain a deeper understanding of the relationship between mutations and the affinity between Spike-Ace2 by not only exploring other variants, but also incorporating various machine learning methods.

REFERENCES

[1] A. C. D. Sanches, A. L. R. de Ávila, L. B. Alves, and S. Giuliatti, "Machine Learning in the Identification of Key Residues of Variants and Polymorphisms in the Interaction of ACE2 Proteins with Spike of SARS-CoV2", pp. 7-11, Mar. 2023 [BIOTECHNO 2023, The Fifteenth International Conference on Bioinformatics, Biocomputational Systems and Biotechnologies].

[2] World Health Organization. WHO Director-General's Opening Remarks at the Media Briefing on COVID-19 - 11 March 2020. [Online]. Available from: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020/ (Accessed: 04 December 2023).

[3] Worldometers. COVID-19 CORONAVIRUS PANDEMIC. [Online]. Available at: https://www.worldometers.info/coronavirus (Accessed: 04 December 2023).

[4] World Health Organization. Tracking SARS-CoV-2 variants. [Online]. Available at: https://www.who.int/activities/tracking-SARS-CoV-2-variants (Accessed: 04 December 2023).

[5] S. Choudhary, K. Sreenivasulu, P. Mitra, S. Misra, and P. Sharma, "Role of Genetic Variants and Gene Expression in the Susceptibility and Severity of COVID-19," Annals of Laboratory Medicine, vol. 41(2), pp. 129–138, Mar. 2021, doi: 10.3343/alm.2021.41.2.129.

[6] J. Zepeda-Cervantes et al., "Implications of the Immune Polymorphisms of the Host and the Genetic Variability of SARS-CoV-2 in the Development of COVID-19," Viruses, vol. 14(1), pp. 1–34, Jan. 2022, doi:10.3390/v14010094.

[7] R. Peng, L. -A. Wu, Q. Wang, J. Qi, and G. F. Gao, "Cell entry by SARS-CoV-2," Trends in Biochemical Sciences, vol. 46(10), pp. 848–860, Oct. 2021, doi: 10.1016/j.tibs.2021.06.001.

[8] O. Fleetwood, M. A. Kasimova, A. M. Westerlund, and L. Delemotte, "Molecular Insights from Conformational Ensembles via Machine Learning," Biophysical Journal, vol. 118(3), pp. 765–780, Feb. 2020, doi:10.1016/j.bpj.2019.12.016.

[9] A. Pavdola et al., "Machine Learning Reveals the Critical Interactions for SARS-CoV-2 Spike Protein Binding to Ace2", J. Phys. Chem. Lett., vol. 12(13), pp. 5494–5502, Jun. 2021, doi:10.1021/acs.jpclett.1c01494.

[10] H. M. Berman et al., "The Protein Data Bank," Nucleic Acids Research, vol. 28(1), pp. 235–242, Jan. 2000, doi:10.1093/nar/28.1.235.

[11] A. Sali and T. L. Blundell, "Comparative protein modelling by satisfaction of spatial restraints," J. Mol. Biol., vol. 234(1), pp. 779–815, Dec. 1993, doi:10.1006/jmbi.1993.1626.

[12] E. F. Petersen et al., "UCSF Chimera—A visualization system for exploratory research and analysis", Journal of Computational Chemistry, vol. 25(13), pp. 1605-1612, Oct. 2004, doi:10.1002/jcc.20084.

[13] D. V. D. Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, "GROMACS: Fast, flexible, and free", Journal of Computational Chemistry, vol. 26(16), pp. 1701-1718, Dec. 2005, doi:10.1002/jcc.20291.

[14] J. Huang and A. D. MacKerell, "CHARMM36 all‐atom additive protein force field: Valida-tion based on comparison to NMR data," J. Comput. Chem., vol. 34(1), pp. 2135-2145, Sep. 2013, doi:10.1002/jcc.23354.

[15] R. Kumari and R. Kumar, "Open Source Drug Discovery Consortium et al. g_mmpbsa—A GROMACS Tool for High-Throughput MM-PBSA Calculations", Journal of Chemical Information Model, v. 54(7), pp. 1951-1962, May. 2014, doi:10.1021/ci500020m.

[16] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of machine learning research, vol. 12, pp. 2825–2830, 2011, doi:10.48550/arXiv.1201.0490.

[17] W. McKinney, "Data Structures for Statistical Computing in Python," Proceedings of the Ninth Python in Science Conference (PPSC 2010), pp. 56–61, doi:10.25080/Majora-92bf1922-00a.

[18] C. R. Harris et al., "Array programming with NumPy," Nature, vol. 585, pp. 357–362, 2020, doi:10.1038/s41586-020-2649-2.

[19] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," The Third International Conference for Learning Representations (ICLR 2015), SAN DIEGO, Jan. 2017, doi:10.48550/arXiv.1412.6980.

[20] G. S. Montavon, S. Bach, A. Binder, W. Samek, and K. -R. Müller, "Explaining nonlinear classification decisions with deep Taylor decomposition," Pattern Recognition. vol. 65, pp. 211–222, May. 2017, doi:10.1016/j.patcog.2016.11.008.

[21] A. Bornot, C. Etchebest, and A. G. de Brevern, "Predicting protein flexibility through the prediction of local structures", Proteins, vol. 79(3), pp. 839-852, Mar. 2011, doi:10.1002/prot.22922.

[22] S. Kumar, T. S. Thambiraja, K. Karuppanan, and G. Subramaniam, "Omicron and Delta variant of SARS-CoV-2: A comparative computational study of Spike protein," Journal

of Medical Virology, vol. 94(4), pp. 1641-1649, Apr. 2022, doi:10.1002/jmv.27526.

[23] X. Duan et al., "A non-Ace2-blocking neutralizing antibody against Omicron-included SARS-CoV-2 variants", Signal Transduction and Targeted Therapy, vol. 7(23), pp. 1-3, Jan. 2022, doi:10.1038/s41392-022-00879-2.

[24] T. B. Mahmood et al., "Missense mutations in Spike protein of SARS-CoV-2 delta variant contribute to the alteration in viral structure and interaction with hAce2 receptor", Immun Inflamm Dis., vol. 10(9), Sep. 2022, doi:10.1002/iid3.683.

[25] X. Deng et al., "Transmission, infectivity, and neutralization of a Spike L452R SARS-CoV-2 variant.", Cell., vol 184(13), pp. 3426-3437, Jun. 2021, doi:10.1016/j.cell.2021.04.025.

[26] S. D. Giacomo, D. Mercatelli, A. Rakhimov, and F. M. Giorgi, "Preliminary report on severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) Spike mutation T478K", J Med Virol., vol. 93(9), pp. 5638-5643, Sep. 2021, doi:10.1002/jmv.27062.

[27] B. Jawad, P. Adhikari, R. Podgornik, and W. -Y. Ching, "Binding Interactions between RBD of Spike-Protein and Human Ace2 in Omicron variant", bioRxiv 2022.02.10.480009, 2022, doi:10.1101/2022.02.10.480009.

[28] A. Ali and R. Vijayan, "Dynamics of the Ace2–SARS-CoV-2/SARS-CoV Spike protein interface reveal unique mechanisms", Sci Rep, vol. 10(14214), pp. 1-12, Aug. 2020, doi:10.1038/s41598-020-71188-3.

[29] J. de Andrade, P. F. B. Gonçalves, and P. A. Netz, "Why Does the Novel Coronavirus Spike Protein Interact so Strongly with the Human Ace2? A Thermodynamic Answer", Chembiochem, vol. 22(5), pp. 865-875, Mar. 2021, doi:10.1002/cbic.202000455.

[30] K. Suryamohan et al., "Human Ace2 receptor polymorphisms and altered susceptibility to SARS-CoV-2," Communications Biology, vol. 4(1), pp. 1–11, Apr. 2021, doi:10.1038/s42003-021-02030-3.

[31] W. Zhang et al., "Structural basis for mouse receptor recognition by SARS-CoV-2 omicron variant", PNAS, vol. 119(44), pp. e2206509119, Oct. 2022, doi: 10.1073/pnas.220650911.

[32] W. B. Wang et al., "E484K mutation in SARS-CoV-2 RBD enhances binding affinity with hAce2 but reduces interactions with neutralizing antibodies and nanobodies: Binding free energy calculation studies", J Mol Graph Model, vol. 109(108035), pp. 1-15, Dec. 2021, doi:10.1016/j.jmgm.2021.108035.

[33] J. A. Rodríguez et al., "Computational modeling of the effect of five mutations on the structure of the Ace2 receptor and their correlation with infectivity and virulence of some emerged variants of SARS-CoV-2 suggests mechanisms of binding affinity dysregulation," vol. 368, pp. 110244–110244, Dec. 2022, doi:10.1016/j.cbi.2022.110.