Singular Interface Dynamics of the SARS-CoV-2 Delta Variant Explained with Contact Perturbation Analysis

Aria Gheeraert, Laurent Vuillon, Laurent Chaloin, Olivier Moncorgé, Thibaut Very, Serge Perez, Vincent Leroux, Isaure Chauvot de Beauchêne, Dominique Mias-Lucquin, Marie-Dominique Devignes, et al.

To cite this version:

Aria Gheeraert, Laurent Vuillon, Laurent Chaloin, Olivier Moncorgé, Thibaut Very, et al.. Singular Interface Dynamics of the SARS-CoV-2 Delta Variant Explained with Contact Perturbation Analysis. Journal of Chemical Information and Modeling, 2022, 62 (12), pp.3107-3122. 10.1021/acs.jcim.2c00350 . hal-03708020

HAL Id: hal-03708020
https://inria.hal.science/hal-03708020v1

Submitted on 7 Nov 2022

Singular Interface Dynamics of the SARS-CoV-2 Delta Variant Explained with Contact Perturbation Analysis

Aria Gheeraert,[†,‡] Laurent Vuillon,[†] Laurent Chaloin,[¶] Olivier Moncorgé,[¶] Thibaut Very,[§] Serge Perez,[‖] Vincent Leroux,[⊥] Isaure Chauvot de Beauchêne,[⊥] Dominique Mias-Lucquin,[⊥] Marie-Dominique Devignes,[⊥] Ivan Rivalta,[*,‡,#] and Bernard Maigret[*,⊥]

†*Laboratory of Mathematics (LAMA), CNRS, University of Savoie Mont Blanc, France*

‡*Dipartimento di Chimica Industriale "Toso Montanari", Università di Bologna, Bologna, Italy*

¶*Infectious disease Research Institute of Montpellier (IRIM), CNRS, University of Montpellier, France*

§*CNRS - IDRIS, rue John von Neumann BP 167 91403 Orsay cedex - France*

‖*University Grenoble Alpes, CNRS, CERMAV, 38000 Grenoble, France*

⊥*University of Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France*

#*Laboratory of Theoretical Chemistry, ENS Lyon, CNRS, Lyon, France*

E-mail: i.rivalta@unibo.it; bernard.maigret@loria.fr

Abstract

Emerging SARS-CoV-2 variants raises concerns about our ability to withstand the Covid-19 pandemic and for this, understanding mechanistic differences of those variants is crucial. In this study, we investigate disparities between the SARS-CoV-2 wild-type and five variants that emerged in late 2020 focusing on the Spike RBD/ACE2 monomer complex and dynamics using existing X-Ray structures and Molecular Dynamics simulations. Dihedral angle PCA shows strong similarities inside the Spike RBD dynamics in the *Alpha*, *Beta*, *Gamma* and *Delta* variants in contrast with the WT and *Epsilon*. Dynamical perturbation networks shows that differences arise in interface dynamics between the Alpha, Beta and Gamma variant relative to the Delta variant but that those effects are not directly imputable to the L452R and T478K mutations. Finally contact principal component analysis shows how the L452R and T478K mutations acts synergistically on neighboring residues to provoke these drastic changes.

Introduction

The SARS-CoV-2 virus, associated to the Covid-19 pandemic, has spread all over the world by first infecting human pulmonary cells. This critical step is achieved through specific interactions between the homotrimeric transmembrane Spike glycoprotein (S protein, with 1,273 residues in each monomer) and human angiotensin-converting enzyme-2 receptors (ACE2).[1,2] This attachment to cells is specifically mediated by the "receptor binding domain" (RBD, residues 319-541) of the Spike that binds with high affinity the N-terminal helix of ACE2,[3,4] allowing subsequent conformational changes and fusion between cell and viral membranes. As in many other viral infectious diseases, the emergence of mutant strains (or variants) ineluctably has arisen due to its zoonotic origin, interspecies transmission and human host adaptation. As the main important step in cell infection is the recognition of the specific ACE2 receptor, mutations occurring in Spike protein may confer increased or decreased infectivity potential, contributing to changes in transmission rates. With the rapid emergence of variants of concern (VOC) that quickly spread worldwide, the characteristics of viral transmission, disease severity and neutralization susceptibility have been compromised. The first VOC was identified in the UK in late December 2020 (*Alpha* variant / B.1.1.7 lineage).

While another variant (*Beta*, B.1.351) emerged independently in South Africa, new variants arose in Brazil (*Gamma*, P.1), in California (*Epsilon*, B.1.427/B.1.429) and finally in India (*Delta*/*Kappa*, B.1.617.1/2/3). The Alpha and Epsilon variant have been de-escalated as threat in summer 2021. In November 2021, the latest VOC (*Omicron*, B.1.1.529) was first detected in South Africa and has already spread to multiple countries and is now the current dominant form. Prior to this, the Delta variant was dominant for almost an year. The mechanisms by which these mutations modulate the infectivity or the severity of the disease are not fully understood, and only predictions can be drawn from phylogenetic studies[5] or binding free-energy calculations.[6] Focusing on the first step of viral infection or cell entry, several mutations encountered in the spike RBD are commonly shared by most variants, like N501Y or L452R. On the other hand, some mutations are more distinct, like T478K, which was exclusive to *Delta* prior to the discovery of the *Omicron* variant. The physicochemical interactions between hydrophobic and charged residues might greatly alter the recognition phase or the binding affinity between RBD and ACE2 receptors. For instance, the mutation N501T has been already shown to reduce the affinity of host ACE2 protein and S protein *in vitro*.[7] Here, we report an extensive investigation of the interaction of the Spike RBD domain with its human ACE2 receptor at the atomistic level, for the original SARS-CoV-2 virus as well as its five variants that emerged in late 2020, as detailed in Table 1.

To this aim, we focus on the analysis of the primary molecular interactions between Spike and ACE2 based on experimental structural data available from the Protein data bank (PDB). First, we investigate contact changes between the available X-ray structures[3,18] Wild-Type (WT) and the *Alpha*, *Beta*, *Gamma* variants, at 2.85, 2.63 and 2.80 Å, respectively. Here, we did not compare those results with available Cryo-EM structures of the *Delta* and *Epsilon* variants[19] because the involve structures resolved at a lower atomic resolution that do not allow appropriate computations of atomic contacts (i.e. > 3 Å). Then, we model all variants (using a common modeling procedure based on the WT structure with highest resolution), and we performed molecular dynamics (MD) simulations of the monomeric form

Table 1: SARS-CoV-2 variants investigated in the present work. The epidemiological status is as reported by the European Center for Disease Prevention and Control (ECDC) as of 15 December 2021 (https://www.ecdc.europa.eu/en/covid-19/variants-concern). Mutations of interest found in Spike RBD compared to the WT SARS-CoV-2 strain are depicted in bold. *DE: De-escalated

WHO label (Lineage)	Status	First detected	Spike mutations	Impact on transmissibility	Impact on immunity	Impact on severity	Transmission in EU
Alpha (B.1.1.7)	DE*	UK (September 2020)	**N501Y**, D614G, P681H	Yes[8]	No	Yes[9,10]	Low
Beta (B.1.351)	VOC	SA (September 2020)	**K417N**, **E484K**, **N501Y**, D614G, A701V	Yes[11]	Yes[12,13]	Yes[9]	Medium
Gamma (P.1)	VOC	Brazil (December 2020)	**K417T**, **E484K**, **N501Y**, D614G, H655Y	Yes[14]	Yes[15]	Yes[9]	Medium
Delta (B.1.617.2)	VOC	India (December 2020)	**L452R**, **T478K**, D614G, P681R	Yes[16]	Yes[16]	Yes[16]	High
Epsilon (B.1.427/ B.1.429)	DE*	USA (September 2020)	**L452R**, D614G	Unclear[17]	Yes[17]	No	Very low

(1 unit of each protein) of various Spike-ACE2 systems. Thus, we performed the analysis of the primary molecular interactions between Spike and ACE2 focusing on the effect of the different mutations on the atomic contacts at the interface and the corresponding binding dynamics. This information, indeed, is not directly accessible from the crystallographic structural models available in the PDB ($>$ 200 X-ray or CryoEM-derived structures) and requires atomistic simulations. We adopted several tools to analyse the MD trajectories and to cross compare them, including dihedral angle principal components analysis (dPCA),[20] static and dynamical perturbation contact networks (PCN and DPCN, respectively) and contact principal component analysis (cPCA).[?] The dPCA that the different mutations triggers similar rearrangements inside the spike RBD in the *Alpha*, *Beta*, *Gamma* and *Delta* variant that are not fully reproduced in the *Epsilon* variant. Dynamical perturbation contact networks shows that drastic differences in interface dynamics arise between the *Delta* variant and the *Alpha*-to-*Gamma* group, despite these changes relate to mutations (L452R and T478K) that involve residues far from the interface. Finally, using cPCA, we show how synergistic effects of L452R and T478K mutations in *Delta* trigger a pattern of specific contact rearrangements that strongly affect the RBD/ACE2 interface. This knowledge on the initial molecular mechanisms triggered by the Spike-ACE2 association provides a fundamental understanding of this critical aspect of viral infection, and may be very valuable for the rational design of antiviral therapies.

Materials and Methods

3D Models building and MD simulations

RBD/ACE2 wild type and mutants complexes

Several similar structures of the RBD/ACE2 wild type human monomer-monomer complex are available in the PDB database[1,3,7] (see Figure 1 below) and we used the one with the highest resolution (2.45 Å): 6M0J.[3]

The Visual Molecular Dynamics program (VMD)[21] was used to prepare the structural

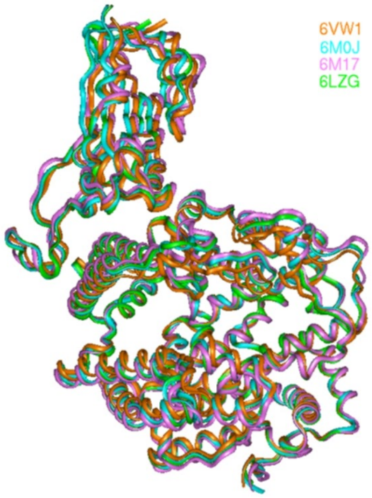

Figure 1: Superposition of several RBD/ACE2 complexes found in the PDB

models starting from the PDB structures, including the required point mutations. Molecular dynamics (MD) simulations were performed using the NAMD package[22] in conjunction with the recent CHARMM36 force field.[23] Six RBD/ACE2 complexes were considered in the present work: the WT and five variants among the most infectious strains (*Alpha* B.1.1.7, *Beta* B.1.351, *Gamma* P.1, *Delta* B.1.617.2 and *Epsilon* B.1.427 variant). Each protein-protein complex was placed in a TIP3P[24] water explicit solvent box of 150 $Å^3$ with periodic boundary conditions to simulate the biological environment realistically. Next, Na+ ions were added for ensuring neutrality of the periodic box. Each system was firstly energy minimized performing 64,000 steps of conjugate gradients, next equilibrated (10 ns MD simulation) and a trajectory of 1 μs was then produced. The simulations were carried out in the isobaric-isothermal ensemble, maintaining constant pressure and temperature at 1 atm and 300K, respectively, by means of Langevin dynamics and Langevin piston approaches as implemented in NAMD. The equation of motion was integrated every fs, using the r-RESPA algorithm[25] to update short and long-range contributions at different frequencies. Long-range electrostatic interactions were treated using the particle-mesh Ewald approach.[26]

Every ps, one frame was saved from the trajectory file, leading to a total of 1,000,000 frames for further analysis.

MD Analysis tools

Root mean square deviation

Root mean square deviation of atomic positions is a first rough indicator of simulation convergence. First, we align trajectories with respect to their initial conformation by minimizing the RMSD of backbone atomic positions. Then we report minimal RMSD fluctuations over time. Since, in our models, the spike RBD contains 229 residues and the ACE2 protein 603, it is possible that averaging the RMSD on the global ACE2/RBD complex hides destabilization occuring specifically by mutations in the RBD. To assess more directly possible effects of mutations, we also computed the RMSD of backbone atomic positions restricted either to the RBD (excluding terminal segments, residues S325-N540) or to the Receptor Binding Motif (RBM, residues S438-Q506) where most mutations are located.

Dihedral angle principal component analysis

Principal component analysis (PCA)[27–35] of MD simulations is a general method to extract essential motions of a system and to reduce the high-dimensional evolution a proteic system in a low-dimension landscape. In PCA, the feature choice is crucial and there has been an incentive to use internal coordinates like dihedral angles[20] over external coordinates (e.g. cartesian coordinates).[36–38] In this formulation, for each frame we compute $2N$ dihedral angles and linearize them from the circular space using the transformations:

$$q_{4n} = \sin\phi_n; \quad q4n+1 = \cos\phi_n; \quad q4n+2 = \sin\psi_n; \quad q4n+3 = \cos\psi_n \tag{1}$$

with $n = 1, \ldots, N$ corresponding to the N pairs of consecutive residues from which dihedral angles are considered (in practice $= N_{\text{residues}} - N_{\text{chains}}$). In this study, we accounted for all

ϕ and ψ backbone dihedral angles. Since RBD variants only show single point mutations, the considered models have all the same number of backbone dihedral angles and can be compared straightforwardly. An observation matrix $Q_{i,j}$ of size ($N_{\text{frames}} \times 2N$) is constructed, where the columns are all linearizations of ϕ and ψ dihedral angles and the rows all possible observation states (10,000 frames for the WT and each variant so 60,000 frames in total). The scikit-learn[39] implementation of PCA decomposition to get the principal components (PCs) was used. Restricting to the two first eigenvectors, they can be used to obtain the free energy-landscape of the system:

$$G(\text{PC1}, \text{PC2}) = -k_B T \left[\ln P(v_1, v_2) - \ln P_{\text{max}}\right] \tag{2}$$

Here $P(v1, v2)$ is the probability distribution obtained from a bivariate kernel density estimate,[40,41] which is subtracted to ensure that $\Delta\text{G} = 0$ for the lowest free energy minimum. Then the influence of the nth consecutive pair of residue in a component i is then expressed as the sum of the squares of the influence of its features:

$$I_{i,n} = \sum_{j=n}^{n+3} v_{i,j}^2 \tag{3}$$

where v_i is the eigenvector corresponding to component i and $v_{i,j}$ the coefficient corresponding to feature $q_{i,j}$.

Ward's minimum variance method

Considering that the dPCA is built on maximization of variance property, in order to find clusters of frames in the highest density regions of the projection, it is meaningful to group together minimum variance regions. Thus, Ward's minimum variance method[42] has been used to build a hierarchical clustering of the frames in the projected space. We then measure the discrete acceleration of the height of each consecutive cluster and we set the optimal number of clusters as the one that maximizes this acceleration. The acceleration on the x-axis is shifted so that the initial acceleration value is for a number of clusters equal to two. The ensuing clustering of frames allows to differentiate regions with the highest density in the system energy-landscape. Ward's minimum variance method also provides a good way to detect key moments in a given simulation where the system undergoes large dynamical changes.

Perturbation contact network analysis

Contact networks represent a protein as a collection of nodes, i.e. residues, that are connected by edges if those residues satisfy a contact condition. Here, in line with our previous works,[43–45] the contact condition is achieved if at least one heavy atom from a residue is at a distance below 5 Å from another heavy atom in another residue. Edges between residues are then weighted by the total number of atomic contact pairs that satisfy this contact condition. Individual contact networks can be obtained from experimental PDB structures or from frames of MD simulations. "Static" contact networks are derived from a single experimental structure while time-averaged networks of MD simulations correspond to dynamical contact networks. Then, in order to compare two contact networks (whether static or dynamical) and highlight contact differences between these structures we subtract one from the other (formally, we subtract their weighted adjacency matrices). The differences between the two contacts networks are visualized on the 3D model of the protein by assigning colors to the edges of the dynamical perturbation network according to the sign of the edges. Here, when we subtract the WT network from the mutant network, we assign the color red to a positive sign (i.e. stronger contacts in the mutant) and blue to a negative sign (i.e. stronger contacts in the WT). Finally, for visualization purposes, a weight threshold can be applied to select edges kept for display. Here, in line with previous works,[45] using a heavy-atom network, we used an absolute threshold of 5 when explicitly mentioned. Isolated nodes after this process are also pruned to simplify the visualization. The main advantage of such a method is to get a direct and global view of all interactions resulting from chain motions and to allow the detection of subtle movements including those occurring in loops.

Contact principal component analysis and clustering

All contacts belonging to different sets of simulations are reported in a matrix of size $N_{\text{frames}} \times N_{\text{contacts}}$ are reported. If a contact is not present in a frame, its value in the contact matrix is reported as zero. Using this approximation, the number of features is not fixed prior to computations and may vary. Then, the same methodology as described previously for dihedral angles is applied to get a projection of the frames in the contact feature space. In this case the influence of each contact in the PCA is directly given by the eigenvectors. Finally, we cluster the frames of each trajectory using Ward's minimum variance method described above.

Results

Static perturbation contact analysis

Recently available structures of the *Alpha*, *Beta* and *Gamma* variant RBD in complex with the ACE2 protein[18] give precious molecular basis for the understanding of altered binding in emerging variants. In Figure 2, we report the static perturbation contact network (PCN) between the RBD/ACE2 complex from the *Alpha Beta*, and *Gamma* variants (respectively PDB: 7EKF, 7EKG, 7EKC) and the WT (PDB: 6M0J), showing the main difference in atomic contacts deducible from X-ray experiments. Focusing at the WT, the interface between the spike RBD and the ACE2 involves various secondary structures elements in the spike RBD. First, in the $\alpha3$ helix, residue K417 is in contact with residue D30 located in the $\alpha1$ helix of the ACE2 receptor. Then, the $\alpha4$-$\beta5$ loop (residues D442-Y451) has a few contacts with the $\alpha1$ helix of ACE2 (i.e. G446-Q42 and Y449-D38). In the $\beta5$ sheet (residues L452-R454), residue Y453 is in contact with H34 of the ACE2 $\alpha1$ helix. The $\beta5$-$\beta6$ loop (residues L455-F490) is also mainly in contact with the $\alpha1$ helix (L455-H34, F456-T27,

N487-Q24, Y489-F28, Y489-T27, Y489-K31) but some residues are also interacting with the $\alpha2$ helix of the ACE2 receptor (N487-Y83, F486-L79, F486-Y83). In the $\beta6$ sheet (residues P491-Q493), residue Q493 is in contact with H34 and E35 of ACE2 $\alpha1$ helix. The nearby $\beta6$-$\alpha5$ loop of RBD (residue S494-Y505) is also interacting with ACE2 $\alpha1$ and with the β-turn (G352-D355), the most relevant contacts being: Q498-Y41, Q498-Q42, N501-Y41, N501-K353, Y505-K353. The largest number of atomic contacts (i.e 43 atomic pairs) in the WT is found for the interaction Y505-K353 while the N501-K353 and Q498-Y41 contacts are tied second (with 25 pairs). Among all mutated residues involved in the variants studied here, only K417 and N501 have a significant contact across the interface (< 5 atomic contacts) in the WT. It has to be noted that RBD residue E484 also possess a minimal contact (1 atomic pair) with K31 in the ACE2 $\alpha1$ helix.

In the *Alpha* variant, which contains only the N501Y mutation, the main contact changes are directly associated with this residue, featuring an increase in contact between the Y501-Y41 and Y501-K353 pairs (+14 and +11 atomic pairs, respectively). These increase in contacts are partially compensated by some contact losses, including those of the Q498-Q42 (-7) and Y505-Q37 (-5) interactions. Interestingly, far from the mutation spot, there is also increase in contact with H34 in the ACE2 $\alpha1$ helix associated to the Q493-H34 and the Y453-H34 interactions (+13 and +10 atomic pairs respectively) and some decrease of contact with the ACE2 $\alpha2$ helix, involving F486-L79 (-5) and F486-Y83 (-4). Overall, the increase in number of atomic contacts of *Alpha* variant with respect to WT is about 2%.

In the *Beta* variant, the same direct influence of the N501Y mutation are observed around residue N501. As expected, the K417N mutation breaks the K417-D30 salt-bridge and contact losses are observed for K417-D30 (-7) and for nearby contacts: Q493-H34 (-7), Q493-E35(-12) and L455-H34(-6) pairs. A slight increase in contacts for the Q493-K31 pair (+8) partially compensates this effect. The other mutation, i.e. E484K, breaks the weak E484-K31 contact (-1). Overall, the *Beta* variant feature a loss of about a 5% of contacts with respect to WT.

Finally, the *Gamma* variant is very similar to *Beta* but there the intensification of the Y501-Y41 contact is further magnified (15 atomic contacts in the WT, 26 in *Alpha*, 28 in *Beta* and 33 in *Gamma*) while contacts losses due to the loss of the K417-D30 salt bridge (T417-D30 also loses the 7 atomic contacts) are mitigated: only the Q493-E35 pair (-5) undergoes contact loss. Two other inter-residues interactions show some indirect effects of those mutations at the interface, i.e. F456-K31 (+6) and F486-L79 (+6). Similar to the *Alpha*, the *Gamma* variant features a ca. 1% of contact increase with respect to WT.

More general trends of intra-domains contact perturbations can be observed in the static PCN analysis, indicating that ACE2 contacts are more affected by mutations than RBD ones for the *Alpha*-to-*Gamma* variants and, overall, the *Gamma* variant features larger perturbations than the other two. The valuable information available from this static PCN analysis is however lacking dynamical effects that are going to be characterized in the following sections, where we also consider the comparison with the *Delta* and *Epsilon* variants that are lacking crystallographic structures.

Dihedral angle principal component analysis

We performed microsecond MD simulations on the WT and its five variants *Alpha*-to-*Epsilon* to characterize the effects of mutations on the RBD and ACE2 dynamics. RMSD analysis of these MD trajectories (Figures S1-S3) indicated that all systems equilibrated within 200 ns after the pre-equilibration steps, including the domains where most mutations are present, i.e. RBD and RBM. The dPCA has been initially performed on the whole (1 microsecond) MD simulation of each system (see Figure S4 in the SI). In all simulations, the PC1 and PC2 values undergo drastic adjustments between 200 and 600 ns. The latest major shift is attained in the *Gamma* variant at 600 ns and thereafter PC1 and PC2 values are stable. Thus, here and in all the remaining analysis of this work, we focus on the frames of the last 400 nanoseconds for all MD simulations (employing then $N_{\text{features}} = 722$ and $N_{\text{frames}} = 24,000$).

When representing MD frames in a PC1 vs PC2 plane, as depicted in Figure 3, the WT

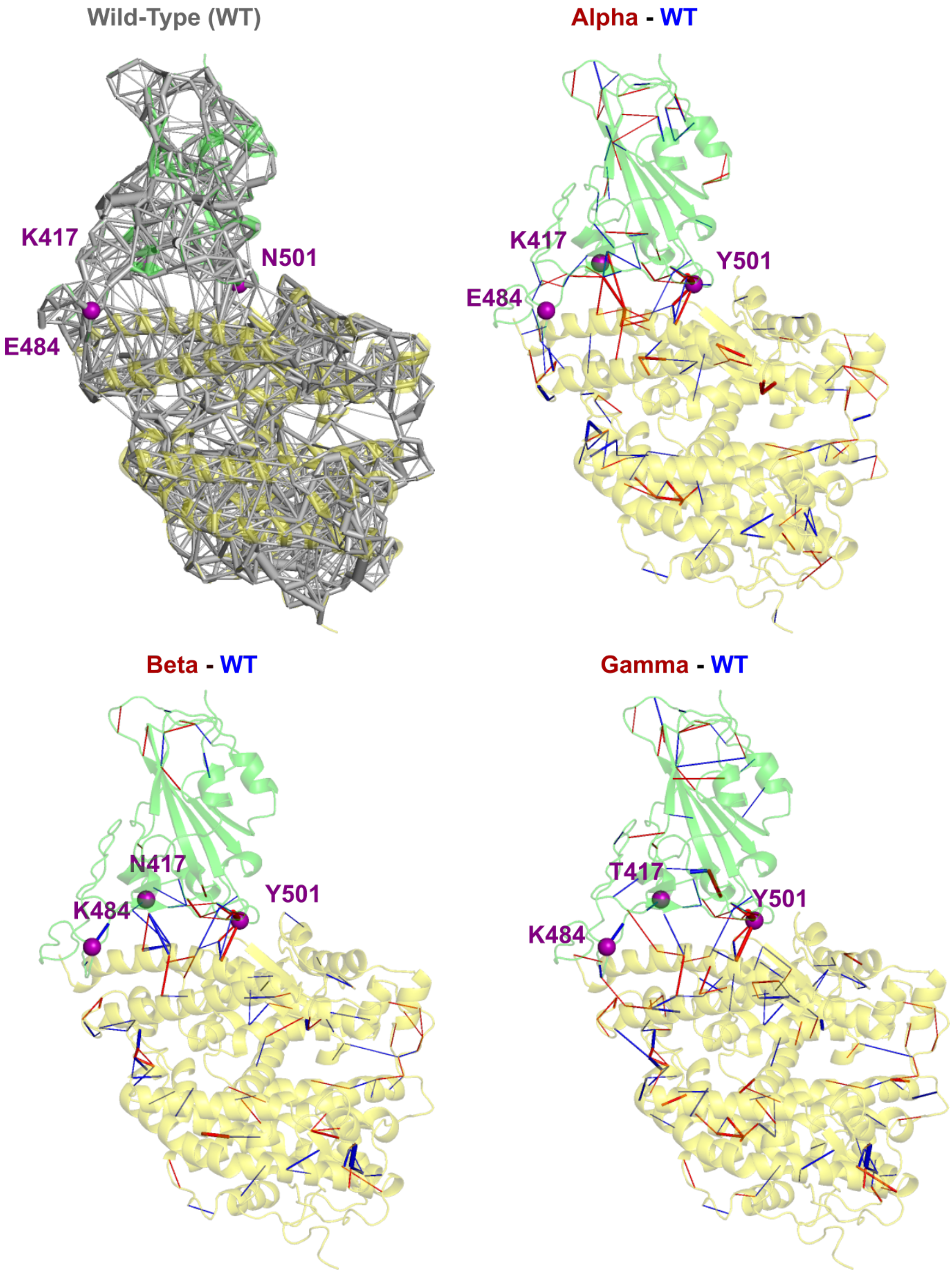

Figure 2: Static perturbation network at threshold 5 between *Alpha* (PDB: 7EKF), *Beta* (PDB: 7EKG), *Gamma* (PDB: 7EKC) and the WT (PDB: 6M0J)

and *Epsilon* systems are both isolated (PC1>0 and PC2<0 for WT; PC1>0 and PC2>0 for *Epsilon*) from *Alpha*-to-*Delta* variants that are grouped together (PC1<0, PC2≈0). This grouping of *Alpha*-to-*Delta* variants as function of the first two dPCA components suggests that different mutations might have similar effects on the RBD motion wit h respect to that

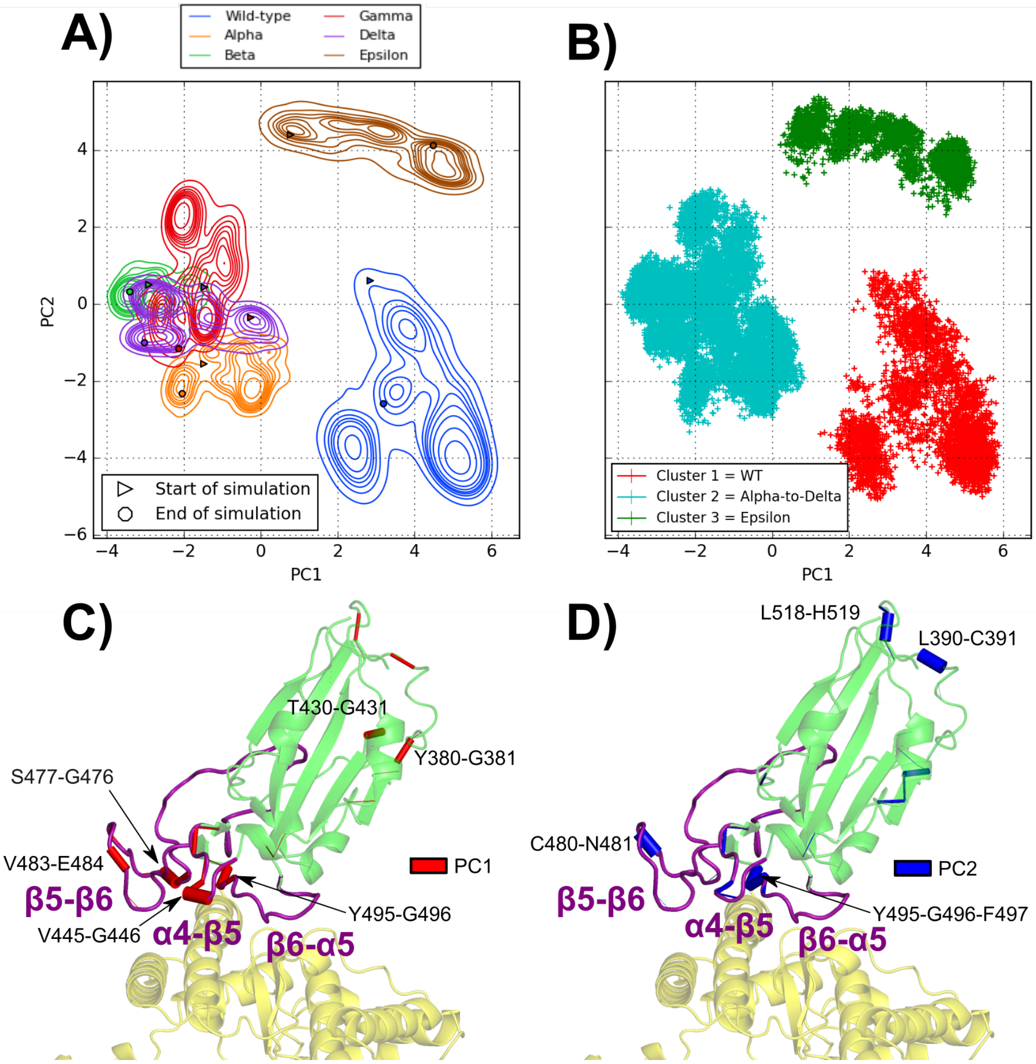

Figure 3: Projection of the frames corresponding to the final 400ns of simulation for the six studied complexes in the two dPCA eigenvector dimensions with (A) contour plots representing a kernel density estimate of the population of each complex, (B) scatter plot representing the three main clusters obtained through Ward's minimum variance method. Representation of the influence (as cylinders with a width proportional to the influence) of each dihedral angle in the PC1 (C) and PC2 (D) eigenvectors on the spike-RBD (green)/ACE2 (yellow) complex. The $\alpha4$-$\beta5$, $\beta5$-$\beta6$ and $\beta6$-$\alpha5$ loops are highlighted in purple.

of WT (see time evolutions of PC1 and PC2 in Figure S5D-E in the SI). In fact, the *Delta* variant does not share mutations with *Alpha*, *Beta* and *Gamma* that, instead, all have in common the N501Y mutation. Notably, the *Epsilon* variant, despite sharing the L452R mutation with *Delta*, is separated from it (see also Fig. S5D-E). The dPCA results indicate that the PC1 (i.e. the largest variance axis) discriminates the *Alpha*-to-*Delta* group from both the WT and the *Epsilon* variant. Looking at the main conformational changes in the MD simulations, one can realize that the motion relating WT and *Epsilon* (along PC1) refers to a large displacement of the α4-β5 loop (see Fig. S22 in the SI). On the other hand, the second principal component separates *Epsilon* from all the other systems mainly because they feature different fluctuations of the β5'-β6' loop (see Fig. S23 in the SI). Ward's minimum variance method quantitatively confirms this behavior, showing an optimal number of clusters (see Fig. S4 in the SI) equal to three, corresponding to the WT, *Epsilon* and *Alpha*-to-*Delta* groups. In Figure 3, the residue pairs with the most influence on the RBD dynamics are reported. The vast majority of these residues are located in three loops belonging to the RBM (438-506): α4-β5 (residues L455-F490), β5-β6 (residues L455-F490), β6-α5 (residue S494-Y505). It's interesting to note that the α4-β5 and β6-α5 loops are in contact and contain respectively mutations L452R (*Delta* and *Epsilon* variants) and N501Y (*Alpha*, *Beta* and *Gamma* variants). The time-evolution of the V483-E484 dihedral angles (see Fig. S5 in the SI) actually shows that their fluctuations are analogous in variants with (*Beta* and *Gamma*) or without (*Alpha*, *Delta*, *Epsilon*) the E484K mutation. On the other hand, in the WT, these dihedrals have a different behavior, i.e. featuring larger fluctuations and significant shifts in the microsecond simulations. This suggests that, in fact, despite the V483-E484 dihedral angle is involved in the main conformation motions of the RBD, the E484K mutation is not alone responsible for alterations of the RBM structure and motion. The above analysis of critical dihedral angles is quite useful, then, to understand the dynamics of the RBD upon mutations and to characterize some similarities and differences among various variants. However, dPCA does not provide an atomistic picture of the ACE2 and Spike RBD proteins responses to mutations. In order to recover this important information, an analysis of atomic contacts is reported in the next section, with a focus on the ACE2/RBD interface.

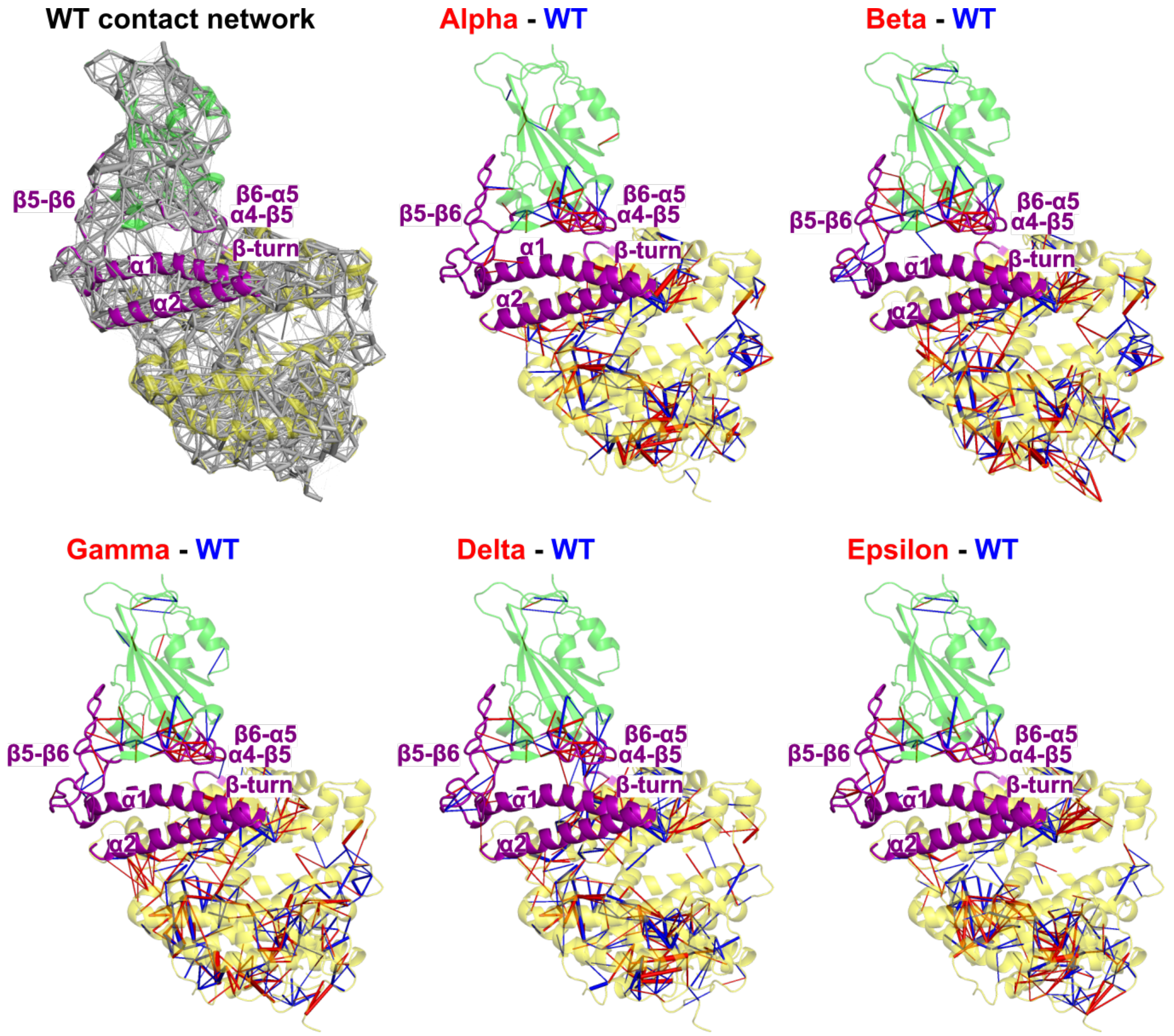

Figure 4: Complete Perturbation Network between each variant and the WT. The spike-RBD(green) / ACE2(yellow) complex is represented in cartoon. Stronger contacts in the WT are represented by a blue edge and in the variant in red. Edge width is proportional to their weight.

Dynamical perturbation contact network analysis

The dynamical contact network of the WT simulation and dynamical perturbation contact network (DPCN) between variants and the WT are reported in Figure 4. At first glance, the resemblance between DPCN from *Alpha*-to-*Delta* simulations is striking. Inside the Spike RBD, there is one main patch of contact changes located between the $\alpha 4$-$\beta 5$ and $\beta 6$-$\alpha 5$ loops that is present in the *Alpha*-to-*Delta* variant while a similar (but not identical) patch exists in the *Epsilon* variant. Interestingly, parts of the RBD located farther from the interface with ACE2 appear significantly less affected by mutations. The interface between the two proteins displays some contact changes but, with the notable exception of *Delta*, these are of a lesser magnitude (i.e. smaller number of total atomic contacts for each residue pair) than internal contacts perturbations in the RBD and in the ACE2 receptor. Surprisingly, the ACE2 receptor is subject to much more contact changes than the RBD upon mutations and some resemblance between contact perturbations can be observed between the five variants. Looking at the propagation of perturbations within the ACE2 receptor, from the RBD interface to the opposite side of the ectodomain, one could speculate that, upon mutations in RBD, the binding of these five Spike variants might eventually trigger a response of the ACE2 receptor that significantly differs from that of the WT.

Notably, when considering the total number of average contacts at the interface in the last 400 ns of MD simulations (see Figure S7 in the SI), all variants feature less atomic contacts at the interface than the WT. In particular, the interface between the ACE2 receptor and the *Alpha* and *Beta* variant shows a decrease of 12% in atomic contacts, the *Gamma* interface a decrease of 11%, while the *Delta* and *Epsilon* interfaces decrease by 4%. In Figure 5, a close view of the DPCN near the ACE2/RBD interface is reported along with the list of contact pairs involved (Figure 4B). The Spike RBD binds to three main areas of the ACE2 receptor, two helices, i.e. $\alpha 1$ (residues T20 to Y41), and $\alpha 2$ (mainly residues L79, M82 and Y83), and a β-turn (residues G352-D355). Among the WT residues mutated in the five variants, which are all located close to the RBD/ACE2 interface, only residues K417 and N501 are involved in the interface contacts during the MD simulation of WT, i.e. possessing (in average) > 5 atomic contacts with ACE2. Despite only two mutated residues are directly involved in the interface contacts, other atomic contacts at the interface are indirectly affected by mutations. Here, we describe the direct and indirect contact perturbations upon mutations in the five variants. As shown in Figure 5A, the *Delta* is certainly the variant that features the largest number of interface contact perturbations despite, as described below, its mutations are not

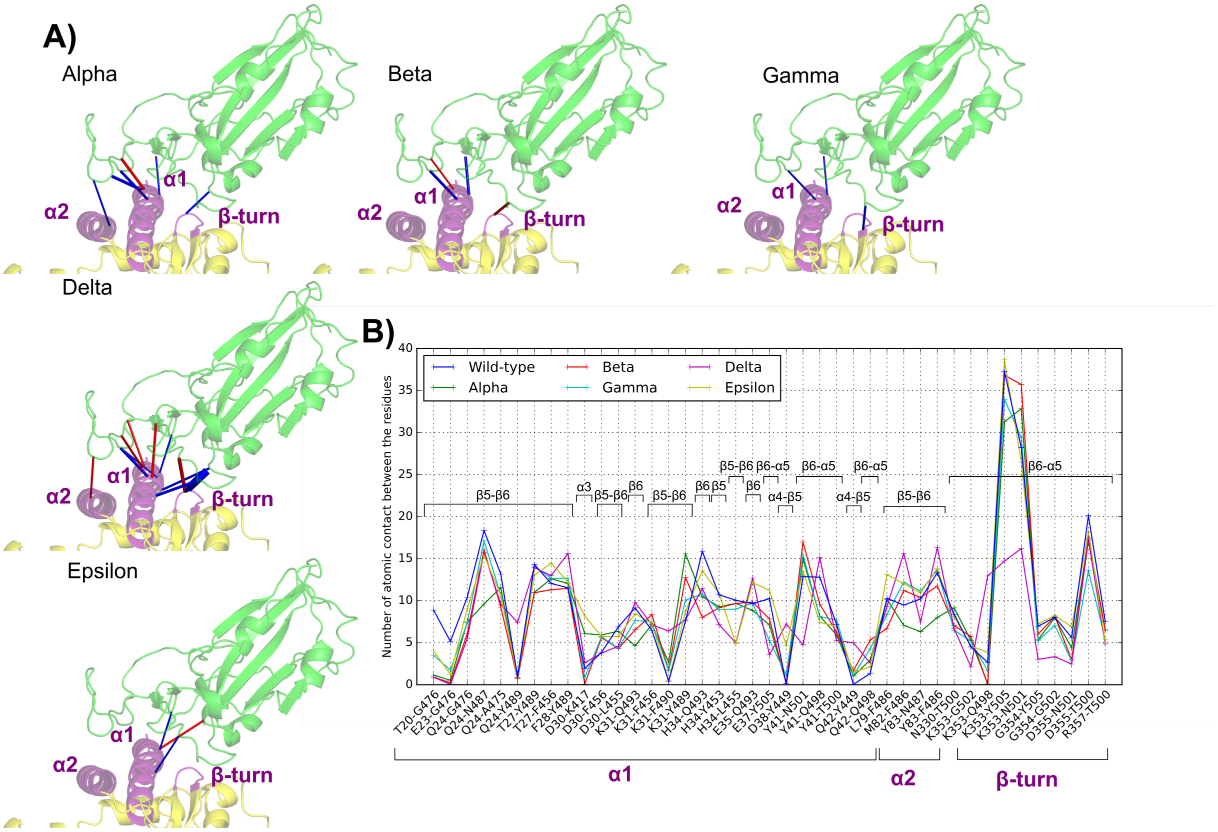

Figure 5: (A) Perturbation networks using a threshold value of 5 between the WT RDB (green)/ACE2 (yellow) complex and its mutants (*Alpha*, *Beta*, *Gamma*, *Delta* and *Epsilon*). Stronger contacts in the WT are represented by a blue edge and in the variant in red. Edges width is proportional to their weight and visualization factor the same for each variant. (B) Average number of interresidual atomic contacts in all pairs at the interface (labeled in the WT residue name) with more than 5 contacts in at least one simulation.

directly involved in interface contacts.

In the WT, K353 residue in a β-turn of the ACE2 belongs to a dense interface contact network with the $\beta6$-$\alpha5$ loop of RBD (see main peak in Figure 5B), involving the K353-N501 and K353-Y505 interactions. Notably, in the *Delta* variant, despite N501 is conserved, these two contacts are disrupted and a new interface interaction is established between K353 and Q498. In the other variants, the K353-Y505 contact remains stable, but the K353-N501 interaction (stable in *Epsilon*) becomes slightly stronger in all N501Y variants, as a consequence of the π-cation formation mentioned above. Indeed, as discussed in the static PCN analysis, the K353-Y501 π-cation formation in the *Alpha*, *Beta* and *Gamma* variants is accompanied by that of a T-shaped π-stacking interaction between Y501 and Y41, located at the $\alpha1$ of ACE2. In contrast, in the *Delta* variant, the Y41-N501 contact is substituted by a stronger Y41-Q498 interaction.

In all models, the $\alpha2$ helix of ACE2 is in contact with two residues of the RBD $\beta5$-$\beta6$ loop: F486 and N487. With respect to the WT, the *Alpha* variant features a slight decrease of all contacts in this region while *Beta* and *Gamma* remains relatively untouched. The *Delta* variant shows again the most disparities: an increase in the M82-F486 and Y83-F486 contacts and a decrease in the Y83-N487 contact are detected. The proximity of these residues with mutation T478K suggests an indirect effect of this mutation (specific of *Delta* variant) on the ACE2/RBD interface. In the *Epsilon*, just a slight increase in the L79-F486 contact is detected and the rest of contacts remains similar to those of the WT.

The $\alpha1$ helix of ACE2 is in contact with many different secondary structures of the Spike. In particular, contacts with the $\beta5$-$\beta6$ loop of RBD involves the Y489 residue that features interesting contact perturbations upon mutations at the interface with ACE2 $\alpha1$. In fact, Y489 strengthens the contact with residue F28 in WT while it establishes a new contact with residue Q24 in the *Delta* variant. In other variants, instead, the Y489-K31 is strengthened as a consequence of the loss of the weak E484-K31 electrostatic interaction found in the WT. Still, at the $\alpha1$ (nearby K31) D30 establishes a salt bridge with residue K417, another mutation spot. This K417-D30 salt-bridge has been found as a transient contact in MD simulations of *Epsilon* and *Alpha* variants but this interaction is never observed in the *Beta* and *Gamma* trajectories, featuring the K417N and K417T mutations, respectively. Surprisingly, in *Delta* and WT, without K417 mutation, this salt-bridge is also broken during the dynamics. While in the available X-ray structures (WT and *Alpha*) the K417-D30 salt-bridge results to be present, our MD simulations suggests that this interaction might be actually weak and prone to rupture.

The *Alpha*-to-*Gamma* dynamics reproduce the main interface perturbation found in all corresponding crystal structures, which is the enhanced interactions between Y501, K353 and Y41. In *Beta* and *Gamma*, the contact loss associated to the K417(N/T)-D30 salt-bridge breaking is also consistent with crystal structures. Interestingly, the WT dynamics shed light about the statistical significance of the K417-D30 interaction, since this salt-bridge features a breaking-formation dynamic even in absence of mutations.

Importantly, Delta mutation spots do not belong to the interface contacts but they evidently have a significant impact at the interface. More generally, studying systematically the indirect effects of mutations is challenging, especially for comparative studies of mutants, and a more general type of analysis pointing at the most significant contact changes in various systems is required to understand why, for instance, the Delta variant features the largest interface perturbations despite the absence of interface mutations.

Contact principal component analysis

The cPCA is used to characterize the overall information on dynamical contacts resulting from MD simulations of WT and RBD variants into its PCs. In particular, we found 9,432 different contacts in the concatenated trajectories of the WT and five variants (considering the last 400 ns for each system). As shown in Figure 6A, the scatter plot of the first two PCs shows how cPCA can cluster frames featuring similar dynamical contacts and, thus, characterize different systems according to that. In contrast to dPCA, here, frames are separated in four main clusters: one with the WT and the *Epsilon* variant (negative PC1 and PC2), one with the *Delta* variant (positive PC2 and negative PC1), one with the *Alpha* variant (positive PC1 and PC2) and one with the *Beta* and *Gamma* variant (positive PC1 and negative PC2). In this representation, positive values of the PC1 separate *Alpha*, *Beta* and *Gamma* from WT, *Delta* and *Epsilon*. The positive values of the PC2, instead, discriminate *Alpha* and *Delta* from *Beta*, *Gamma*, *Epsilon* and the WT. The following PCs, i.e. those referring to smaller eigenvalues than the two largest ones, are associated to specific separations between systems: the third component separates the WT (negative PC3), the

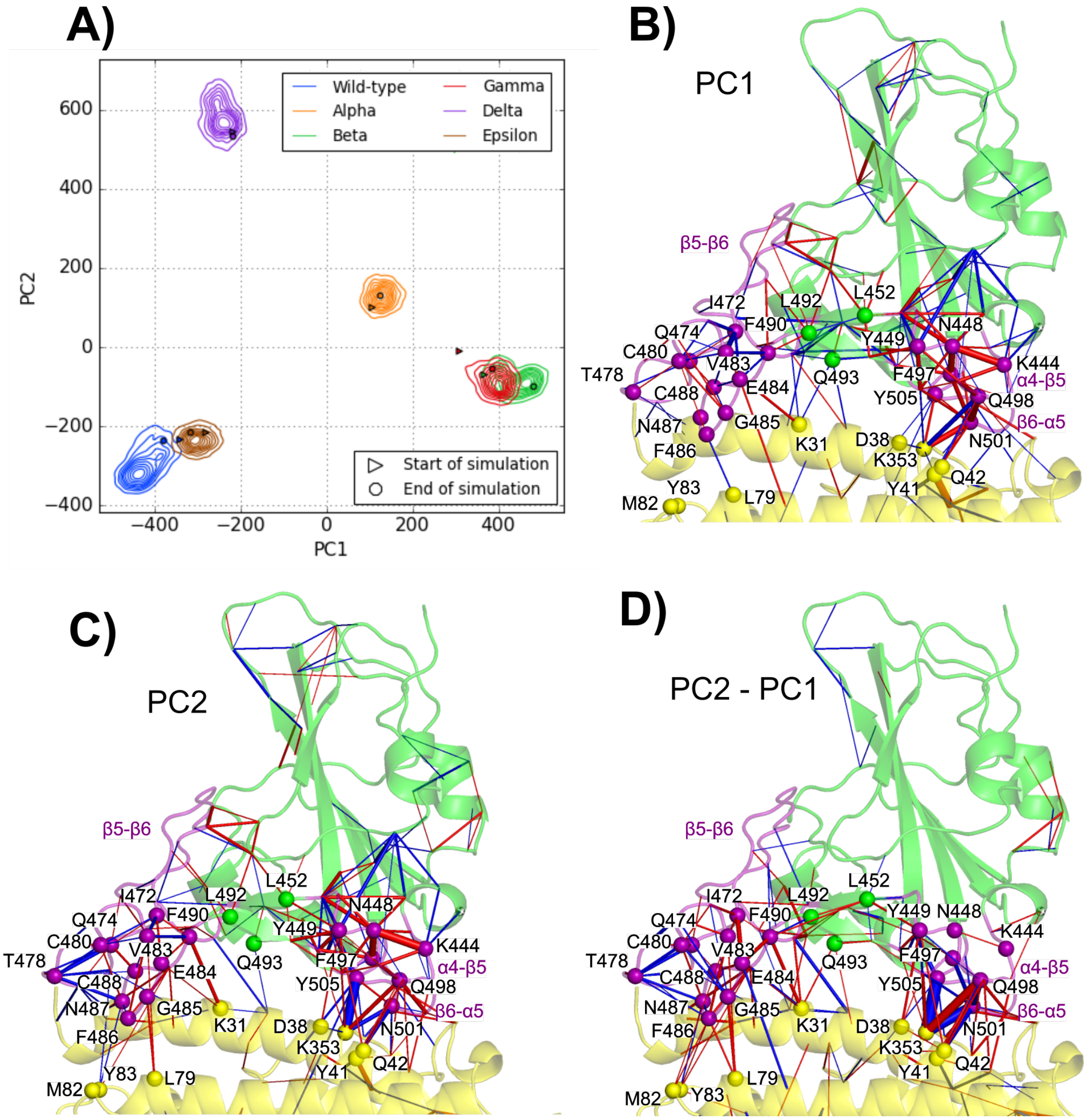

Figure 6: Projection of the frames corresponding to the final 400ns of simulation for the six studied complexes in the two cPCA eigenvector dimensions with (A) terrain lines representing a kernel density estimate of the population of each complex. Network representation of the influence (as cylinders with a width proportional to the influence) of each contact in the PC1 (B) and PC2 (C) and PC2-PC1 (D) eigenvectors projected on the spike-RBD(green) / ACE2(yellow) WT complex. Blue edges show a negative contribution to the principal component while red edges show a positive contribution in the principal component. Contacts with a contribution of less than 1% to the eigenvector were discarded.

Epsilon (positive PC3) from the rest while the fourth one separates *Alpha* (positive PC4) and *Gamma* (negative PC4) from the rest and, finally, the fifth component discriminates between *Alpha* and *Gamma* (negative PC5) from *Beta* (positive PC5) and the rest. Smaller components than PC5 are associated to dynamical contact changes within simulations of each system, e.g. PC6 relates to dynamic contacts occurring in the *Delta* variant. In the dPCA, instead, this kind of clustering associated to each specific system starts with the third principal component. Thus, cPCA provide finer distinctions between the systems under investigation, in terms of dynamical contact changes, with respect to dPCA, especially showing some characteristics of the *Delta* variant.

The representation of PC1 (with positive values for *Alpha*, *Beta* and *Gamma* and negative values for the rest) and PC2 (with positive values for *Delta* and *Alpha* and negative values for the rest) in terms of contact networks near the interface, is depicted in Figure 6B-C. Therefore, in order to better differentiate the *Delta* network from the others, also the PC2-PC1 difference is represented in terms of contact network (see Figure 6D), with PC2-PC1 positive values being associated to number of contacts that are large in *Delta* and small in *Beta* and *Gamma* variants, while viceversa for negative values of PC2-PC1 difference (*Alpha*, *Epsilon* and WT contribute only minimally to this network since PC2-PC1 differences are small in these cases). Here, the analysis of the PC2-PC1 differences provides insights into the link between the two *Delta* mutations (T478K and L452R) and their indirect effects at the interface.

Starting from the T478K mutation (exclusive to *Delta*), located in the RBD β5-β6 loop, we found that this residue is a central hub of negative edges in both the PC2 and the PC2-PC1 networks (see Figure 6C-D), involving contacts with residues Q474, C480, F486, N487 and C488. This indicates that few contacts between those residues are characteristic of the *Delta* variant. In fact, as shown in Figure 7A, a hydrophobic clusters is observed nearby the C480-C488 disulfide bridge in the β5-β6 loop of the WT (and also in all other variants but *Delta*), involving the hydrophobic moieties of Q474 and T478 and residues I472, V483 and F490. Upon T478K mutation, in the *Delta* variant, the insertion of the lysine side chain does not allow for such arrangement and consequently residue K478 is repelled out of the cluster. This loss of interaction in *Delta* is associated with flipping of the C480-C488 bridge

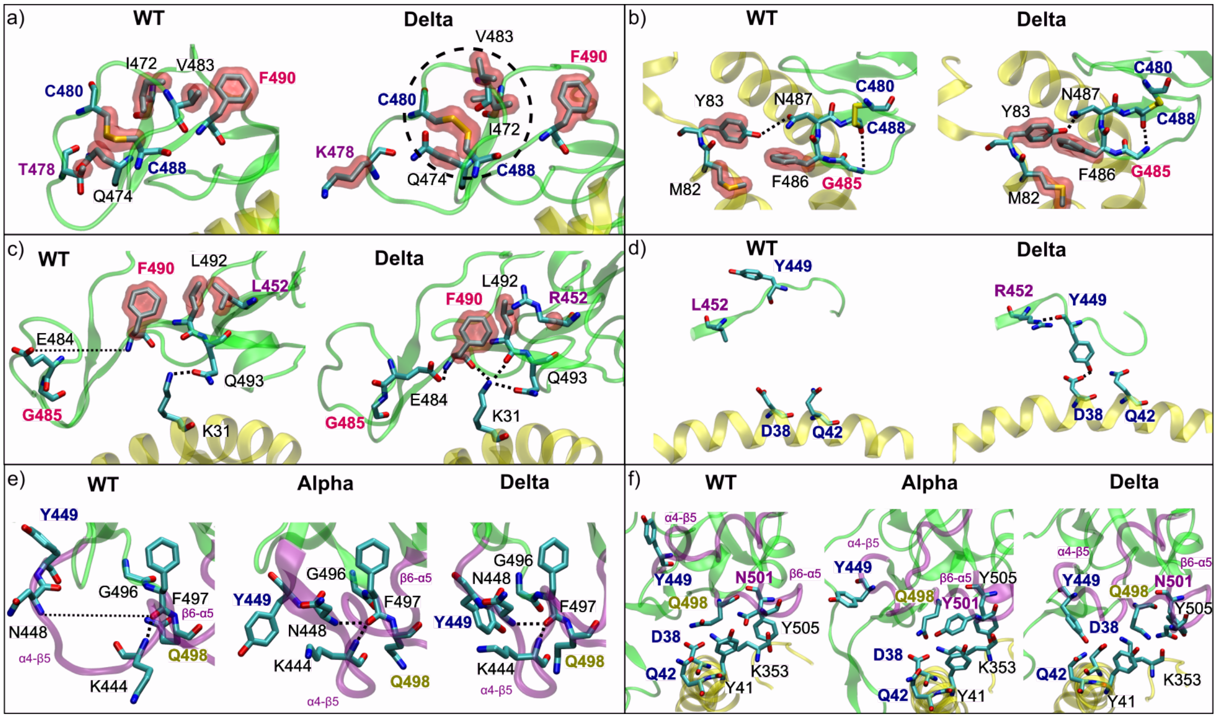

Figure 7: Representative MD snapshots of some contacts in the different models emphasized by contact analysis.

that in turns push residue F490 far from the cluster. As a consequence of this rearrangement of the hydrophobic cluster, a backbone G485(NH)-C488(O) hydrogen bond is stabilized in *Delta*, determining a better folding of the $\beta5$-$\beta6$ loop, as depicted in Fig. 7b. This differently folded structure also inevitably affects the dynamics of residues F486 and N487, which were previously highlighted in the DPCN of *Delta* at the RBD/ACE2 interface. These residues, indeed, show more contacts with M82 and Y83 (located in the ACE2 $\alpha2$ helix) in the *Delta* variant than in the WT. This proves that the T478K mutation is indirectly responsible of the contact increase between the Spike-RBD and the $\alpha2$ helix of ACE2.

The change in folding of the $\beta5$-$\beta6$ loop induced by the T478K mutation *Delta* has, moreover, other indirect effects on the RBD/ACE2 interface that are synergic with the effects of the L452R mutation. In fact, as shown in the PC2-PC1 network in Fig. 7D, the negative edges around the T478K mutation in the RBD $\beta5$-$\beta6$ loop are somehow compensated by the positive edges around residue F490, i.e. the residue repelled out of the hydrophobic cluster in the *Delta* variant. This set of predominantly positive edges involves residues E484 (neighbor of G485), L492, L452 (mutated to R in *Delta*) and K31 across the interface. Indeed, the perturbations from the T478K appear to be connected to those induced by the L452R mutation through residue F490 (in the $\beta5$-$\beta6$-loop) and L492 (in the $\beta6$-sheet). Figure 7c shows that the contemporary T478K and L452R mutations in *Delta* have a significant effect on hydrogen bonding network around the interface residue K31. In particular, the the dynamics of residue F490 is synergistically affected by the two mutations from two different sides: on one side the change in folding of the $\beta5$-$\beta6$ loop upon T478K mutation stabilizes the E484-F490 hydrogen bond while, on the other side, since F490 is also in hydrophobic contact with L492 and L452, upon L452R mutation, the arginine sidechain promotes hydrogen bonding interactions of the L492 and F490 backbones with the K31 sidechain. This finally results into three hydrogen bonds between the NH3+ head of K31 and the sidechain oxygen of Q493 and the backbone oxygens of L492 and F490, which is a characteristic interface arrangement of the *Delta* variant (i.e. in the WT only Q493 is hydrogen-bonded with K31) and it results from the combination of the two T478K and L452R mutations (far from the interface). Here, we should note that the Omicron variant also possess the T478K mutations (same as *Delta*) but in conjunction with the E484A mutation. This latter mutation is somehow surprising since the E484K mutation is very common in Spikes' mutants (e.g. it is found in *Beta*, *Gamma*, *Mu*, *Lambda*, *Eta*, *Theta*) while E484A is exclusive to Omicron. This open to the question of how much the E484A mutation in *Omicron* could influence the effects of the *Delta* T478K mutations, which should be addressed in further studies. It is worth mentioning that the E484K mutation, present in *Beta* and *Gamma* but not in *Alpha*, differentiates these variants in terms of the interface contacts between the ACE2 receptor and the RBD $\beta5$-$\beta6$ loop, involving the network of contacts highlighted in this region by the PC1 and PC2 components. In the PC2-PC1 contact network (see Fig. 7D), residue L452(R) is a bridging node that connects the contact perturbations in the $\beta5$-$\beta6$ loop (described above and involving the T478K mutation) with those of the $\alpha4$-$\beta5$ and $\beta5$-$\beta6$ loops. Residue L452(R) has a positive edge with residue Y449 in this network, meaning that a close L452(R)-Y449 contact is typical of the *Delta* variant. In turn, Y449 displays direct connections with interface residue, featuring positive edges with D38 and Q42 in the ACE2 receptor. Figure 7d shows that, indeed, upon L452R mutation, the arginine sidechain is able to make a hydrogen bond with Y449(O), which promotes a flipping of Y449 sidechain, allowing for the formation of a Y449-D38 interface hydrogen bond that alters the surrounding H-bonding network, involving also Q42. Notably, the perturbations around residue Y449 in the PC2-PC1 network are minimal, as a consequence of the fact that perturbations inside the spike RBD (i.e. in the α4-β5 and β6-α5 loops) are rather similar in the PC1 and PC2 networks (see Fig. 6B,C). In particular, the K444-N448, N448-F497 pairs features a large number of contacts in PC1 and PC2 but they virtually vanish in the PC2-PC1 network, indicating that rearrangements of contacts in this region are significant in all variants but somehow differ from *Epsilon* that is more similar to WT, in line with the DPCN results depicted in Figure 4. At the same time, the largest PC2-PC1 differences are found at the interface between these spike RBD loops and the ACE2 receptor. Here, in the DPCN interface analysis we highlighted the role of residues Q498, N501, Y505 in the β6-α5 loop in contact with Y41 in the α1 helix and K353 in the β-turn. Figure 7e shows how, in the *Delta* variant, upon the Y449 flipping mentioned above, the backbone N448(NH)-F497(O) hydrogen bond adds up to the preexisting K444(NH)-F497(O) one. Very interestingly, the very same two hydrogen bonds are also formed in the *Alpha* variant, featuring the sole N501Y mutation. This indicated that such single mutation int the *Alpha* RBD creates a H-bonding network in the α4-β5 and β6-α5 loops of RBD similar to that produced by the indirect effects of the L452R mutation in the *Delta* variant (via residue Y449). Notably, these contact changes at the RBD common to both L452R and N501Y mutations in *Delta* and *Alpha*, respectively, have not an effect on the interface contacts. In fact, as shown in Figure 7f, the WT and *Alpha* interfaces involve interactions between the same residues (i.e. D38, Y41, Q42, K353, Q498, N501(Y), Y505) despite the presence of the N501Y mutation, which only changes the type of some interactions (most notably the Y41-N501 π-polar interaction is promoted to a Y41-Y501 π-π interaction). On the other hand, the interface in the *Delta* variant largely differs from those of the WT and *Alpha* since the involved residues now include Y449 instead of N501 and Y505. Interestingly, this shows how the indirect effect of *Delta* L452R mutation on the interface contacts, via the Y449 residue and the Y449-D38 interaction (see Figure 7d), has a large impact on the RBD/ACE2 interface as previously mentioned in the DPCN analysis, see Fig. 5A. As a result of the L452R mutation in *Delta* variant, thus, the formation of the R452-Y449 interaction is associated to structural rearrangements of the α4-β5 and β6-α5 loops that modify the interface contacts by including the Y449-D38 hydrogen bond and substituting the N501-Y41 interaction with the Q498-Y41 hydrogen bond, pushing away residues N501 and Y505 from the interface (breaking their contacts with residue K353). As evident from Figure 6D, in fact, these interface changes are the most prominent in the PC2-PC1 network and represent the long distance effects of the L452R mutation on the RBD/ACE2 interface.

Conclusions

In this study, we first analyzed the (static) networks of atomic contacts between the Spike RBD protein and the ACE2 human receptor based on the available crystallographic structures of the *Alpha*-to-*Gamma* variants of SARS-CoV-2, capturing the contact changes with respect to the WT and thus perturbations due to RBD mutations. Then, in order to account for dynamical effects of RBD mutations on Spike/ACE2 interface contacts, microsecond MD simulations have been performed on the WT and the *Alpha*-to-*Epsilon* variants. Various tools for MD trajectories analysis have been used to recover the main similarities and differences between various Spike RBD variants interacting with the human ACE2 receptor.

First, the analysis of protein essential motions based on backbone dihedral angles, namely dPCA, allowed recognizing mobile RBD regions whose dynamics is altered by mutations. The first principal components of backbone dihedral angles is associated with motions in the α4-β5 loop while the second principal components is associated with motions in the β5-β6 loop. Considering these essential motions, three distinct behaviour have been observed for the various MD simulations: on one side, a cluster involving the *Alpha*, *Beta*, *Gamma* and *Delta* variants features a tight $\alpha4$-$\beta5$ loop and a flexible $\beta5$-$\beta6$ loop; on the other hand, the WT features a flexible $\alpha4$-$\beta5$ loop and a tight $\beta5$-$\beta6$ loop; while for the *Epsilon* variant, the tightest $\beta5$-$\beta6$ loop was observed along with a partially flexible $\alpha4$-$\beta5$ loop. Interestingly, this clustering correlates with the impact in transmissibility and severity of the SARS-CoV-2 disease in the studied variants. These results suggest that the L452R and N501Y mutation have closely related effects on RBD motions near the interface. However, as evidenced by the dPCA of the *Epsilon* variant, these motions are not fully reproduced in the absence of the T478K mutation, which indicates an interdependence between these mutations. Still, the dPCA analysis did not allow to differentiate the *Delta* variant, the dominating one in most of the 2021 year, from the others.

Then, we were able to recover some specificity of the *Delta* variant by studying the dynamical perturbation contact network, with a focus on the RBD/ACE2 interface. The comparisons between WT atomic contact network with those of *Alpha*-to-*Epsilon* variants showed many similarities among the *Alpha*-to-*Gamma* variants that share the N501(Y) mutation, which promotes specific perturbations for the interface contacts of Y501 with K353 and Y41 residues, while the rest of interface contacts remains essentially preserved. By contrast, in the *Delta* variant, significant contact changes at the interface have been found despite the absence of interface mutations. Indeed, all interface contact changes in *Delta* cannot be directly attributed to the T478K and L452R mutations that must have indirect (but large) effects on the interface.

The subsequent cPCA analysis shed, finally, light on the propagation of contact perturbations induced by the T478K and L452R mutations in the *Delta* variant. This analysis showed that the T478K mutation alters the contacts of a hydrophobic cluster (involving residues Q474, T478, I472, V483 and F490) around the C480-C488 disulfide bridge inside the $\beta5$-$\beta6$ loop of the RBD and promotes the formation of a G485-C488 backbone hydrogen bond. In turn, this rearrangement affects the position of residue F486 and N487 that increase their interface contacts with the $\alpha 2$ helix of ACE2. At the same time, in the WT residue F490, L492 and L452 are involved in another hydrophobic cluster that upon L452R mutations adjusted because of both the presence of residue R452 and the alteration of F490 contacts due to the T478K mutation. In turn, residue F490 and L492 create a triple hydrogen bond at the interface of *Delta* with residue K31, which was H-bonded just to residue Q493 in the WT. Since it belongs to both the aforementioned T478- and L452-related hydrophobic clusters, residue F490 resulted to be central for the propagation of contacts changes due to the simultaneous T478K and L452R mutations that result to cooperate in inducing the interface perturbations found in *Delta*.

Our results highlight the singular mechanism of action of the mutations in the *Delta* variant that could eventually explain why it dominated over preceding variants. Moreover, since the recent *Omicron* variant possess the same T478K mutation but in conjunction with the E484A one, it remains to elucidate if a synergistic long-range effect of multiple mutations like that found here for the *Delta* variant is also operating for the currently dominating *Omicron*.

Acknowledgement

This work was granted access to the HPC resources of IDRIS under the allocation 2021-AP010712548 made by GENCI.

Experiments presented in this paper were carried out using the Grid'5000 testbed, supported by a scientific interest group hosted by Inria and including CNRS, RENATER and several Universities as well as other organizations (see `https://www.grid5000.fr`).

AG and LV thank the support of CNRS the 80prime and MITI programs and acknowledge the support of the Institut Rhônalpin des systèmes complexes, IXXI-ENS-Lyon, Lyon, France.

AG and acknowledge the support of the Federation of European Biochemical Societies for its Short-Term Fellowship and the use of the “Pôle Scientifique de Modélisation Numérique”

(PSMN) at the École Normale Supérieure de Lyon, France.

AG thanks Federica Maschietto for fruitful discussions about Ward's minimum variance method.

Supporting Information Available

RMSD fluctuations, dPCA and cPCA on the complete trajectory, time evolution of key distances and dihedral angles, cartesian coordinates PCA, cPCA components up to 8, time evolution of interface loops flexibility.

References

(1) Wang, Q.; Zhang, Y.; Wu, L.; Niu, S.; Song, C.; Zhang, Z.; Lu, G.; Qiao, C.; Hu, Y.; Yuen, K.-Y., et al. Structural and functional basis of SARS-CoV-2 entry by using human ACE2. *Cell* **2020**, *181*, 894–904.

(2) Walls, A. C.; Park, Y.-J.; Tortorici, M. A.; Wall, A.; McGuire, A. T.; Veesler, D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* **2020**, *181*, 281–292.

(3) Lan, J.; Ge, J.; Yu, J.; Shan, S.; Zhou, H.; Fan, S.; Zhang, Q.; Shi, X.; Wang, Q.; Zhang, L., et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **2020**, *581*, 215–220.

(4) 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.

(5) Castillo, A. E.; Parra, B.; Tapia, P.; Acevedo, A.; Lagos, J.; Andrade, W.; Arata, L.; Leal, G.; Barra, G.; Tambley, C., et al. Phylogenetic analysis of the first four SARS-CoV-2 cases in Chile. *J. Med. Virol.* **2020**, *92*, 1562–1566.

(6) Chen, J.; Wang, R.; Wang, M.; Wei, G.-W. Mutations strengthened SARS-CoV-2 infectivity. *J. Mol. Biol.* **2020**, *432*, 5212–5226.

(7) Shang, J.; Ye, G.; Shi, K.; Wan, Y.; Luo, C.; Aihara, H.; Geng, Q.; Auerbach, A.; Li, F. Structural basis of receptor recognition by SARS-CoV-2. *Nature* **2020**, *581*, 221–224.

(8) Davies, N. G.; Abbott, S.; Barnard, R. C.; Jarvis, C. I.; Kucharski, A. J.; Munday, J. D.; Pearson, C. A.; Russell, T. W.; Tully, D. C.; Washburne, A. D., et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B. 1.1. 7 in England. *Science* **2021**, *372*.

(9) Funk, T.; Pharris, A.; Spiteri, G.; Bundle, N.; Melidou, A.; Carr, M.; Gonzalez, G.; Garcia-Leon, A.; Crispie, F.; O'Connor, L., et al. Characteristics of SARS-CoV-2 variants of concern B. 1.1. 7, B. 1.351 or P. 1: data from seven EU/EEA countries, weeks 38/2020 to 10/2021. *Eurosurveillance* **2021**, *26*, 2100348.

(10) Davies, N. G.; Jarvis, C. I.; Edmunds, W. J.; Jewell, N. P.; Diaz-Ordaz, K.; Keogh, R. H. Increased mortality in community-tested cases of SARS-CoV-2 lineage B. 1.1. 7. *Nature* **2021**, *593*, 270–274.

(11) Tegally, H.; Wilkinson, E.; Giovanetti, M.; Iranzadeh, A.; Fonseca, V.; Giandhari, J.; Doolabh, D.; Pillay, S.; San, E. J.; Msomi, N., et al. Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature* **2021**, *592*, 438–443.

(12) Cele, S.; Gazy, I.; Jackson, L.; Hwa, S.-H.; Tegally, H.; Lustig, G.; Giandhari, J.; Pillay, S.; Wilkinson, E.; Naidoo, Y., et al. Escape of SARS-CoV-2 501Y. V2 from neutralization by convalescent plasma. *Nature* **2021**, *593*, 142–146.

(13) Madhi, S. A.; Baillie, V.; Cutland, C. L.; Voysey, M.; Koen, A. L.; Fairlie, L.; Padayachee, S. D.; Dheda, K.; Barnabas, S. L.; Bhorat, Q. E., et al. Efficacy of the ChAdOx1 nCoV-19 Covid-19 vaccine against the B. 1.351 variant. *N. Engl. J. Med.* **2021**, *384*, 1885–1898.

(14) Faria, N. R.; Mellan, T. A.; Whittaker, C.; Claro, I. M.; Candido, D. d. S.; Mishra, S.; Crispim, M. A.; Sales, F. C.; Hawryluk, I.; McCrone, J. T., et al. Genomics and epidemiology of the P. 1 SARS-CoV-2 lineage in Manaus, Brazil. *Science* **2021**, *372*, 815–821.

(15) Dejnirattisai, W.; Zhou, D.; Supasa, P.; Liu, C.; Mentzer, A. J.; Ginn, H. M.; Zhao, Y.; Duyvesteyn, H. M.; Tuekprakhon, A.; Nutalai, R., et al. Antibody evasion by the P. 1 strain of SARS-CoV-2. *Cell* **2021**, *184*, 2939–2954.

(16) Sheikh, A.; McMenamin, J.; Taylor, B.; Robertson, C. SARS-CoV-2 Delta VOC in Scotland: demographics, risk of hospital admission, and vaccine effectiveness. *Lancet* **2021**,

(17) Deng, X.; Garcia-Knight, M. A.; Khalid, M. M.; Servellita, V.; Wang, C.; Morris, M. K.; Sotomayor-González, A.; Glasner, D. R.; Reyes, K. R.; Gliwa, A. S., et al. Transmission, infectivity, and antibody neutralization of an emerging SARS-CoV-2 variant in California carrying a L452R spike protein mutation. *MedRxiv* **2021**,

(18) Han, P.; Su, C.; Zhang, Y.; Bai, C.; Zheng, A.; Qiao, C.; Wang, Q.; Niu, S.; Chen, Q.; Zhang, Y., et al. Molecular insights into receptor binding of recent emerging SARS-CoV-2 variants. *Nat. Commun.* **2021**, *12*, 1–9.

(19) Mannar, D.; Saville, J. W.; Zhu, X.; Srivastava, S. S.; Berezuk, A. M.; Zhou, S.; Tuttle, K. S.; Kim, A.; Li, W.; Dimitrov, D. S., et al. Structural analysis of receptor binding domain mutations in SARS-CoV-2 variants of concern that modulate ACE2 and antibody binding. *Cell Rep.* **2021**, *37*, 110156.

(20) Mu, Y.; Nguyen, P. H.; Stock, G. Energy landscape of a small peptide revealed by dihedral angle principal component analysis. *Proteins: Struct., Funct., Bioinf.* **2005**, *58*, 45–52.

(21) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.

(22) Phillips, J. C.; Hardy, D. J.; Maia, J. D.; Stone, J. E.; Ribeiro, J. V.; Bernardi, R. C.; Buch, R.; Fiorin, G.; Hénin, J.; Jiang, W., et al. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J. Chem. Phys.* **2020**, *153*, 044130.

(23) Huang, J.; MacKerell Jr, A. D. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* **2013**, *34*, 2135–2145.

(24) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(25) Tuckerman, M.; Berne, B. J.; Martyna, G. J. Reversible multiple time scale molecular dynamics. *J. Chem. Phys.* **1992**, *97*, 1990–2001.

(26) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N log (N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(27) Kitao, A.; Hirata, F.; Gō, N. The effects of solvent on the conformation and the collective motions of protein: normal mode analysis and molecular dynamics simulations of melittin in water and in vacuum. *Chem. Phys.* **1991**, *158*, 447–472.

(28) Hayward, S.; Go, N. Collective variable description of native protein dynamics. *Annu. Rev. Phys. Chem.* **1995**, *46*, 223–250.

(29) Balsera, M. A.; Wriggers, W.; Oono, Y.; Schulten, K. Principal component analysis and long time protein dynamics. *J. Phys. Chem.* **1996**, *100*, 2567–2572.

(30) Kitao, A.; Hayward, S.; Go, N. Energy landscape of a native protein: Jumping-among-minima model. *Proteins: Struct., Funct., Bioinf.* **1998**, *33*, 496–517.

(31) Hess, B. Similarities between principal components of protein dynamics and random diffusion. *Phys. Rev. E* **2000**, *62*, 8438.

(32) Hess, B. Convergence of sampling in protein simulations. *Phys. Rev. E* **2002**, *65*, 031910.

(33) Tournier, A. L.; Smith, J. C. Principal components of the protein dynamical transition. *Phys. Rev. Lett.* **2003**, *91*, 208106.

(34) Lange, O. F.; Grubmüller, H. Full correlation analysis of conformational protein dynamics. *Proteins: Struct., Funct., Bioinf.* **2008**, *70*, 1294–1312.

(35) David, C. C.; Jacobs, D. J. *Protein dynamics*; Springer, 2014; pp 193–226.

(36) Altis, A.; Otten, M.; Nguyen, P. H.; Hegger, R.; Stock, G. Construction of the free energy landscape of biomolecules via dihedral angle principal component analysis. *J. Chem. Phys.* **2008**, *128*, 06B620.

(37) Sittel, F.; Jain, A.; Stock, G. Principal component analysis of molecular dynamics: On the use of Cartesian vs. internal coordinates. *J. Chem. Phys.* **2014**, *141*, 07B605_1.

(38) Jain, A.; Stock, G. Hierarchical folding free energy landscape of HP35 revealed by most probable path clustering. *J. Phys. Chem. B* **2014**, *118*, 7750–7760.

(39) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V., et al. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research* **2011**, *12*, 2825–2830.

(40) Parzen, E. On estimation of a probability density function and mode. *Ann. Math. Stat.* **1962**, *33*, 1065–1076.

(41) Davis, R. A.; Lii, K.-S.; Politis, D. N. *Selected Works of Murray Rosenblatt*; Springer, 2011; pp 95–100.

(42) Ward Jr, J. H. Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* **1963**, *58*, 236–244.

(43) Vuillon, L.; Lesieur, C. From local to global changes in proteins: a network view. *Curr. Opin. Struct. Biol.* **2015**, *31*, 1–8.

(44) Dorantes-Gilardi, R.; Bourgeat, L.; Pacini, L.; Vuillon, L.; Lesieur, C. In proteins, the structural responses of a position to mutation rely on the Goldilocks principle: not too many links, not too few. *Phys. Chem. Chem. Phys.* **2018**, *20*, 25399–25410.

(45) Gheeraert, A.; Pacini, L.; Batista, V. S.; Vuillon, L.; Lesieur, C.; Rivalta, I. Exploring allosteric pathways of a v-type enzyme with dynamical perturbation networks. *J. Phys. Chem. B* **2019**, *123*, 3452–3461.

TOC Graphic