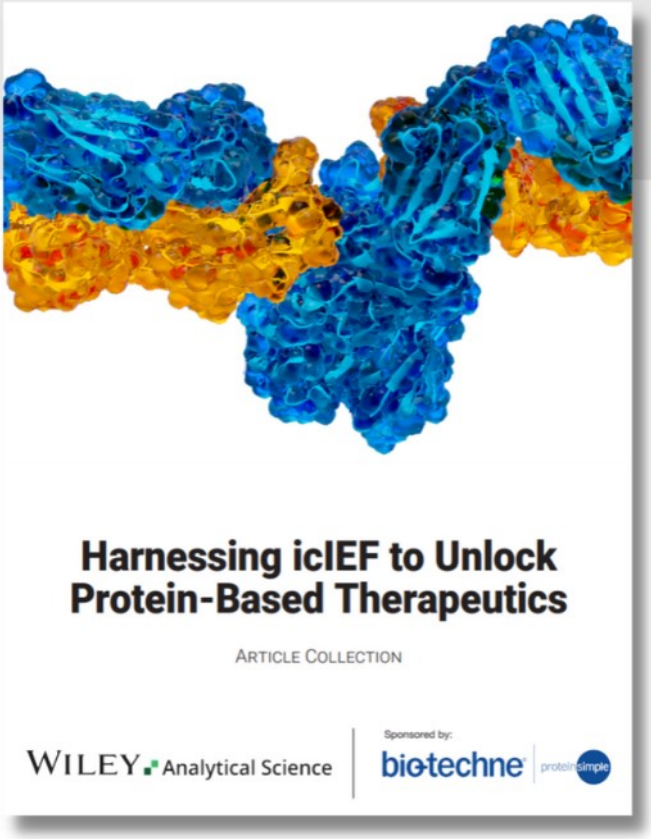
Harnessing icIEF to Unlock Protein-Based Therapeutics



Read the new Article Collection

Keep up to date with the latest developments in biotherapeutics and the range of treatments for various diseases with our latest article collection. Find out how imaged cIEF (icIEF) technique is essential for quality control and analytical development of these drugs, as it accurately determines the surface charge of lipid nanoparticles and the charge heterogeneity of proteins and antibodies.

This article collection aims to provide you with more information on these techniques and technologies, helping you further your research in this field.

bio-techne® | protein simple

WILEY

Check for updates

Received: 27 September 2022 | Revised: 20 March 2023 | Accepted: 1 April 2023

DOI: 10.1002/prot.26497

RESEARCH ARTICLE

PROTEINS WILEY

Dynamical changes of SARS-CoV-2 spike variants in the highly immunogenic regions impact the viral antibodies escaping

Lorenzo Di Rienzo[1] | Mattia Miotto[1] | Fausta Desantis[1,2] | Greta Grassmann[3] | Giancarlo Ruocco[1,4] | Edoardo Milanetti[1,4]

[1]Center for Life Nano-& Neuro-Science, Istituto Italiano di Tecnologia, Rome, Italy

[2]The Open University Affiliated Research Centre at Istituto Italiano di Tecnologia, Genoa, Italy

[3]Department of Biochemical Sciences "Alessandro Rossi Fanelli", Sapienza University of Rome, Rome, Italy

[4]Department of Physics, Sapienza University of Rome, Rome, Italy

Correspondence
Lorenzo Di Rienzo, Center for Life Nano-& Neuro-Science, Istituto Italiano di Tecnologia, Viale Regina Elena 291, 00161 Rome, Italy.
Email: lorenzo.dirienzo@iit.it

Funding information
European Research Council Synergy grant ASTRA, Grant/Award Number: 855923

Abstract

The prolonged circulation of the SARS-CoV-2 virus resulted in the emergence of several viral variants, with different spreading features. Moreover, the increased number of recovered and/or vaccinated people introduced a selective pressure toward variants able to evade the immune system, developed against the former viral versions. This process results in reinfections. Aiming to study the latter process, we first collected a large structural dataset of antibodies in complex with the original version of SARS-CoV-2 Spike protein. We characterized the peculiarities of such antibodies population with respect to a control dataset of antibody-protein complexes, highlighting some statistically significant differences between these two sets of antibodies. Thus, moving our attention to the Spike side of the complexes, we identify the Spike region most prone to interaction with antibodies, describing in detail also the energetic mechanisms used by antibodies to recognize different epitopes. In this framework, fast protocols able to assess the effect of novel mutations on the cohort of developed antibodies would help establish the impact of the variants on the population. Performing a molecular dynamics simulation of the trimeric form of the SARS-CoV-2 Spike protein for the wild type and two variants of concern, that is, the Delta and Omicron variants, we described the physicochemical features and the conformational changes experienced locally by the variants with respect to the original version. Hence, combining the dynamical information with the structural study on the antibody-spike dataset, we quantitatively explain why the Omicron variant has a higher capability of escaping the immune system than the Delta variant, due to the higher conformational variability of the most immunogenic regions. Overall, our results shed light on the molecular mechanism behind the different responses the SARS-CoV-2 variants display against the immune response induced by either vaccines or previous infections. Moreover, our analysis proposes an approach that can be easily extended to both other SARS-CoV-2 variants or different molecular systems.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
© 2023 The Authors. *Proteins: Structure, Function, and Bioinformatics* published by Wiley Periodicals LLC.

1 | INTRODUCTION

Since late 2019, the Coronavirus Disease 2019 (COVID-19), a condition involving the human respiratory system, has been causing a worldwide pandemic.[1,2] The causative agent, the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), belongs to the family of beta coronaviruses and represents the third highly pathogenic coronavirus with a zoonotic origin that emerged in humans causing respiratory illness.[3,4] Despite the social distancing measures introduced several times around the world, COVID-19 infection is still largely present and has caused, according to World Health Organization on August 18, 2022, over 600 million cases and 6 million deaths.

The SARS-CoV-2 cell entry mechanism relies on the viral Spike (S) glycoprotein, which forms a homotrimeric structure that contacts receptors on the cell surface.[5,6] Since S protrudes from the viral capsid exposing itself to the external environment, unsurprisingly this protein is the human immune system's main target.[7] In addition, almost all the designed vaccines, both protein, and mRNA based, have focused on it.[8,9] Indeed, antibodies (also called immunoglobulins) are Y-shaped molecules that play a key role in the immune system, used to identify any nonself pathogens, such as SARS-CoV-2. The two tips of the molecule host two twin antigen binding sites, each one formed by the pairing of two chains, called heavy and light chains. Each antigen binding site is formed by six hyper-variable loops of complementary determining regions (CDRs), three being part of the heavy and three of the light chain. It is important to note that the capability of the antibodies to recognize virtually any nonself antigen is due to the high sequence variability they experience in CDRs, while the global architecture of the molecule is conserved.[10–12]

In light of these considerations, a very large portion of the antibodies elicited against SARS-CoV-2, both caused by infection or vaccination, actually target S protein.[13] In particular, structural studies have shown that the antibodies preferentially bind to specific regions on S protein,[14] located in the N-terminal domain (NTD), the region apt to bindings with sialoside molecules,[15] or in the receptor binding domain (RBD), the region where S contacts its main cellular receptor, ACE2.[16,17] As for the other coronaviruses, mutations in the SARS-CoV-2 genetic code randomly occur in viral replication, where the ones that increase the fitness are preserved giving rise to new variants.[18–22] For instance, concerning the original line of the SARS-CoV-2, one of the first registered mutations regards the amino acid substitution D614G in S protein. Established in March 2020, this mutation allowed the Spike RBD to assume a conformation more suitable for binding ACE2 and rapidly became dominant.[23–25] Indeed, RNA viruses are characterized by a low replicative fidelity: this allows the adaptation to different environments and evolutive pressure, in turn enabling them to escape the host immunity.[18,26] In this scenario, during the spreading pandemic several SARS-CoV-2 variants have emerged: some of them, such as Delta (B.1.612.2) and Omicron (B.1.1.529) variants, have been defined as "variants of concerns" (VOC) by the World Health Organization.[27,28] In this framework, one of the major issues of the emerging variants is the escaping ability from the immune system, which has developed antibodies against a different version of the virus.[29,30] Indeed, the initial infections and the vaccination campaign have originated an immunity against the original version of the S protein. This protection can be endangered if the emerged variants are characterized by many mutations on the S protein, especially if these mutations sensibly alter the physical–chemical properties of antibody-targeted S regions.[30–32]

Interestingly, in the past years, several computational approaches have been attempted to predict the structural determinants of the S-antibody recognition and the effects that mutations can have on them. For instance, molecular dynamics studies highlighted that epitope regions on S are characterized by low-intensity energetic coupling with the rest of the structure[33] and that the RBD rigidity can explain the increased affinity of this virus compared to SARS-CoV.[34] Molecular dynamics have been also utilized to predict the molecular mechanisms driving the virulence of emerging SARS-CoV-2 variants[35] and to understand how the epitope regions were affected by these variants.[36] In addition to this, integrating a very large amount of heterogeneous data, Wang et al. discuss the viral evolution mechanism and forecast the next possible vaccine-breakthrough variants.[30] These data are then used in an artificial intelligence model, proving its effectiveness by comparing its prediction with the actual data about the emerging variants in the past year.[37,38]

In this work, we explore from a structural point of view the antibody-S interactions, identifying the main peculiarities of such bindings. We also relate them with the main differences S protein registers when Delta and Omicron variants are considered.

First, we collected two structural datasets of antibodies in interaction with their antigens. The first one, termed *Spike Dataset*, collects 297 nonredundant complex structures involving antibodies recognizing SARS-CoV-2 S protein; the second one, *General dataset*, is composed of 684 nonredundant complexes of antibodies binding nonspecific proteins. By analyzing these two datasets in terms of sequence and structure, we highlighted some differences in S-binding antibodies, in terms of sequence, CDR length, and antigen-contacting residues.

Thus, examining the Spike dataset we identified the S region most frequently recognized by antibodies. Indeed, we associated each S residue with an *immunogenicity index*, reflecting the number of complexes whose residues are in interaction with the antibody. Moreover, such analysis allowed us to define three classes of anti-S antibodies, according to the S region they bind. Calculating the nonbonded van der Walls and Coulomb energies at the molecular interface, we underlined some differences in the binding mechanisms between these three classes.

Hence, we asked how the mutations in the S sequence carried by the variants can impact this scenario, locally modifying the characteristics of the exposed regions of the S protein. In particular, we selected Delta and Omicron variants and we studied through molecular dynamics simulations the trimeric form of both the original and the two variants of S protein. Adopting a molecular dynamics approach allowed us to consider, beyond the short-range effects of the mutations, also the dynamics, and the long-range protein behavior.

Along the trajectory, we investigated which regions experience the highest difference between the mutated and the original S

10970134, 2023, 8, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/prot.26497 by University Di Roma La Sapienza, Wiley Online Library on [28/11/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

protein, in terms of shape similarity and hydropathy. To study the shape we adopted a method we recently developed employing the Zernike polynomials formalism.[39] In this method, each protein surface region is associated with an ordered set of numerical descriptors, defining the shape geometry of that molecular patch. Such characterization, independent of the orientation of the protein, permits an easy comparison between molecular patches and has already proven its efficacy in the evaluation of patch similarity or complementarity.[40–49] In addition to this, we studied the local changes in chemical properties of S surface regions. Using the residue hydropathy scale introduced by Di Rienzo et al.[50] each surface point can be indeed labeled with the value of hydrophobicity of the residues generating it. The hydrophobicity index of a patch is then defined as the mean of the values associated with the points included in that surface region.

Starting from the shape and hydrophobicity characterizations, we defined two *variability indexes* reflecting the changeability of each Spike residue. These indexes, which can be computed independently for Delta or Omicron variants, allowed us to compare the chemicophysical properties in the variants with respect to the original S protein. Lastly, we related the variability indexes with the immunogenicity index, to give further insights into the escaping ability these two variants exhibit against antibodies elicited against the original version of S protein.

2 | RESULTS AND DISCUSSION

2.1 | Sequence and structural analysis of anti-spike antibodies

In this section, we analyze the peculiar traits characterizing anti-S protein antibodies.

Starting from the very first phases of the pandemic, the interaction between Spike and antibodies has been studied. In this work,[51] polyclonal immunoglobulins from COVID-19 convalescent individuals have been characterized, providing some preliminary considerations regarding the antibodies side of the interface. Thus, the effort of the scientific community has been mainly focused on the description of the most immunogenic regions of the Spike protein, and how the variant's amino acid substitutions can impact the binding with known antibodies.[52–55] A complete overview of the various antibody-spike complexes experimentally determined has been furnished in this recent paper, with particular attention to the epitope region the antibodies are bound to.[56]

However, we focused on a different aspect. We study antibody sequences (using the Chotia numbering scheme[10,11]), the number of residues composing CDRs, and the position of the antigen-contacting residues. To highlight eventual originalities, we select two distinct structural datasets composed of experimental structures of antibody–antigen complexes: in the *Spike dataset*, 298 antibody-S complexes are present while in the *General dataset*, we gather 684 antibody–antigen structures, to work as a control.

The results of these analyses are shown in Figure 1.

First, we investigate if the sequences of anti-Spike antibodies are statistically different with respect to the control sequences (antibodies binding proteins different from SARS-CoV-2 Spike). In particular, Figure 1. A shows the logo representations of both the heavy and the light antibodies chain belonging to the *Spike dataset*, obtained with a multiple sequence alignment and the WebLogo application.[57] For each residue in the antibody sequence, the probability to find the indicated amino acid in that position is indicated by the height of the corresponding letter. In Figure 1B, we report the positions where the differences with the control case are more evident. In particular, focusing on the positions populated almost in 75% of the antibodies (both in the *Spike dataset* and in the *General dataset*) we select the cases where it occurs a 25% difference in an amino acid frequency between the two datasets. It turns out that some important differences exist in positions H5, H40, H60, H83, and L43: compared to the control case, the anti-S antibodies use with more probability ALA in H40, H60, and L43 while they prefer VAL and ARG in H5 and H83, respectively.

Next, we focus our attention on the length of antibodies CDRs. For each of the six loops, we build the histogram regarding the number of residues composing the loop, using separately the *Spike dataset* and the *General dataset*. We thus compare the loop length distributions using the Kolmogorov–Smirnov statistical test. We report in Figure 1C the loops that result to be different (p-value $<$ .01): it can be noted that anti-spike antibodies typically employ an H2 CDR shorter than the control, while both L1 and L3 are characterized by a higher number of residues.

Lastly, we identify the antibody residues that are more involved in the interaction with the antigen (intuitively, in the *Spike dataset* the antigen is always the S protein). In our approach, an antibody residue is in contact with the antigen if its CA atom is closer than 8 Å to any antigen CA atom. In Figure 1D, we report the result of this analysis. In the left bar plots, we report the probability of antigen interaction regarding each antibody position, as obtained using the *Spike dataset*. As expected, both for heavy (upper panel) and light (lower panel) chains, the interaction with the antigen is mediated by the CDRs, as three well-separated peaks emerge from the plot. Interestingly, in the right panels, we show the comparison between the results in the *Spike dataset* and the *General dataset*: the residues with the most marked differences are reported in the upper and the lower panel for heavy and light chains, respectively. It has to be noted that anti-S protein antibodies use H1 residues to contact the antigen with more frequency than in the control case, while H3 is less preferred. Analogously, the interaction involving L1 residues is more common in anti-S antibodies than in the general case.

2.2 | Preferential epitopes and energy of spike-antibodies interaction

Here, we focus our attention on the S protein side of the complexes, investigating where the S regions most prone to antibody binding are located and which the energetic mechanisms for the recognition are.

10970134, 2023, 8, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/prot.26497 by University Di Roma La Sapienza, Wiley Online Library on [28/11/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

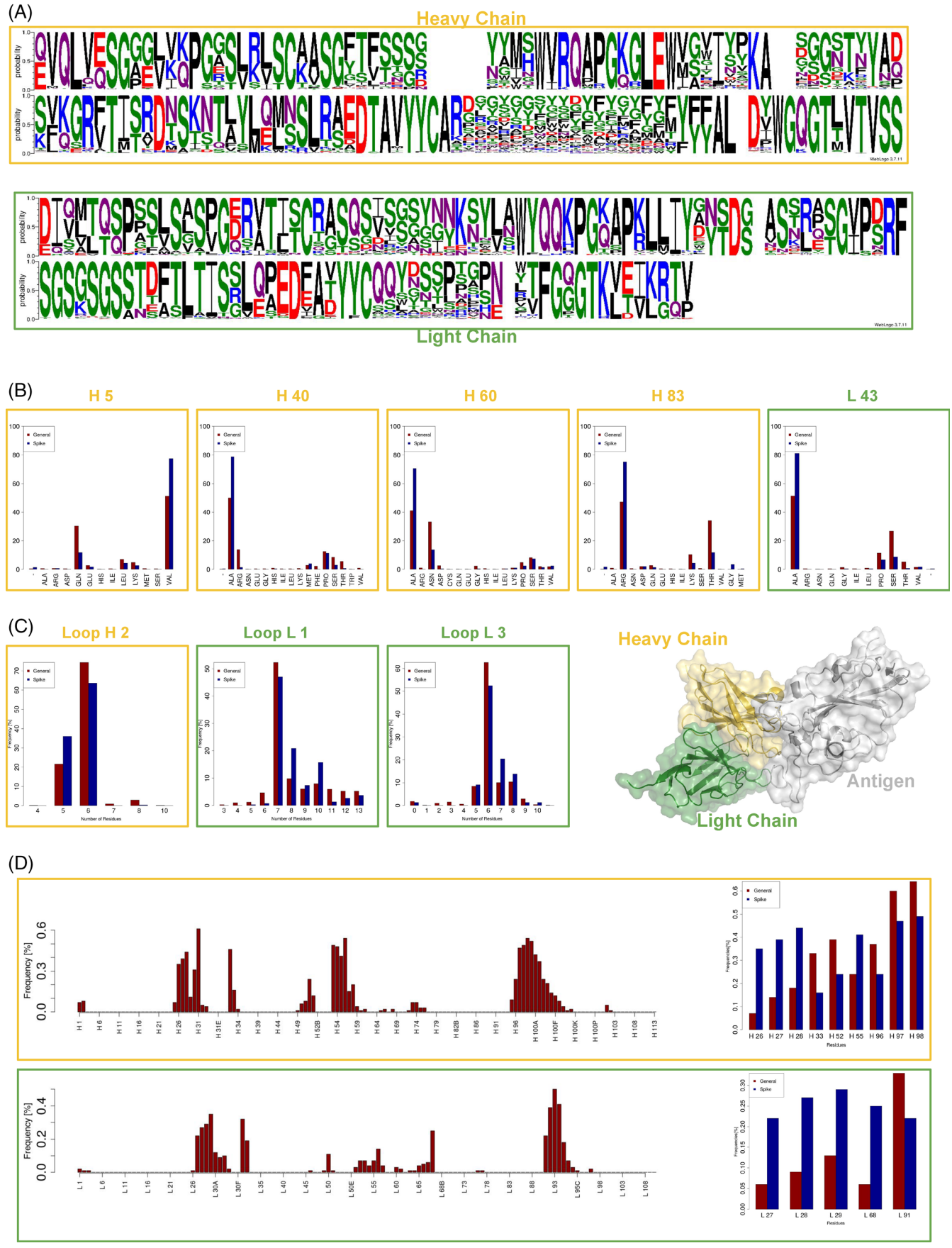


FIGURE 1 Legend on next page.

10970134, 2023, 8, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/prot.26497 by University Di Roma La Sapienza, Wiley Online Library on [28/11/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

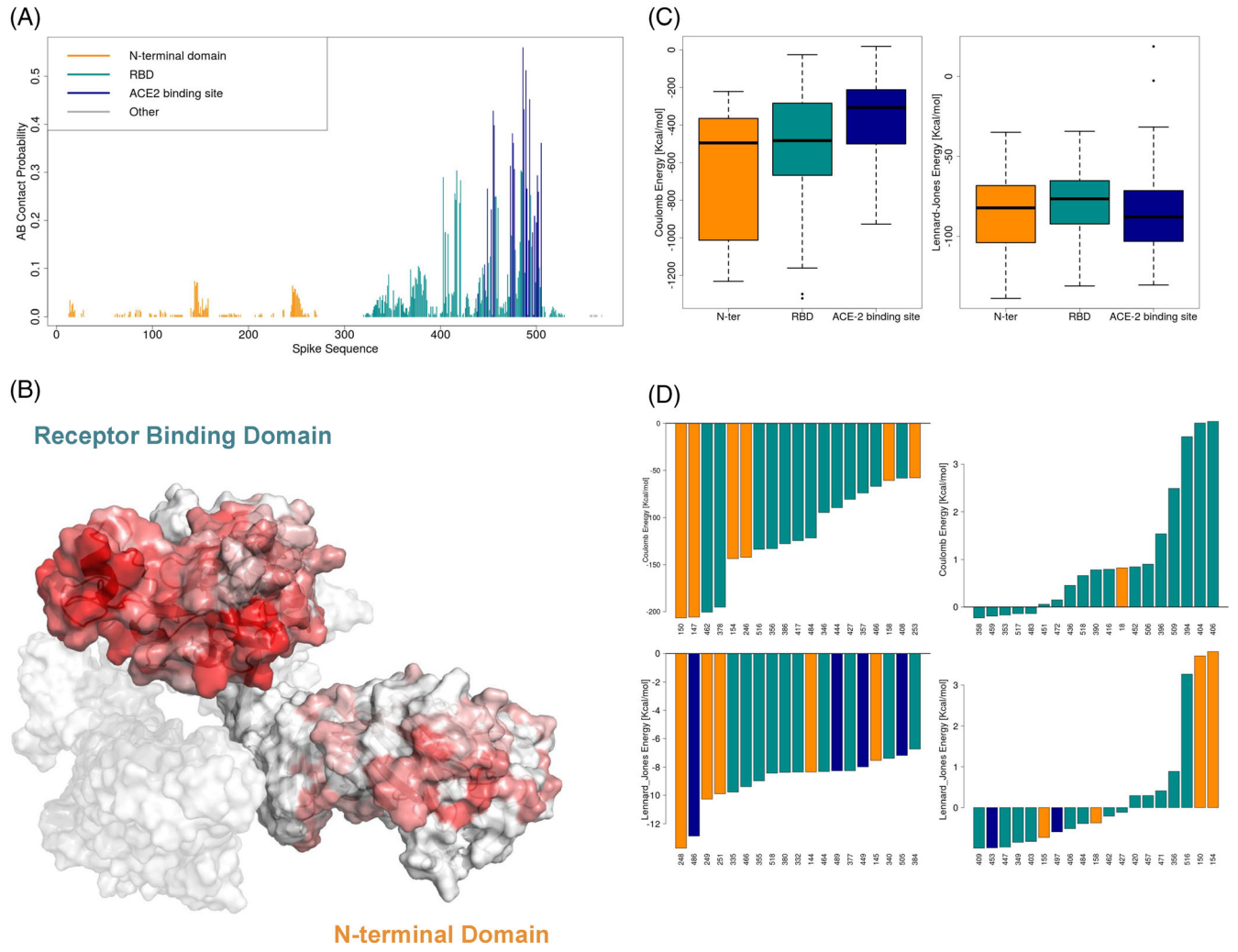


FIGURE 2 Analysis of the spike-antibodies interactions. (A) Frequency with which each Spike residue interacts with an antibody in the *Spike dataset*. (B) Molecular representation of the Spike protein, colored according to the frequencies reported in panel (A). (C) Boxplots of the intermolecular energies distributions for the three classes of antibody-Spike complexes described in the main text. The right panel accounts for Coulombic interactions and the left panel for Lennard-Jones ones. (D) Residues characterized by the most favorable or unfavorable mean strength of the Coulombic (upper panels) and Lennard-Jones (lower panels) interactions.

Working on the *Spike dataset*, we define the intermolecular contact as in the previous section (i.e., an S residue is in contact with the antibody if its CA atom is closer than 8 Åto any CA of the antibody). Figure 2A shows the interaction frequency computed for each residue considering all the complexes in the dataset. The bars in yellow regard residues from the S NTD, whereas the bars in cyan refer to the ones in the RBD: among these residues, we depict in blue the ones involved in ACE2 recognition. This frequency determines the residue *immunogenicity index*, represented in Figure 2B, where the higher the red intensity, the higher the residue immunogenicity.

These results allow us to classify the antibody-Spike complexes in three categories, according to the S region used for the recognition: N-ter (27 complexes, 9% of the dataset), RBD (antibodies bound to RBD but not in the ACE2 binding site, 77 complexes, 26% of the dataset), ACE2 BS (antibodies whose epitope residues are at least 25% in common with ACE2 binding site, 188 complexes, 64% of the

FIGURE 1 Analysis of the antibodies structural datasets. In the central panel, we insert the molecular representation of an antibody–antigen interaction. (A) Variable domain sequence logo representation of antibodies in the *Spike dataset*. (B) Histograms of amino acid occurrences in the antibody position characterized by the most evident difference between the *Spike dataset* and the *General dataset*. (C) Histograms of the distributions of the number of residues composing CDRs. The three selected CDRs are characterized by a statistically relevant difference in the two datasets, according to the Kolmogorov–Smirnov test. (D) Position probability of antigen interaction, considering the heavy (upper panel) and light (lower panel) chain. On the left, the frequency with which each position is seen in interaction with the antigen within the *Spike dataset*. On the right, the frequency calculated in the Spike and the *General dataset* for the positions with the most marked differences.

10970134, 2023, 8, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/prot.26497 by University Di Roma La Sapienza, Wiley Online Library on [28/11/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

complexes). Therefore, it emerges that in most known cases antibodies recognize Spike epitopes overlapping with the ACE2 binding site. Nevertheless, a non-negligible part of the antibodies exerts their activity by binding Spike in other regions that are part of the RDB or the NTD.

It has to be noted that in previous years various works have provided insightful and interesting classifications of Spike-binding antibodies depending on their binding properties.[58] In particular, these experimental papers, typically based on tens of antibody structures, cataloged antibodies depending on the Spike regions they bind, observing immunogenic regions on the RBD,[7,59,60] NTD,[7,61] and S2.[7,62] Here, we provide a comprehensive study based on 297 antibody-Spike complexes, where the diversity between antibodies is ensured by modulating the sequence identity.

Thus, we investigate if, from an energetic point of view, the mechanisms of interaction differ among these classes. To achieve this purpose, for each antibody-spike complex we calculate the Coulombic and Lennard-Jones intermolecular interaction energies between all couples of residues closer than 12 Å (see "Materials and Methods" Section for details), as we did in our previous works.[63,64] We report the distribution of the interface energies in the boxplots in Figure 2C, where the antibodies are separated into the three categories described above. It is worth commenting that the N-ter and RBD antibodies are characterized by a higher number of strong favorable Coulombic energies than the ACE2 BS ones (left panel). On the other hand, the three classes of antibody-Spike structures do not differ in what concerns the Lennard-Jones energies.

Lastly, we study which residues are on average responsible for the energy at the interaction. We define the *residue strength* as the sum of the energies of all the intermolecular interactions involving that residue. In Figure 2D, we show the residues characterized by the most favorable (right panels) and unfavorable (left panels) mean strength values, both for Coulombic (upper panels) and Lennard-Jones (lower panel) energies. This analysis is conducted on the residues seen in contact with the antibody at least five times. As expected, residues with high mean Coulombic strength are in the NTD or RBD (as indicated in Figure 2A by the yellow and cyan bars respectively). In addition, ACE2 binding site residues are characterized by a good mean Lennard-Jones strength (lower left panel, blue bars), while NTD and RBD residues can have strong unfavorable Lennard-Jones strength (lower right panel, cyan and yellow bars). The results of the energetic analyses might help in the design of effective antibodies against the virus and its future variants on specific regions.

2.3 | Analysis of the spike molecular dynamics simulations

All the above analyses, including the structural characteristics study of the interaction with antibodies, were conducted on the original S protein. Thus, to investigate the impact of SARS-CoV-2 variants on the immune response generated against the original S protein, it is necessary to understand how much the antibody-targeted S regions are different in the viral variants. With this aim, we select the original Spike (hereafter referred to as wild type) and two well-known VOC, that is, the Delta and Omicron variants, and we perform a 100-ns long molecular dynamics simulation of the trimeric form of all these three Spike variants. Indeed, the differences in the local dynamical behavior the variants experience with respect to the wild type can give us insights into the persistence of bindings with antibodies.

We then compared the three variants Spike molecular dynamics simulation. We use the root mean square deviation (RMSD) observable, calculated over some different portions of the S molecule. The results of this analysis are reported in Figure 3A. The upper left panel regards the whole proteins: after a short equilibration time, all three proteins reach equilibrium with a similar displacement from the initial configuration, highlighting overall comparable stability. In the next three panels (Figure 3B–D) we report the RMSD obtained locally over the NTD, the RBD, and the ACE2 binding site. As evident from the upper right panel, the NTD in all the cases is very mobile: however, it is worth noting that both variants show a behavior more unstable than the wild type. Interestingly, ACE2 appears to present an opposite tendency. Indeed, while the stability of the RBD as a whole is comparable in the three simulations (lower left panel), the mutations seem to lower the RMSD and stabilize the ACE2 binding site (lower right panel), especially for the Omicron case. Moreover, we consider the root mean square fluctuation (RMSF), calculated over each residue of the NTD and RBD in all three molecular simulations. The results can be found in Figure 3B, where the upper, the central, and the bottom panel regard the wild type, the Delta, and the Omicron S protein, respectively. It should be remarked that the variants that improved mobility of the NTD previously discussed have different origins in Delta and Omicron. Indeed, by comparing the three plots it becomes evident that the RMSF of residues around 240 is responsible for the higher mobility of the Delta S protein. Conversely, the omicron RMSF results show very high mobility in the residues around 140, while the peak around 240 disappears. However, the differences in the RBD are less evident: this notwithstanding, looking at the ACE2 binding site residues (blue bars), a decrease in the RMSF of such residues emerges.

It has to be noted that, as known, SARS-CoV-2 Spike protein is mostly covered by glycans and they play an essential role in various aspects of the Spike structure and dynamics. In fact, besides the shielding role shared with other fusion proteins, many authors have hypothesized that they have a functional role in binding with ACE2 or in maintaining a stable conformation of Spike itself.[65] However, some papers underline that glycans' presence influences the dynamical behavior of the Spike protein on a very long time scale (at least tens of microseconds),[66,67] while on shorter time scales, such as the ones we investigated in this work, it seems that their influence is lower.[68] In addition, we focused on characterizing the binding mechanism between Spike and antibodies: it has been shown that antibody recognition can occur mainly in Spike regions where the glycan shield is not present, thus we considered that the absence of glycans in the Spike simulations was not expected to significantly affect our findings, while allowed for a relevant decrease of the computational cost. In this

10970134, 2023, 8, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/prot.26497 by University Di Roma La Sapienza, Wiley Online Library on [28/11/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

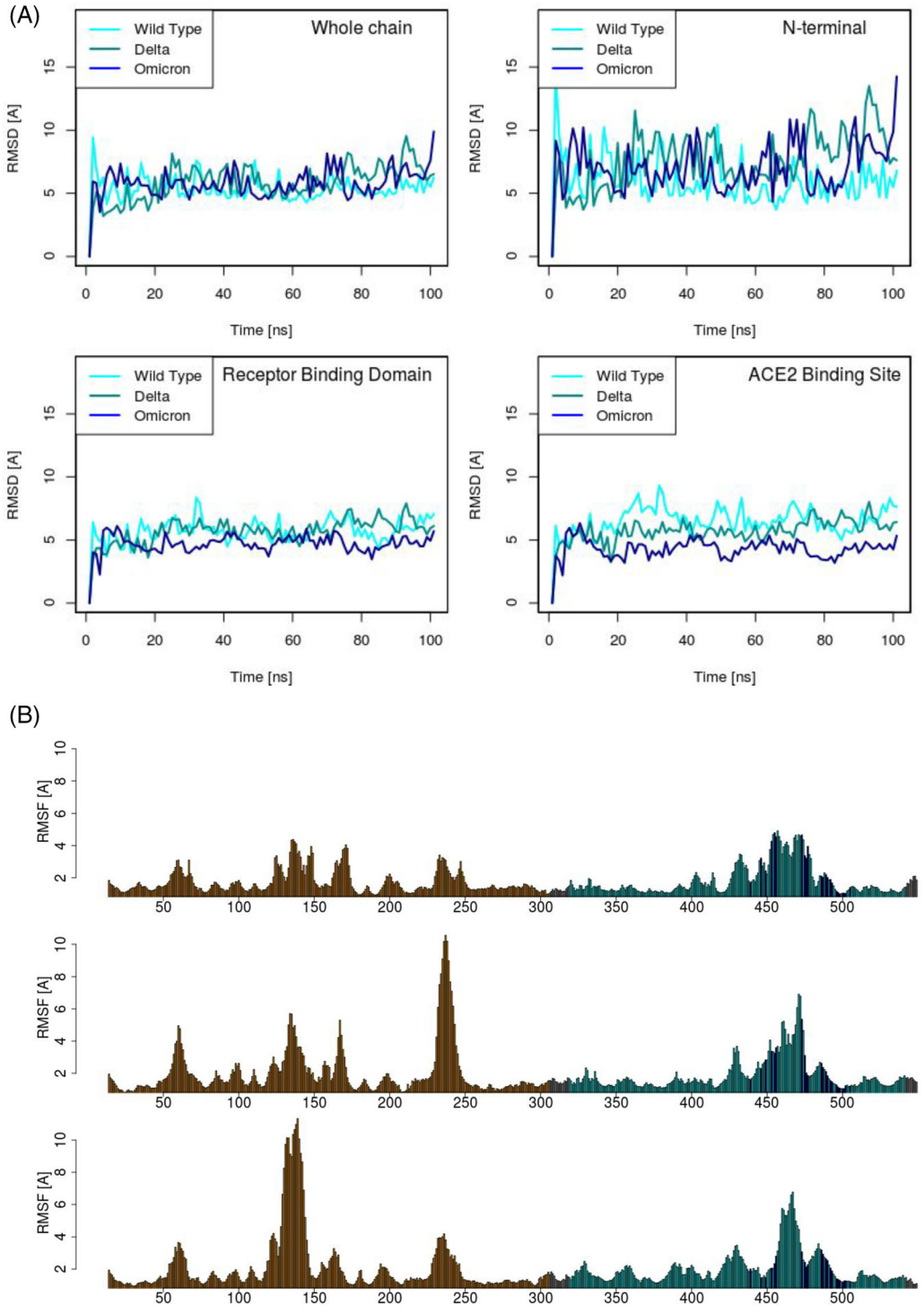


FIGURE 3 Analysis of the molecular dynamics simulations. (A) root mean square deviation (RMSD) as a function of time for the Spike protein of the wild type (light blue) and the Delta (blue) and Omicron (dark blue) variants. The RMSD is computed for four different portions of the molecule: from left to right and from top to bottom the whole protein, the N-terminal domain, the receptor binding domain, and the ACE2 binding site are considered. (B) Root mean square fluctuations of the Spike residues of the N-teminal domain and the receptor binding domain. The top, central, and bottom panels report the results for wild type, Delta, and Omicron, respectively. Residues from a different portion of the Spike are differently colored: orange for the N-terminal, cyan for the receptor binding domain, and blue for the ACE2 binding site.

panorama, we performed two additional 250 ns-long molecular dynamics simulations of the Spike S1 domain, in its WT sequence, one with the glycans and one without them. Overall, the results of these analyses represent a convincing indication that residue mobility on a short time scale is not strongly affected by the presence/absence of glycans (See Data S1).

10970134, 2023, 8, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/prot.26497 by University Di Roma La Sapienza, Wiley Online Library on [28/11/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

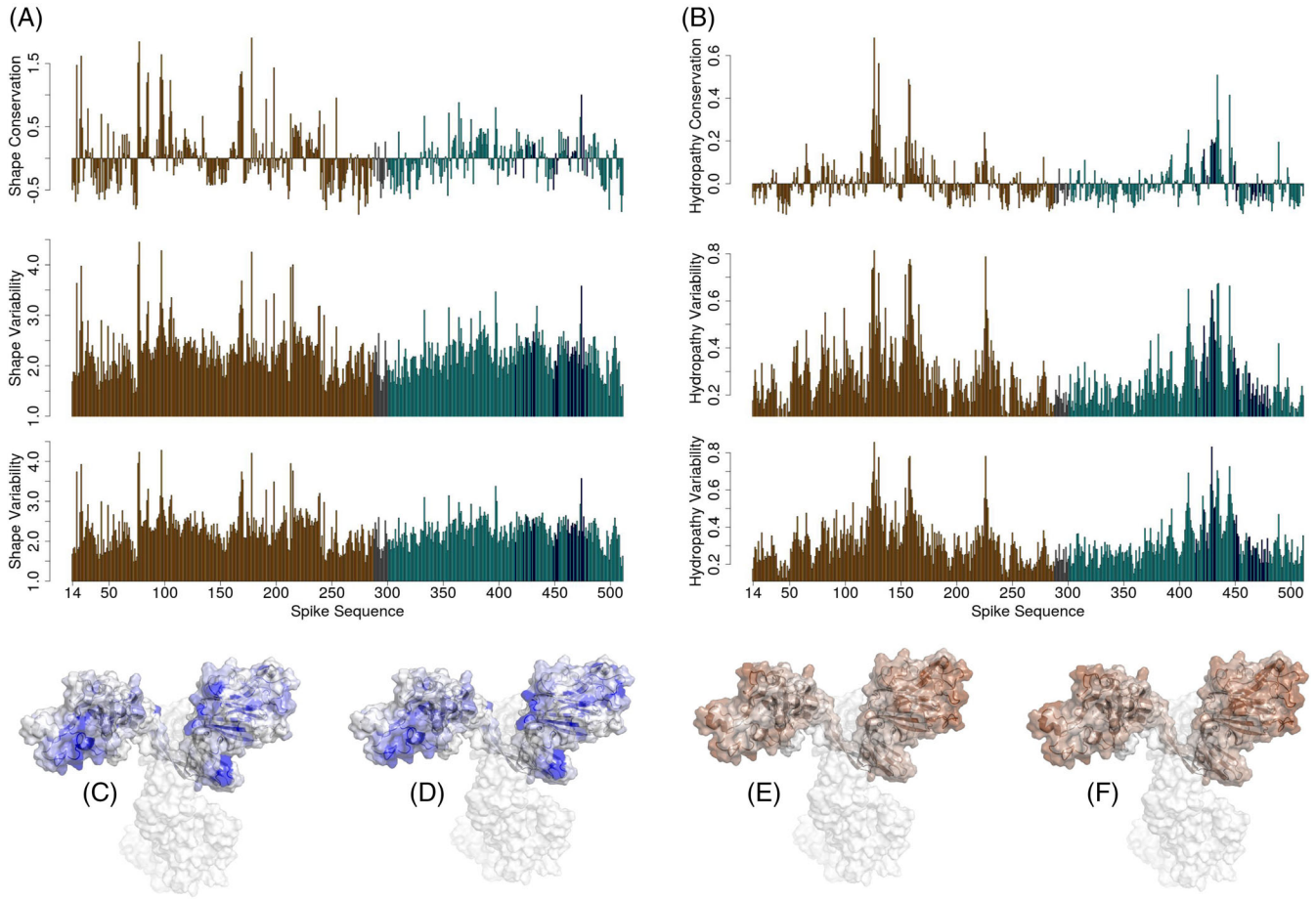


FIGURE 4 Physico-chemical analysis of the changes exhibited by the variants Spike protein. (A) Shape conservation index for all the residues of the wild type (top), Delta (middle), and Omicron (bottom) spike protein obtained as the average of the Zernike distances of the molecular surface portion centered around each spike residue during the molecular dynamics simulation. (B) Same as in panel (A) but considering the hydropathic index (see Methods). (C) Molecular representation of the Delta variants, where the intensity of the blue color is determined by the shape changes with respect to the wild type. (D) Same as in panel (C) but for the Omicron variant. (E), (F) Same as in panels (C) and (D) but for the hydrophatic index.

2.4 | Modeling the physico-chemical changes in SARS-CoV-2 variants and their importance for antibody recognition

To further analyze the local differences in the S protein induced by the variants mutations, we introduce two descriptors to quantitatively characterize the local shape and hydrophobicity. For this purpose, we build for each frame of the molecular dynamics simulations the corresponding molecular surface, using the DMS software.[69] We then exhaustively sample the molecular surface, selecting 10% of its points. Each point is used to determine a patch, defined as the set of molecular surface points closer than a threshold to that point. Finally, we assign each patch to the residue generating its center.

The shape of each patch can be thus characterized in terms of 2D Zernike descriptors, according to a method we recently developed[39]: the geometrical features of the patch are summarized in an ordered set of numerical descriptors, allowing an easy patch-to-patch comparison applying a standard Euclidean metrics between their descriptors. Moreover, we characterize also the hydrophobicity of a patch: using a residue molecular-dynamics-based hydrophobicity scale we recently published,[50] the hydrophobicity of a patch is defined as the weighted mean hydrophobicity of the residues generating its points (see "Materials and Methods" for details).

In Figure 4A, we report the results of the shape analysis conducted over the three molecular dynamics simulations of the wild-type and variants of S protein. The top panel concerns the wild-type simulation and highlights which regions on S show the highest shape variability. To measure the shape variability of a patch, we start by calculating its Zernike descriptors for all the frames of the simulation and computing the distances between all these descriptions. The average of such distances quantifies the shape variability this patch experiences along the simulation. Then, we perform a mean on the patches relative to the same residue. Lastly, we standardize our results by subtracting to each residue value the global mean. In this way, we can identify which regions are more variable in shape: a positive high value means a high variability while a negative high value indicates

10970134, 2023, 8, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/prot.26497 by University Di Roma La Sapienza, Wiley Online Library on [28/11/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

high shape conservation. Looking at the top panel of Figure 4A we confirm the relative instability of the wild-type NTD, overall characterized by a high number of positive peaks. Interestingly, one of the most variable regions in the RBD includes some residues responsible for ACE2 binding.

The central and the bottom panels are in Figure 4A represents the shape variability, with respect to the wild type, exhibited by the Delta and Omicron S protein variants, respectively. To obtain these graphs, we characterize each patch of each frame with the Zernike descriptors. As a next step, we calculate the distances between its shape description in all the frames of the considered variant simulation and its shape description in all the frames of the wild-type simulation. In this way, we define a measure of the shape changes between each variant and the wild type. We call this measure the *shape variability index*. Intuitively, the regions mostly characterized by large shape changes induced by the mutations should be the most unstable in the wild-type case: this is true for both the Delta variant (central panel, Pearson correlation coefficient of 0.90) and the Omicron case (bottom panel, Pearson correlation coefficient of 0.92).

The same analysis is performed in terms of hydrophobicity in Figure 4B: the peculiarity here is that the patch hydrophobicity is summarized in a single numerical value. The top panel regards the wild-type Spike protein simulation, where a high positive value indicates a residue whose patches tend to modify its hydrophobic behavior during the simulation while high hydrophobicity conservation is shown by residues characterized by negative values. Here, even if the lowest conservation is still in the NTD, the exposed regions are equally characterized by low conservation. As in the previous shape analysis, the central and bottom panels of Figure 4B measure the changes exhibited by the variant forms of S with respect to the wild type. Symmetrically, we call this residue-level measure the *hydrophobicity variability index* of either Delta or Omicron.

To offer a more understandable representation of these results, in Figure 4C, D, we report a molecular representation of the two variants, where the intensity of the blue color is determined by the shape changes with respect to the wild-type. Analogously, in Figure 4E, F the Spike protein of the two variants is colored according to hydrophobicity changes: the higher the intensity of brown the higher the difference experienced in that region by the variant.

2.5 | Relationship between physico-chemical changes and immunogenicity

In the previous sections, we characterized the S protein regions both by looking at immunogenicity, (i.e., the frequency with which antibodies bind that region), and variability (i.e., the changes that that region experiences due to the variants mutations). Therefore, each S residue is now described by several indexes: an *immunogenicity index* reflecting how many times it was experimentally observed in interaction with an antibody; a *shape variability index* indicating how much the molecular regions surrounding the antibody binding site change their shape because of the mutations in the variants; a *hydrophobicity variability index* summarizing the chemical changes in hydrophobicity of the patches around the antibody binding site when the variants are considered.

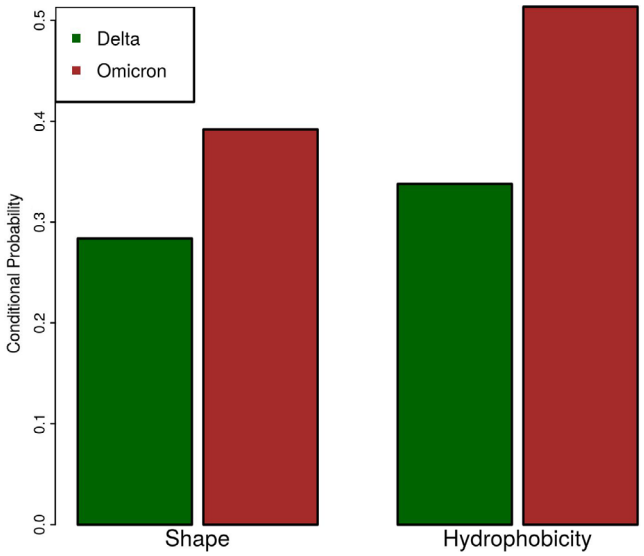


FIGURE 5 Conditional probability analysis. Conditional probabilities of finding a residue with a high variance of shape or hydrophatic indexes given that such residue shows high immunogenicity. Green bars display the conditional probabilities computed using Delta spike residues, while red bars are obtained considering Omicron spike residues.

To relate these quantities, we adopt a conditional probability approach. Indeed, if the variant's physico-chemical changes occur in regions characterized by a high antibody binding frequency, it is more likely that the antibodies generated against the wild-type version of S protein cannot recognize anymore the Spike mutated version. Therefore, we want to know the probability of a residue having surface regions highly variable in shape and/or hydrophobicity, *conditioned* to its high immunogenicity.

To do this, we binary classify residues in strongly immunogenic and weakly immunogenic, highly shape-variable and lowly shape-variable (for both delta and omicron), highly hydrophobic variable and lowly hydrophobic variable (for both Delta and Omicron). Therefore, we can define the conditional probabilities:

$$\begin{aligned} P_S^{\delta} &= P\left(V_S^{\delta} | I\right) = \frac{P\left(V_S^{\delta} \cap I\right)}{P(I)} \\ P_S^{o} &= P\left(V_S^{o} | I\right) = \frac{P\left(V_S^{o} \cap I\right)}{P(I)} \\ P_H^{\delta} &= P\left(V_H^{\delta} | I\right) = \frac{P\left(V_H^{\delta} \cap I\right)}{P(I)} \\ P_H^{o} &= P\left(V_H^{o} | I\right) = \frac{P\left(V_H^{o} \cap I\right)}{P(I)}, \end{aligned} \tag{1}$$

where the subscripts S or H mean shape and hydrophobicity, respectively. δ and o superscripts stand for the results relative to the two corresponding variants. V indicates the *highly variable* state of a

10970134, 2023, 8, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/prot.26497 by University Di Roma La Sapienza, Wiley Online Library on [28/11/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

residue, while *I* indicates the classification as *highly immunogenic* of a residue.

We show such probabilities in Figure 5. It emerges that the conditional probabilities obtained by the Omicron variant are higher than the ones of the Delta, both in terms of shape and hydrophobicity. This means the S protein of the Omicron variant is more different from the wild type in the antibody interested regions than the Delta variant: this could imply a more pronounced escape capability from antibodies elicited against the wild-type S protein, as confirmed by literature.[70,71]

3 | CONCLUSIONS

After more than 2 years of the SARS-CoV-2 pandemic worldwide spreading, the onset of viral variants still represents a dangerous issue for public health. Indeed, even if the vaccination campaign and the previous infections have generated an immunization against the original version, the mutations acquired by such variants can virtually confer to the virus the capability to escape the immune system. From this point of view, the new vaccines, updated with some VOC, can represent an effective way to contain the phenomenon.

To quantify this effect we preliminary studied the peculiarities of the interaction between antibodies and the original version of the SARS-CoV-2 Spike protein, which is the main target of the antiviral antibodies both in case of infection or vaccination. Interestingly, we identified some features of these antibodies that are statistically different from the wider population of generic protein-binding antibodies. Moreover, we identified the immunogenic regions of the Spike protein, adopting the frequency of interaction with antibodies as a proxy.

Thus we selected two important SARS-CoV-2 variants of concern, Delta and Omicron, to investigate how their Spike protein mutations impact the interaction with the antibodies elicited against the original version of Spike. The molecular dynamics we performed, we simulated the trimeric form of original and variants Spike protein, allowing us to evaluate the long-range effects of the mutations and to obtain a sample of the possible conformations that Spike can assume in all the variants.

Studying the intensity of the geometrical and chemical changes experienced by the two variants, we identified in both cases the Spike regions most affected by the mutations. We, therefore, obtained insights on the physicochemical variability of Spike regions using molecular dynamics data, while we reaped information about the immunogenicity of such regions using experimental complexes.

It is already known that Omicron can escape antibodies better than the Delta variant. Correlating the information of our two approaches we suggested a possible explanation: even if the overall variability between the two variants is comparable, Omicron presents a higher variability in the highly immunogenic regions compared to Delta.

In addition, it is worth noting the general validity of our approach, which could be applied both to new SARS-CoV-2 variants of concern and on other viral pathogens.

4 | MATERIALS AND METHODS

4.1 | Datasets

The *Spike dataset* was built using CoV-AbDab[72] and the *General dataset* with SabDab.[73] Separately for both datasets, we selected the antibody complexes with a level of redundancy lower than 90% using cd-hit.[74] All the structures are renumbered according to Chotia numbering scheme[10,11] with an *in-house* Python script.

The sequences of the original, Delta and Omicron Spike protein have been taken from the GitHub repository of a recent work.[75] We then modeled the three trimeric structures with Swiss Model,[76] using the structure deposited in Protein Data Bank[77] with the code 6vxx as a template.

The Spike NTD comprises residues 1–305. The Spike RBD is composed of residues 319–541. The residues forming the ACE2 binding sites, as defined here,[44] are 439, 446, 449, 453, 455, 456, 473, 475, 476, 477, 486, 487, 489, 490, 492, 493, 496, 497, 498, 500, 501, 502, 505.

4.2 | Nonbonded energy calculation

The partial charges were assigned to atoms using the PDB2PQR software,[78] with standard options. Before the proper energy calculation, the structures were minimized with Gromacs 2020.6.[79]

To compute intermolecular interactions we used the parameters of the CHARMM force field.[80] In particular, given two atoms, *l* and *m*, with partial charges q_l and q_m, the Coulombic interaction between them is defined as:

$$E^{C}_{lm} = \frac{1}{4\pi\varepsilon_0}\frac{q_l q_m}{r_{lm}}, \tag{2}$$

where r_{lm} is the distance between the two atoms, and ε_0 is the vacuum permittivity.

The Lennard-Jones potential is defined as in the following equation:

$$E^{\mathrm{LJ}}_{lm} = \sqrt{\varepsilon_l \varepsilon_m}\left[\left(\frac{R_{\min}{}^{l} + R^{m}_{\min}}{r_{lm}}\right)^{12} - 2\left(\frac{R_{\min}{}^{l} + R^{m}_{\min}}{r_{lm}}\right)^{6}\right], \tag{3}$$

where ε_l and ε_m are the potential well depths for *l* and *m*, respectively. $R^{l}_{\min}$ and $R^{m}_{\min}$ represent potential minima distances.

Summing over all the atoms pairings, the total interaction energy between residue *i* and residue *j* is:

$$E^{X}_{AA_{ij}} = \sum_{l=1}^{N^{i}_{\mathrm{atom}}}\sum_{m=1}^{N^{j}_{\mathrm{atom}}} E^{X}_{lm}, \tag{4}$$

where *X* stands for Coulombic ($X = C$) or Lennard-Jones ($X = \mathrm{LJ}$).

10970134, 2023, 8, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/prot.26497 by University Di Roma La Sapienza, Wiley Online Library on [28/11/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

4.3 | Molecular dynamics simulations

The simulations of the Spike trimers were performed using Gromacs 2019.6,[79] using the the CHARMM-36 force field.[81] Proteins were placed in a dodecahedric simulative box, with periodic boundary conditions. We used the TIP3P model for water molecules.[82] In all the systems, all protein atoms were at least at a distance of 1.1 nm from the box borders. The minimizations were performed with the steepest descent algorithm. Next, a two-step thermalization of the system was run in NVT and NPT environments each for 0.1 ns at 2 fs time-step. Using the v-rescale thermostat, the temperature was kept constant at 300 K. In the production runs of 100 ns, the pressure was set at 1 bar with the Parrinello–Rahman barostat.[83] We adopted the LINCS algorithm[84] to constrain bonds involving hydrogen atoms. Short-range nonbonded interactions were evaluated with a cutoff of 12Å. The Particle Mesh Ewald method[85] was adopted for the long-range electrostatic interactions.

4.4 | Patches definition

All the molecular surfaces used in this work have been calculated using the DMS software with standard parameters.[69]

The center of the patches have been defined using the starting structure of the Spike protein original version, sampling one point per Å^2 from the molecular surface of such structure. Each of the resulting 27 179 points has been used to build a patch. In the starting structure of the wild-type Spike protein, a patch is defined as the set of molecular surface points closer than 6 Å to the patch center. To determine the patch centers in all the other simulation frames and for the variants, we super-positioned each structure with the starting structure of the original Spike. The points closest to the ones selected on this original version were taken as the patches center of that structure. The patch was then constructed using the same threshold of 6 Å.

4.5 | Zernike descriptors

The points composing a patch can be projected with a conical symmetry onto a plane, in a way able to maintain the geometrically relevant information.[39] Therefore, each patch can be summarized as a 2D function $f(r,\phi)$ defined in the unitary circle (region $r < 1$). Therefore it can be expanded in the Zernike polynomials basis

$$f(r,\phi) = \sum_{n=0}^{\infty} \sum_{m=0}^{m=n} c_{nm} Z_{nm}, \tag{5}$$

where

$$\begin{aligned} c_{nm} &= \frac{(n+1)}{\pi} \langle Z_{nm} | f \rangle = \\ &= \frac{(n+1)}{\pi} \int_0^1 dr r \int_0^{2\pi} d\phi Z_{nm}^*(r,\phi) f(r,\phi) \end{aligned} \tag{6}$$

are the Zernike moments, the expansion coefficients. $Z_{nm}(r,\phi)$ are the Zernike polynomials, made by a radial and an angular factor:

$$Z_{nm} = R_{nm}(r) e^{im\phi}. \tag{7}$$

The radius dependence, given n and m, is expressed by the following expression:

$$R_{nm}(r) = \sum_{k=0}^{\frac{n-m}{2}} \frac{(-1)^k (n-k)!}{k! \left(\frac{n+m}{2}-k\right)! \left(\frac{n-m}{2}-k\right)!} r^{n-2k}. \tag{8}$$

For each couple of polynomials, it holds:

$$\langle Z_{nm} | Z_{n'm'} \rangle = \frac{\pi}{(n+1)} \delta_{nn'} \delta_{mm'}. \tag{9}$$

Therefore, the set of polynomials forms a basis. The knowledge of all the coefficients $\{c_{nm}\}$ permits the description of the original function, while the detail level of the description is determined by the order of expansion, $N = \max(n)$.

The modulus of a coefficient ($z_{nm} = | c_{nm} |$) does not depend on the phase, being invariant if we perform a rotation around the origin. The z_{nm} is the Zernike invariant descriptors.

The shape similarity between two patches is, therefore, studied by comparing their Zernike invariants. In particular, we measured the similarity between patch i and j as the Euclidean distance between their invariant vectors. We adopted $N = 20$, therefore dealing with 121 invariant descriptors for each patch.

4.6 | Hydropathy of patches

Each point of a patch has been generated from one residue. Each amino acid is characterized by a hydrophobic value.[50] Therefore each patch point can be associated to the hydrophobic value of the residue generating it. The hydrophobicity of a patch is the mean of all the patch points hydrophobicity.

AUTHOR CONTRIBUTIONS

Lorenzo Di Rienzo: Conceptualization; investigation; writing – original draft; methodology; writing – review and editing; software; formal analysis; data curation. **Mattia Miotto:** Conceptualization; methodology; software; writing – review and editing. **Fausta Desantis:** Investigation; formal analysis; writing – review and editing. **Greta Grassmann:** Investigation; formal analysis; writing – review and editing. **Giancarlo Ruocco:** Funding acquisition; writing – review and editing; supervision. **Edoardo Milanetti:** Conceptualization; methodology; supervision; investigation; writing – review and editing.

ACKNOWLEDGMENT

The research leading to these results has been also supported by European Research Council Synergy grant ASTRA (no. 855923). Open

10970134, 2023, 8, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/prot.26497 by University Di Roma La Sapienza, Wiley Online Library on [28/11/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Access Funding provided by Istituto Italiano di Tecnologia within the CRUI-CARE Agreement.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1002/prot.26497.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are all available in Protein Data Bank at https://www.rcsb.org/. The list of all the structures used in this study is available from the corresponding author.

ORCID

Lorenzo Di Rienzo https://orcid.org/0000-0003-4114-5049

REFERENCES

1. Zhou P, Yang X-L, Wang X-G, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020;579:270-273.
2. Abdool Karim SS, de Oliveira T. New SARS-CoV-2 variants—clinical, public health, and vaccine implications. *New Engl J Med*. 2021;384:1866-1868.
3. Hu B, Guo H, Zhou P, Shi Z-L. Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev Microbiol*. 2021;19:141-154.
4. Cui J, Li F, Shi Z-L. Origin and evolution of pathogenic coronaviruses. *Nat Rev Microbiol*. 2019;17:181-192.
5. Turoňová B, Sikora M, Schürmann C, et al. In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. *Science*. 2020;370:203-208.
6. Tortorici MA, Veesler D. *Advances in Virus Research*. Vol 105. Elsevier; 2019:93-116.
7. Piccoli L, Park Y-J, Tortorici MA, et al. Mapping neutralizing and Immunodominant sites on the SARS-CoV-2 spike receptor-binding domain by structure-guided high-resolution serology. *Cell*. 2020;183:1024-1042.e21.
8. Xia X. Domains and functions of spike protein in SARS-Cov-2 in the context of vaccine design. *Viruses*. 2021;13:109.
9. Martnez-Flores D, Zepeda-Cervantes J, Cruz-Reséndiz A, Aguirre-Sampieri S, Sampieri A, Vaca L. SARS-CoV-2 vaccines based on the spike glycoprotein and implications of new viral variants. *Front Immunol*. 2021;12:701501.
10. Chothia C, Lesk AM. Canonical structures for the hypervariable regions of immunoglobulins. *J Mol Biol*. 1987;196:901-917.
11. Chothia C, Lesk AM, Tramontano A, et al. Conformations of immunoglobulin hypervariable regions. *Nature*. 1989;342:877-883.
12. Tramontano A, Chothia C, Lesk AM. Framework residue 71 is a major determinant of the position and conformation of the second hypervariable region in the VH domains of immunoglobulins. *J Mol Biol*. 1990;215:175-182.
13. Pang NY-L, Pang AS-R, Chow VT, Wang D-Y. Understanding neutralising antibodies against SARS-CoV-2 and their implications in clinical practice. *Mil Med Res*. 2021;8:1.
14. Xiaojie S, Yu L, Guang Y, Min Q, et al. Neutralizing antibodies targeting SARS-CoV-2 spike protein. *Stem Cell Res*. 2021;50:102125.
15. Milanetti E, Miotto M, Di Rienzo L, et al. In-Silico Evidence for a Two Receptor Based Strategy of SARS-CoV-2. *Front Mol Biosci*. 2021;8:690655.
16. Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science*. 2020;367:1444-1448.
17. Cerutti G, Guo Y, Zhou T, et al. Potent SARS-CoV-2 neutralizing antibodies directed against spike N-terminal domain target a single supersite. *Cell Host Microbe*. 2021;29:819-833.e7.
18. Domingo E, Holland J. RNA virus mutations and fitness for survival. *Annu Rev Microbiol*. 1997;51:151-178.
19. Miotto M, Monacelli L. Genome heterogeneity drives the evolution of species. *Phys Rev Res*. 2020;2:043026. doi:10.1103/PhysRevResearch.2.043026
20. Duchene S, Featherstone L, Haritopoulou-Sinanidou M, Rambaut A, Lemey P, Baele G. Temporal signal and the phylodynamic threshold of SARS-CoV-2. *Virus Evolution*. 2020;6:veaa061.
21. Portelli S, Olshansky M, Rodrigues CH, et al. Exploring the structural distribution of genetic variation in SARS-CoV-2 with the COVID-3D online resource. *Nat Genet*. 2020;52:999-1001.
22. Miotto M, Rienzo LD, Gosti G, et al. Inferring the stabilization effects of SARS-CoV-2 variants on the binding with ACE2 receptor. *Commun Biol*. 2022;5:20221.
23. Trucchi E, Gratton P, Mafessoni F, et al. Population dynamics and structural effects at short and long range support the hypothesis of the selective advantage of the G614 SARS-CoV-2 Spike variant. *Mol Biol Evol*. 2021;38(5):1966-1979.
24. Baric RS. Emergence of a highly fit SARS-CoV-2 variant. *New Engl J Med*. 2020;383:2684-2686.
25. Zhang J, Cai Y, Xiao T, et al. Structural impact on SARS-CoV-2 spike protein by D614G substitution. *Science*. 2021;372(6541):525-530.
26. Mittal A, Khattri A, Verma V. Structural and antigenic variations in the spike protein of emerging SARS-CoV-2 variants. *PLoS Pathog*. 2022;18:e1010260.
27. Callaway E et al. Delta coronavirus variant: scientists brace for impact. *Nature*. 2021;595:17-18.
28. Thakur V, Ratho RK. OMICRON (B.1.1.529): a new SARS-CoV-2 variant of concern mounting worldwide fear. *J Med Virol*. 2022;94:1821-1824.
29. Weisblum Y, Schmidt F, Zhang F, et al. Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. *Elife*. 2020;9:e61312.
30. Wang R, Chen J, Hozumi Y, Yin C, Wei G-W. Emerging vaccine-breakthrough SARS-CoV-2 variants. *ACS Infect Dis*. 2022;8:546-556.
31. Plante JA, Mitchell BM, Plante KS, Debbink K, Weaver SC, Menachery VD. The variant gambit: COVID-19's next move. *Cell Host Microbe*. 2021;29:508-515.
32. Harvey WT, Carabelli AM, Jackson B, et al. SARS-CoV-2 variants, spike mutations and immune escape. *Nat Rev Microbiol*. 2021;19:409-424.
33. Serapian SA, Marchetti F, Triveri A, et al. The answer lies in the energy: how simple atomistic molecular dynamics simulations may hold the key to epitope prediction on the fully glycosylated SARS-CoV-2 spike protein. *J Phys Chem Lett*. 2020;11:8084-8093.
34. Spinello A, Saltalamacchia A, Magistrato A. Is the rigidity of SARS-CoV-2 spike receptor-binding motif the Hallmark for its enhanced infectivity? Insights from all-atom simulations. *J Phys Chem Lett*. 2020;11:4785-4790.
35. Spinello A, Saltalamacchia A, Borisek J, Magistrato A. Allosteric cross-talk among Spike's receptor-binding domain mutations of the SARS-CoV-2 south African variant triggers an effective hijacking of human cell receptor. *J Phys Chem Lett*. 2021;12:5987-5993.
36. Triveri A, Serapian SA, Marchetti F, et al. SARS-CoV-2 spike protein mutations and escape from antibodies: a computational model of epitope loss in variants of concern. *J Chem Inf Model*. 2021;61:4687-4700.
37. Chen J, Wang R, Hozumi Y, et al. Emerging dominant SARS-CoV-2 variants. *J Chem Inf Model*. 2022;63(1):335-342.
38. Chen J, Wei G-W. Omicron BA.2 (B.1.1.529.2): high potential for becoming the next dominant variant. *J Phys Chem Lett*. 2022;13:3840-3849.

10970134, 2023, 8, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/prot.26497 by University Di Roma La Sapienza, Wiley Online Library on [28/11/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

39. Milanetti E, Miotto M, Di Rienzo L, Monti M, Gosti G, Ruocco G. 2D Zernike polynomial expansion: finding the protein-protein binding regions. *Comput Struct Biotechnol J*. 2021;19:29-36.
40. Venkatraman V, Yang YD, Sael L, Kihara D. Protein–protein docking using region-based 3D Zernike descriptors. *BMC Bioinformatics*. 2009; 10:407.
41. Daberdaku S, Ferrari C. Exploring the potential of 3D Zernike descriptors and SVM for protein–protein interface prediction. *BMC Bioinformatics*. 2018;19:35.
42. Sandomenico A, Di Rienzo L, Calvanese L, et al. Insights into the interaction mechanism of DTP3 with MKK7 by using STD-NMR and computational approaches. *Biomedicine*. 2020;9:20.
43. Di Rienzo L, De Flaviis L, Ruocco G, Folli V, Milanetti E. Binding site identification of G protein-coupled receptors through a 3D Zernike polynomials-based method: application to C. elegans olfactory receptors. *J Comput Aided Mol Des*. 2022;36:11-24.
44. Di Rienzo L, Monti M, Milanetti E, et al. Computational optimization of angiotensin-converting enzyme 2 for SARS-CoV-2 Spike molecular recognition. *Comput Struct Biotechnol J*. 2021;19:3006.
45. Di Rienzo L, Milanetti E, Testi C, et al. A novel strategy for molecular interfaces optimization: the case of ferritin-transferrin receptor interaction. *Comput Struct Biotechnol J*. 2020;18:2678-2686.
46. Daberdaku S, Ferrari C. Antibody interface prediction with 3D Zernike descriptors and SVM. *Bioinformatics*. 2019;35:1870-1876.
47. Di Rienzo L, Milanetti E, Ruocco G, Lepore R. Quantitative description of surface complementarity of antibody-antigen interfaces. *Front Mol Biosci*. 2021;8:749784.
48. Sael L, Li B, La D, et al. Fast protein tertiary structure retrieval based on global surface shape similarity. *Proteins: Struct Funct Bioinformatics*. 2008;72:1259-1273.
49. Grassmann G, Miotto M, Di Rienzo L, et al. A computational approach to investigate TDP-43 RNA-recognition motif 2 C-terminal fragments aggregation in amyotrophic lateral sclerosis. *Biomolecules*. 2021;11: 1905.
50. Di Rienzo L, Miotto M, Bò L, Ruocco G, Raimondo D, Milanetti E. Characterizing Hydropathy of amino acid side chain in a protein environment by investigating the structural changes of water molecules network. *Front Mol Biosci*. 2021;8:626837.
51. Barnes CO, West AP, Huey-Tubman KE, et al. Structures of human antibodies bound to SARS-CoV-2 spike reveal common epitopes and recurrent features of antibodies. *Cell*. 2020;182:828-842.e16.
52. Guérin P, Yahi N, Azzaz F, Chahinian H, Sabatier J-M, Fantini J. Structural dynamics of the SARS-CoV-2 spike protein: a 2-year retrospective analysis of SARS-CoV-2 variants (from Alpha to Omicron) reveals an early divergence between conserved and variable epitopes. *Molecules*. 2022;27:3851.
53. Gan HH, Zinno J, Piano F, Gunsalus KC. Omicron spike protein has a positive electrostatic surface that promotes ACE2 recognition and antibody escape. *Front Virol*. 2022;2. doi:10.3389/fviro.2022.894531
54. Yin R, Guest JD, Taherzadeh G, et al. Structural and energetic profiling of SARS-CoV-2 receptor binding domain antibody recognition and the impact of circulating variants. *PLoS Comput Biol*. 2021;17: e1009380.
55. Bozdaganyan ME, Shaitan KV, Kirpichnikov MP, Sokolova OS, Orekhov PS. Computational analysis of mutations in the receptor-binding domain of SARS-CoV-2 spike and their effects on antibody binding. *Viruses*. 2022;14:295.
56. Liu Y, Arase H. Neutralizing and enhancing antibodies against SARS-CoV-2. *Inflamm Regeneration*. 2022;42:58.
57. Crooks GE, Hon G, Chandonia J-M, Brenner SE. WebLogo: A Sequence Logo Generator: Figure 1. *Genome Res*. 2004;14:1188-1190.
58. Bárcena M, Barnes CO, Beck M, et al. Structural biology in the fight against COVID-19. *Nat Struct Mol Biol*. 2021;28:2-7.
59. Barnes CO, Jette CA, Abernathy ME, et al. SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. *Nature*. 2020;588: 682-687.
60. Shiakolas AR, Kramer KJ, Wrapp D, et al. Cross-reactive coronavirus antibodies with diverse epitope specificities and fc effector functions. *Cell Rep Med*. 2021;2:100313.
61. McCallum M, De Marco A, Lempp FA, et al. N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. *Cell*. 2021;184:2332-2347.e16.
62. Sauer MM, Tortorici MA, Park Y-J, et al. Structural basis for broad coronavirus neutralization. *Nat Struct Mol Biol*. 2021;28:478-486.
63. Miotto M, Di Rienzo L, Corsi P, Ruocco G, Raimondo D, Milanetti E. Simulated epidemics in 3D protein structures to detect functional properties. *J Chem Inf Model*. 2020;60:1884-1891.
64. Desantis F, Miotto M, Di Rienzo L, Milanetti E, Ruocco G. Spatial organization of hydrophobic and charged residues affects protein thermal stability and binding affinity. *Sci Rep*. 2022;12:1.
65. Casalino L, Gaieb Z, Goldsmith JA, et al. Beyond shielding: the roles of Glycans in the SARS-CoV-2 spike protein. *ACS Cent Sci*. 2020;6: 1722-1734.
66. Pang YT, Acharya A, Lynch DL, Pavlova A, Gumbart JC. SARS-CoV-2 spike opening dynamics and energetics reveal the individual roles of glycans and their collective impact. *Commun Biol*. 2022;5:1170.
67. Dodero-Rojas E, Onuchic JN, Whitford PC. Sterically confined rearrangements of SARS-CoV-2 spike protein control cell invasion. *Elife*. 2021;10:e70362.
68. Rath SL, Kumar K. Investigation of the effect of temperature on the structure of SARS-CoV-2 spike protein by molecular dynamics simulations. *Front Mol Biosci*. 2020;7:583523.
69. Richards FM. Areas, volumes, packing, and protein structure. *Annu Rev Biophys Bioeng*. 1977;6:151-176.
70. Hu J, Peng P, Cao X, et al. Increased immune escape of the new SARS-CoV-2 variant of concern omicron. *Cell Mol Immunol*. 2022;19: 293-295.
71. Planas D, Saunders N, Maes P, et al. Considerable escape of SARS-CoV-2 omicron to antibody neutralization. *Nature*. 2022;602:671-675.
72. Raybould MI, Kovaltsuk A, Marks C, Deane CM. CoV-AbDab: the coronavirus antibody database. *Bioinformatics*. 2021;37:734-735.
73. Dunbar J, Krawczyk K, Leem J, et al. SAbDab: the structural antibody database. *Nucleic Acids Res*. 2014;42:D1140-D1146.
74. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 2006; 22:1658-1659.
75. Fatihi S, Rathore S, Pathak AK, et al. A rigorous framework for detecting SARS-CoV-2 spike protein mutational ensemble from genomic and structural features. *Curr Res Struct Biol*. 2021;3:290-300.
76. Waterhouse A, Bertoni M, Bienert S, et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*. 2018;46:W296-W303.
77. Burley SK, Berman HM, Bhikadiya C, et al. RCSB protein data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids Res*. 2019;47:D464-D474.
78. Dolinsky TJ, Czodrowski P, Li H, et al. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res*. 2007;35:W522-W525.
79. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS: fast, flexible, and free. *J Comput Chem*. 2005;26:1701-1718.
80. Vanommeslaeghe K, Hatcher E, Acharya C, et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem*. 2010;31:671-690.
81. Brooks BR, Brooks CL, Mackerell AD, et al. CHARMM: the biomolecular simulation program. *J Comput Chem*. 2009;30:1545-1614.

10970134, 2023, 8, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/prot.26497 by University Di Roma La Sapienza, Wiley Online Library on [28/11/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

82. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys*. 1983;79:926-935.
83. Parrinello M, Rahman A. Crystal structure and pair potentials: a molecular-dynamics study. *Phys Rev Lett*. 1980;45:1196-1199.
84. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: a linear constraint solver for molecular simulations. *J Comput Chem*. 1997;18:1463-1472.
85. Cheatham TEI, Miller JL, Fox T, Darden TA, Kollman PA. Molecular dynamics simulations on solvated biomolecular systems: the particle mesh Ewald method leads to stable trajectories of DNA, RNA, and proteins. *J Am Chem Soc*. 1995;117:4193-4194.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Di Rienzo L, Miotto M, Desantis F, Grassmann G, Ruocco G, Milanetti E. Dynamical changes of SARS-CoV-2 spike variants in the highly immunogenic regions impact the viral antibodies escaping. *Proteins*. 2023;91(8):1116-1129. doi:10.1002/prot.26497

10970134, 2023, 8, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/prot.26497 by University Di Roma La Sapienza, Wiley Online Library on [28/11/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License