

Article

D936Y and Other Mutations in the Fusion Core of the SARS-CoV-2 Spike Protein Heptad Repeat 1: Frequency, Geographical Distribution, and Structural Effect

Romina Oliva ^{1,*} , Abdul Rajjak Shaikh ² , Andrea Petta ³, Anna Vangone ⁴ and Luigi Cavallo ²

¹ Department of Sciences and Technologies, University Parthenope of Naples, Centro Direzionale Isola C4, I-80143 Naples, Italy

² Kaust Catalysis Center, Physical Sciences and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia; abdulrajjak.shaikh@kaust.edu.sa (A.R.S.); luigi.cavallo@kaust.edu.sa (L.C.)

³ Dipartimento di Informatica ed Applicazioni, University of Salerno, Via Papa Paolo Giovanni II, I-84048 Fisciano, Italy; anpetta@unisa.it

⁴ Roche Innovation Center Munich, Pharma Research and Early Development, Large Molecule Research, Nonnenwald 2, 82377 Penzberg, Germany; anna.vangone@roche.com

* Correspondence: romina.oliva@uniparthenope.it

Abstract: The crown of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is constituted by its spike (S) glycoprotein. S protein mediates the SARS-CoV-2 entry into the host cells. The “fusion core” of the heptad repeat 1 (HR1) on S plays a crucial role in the virus infectivity, as it is part of a key membrane fusion architecture. While SARS-CoV-2 was becoming a global threat, scientists have been accumulating data on the virus at an impressive pace, both in terms of genomic sequences and of three-dimensional structures. On 15 February 2021, from the SARS-CoV-2 genomic sequences in the GISAID resource, we collected 415,673 complete S protein sequences and identified all the mutations occurring in the HR1 fusion core. This is a 21-residue segment, which, in the post-fusion conformation of the protein, gives many strong interactions with the heptad repeat 2, bringing viral and cellular membranes in proximity for fusion. We investigated the frequency and structural effect of novel mutations accumulated over time in such a crucial region for the virus infectivity. Three mutations were quite frequent, occurring in over 0.1% of the total sequences. These were S929T, D936Y, and S949F, all in the N-terminal half of the HR1 fusion core segment and particularly spread in Europe and USA. The most frequent of them, D936Y, was present in 17% of sequences from Finland and 12% of sequences from Sweden. In the post-fusion conformation of the unmutated S protein, D936 is involved in an inter-monomer salt bridge with R1185. We investigated the effect of the D936Y mutation on the pre-fusion and post-fusion state of the protein by using molecular dynamics, showing how it especially affects the latter one.

Keywords: COVID-19; spike protein; mutations; molecular dynamics; infectivity



Citation: Oliva, R.; Shaikh, A.R.; Petta, A.; Vangone, A.; Cavallo, L. D936Y and Other Mutations in the Fusion Core of the SARS-CoV-2 Spike Protein Heptad Repeat 1: Frequency, Geographical Distribution, and Structural Effect. *Molecules* **2021**, *26*, 2622. <https://doi.org/10.3390/molecules26092622>

Academic Editors: Angelo Facchiano and Anna Marabotti

Received: 17 March 2021

Accepted: 12 April 2021

Published: 30 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Coronavirus Disease 2019 (COVID-19) is caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which is also referred to as human coronavirus 2019 (hCoV-2). SARS-CoV-2 is a novel virus belonging to the β genus coronaviruses, which also include two highly pathogenic human viruses identified in the last two decades, known as the severe acute respiratory syndrome coronavirus (SARS-CoV) and the Middle East respiratory syndrome coronavirus (MERS-CoV) [1–3].

Coronaviruses are named after the protruding spike (S) glycoproteins on their envelope, giving a crown shape to the virions [4]. Of the four structural proteins of coronaviruses: S, envelope (E), membrane (M), and nucleocapsid (N), the S protein is the one playing a key role in mediating the viral entry into the host cells [5–7], making it

one of the main targets for the development of therapeutic drugs and vaccines [8–14]. Comprised of two functional subunits, S1 and S2, it first binds to a host receptor through the receptor-binding domain (RBD) in the S1 subunit and then fuses the viral and host membranes through the S2 subunit [7,15]. In the pre-fusion conformation, the SARS-CoV-2 S protein forms homotrimers protruding from the viral surface, where its RBD binds to the angiotensin-converting enzyme 2 ACE2 receptor on the host cell surface [1] (like the SARS-CoV homolog [16], and differently from MERS-CoV S, which recognizes a different receptor, the dipeptidyl peptidase 4 [17]. Receptor binding and proteolytic processing by cellular proteases then cause S1 to dissociate and S2 to undergo large-scale conformational changes toward a stable structure, bringing viral and cellular membranes into close proximity for fusion and infection [7,15,18].

While the outbreak of COVID-19 was rapidly spreading all over the world, affecting millions of people and becoming a global threat, laboratories worldwide promptly started to sequence a large number of SARS-CoV-2 genomes. All the available genomic data is accessible through the Global Initiative on Sharing All Influenza Data (GISAID) website, which is an invaluable open access resource [19,20]. Simultaneously, crucial structural knowledge has been achieved on SARS-CoV-2, especially regarding the S protein. On 15 March 2021, \approx 280 experimental 3D structures of the SARS-CoV-19 S protein were available from the Protein Data Bank (PDB) [21]. These include the structure of the SARS-CoV-2 S protein in the pre-fusion conformation, also bound to the ACE2 receptor [22–28], and of the post-fusion core of its S2 subunit in the post-fusion conformation [29].

After the proteolytic processing, in the post-fusion conformation, the S protein HR1 and HR2 motifs interact with each other to form a six-helix bundle (6-HB), which promotes initiation of the viral and cellular membranes fusion [7,15,18]. The 6-HB formation is a conserved and critical mechanism for viral fusion and entry, shared by all coronaviruses. The HR1 “fusion core” is named after its role in giving many interactions with HR2 in the post-fusion conformation, thus, playing a key role in the virus infectivity, and being a target for inhibitors of the SARS-CoV-2 fusion capacity [29–31]. On these bases, we decided to investigate the frequency and possible structural effect of the mutations accumulated over time in this crucial, functional motif.

On 15 February 2021, genomic sequences of SARS-CoV-2 available from GISAID had overcome 550,000. A search in such a resource showed a variable occurrence of mutations at different positions of the HR1 fusion core, with mutant D936Y being the most frequent, with 1296 occurrences, particularly in some European countries, especially Finland and Sweden. Among the identified HR1 mutants, it is also the mutant expected to have the most significant structural consequences, as D936 is involved in an inter-monomer salt bridge in the post-fusion assembly. Therefore, we performed a comparative study of the wild-type S protein and the D936Y mutant, both in their pre-fusion and post-fusion conformations by molecular dynamics (MD). Results of the MD simulations helped us illuminate the effect of the mutation on the protein structure and dynamics.

We also investigated the structural basis, both in the pre-fusion and post-fusion conformation as well as sequencing dates and geographical distribution of the other two most frequent HR1 mutations, S939F and S929T, with 1108 and 467 occurrences, respectively. For the pre-fusion conformation, a structure with one RBD in the up and two RBDs in the down position was considered, which has been recently proposed, based on molecular dynamics analyses [32], to be significantly more stable than the structure with two RBD in the up position, and, therefore, likely to be the protein state observed prior to interacting with the cell surface [33].

Considering the impressive pace at which new SARS-CoV-2 sequences are obtained and collected, we have also set up a web application providing a periodic update of mutations in the S protein HR1 fusion core (at <https://www.molnac.unisa.it/BioTools/cov2smt/index.php>) (accessed on 18 April 2021).

2. Results

2.1. Identification of the HR1 “Fusion Core” Mutations

The HR1 of coronaviruses S proteins undergoes one of the most notable rearrangements within the protein between the pre-fusion and post-fusion conformations. In the post-fusion conformation, it experiences a refolding of the pre-fusion multiple helices and intervening regions into a single continuous helix (Figure 1). As already mentioned, three of these long helices then form a 6HB with three HR2 helical motifs [18,29,30]. The HR1 and its “fusion core” particularly play a crucial role in the virus infectivity.

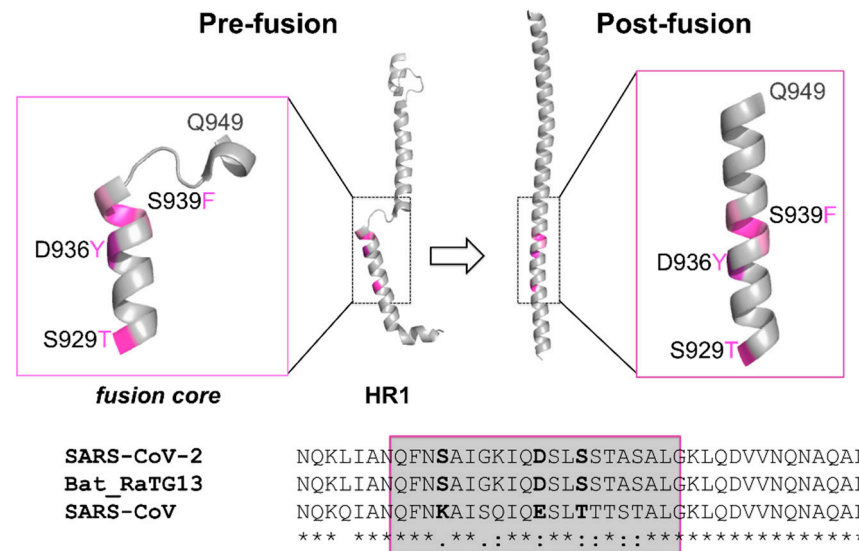


Figure 1. Structural and sequence location of the reported mutations. Top: Cartoon representation of SARS-CoV-2 S protein HR1 and its fusion core (insets) in the pre-fusion and post-fusion conformations (PDB IDs: 6VSB and 6LXT). Discussed mutations are colored in purple and labelled. Q949, at the end of the fusion core, is also labeled. Bottom: Sequence alignment of the HR1 fusion core (framed) and 10 residues up-stream and down-stream in the S protein of SARS-CoV-2, bat coronavirus RaTG13 (protein_ID: QHR63300.2), and SARS-CoV (protein_ID: AAP13441.1).

On 15 February 2021, we downloaded all the SARS-CoV-2 genomic sequences from the GISAID resource, extracted from them 415,673 complete S protein sequences, and identified all the point mutations occurring in the S929-Q949 region (see Methods). The identified mutations with the relative number of occurrences are reported in Table S1. Only the most frequent mutations are reported in Table 1.

Table 1. Occurrences of most frequent mutations on the HR1 “fusion core” on 15 February 2021.

# S Protein Sequences	S929T	D936Y	S939F
415,673	467	1296	1108

Most of the positions, such as 931, 933 to 935, 937, 944-945, and 948-949, were virtually unaffected by mutational events, with a maximum mutation rate of 0.005%. Positions 930, 937, 941, and 946-947 were also little affected, with a mutation rate below 0.02%. Positions 932, 938, 940, 942-943 had a mutation rate between 0.025% and 0.055%. Positions featuring the higher number of mutations were 929, 936, and 939, which are all located in the N-terminal half of the HR1 fusion core and featuring a mutation rate above 0.1%. Starting from position 929, S929 was found to mutate to threonine in 467 sequences and to asparagine/arginine/glycine in 60/5/1 sequences.

As for position 936, D936 was mutated to tyrosine in 1296 sequences and to asparagine/histidine/valine/glycine/glutamate/glutamine/alanine/serine in 148/125/44/24/17/3/1/1 sequences.

Finally, S939 was mutated to phenylalanine in 1108 sequences and to tyrosine/leucine/alanine in 5/3/1 sequences.

2.2. Geographical Distribution of the HR1 “Fusion Core” of Most Frequent Mutations

The geographical distribution, per country, of the investigated mutations is reported in Figure 2. The first occurrence of the S929T mutation was deposited in GISAID on 18 April 2020, which is sequenced in Canada. On 15 February, however, the large majority of its occurrences was reported from England (440 over 467, corresponding to 94%). The remaining 27 occurrences were also mostly sequenced in Europe, with only 5 overall occurrences from USA and Canada, 2 from Australia, and 1 from South Africa.

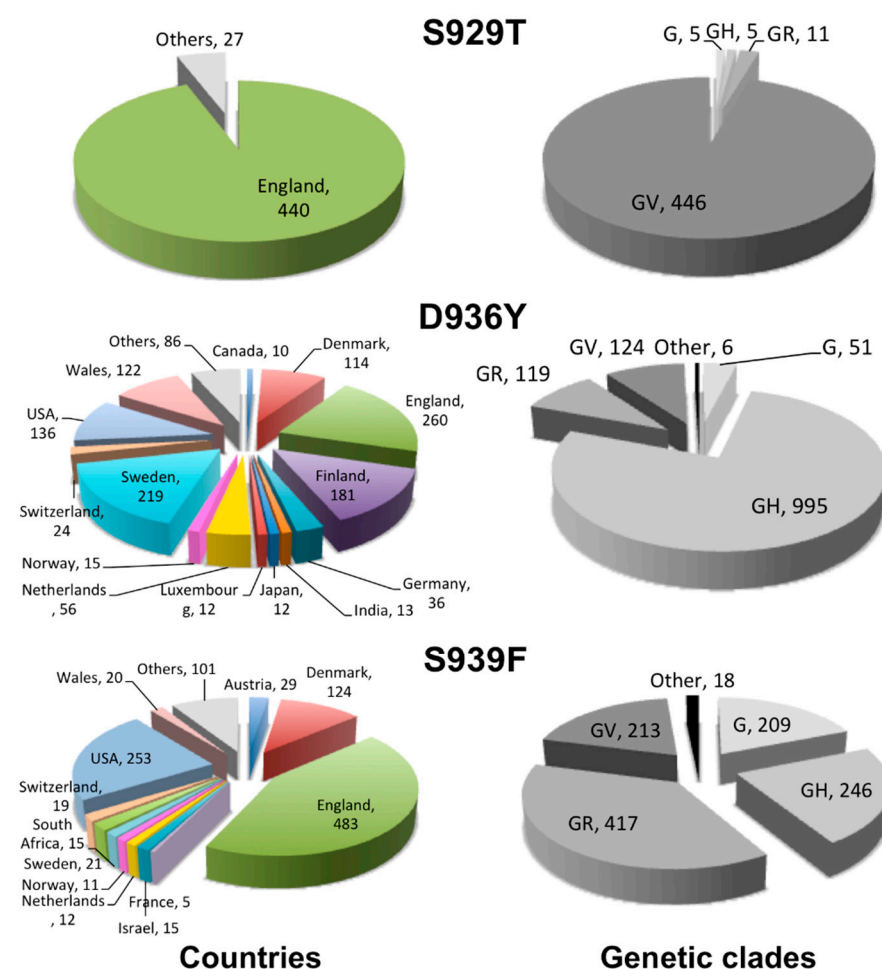


Figure 2. Countries and genetic clades. Pie chart visualization of the geographical distribution (left panel) and phylogenetic classification (right panel) of sequences presenting the S929T, D936Y, and S939F mutations.

The first occurrence of the D936Y mutation was, instead, deposited in GISAID on 8 March 2020, which was sequenced in Sweden. On 15 February 2021, occurrences have been reported from 48 countries. However, Sweden confirms itself as the country with the higher occurrences of such a mutation (in 219 sequences, representing 17% of the total). The other four European countries contributed, together with Sweden, 60% of all the occurrences. These countries are England, Finland, Wales, and Denmark, and reported 260 (20%), 181 (14%), 122 (9.4%), and 114 (8.8%) occurrences, respectively. USA also contributed

a significant number of occurrences (136 occurrences or 10%). The remaining 30% of occurrences were mainly sequenced in European countries, the Netherlands (56), Germany (36), Switzerland (24), Norway (15), Luxembourg (12), Scotland (5), Austria (5), and others, as well as in India (13), Japan (12), Canada (10), Mexico (7), Singapore (5), etc. (for a complete list, see the web site: <https://www.molnac.unisa.it/BioTools/cov2smt/index.php>) (accessed on 28 April 2021).

Notably, the total number of occurrences of the D936Y mutation amounted to 17% of all the 1089 sequences available from Finland and to 12% of all the 1768 sequences available from Sweden.

The first occurrence of the S939F mutation was deposited in GISAID on 25 February 2020 from the United Arab Emirates. On 15 February 2021, it was spread in 44 countries, especially western ones. Three countries represented together 66% of all the occurrences. These countries are England, USA, and Denmark, having reported 483 (37%), 253 (20%), and 124 (9.6%) occurrences. Over 10 occurrences of the mutation were also reported from other European countries: Austria (29), Sweden (21), Wales (20), Switzerland (19), the Netherlands (12), and Norway (11), but also from Israel (15) and South Africa (15). Two more occurrences of the mutation have been reported from the United Arab Emirates between May and June 2020.

2.3. Clade Association of the HR1 “Fusion Core” of Most Frequent Mutations

The distribution of the mutations in high-level phylogenetic groupings, or genetic clades, is plotted in Figure 2. As a reminder, the G/GH/GR/GV clades are among the latest out of eight genetic clades reported in GISAID (S, L, V, G, GH, GR, GV, GRY) [34]. The G clade carries the D614G mutation, now globally dominant, accompanied by other mutations upstream the S protein gene (C241T, C3037T). In addition, the GH clade presents the NS3-Q57H mutation, the GR clade presents the N-G204R mutations, and GV clade presents the S-A222V mutation.

The three reported mutations HR1 are clearly associated with the late G/GH/GR/GV clades. In particular, S929T is mainly associated with the GV clade and D936Y is mainly associated with the GH clade, while S939F is roughly equally associated with the GR, GH, GV, and G clades.

2.4. Sequence Conservation among Similar Viruses

All the amino acids in the three positions more prone to mutation in the SARS-CoV-2 S protein HR1 fusion core are conserved in the bat coronavirus RaTG13 S protein (sharing an overall sequence identity of 97% with SARS-CoV-2 S protein), while all of them are mutated in the SARS-CoV S protein (overall, 76% sequence identical to the SARS-CoV-2 homolog) (see Figure 1). In particular, S929 is a lysine in SARS-CoV, while D936 is substituted by a glutamate and S939 by a threonine. It has been proposed that the SARS-CoV-2 HR1 mutations as compared to SARS-CoV may be associated with enhanced interactions with HR2, further stabilizing the 6-HB structure and maybe leading to increased infectivity of the virus [29]. In this context, it is noteworthy that the point mutations we are discussing did not restore the corresponding SARS-CoV amino acid.

2.5. Effect of the Mutations on the Protein Pre-Fusion Conformation

In the pre-fusion conformation, the most mutated positions are located on the second of four non-coaxial helical segments composing the HR1 (Figure 1). They are all exposed to the solvent (Table 2), and can be modelled as larger residues without causing any structural strain (see Figure 3). These mutations are not expected to cause relevant changes in the pre-fusion structure. However, they could have a destabilizing effect as a consequence of posing large aromatic residues, at positions 936 and 939, in direct contact with the solvent instead of a charged aspartate or polar serine residue.

Table 2. Solvent accessibility of mutated residues in the pre-fusion and post-fusion conformations.

Amino Acid	Pre-Fusion	Post-Fusion
T929	exposed	partly buried (18.6%) ^a
Y936	exposed	partly buried (19.0%)
F939	exposed	exposed

^a Percentage of buried surface upon complex formation.

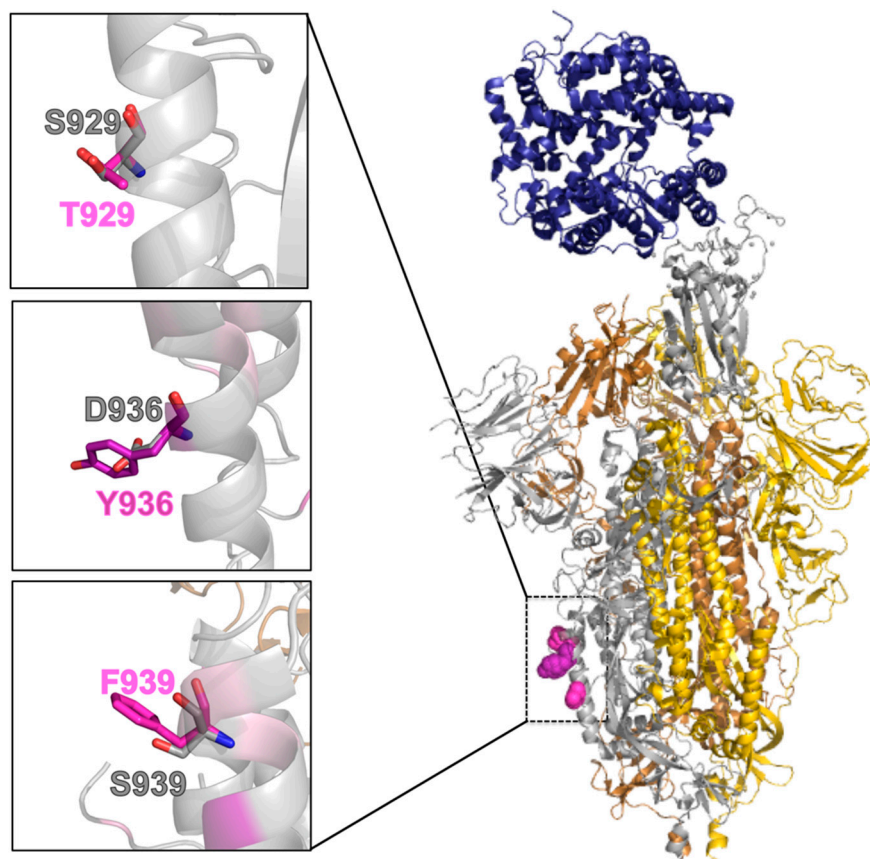


Figure 3. Mutants in the pre-fusion conformation. Right: Cartoon representation of the SARS-CoV-2 S protein in its pre-fusion trimeric conformation (the three monomers are colored in silver, gold, and copper, PDB ID: 6VSB), with the structure of the RBD bound to the ACE2 receptor (in blue, PDB ID: 6M0J) superimposed on its chain A. The most frequent mutations in the HR1 fusion core in GISAID on 15 February are colored purple and shown as a “dots” representation for chain A. Left: Focus on the structural context of each wild-type residue (silver sticks) and corresponding mutant (purple sticks).

2.6. Effect of the Mutations on the Protein Post-Fusion Conformation

When looking at the post-fusion conformation of the SARS-CoV-2 spike protein S2 subunit, these mutations appear more revealing. Two of the wild-type residues, S929 and D936, are engaged in side-chain to side-chain H-bonds with the HR2 segment of an adjacent monomer. In particular, S929 and D936 (HR1 on Chain A) are H-bonded to S1196 and R1185, respectively (HR2 on Chain C, Figure 4). Mutation of S929 to threonine does not cause the loss of the inter-monomer H-bond (Figure 4), while a mutation of D936 to tyrosine, does. The H-bond between D936 and R1185 is actually a salt bridge (estimated to contribute an additional 3–5 kcal/mol to the free energy of protein stability as compared to a neutral H-bond [35]).

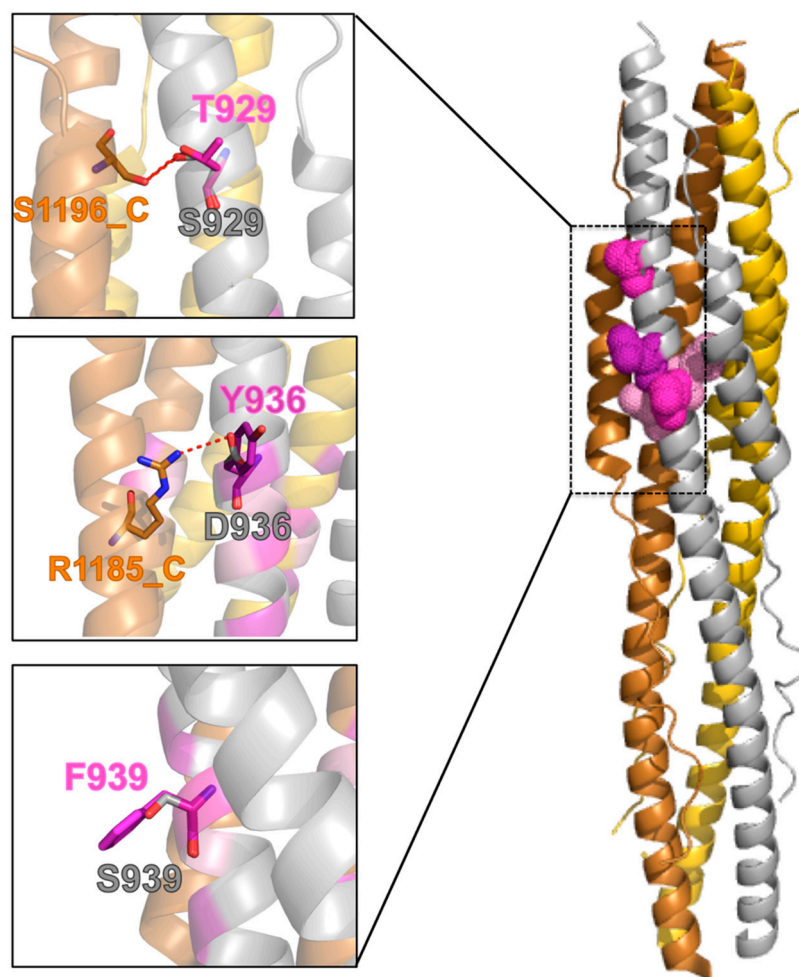


Figure 4. Mutants in the post-fusion conformation. Right: Cartoon representation of the SARS-CoV-2 S protein in its post-fusion trimeric conformation (the three monomers are colored in silver, gold, and copper, PDB ID: 6LXT). The color code is the same in Figure 3. Mutations in the HR1 fusion core are shown in a “dots” representation for chain A. Left: Focus on the structural context of each wild-type residue (silver sticks) and corresponding mutant (purple sticks). H-bonds are shown as red, dashed lines.

Of the remaining most frequent mutations, S939F is completely exposed to the solvent and, therefore, like in the pre-fusion conformation, expected to act unfavorably on the protein solvation energy.

2.7. Molecular Dynamics Analysis

When comparing the effect of the mutations on the pre-fusion and post-fusion structures, it emerges that the D936Y mutation is the one expected to have the greatest structural impact. Since it is also the most frequent mutation occurring on the fusion core of S HR1, we decided to further analyze the effect of such a mutation on the structure and dynamics of the SARS-CoV-2 S protein. To this aim, three 0.5- μ s long MD simulation replicates were run on the mutant and the wild-type protein, both in their pre-fusion and post-fusion conformations, for a total of 6 μ s. We recall in the following the main findings of the MD analysis, while details are reported in the Supplementary Information text and in Figures S1–S12 and Tables S2 and S3.

Both the wild-type and mutant conformations were stable during the whole dynamics, in the pre-fusion and post-fusion conformations, with maximal root mean square deviation (rmsd) values on the C α atoms not exceeding 3.5 Å from the initial structure (Figures S1 and S7). The difference in the rmsd values between the wild-type protein and the

D936Y mutant (Figure 5a) is negligible for the pre-fusion conformation, $0.05 (\pm 0.1) \text{ \AA}$. In the post-fusion conformation, the average rmsd is instead higher, by $0.38 \text{ \AA} (\pm 0.2)$, for the mutant, which seems to acquire some flexibility. The total number of inter-monomer H-bonds from the wild-type to the mutant decreased more in the post-fusion conformation, $-1.8 (\pm 1.1)$, than in the pre-fusion one, $-0.9 (\pm 1.3)$. As we expected, in order for these lost H-bonds to be the inter-monomer D936-R1185 salt bridges we discussed before, we monitored the H-bond distances between D/Y936 and R1185 over time (Figure 5b,c). The minimum distance between the nitrogen atoms of the arginine guanidinium group and the oxygens of the aspartate carboxylate or the hydroxyl oxygen of the mutated tyrosine is reported for each trimer interface. In case of the wild-type, the minimum H-bond distance is $3.32 (\pm 0.7) \text{ \AA}$ and $3.62 (\pm 1.0) \text{ \AA}$ for two interfaces, with distances being within 3.5 \AA in 70% and 57% of frames, respectively. Therefore, these two H-bonds are largely maintained over time. For the third interface, the average distance is instead $6.48 (1.1) \text{ \AA}$, with only 1% of the frames within 3.5 \AA . This is consistent with the reference X-ray structure, where D936 and R1185 on the adjacent monomer are at an H-bond distance for two interfaces, and are, instead, 4.71 \AA apart on the third interface. In case of the mutants, the average distances are all around 4 \AA (3.96 ± 0.7 , 4.29 ± 0.9 and $4.23 \pm 0.9 \text{ \AA}$ for each interface), with the total frames featuring a distance within 3.5 \AA amounting to only 22%. This correlates with the loss of ≈ 2 H-bonds in the mutant conformation. However, it is worth it to remind here that, due to its strong electrostatic nature, a stabilizing interaction between D936 and R1185 is maintained above the classical threshold for an H-bond distance [36].

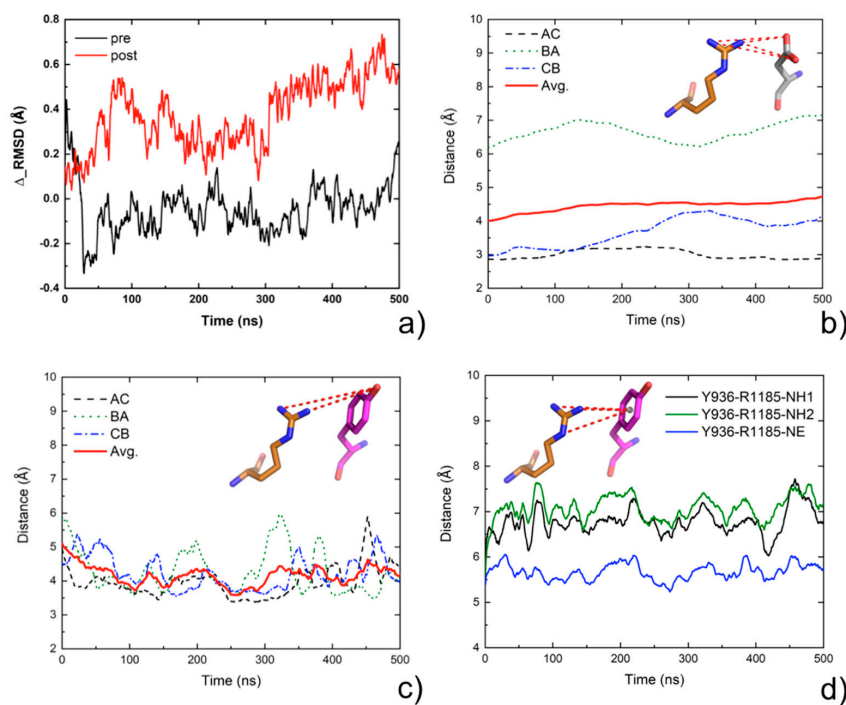


Figure 5. Comparative MD analysis of the Wuhan reference S protein and the D936Y mutant. (a) rmsd difference (rmsd) between the wild-type and the mutant in the pre-fusion (black) and post-fusion (red) conformation, averaged over the three independent 500-ns simulations per system. (b) Wild-type: minimum distance over time between the carboxylate oxygens of D936 and the guanidinium nitrogens of R1185 on the adjacent monomer, averaged over the three independent simulations per system. Values per single trimer interfaces are plotted as dashed lines while the average values over the three interfaces are plotted as a continuous red line. (c) Mutant: minimum distance over time between the hydroxyl oxygen of Y936 and the guanidinium nitrogen of R1185 on the adjacent monomer, averaged over the three independent simulations per system. Values per single trimer interfaces are plotted as dashed lines while the average values over the three interfaces are plotted as a continuous red line. (d) Mutant: distances over time between the center of mass of the Y936 aromatic ring and the guanidinium nitrogen of R1185 on the adjacent monomer, averaged over the three independent simulations and the three interfaces.

Since an arginine can involve a tyrosine in a cation- π interaction, we also monitored the minimum distance between the nitrogen atoms of the R1185 guanidinium group and the center of mass of the Y936 aromatic ring (Figure 5d). Average values are in the 6–7 Å range and never drop below 4.3 Å, which is considered a reasonable cutoff distance for establishing a cation- π interaction [37]. Therefore, the above analysis ruled out the possibility of having a cation- π interaction between these two residues.

Finally, we followed the buried surface area over the simulation time within the MDcons approach finding the post-fusion assembly to be, overall, more compact (i.e., featuring a moderately higher buried surface area upon complex formation) for the wild-type system, as compared to the D936Y mutant (see Figure S13).

3. Discussion

We monitored the mutations accumulated over time on the SARS-CoV-2 S protein HR1 fusion core, and a key structural and functional motif for the virus infectivity, using GISAID as the resource of genomic sequences. The SARS-CoV-2 HR1 fusion core differs in several positions from that of SARS-CoV and its peculiarity has been associated with the higher infectivity of the virus [29]. On 15 February 2021, D936Y was the most frequent mutation on the HR1 fusion core, followed by S939F and S929T. Notably, most of the HR1 fusion core positions are virtually unaffected by mutational events, while all three most-frequent mutations are located on the second of four non-coaxial helical segments composing the HR1. In the pre-fusion conformation, two of these mutations result in large aromatic residues of a tyrosine and a phenylalanine. Such mutations, mainly localized in Europe and USA, are quite late ones, emerging starting from the end of February 2020, and are associated with the late G/GH/GR/GV clades, implying that they co-exist with the globally dominant D614G mutation.

D936Y was the most frequent among the HR1 fusion core mutations on 25 February 2021. While the geographical distribution of S929T, mostly from England, and of S939F, mostly from England, USA, and Denmark, may reflect the higher contribution of these countries to the genomic sequencing of SARS-CoV-2 (the three countries together covered roughly two-thirds of the sequences in GISAID on 15 February), D936Y was widespread. Besides the above countries, in Scandinavia and especially in Finland and Sweden, it represents 17% and 12%, respectively, of all the sequences available from these countries.

We investigated the structural basis of such mutations, finding out that the D936Y mutation is the one expected to have the greatest structural impact. Therefore, we analyzed the effect of such a mutation by molecular dynamics, showing that it causes the loss of a strong inter-monomer salt bridge in the post-fusion conformation of the S protein and introduces some flexibility in it, resulting in an overall slightly reduced compactness of the assembly.

Experimental testing of the D936Y mutation, within a study comprising over 100 S protein variants or glycosylation site modifications [38], has shown a significant decrease of infectivity as compared to the Wuhan reference strain [1] when it was the only variant. It demonstrated instead increased infectivity, as compared to the reference strain, when associated with the D614G variant, which was comparable to that of the strain presenting only the D614G mutation. It is worth noticing that, for other frequent variants included in the same study, such as L5F and D839Y, infectivity was virtually unchanged. The structural effect of the D936Y mutation, that we report here, may call for further functional and clinical studies to clarify its possible consequences on the SARS-CoV-2 virulence.

An up-to-date count of the above mutations is provided at: <https://www.molnac.unisa.it/BioTools/cov2smt/index.php> (accessed on 18 April 2021).

4. Methods

4.1. Identification of Mutations

We downloaded the 550,092 genomic sequences available from GISAID on 15 February 2021. From these sequences, we extracted the nucleotide sequences of the spike protein

and translated them to protein sequences with in-house scripts. Nucleotide sequences featuring an internal stop codon or having at least one undefined (“N”) nucleotide were discarded. Sequences annotated as pangolin, bat, or canine were also discarded. The remaining 415,673 protein sequences were further analysed. As a reference system, we used the genomic sequence with GISAID ID: EPI_ISL_402124, isolated and sequenced in Wuhan (Hubei, China) on 30 December 2019 [1]. Then, upon alignment to the reference sequence, we identified point mutations in all the sets of at least two sequences.

The web application was built using standard HTML, php, and python scripts.

4.2. Mutants Modelling and Analysis

Mutants 3D models were built using the `mutate_model` module of the Modeller 9v11 program [39]. This is an automated method for modelling point mutations in protein structures, which includes an optimisation procedure of the mutated residue in its environment, beginning with a conjugate gradients’ minimisation, continuing with molecular dynamics with simulated annealing, and finishing again by conjugate gradients. The used force field is CHARMM-22. For details, see Reference [40]. Models for mutants in the pre-fusion conformation were built starting from the EM structure of the pre-fusion trimeric conformation (PDB ID: 6VSB, resolution 3.46 Å, [22]). Models for mutants in the post-fusion conformation were built starting from the X-ray structure of the S2 subunit fusion core, featuring residues 912-988 and 1164-1202 (PDB ID: 6LXT, resolution 2.90 Å, [29]). Molecular models were analysed and visually inspected with Pymol [41]. The COCOMAPS web server [42] was used to analyse the inter-chain contacts and H-bonds as well as the residues accessibility to the solvent.

4.3. Molecular Dynamics Simulations

Molecular dynamics simulations were carried out for the wild-type S protein and for the D936Y mutant in the pre-fusion and post-fusion conformations, starting from the experimental structures used for modeling the mutants (see above). For the pre-fusion simulations, we used the trimer of the S2 subunit (PDB ID: 6VSB). From S711 to C1146, respectively, 200 residues upstream and \approx 160 residues downstream of HR1. Missing residues between K811 and R815 and between L828 and Q853 were modeled with the GalaxyFill program [43]. The crystal structure of the post fusion core of the protein S2 subunit (PDB ID: 6LXT), featuring residues 912-988, 1164-1202 [29] was used for the post-fusion simulations. For the D936 mutant, models obtained as detailed in the previous section were used.

All the MD simulations were carried out with Gromacs 2018 [44], using the Amber14SB force field [45]. Each protein was inserted into a rectangular box of TIP3P water molecules, setting a minimum distance of 12.0 Å from it to the box sides and neutralizing the solution with Zn^{2+} and Cl^- ions. A minimization was first carried out, followed by isothermal ensemble (NVT) dynamics using a velocity-rescale thermostat [46] for computing positions and velocities of atoms. Then, 2 ns of isothermal-isobaric ensemble (NPT) dynamics was carried out to equilibrate the structure. Periodic boundary conditions were applied in all directions. The production simulations were carried out using an NPT ensemble for 500 ns. The temperature was maintained constant at 300 K using a velocity-rescale thermostat [46] ($\tau_T = 0.1$ ps) and a pressure of 1 bar was maintained using a Parrinello-Rahman barostat [47] ($\tau_P = 2.0$ ps). Electrostatic interactions beyond 1.2 nm were evaluated by the Particle-Mesh-Ewald (PME) method [48]. Bond lengths were constrained with the LINear Constraint Solver algorithm [49]. Trajectories were analyzed using Gromacs 2018 analysis tools.

For the MDcons analyses [50], using a contact-based approach [51,52] for the dynamical characterization of the interface in protein assemblies, 500 snapshots were generated for each system, by writing the coordinates every 1 ns.

Supplementary Materials: Table S1. Number of occurrences of all mutations on the HR1 “fusion core” on 15 February 2021. Table S2. MD analysis data of the wt and D936Y mutant pre-fusion state. Table S3. MD analysis data of the wt and D936Y mutant post-fusion state. Figures S1–S2.

Pre-fusion state: C α RMSD values versus time. Figure S3. Pre-fusion state: RMSF values per residue. Figure S4. Pre-fusion state: Average number of hydrogen bonds versus time. Figure S5. Pre-fusion state: Potential energy versus time. Figure S6. Pre-fusion state: Electrostatic (ELE) and Lennard Jones (LJ) energies versus time. Figures S7–S8. Post-fusion state: C α RMSD values versus time. Figure S9. Post-fusion state. RMSF per residue. Figure S10. Post-fusion state: Average number of hydrogen bonds versus time. Figure S11. Post-fusion state: Potential energy versus time. Figure S12. Post-fusion state: Electrostatic (ELE) and Lennard Jones (LJ) energies versus time. Figure S13. Buried surface area along the MD simulations for the wt and D936Y mutant post-fusion state.

Author Contributions: R.O. conceived the study, participated in its design, carried out the analyses, and drafted the manuscript. A.R.S. performed the MD simulations. A.P. implemented the web application. A.V. participated in the bioinformatics analyses. L.C. participated in the study's design, in the analyses, and in the implementation of the web application. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available in supplementary material.

Acknowledgments: We gratefully acknowledge all the authors from the originating laboratories responsible for obtaining the specimens and the submitting laboratories where genetic sequence data were generated and shared via the GISAID Initiative, on which this research is based. R.O. thanks MIUR-FFABR (Fondo per il Finanziamento Attività Base di Ricerca) for funding. L.C. acknowledges King Abdullah University of Science and Technology (KAUST) for support and the KAUST Supercomputing Laboratory for providing computational resources.

Conflicts of Interest: The authors declare no conflict of interest.

Sample Availability: Not applicable.

References

1. Zhou, P.; Yang, X.L.; Wang, X.G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H.R.; Zhu, Y.; Li, B.; Huang, C.L.; et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **2020**, *579*, 270–273. [[CrossRef](#)] [[PubMed](#)]
2. Wu, F.; Zhao, S.; Yu, B.; Chen, Y.M.; Wang, W.; Song, Z.G.; Hu, Y.; Tao, Z.W.; Tian, J.H.; Pei, Y.Y.; et al. A new coronavirus associated with human respiratory disease in China. *Nature* **2020**, *579*, 265–269. [[CrossRef](#)] [[PubMed](#)]
3. Huang, C.; Wang, Y.; Li, X.; Ren, L.; Zhao, J.; Hu, Y.; Zhang, L.; Fan, G.; Xu, J.; Gu, X.; et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **2020**, *395*, 497–506. [[CrossRef](#)]
4. Li, F. Receptor recognition mechanisms of coronaviruses: A decade of structural studies. *J. Virol.* **2015**, *89*, 1954–1964. [[CrossRef](#)]
5. Belouzard, S.; Chu, V.C.; Whittaker, G.R. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 5871–5876. [[CrossRef](#)]
6. Millet, J.K.; Whittaker, G.R. Host cell entry of Middle East respiratory syndrome coronavirus after two-step, furin-mediated activation of the spike protein. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 15214–15219. [[CrossRef](#)] [[PubMed](#)]
7. Shang, J.; Wan, Y.; Luo, C.; Ye, G.; Geng, Q.; Auerbach, A.; Li, F. Cell entry mechanisms of SARS-CoV-2. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 11727–11734. [[CrossRef](#)]
8. Salvatori, G.; Luberto, L.; Maffei, M.; Aurisicchio, L.; Roscilli, G.; Palombo, F.; Marra, E. SARS-CoV-2 SPIKE PROTEIN: An optimal immunological target for vaccines. *J. Transl. Med.* **2020**, *18*, 222. [[CrossRef](#)]
9. Du, L.; He, Y.; Zhou, Y.; Liu, S.; Zheng, B.J.; Jiang, S. The spike protein of SARS-CoV-a target for vaccine and therapeutic development. *Nat. Rev. Microbiol.* **2009**, *7*, 226–236. [[CrossRef](#)] [[PubMed](#)]
10. Yi, C.; Sun, X.; Ye, J.; Ding, L.; Liu, M.; Yang, Z.; Lu, X.; Zhang, Y.; Ma, L.; Gu, W.; et al. Key residues of the receptor binding motif in the spike protein of SARS-CoV-2 that interact with ACE2 and neutralizing antibodies. *Cell Mol. Immunol.* **2020**, *17*, 621–630. [[CrossRef](#)] [[PubMed](#)]
11. Tian, X.; Li, C.; Huang, A.; Xia, S.; Lu, S.; Shi, Z.; Lu, L.; Jiang, S.; Yang, Z.; Wu, Y.; et al. Potent binding of 2019 novel coronavirus spike protein by a SARS coronavirus-specific human monoclonal antibody. *Emerg. Microbes Infect.* **2020**, *9*, 382–385. [[CrossRef](#)]
12. Wang, C.; Li, W.; Drabek, D.; Okba, N.M.A.; van Haperen, R.; Osterhaus, A.; van Kuppeveld, F.J.M.; Haagmans, B.L.; Grosveld, F.; Bosch, B.J. A human monoclonal antibody blocking SARS-CoV-2 infection. *Nat. Commun.* **2020**, *11*, 2251. [[CrossRef](#)]
13. McKee, D.L.; Sternberg, A.; Stange, U.; Laufer, S.; Naujokat, C. Candidate drugs against SARS-CoV-2 and COVID-19. *Pharmacol. Res.* **2020**, *157*, 104859. [[CrossRef](#)] [[PubMed](#)]
14. Amanat, F.; Krammer, F. SARS-CoV-2 Vaccines: Status Report. *Immunity* **2020**, *52*, 583–589. [[CrossRef](#)]

15. Heald-Sargent, T.; Gallagher, T. Ready, set, fuse! The coronavirus spike protein and acquisition of fusion competence. *Viruses* **2012**, *4*, 557–580. [[CrossRef](#)] [[PubMed](#)]
16. Li, F.; Li, W.; Farzan, M.; Harrison, S.C. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* **2005**, *309*, 1864–1868. [[CrossRef](#)] [[PubMed](#)]
17. Wang, N.; Shi, X.; Jiang, L.; Zhang, S.; Wang, D.; Tong, P.; Guo, D.; Fu, L.; Cui, Y.; Liu, X.; et al. Structure of MERS-CoV spike receptor-binding domain complexed with human receptor DPP4. *Cell Res.* **2013**, *23*, 986–993. [[CrossRef](#)] [[PubMed](#)]
18. Walls, A.C.; Tortorici, M.A.; Snijder, J.; Xiong, X.; Bosch, B.J.; Rey, F.A.; Velesler, D. Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 11157–11162. [[CrossRef](#)]
19. Elbe, S.; Buckland-Merrett, G. Data, disease and diplomacy: GISAID’s innovative contribution to global health. *Glob. Chall* **2017**, *1*, 33–46. [[CrossRef](#)]
20. Shu, Y.; McCauley, J. GISAID: Global initiative on sharing all influenza data—from vision to reality. *Euro Surveill* **2017**, *22*, 30494. [[CrossRef](#)]
21. Berman, H.M.; Battistuz, T.; Bhat, T.N.; Bluhm, W.F.; Bourne, P.E.; Burkhardt, K.; Feng, Z.; Gilliland, G.L.; Iype, L.; Jain, S.; et al. The Protein Data Bank. *Acta Crystallogr. D Biol. Crystallogr.* **2002**, *58*, 899–907. [[CrossRef](#)] [[PubMed](#)]
22. Wrapp, D.; Wang, N.; Corbett, K.S.; Goldsmith, J.A.; Hsieh, C.L.; Abiona, O.; Graham, B.S.; McLellan, J.S. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **2020**, *367*, 1260–1263. [[CrossRef](#)] [[PubMed](#)]
23. Walls, A.C.; Park, Y.J.; Tortorici, M.A.; Wall, A.; McGuire, A.T.; Velesler, D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **2020**, *181*, 281–292.e286. [[CrossRef](#)] [[PubMed](#)]
24. 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. [[CrossRef](#)] [[PubMed](#)]
25. 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. [[CrossRef](#)]
26. 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.e899. [[CrossRef](#)]
27. 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. [[CrossRef](#)]
28. Yuan, M.; Wu, N.C.; Zhu, X.; Lee, C.D.; So, R.T.Y.; Lv, H.; Mok, C.K.P.; Wilson, I.A. A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV. *Science* **2020**, *368*, 630–633. [[CrossRef](#)]
29. Xia, S.; Liu, M.; Wang, C.; Xu, W.; Lan, Q.; Feng, S.; Qi, F.; Bao, L.; Du, L.; Liu, S.; et al. Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion. *Cell Res.* **2020**, *30*, 343–355. [[CrossRef](#)]
30. Xia, S.; Zhu, Y.; Liu, M.; Lan, Q.; Xu, W.; Wu, Y.; Ying, T.; Liu, S.; Shi, Z.; Jiang, S.; et al. Fusion mechanism of 2019-nCoV and fusion inhibitors targeting HR1 domain in spike protein. *Cell Mol. Immunol.* **2020**. [[CrossRef](#)]
31. Zhu, Y.; Yu, D.; Yan, H.; Chong, H.; He, Y. Design of Potent Membrane Fusion Inhibitors against SARS-CoV-2, an Emerging Coronavirus with High Fusogenic Activity. *J. Virol.* **2020**, *94*, e00620–e00635. [[CrossRef](#)] [[PubMed](#)]
32. Poma, A.B.; Li, M.S.; Theodorakis, P.E. Generalization of the elastic network model for the study of large conformational changes in biomolecules. *Phys. Chem. Chem. Phys.* **2018**, *20*, 17020–17028. [[CrossRef](#)] [[PubMed](#)]
33. Moreira, R.A.; Guzman, H.V.; Boopathi, S.; Baker, J.L.; Poma, A.B. Characterization of Structural and Energetic Differences between Conformations of the SARS-CoV-2 Spike Protein. *Materials* **2020**, *13*, 5362. [[CrossRef](#)] [[PubMed](#)]
34. Alm, E.; Broberg, E.K.; Connor, T.; Hodcroft, E.B.; Komissarov, A.B.; Maurer-Stroh, S.; Melidou, A.; Neher, R.A.; O’Toole, Á.; Pereyaslov, D.; et al. Geographical and temporal distribution of SARS-CoV-2 clades in the WHO European Region, January to June 2020. *Euro Surveill. Bull. Eur. Mal. Transm. Eur. Commun. Dis. Bull.* **2020**, *25*, 2001410. [[CrossRef](#)] [[PubMed](#)]
35. Anderson, D.E.; Becktel, W.J.; Dahlquist, F.W. pH-induced denaturation of proteins: A single salt bridge contributes 3–5 kcal/mol to the free energy of folding of T4 lysozyme. *Biochemistry* **1990**, *29*, 2403–2408. [[CrossRef](#)]
36. Debiec, K.T.; Gronenborn, A.M.; Chong, L.T. Evaluating the strength of salt bridges: A comparison of current biomolecular force fields. *J. Phys. Chem B* **2014**, *118*, 6561–6569. [[CrossRef](#)] [[PubMed](#)]
37. Steiner, T.; Koellner, G. Hydrogen bonds with pi-acceptors in proteins: Frequencies and role in stabilizing local 3D structures. *J. Mol. Biol.* **2001**, *305*, 535–557. [[CrossRef](#)]
38. Li, Q.; Wu, J.; Nie, J.; Zhang, L.; Hao, H.; Liu, S.; Zhao, C.; Zhang, Q.; Liu, H.; Nie, L.; et al. The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity. *Cell* **2020**. [[CrossRef](#)]
39. Sali, A.; Blundell, T.L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **1993**, *234*, 779–815. [[CrossRef](#)]
40. Feyfant, E.; Sali, A.; Fiser, A. Modeling mutations in protein structures. *Protein. Sci.* **2007**, *16*, 2030–2041. [[CrossRef](#)]
41. DeLano Scientific, L. 2002. Available online: <http://www.pymol.org> (accessed on 28 April 2021).
42. Vangone, A.; Spinelli, R.; Scarano, V.; Cavallo, L.; Oliva, R. COCOMAPS: A web application to analyze and visualize contacts at the interface of biomolecular complexes. *Bioinformatics* **2011**, *27*, 2915–2916. [[CrossRef](#)] [[PubMed](#)]
43. Coutsias, E.A.; Seok, C.; Jacobson, M.P.; Dill, K.A. A kinematic view of loop closure. *J. Comput. Chem.* **2004**, *25*, 510–528. [[CrossRef](#)] [[PubMed](#)]

44. Abraham, M.J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J.C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. [[CrossRef](#)]
45. Maier, J.A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K.E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713. [[CrossRef](#)]
46. Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*. [[CrossRef](#)]
47. Parrinello, M.; Rahman, A. Polymorphic Transitions in Single-Crystals—a New Molecular-Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182–7190. [[CrossRef](#)]
48. Essmann, U.; Perera, L.; Berkowitz, M.L.; Darden, T.; Lee, H.; Pedersen, L.G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577–8593. [[CrossRef](#)]
49. Hess, B.; Bekker, H.; Berendsen, H.J.C.; Fraaije, J.G.E.M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472. [[CrossRef](#)]
50. Abdel-Azeim, S.; Chermak, E.; Vangone, A.; Oliva, R.; Cavallo, L. MDcons: Intermolecular contact maps as a tool to analyze the interface of protein complexes from molecular dynamics trajectories. *BMC Bioinform.* **2014**, *15*, S1. [[CrossRef](#)]
51. Oliva, R.; Chermak, E.; Cavallo, L. Analysis and Ranking of Protein-Protein Docking Models Using Inter-Residue Contacts and Inter-Molecular Contact Maps. *Molecules* **2015**, *20*, 12045–12060. [[CrossRef](#)]
52. Vangone, A.; Cavallo, L.; Oliva, R. Using a consensus approach based on the conservation of inter-residue contacts to rank CAPRI models. *Proteins* **2013**, *81*, 2210–2220. [[CrossRef](#)] [[PubMed](#)]