



# Exploring the Association Between Sialic Acid and SARS-CoV-2 Spike Protein Through a Molecular Dynamics-Based Approach

Leonardo Bò<sup>1</sup>, Mattia Miotto<sup>1,2</sup>, Lorenzo Di Rienzo<sup>1</sup>, Edoardo Milanetti<sup>1,2\*</sup> and Giancarlo Ruocco<sup>1,2</sup>

<sup>1</sup> Center for Life Nanoscience, Istituto Italiano di Tecnologia, Rome, Italy, <sup>2</sup> Department of Physics, Sapienza University, Rome, Italy

## OPEN ACCESS

### Edited by:

Cláudio M. Soares,  
New University of Lisbon, Portugal

### Reviewed by:

Haibo Yu,  
University of Wollongong, Australia  
Megan L. O'Mara,  
Australian National University,  
Australia

### \*Correspondence:

Edoardo Milanetti  
edoardo.milanetti@uniroma1.it

### Specialty section:

This article was submitted to  
Pharmaceutical Innovation,  
a section of the journal  
Frontiers in Medical Technology

**Received:** 06 October 2020

**Accepted:** 03 December 2020

**Published:** 13 January 2021

### Citation:

Bò L, Miotto M, Di Rienzo L,  
Milanetti E and Ruocco G (2021)  
Exploring the Association Between  
Sialic Acid and SARS-CoV-2 Spike  
Protein Through a Molecular  
Dynamics-Based Approach.  
Front. Med. Technol. 2:614652.  
doi: 10.3389/fmedt.2020.614652

Recent experimental evidence demonstrated the capability of SARS-CoV-2 Spike protein to bind sialic acid molecules, which was a trait not present in SARS-CoV and could shed light on the molecular mechanism used by the virus for the cell invasion. This peculiar feature has been successfully predicted by *in-silico* studies comparing the sequence and structural characteristics that SARS-CoV-2 shares with other sialic acid-binding viruses, like MERS-CoV. Even if the region of the binding has been identified in the N-terminal domain of Spike protein, so far no comprehensive analyses have been carried out on the spike-sialic acid conformations once in the complex. Here, we addressed this aspect performing an extensive molecular dynamics simulation of a system composed of the N-terminal domain of the spike protein and a sialic acid molecule. We observed several short-lived binding events, reconnecting to the avidic nature of the binding, interestingly occurring in the surface Spike region where several insertions are present with respect to the SARS-CoV sequence. Characterizing the bound configurations via a clustering analysis on the Principal Component of the motion, we identified different possible binding conformations and discussed their dynamic and structural properties. In particular, we analyze the correlated motion between the binding residues and the binding effect on the stability of atomic fluctuation, thus proposing regions with high binding propensity with sialic acid.

**Keywords:** sialic acid, SARS- CoV-2, molecular dynamics, molecular binding, entry mechanism

## 1. INTRODUCTION

The epidemic of SARS-CoV-2, a novel strain of  $\beta$ -coronavirus first documented in late 2019 (1, 2), has rapidly become pandemic. Before the appearance of SARS-CoV-2, four other  $\beta$ -coronaviruses were reported to infect humans. In particular, two of them have already caused severe epidemics worldwide: SARS-CoV, which cause the Severe Acute Respiratory Syndrome (SARS) in 2003 (3, 4), and MERS-CoV, the Middle East Respiratory Syndrome (MERS), first isolated in 2012 (5). The other human  $\beta$ -coronaviruses, OC43 and HKU1, are known to circulate in the human population for many years and cause a mild seasonal illness (6). On the contrary, SARS-CoV-2 mainly affects the lower respiratory system, causing pneumonia, similarly to SARS-CoV and MERS-CoV. Besides, the virus has been found also in other regions of the body, like the gastrointestinal system or the

central nervous system where its effects are still under scrutiny (2, 7–9). Moreover, currently no approved vaccines are present although different antiviral compounds have been proposed (10–13).

From a molecular point of view, the Spike protein of CoVs, protruding from the viral membrane, not only plays a crucial role as a fundamental structural protein, but it also is essential for the interaction between CoV systems and host cells (14). Structurally, the spike protein is found in the trimeric complex, each chain composed of two sub-units: S1 and S2. In SARS-CoV, MERS-CoV, and SARS-CoV-2, the Receptor Binding Domain (RBD), located in the S1 domain, is responsible for viruses interaction with receptors on the host cell surface (15). On the other hand, the S2 subunit is responsible for the fusion between the virus and host membrane, causing the viral genome to penetrate the host cells cytoplasm (16).

To explain the entry molecular mechanism of SARS-Cov-2, a great effort is being spent in the study of the interactions at the molecular level between the virus and the host cells. In recent months it has been shown, through *in vivo* experiments, that SARS-CoV-2 uses the human Angiotensin-Converting Enzyme 2 (ACE2) receptor as primary entry channel (17), similarly to what done by SARS-CoV (18, 19).

Interestingly, the interaction with ACE2 involves the C-terminal domain of SARS-CoV-2 Spike protein, whose amino acid sequence is well conserved with respect to SARS-CoV homologous one (20). Conversely, the N-terminal domain presents some insertions, and these additional surface regions could be used by the virus to bind other cell receptors, so constituting an additional cell entry mechanism (17).

It has been actually demonstrated that several coronaviruses employ sialoglycan-based receptors as key component (21). These viruses bind to cell surface components containing N-acetyl-9-O-acetylneuraminic acid and this interaction is essential for the initiation of an infection (22). In particular, the structural details of the interaction between MERS-CoV Spike protein and a sialic acid molecule has been recently published (23). Comparing the structure of SARS-CoV-2 and MERS Spike proteins, we recently demonstrated that the insertions in the N-terminal domain confer to the solvent-exposed SARS-CoV-2 Spike a structure very similar to the region of the MERS-CoV Spike protein known to interact with sialic acid molecules (24).

These findings suggest that also SARS-CoV-2 gained the capability to bind host glycans, and recent computational and experimental literature confirms this hypothesis.

Indeed, while clinical investigations highlight that the replication of the virus does not take place only in the lower respiratory tract (25), Baker et al. (26) experimentally observed the binding between the NTD of the SARS-CoV-2 spike protein and sialic acid molecules using a glyco-nanoparticle platform.

Although no experimental data are available on the location of the binding region, computational works provided predictions of the possible involved residues. In particular, Robson et al. (27) used a new predictive algorithm to assign to each residue a sialic acid binding propensity. Notably, the regions around residues 58–70 and 255–265 on the N-terminal domain of the Spike display high scores toward sialic acid binding. Awasthi et al. (28)

instead used docking techniques and simulations to evaluate the binding affinities of SARS-CoV-2 with different sialic acid molecules. They found that the flexibility of the Leu244-Gly261 region seems to confer the Spike the ability to bind a wide variety of sialosides.

Here, we present a computational structural biology study based on extensive molecular dynamics simulation aimed to further investigate the binding between SARS-CoV-2 Spike N-terminal domain (NTD) and a sialic acid molecule, in order to blindly highlight the spike region most prone to interact with glycans. We firstly perform a molecular dynamics simulation of the trimeric form of the spike protein, underlining how the inter-chain interactions influence very slightly the overall motion of each N-terminal domain. Thus, we run a molecular dynamics simulation with a sialic acid molecule, to blindly investigate the protein surface pocket preferentially contacted by the glycan. Our analyses assign to each spike residue a binding propensity score, based on the time that residue spends in interaction with sialic acid, identifying the sialic acid most preferred residues. Moreover, we executed a Principal Component Analysis on the coordinates of the interacting atoms belonging both to Spike protein and sialic acid molecule when these two molecules are bound, aiming to identify the principal conformations explored by the molecular complexes. We thus identified 5 binding modes occurring between these two molecules, involving five possible regions on the NTD, and we investigated if the presence of the glycan stabilize such regions.

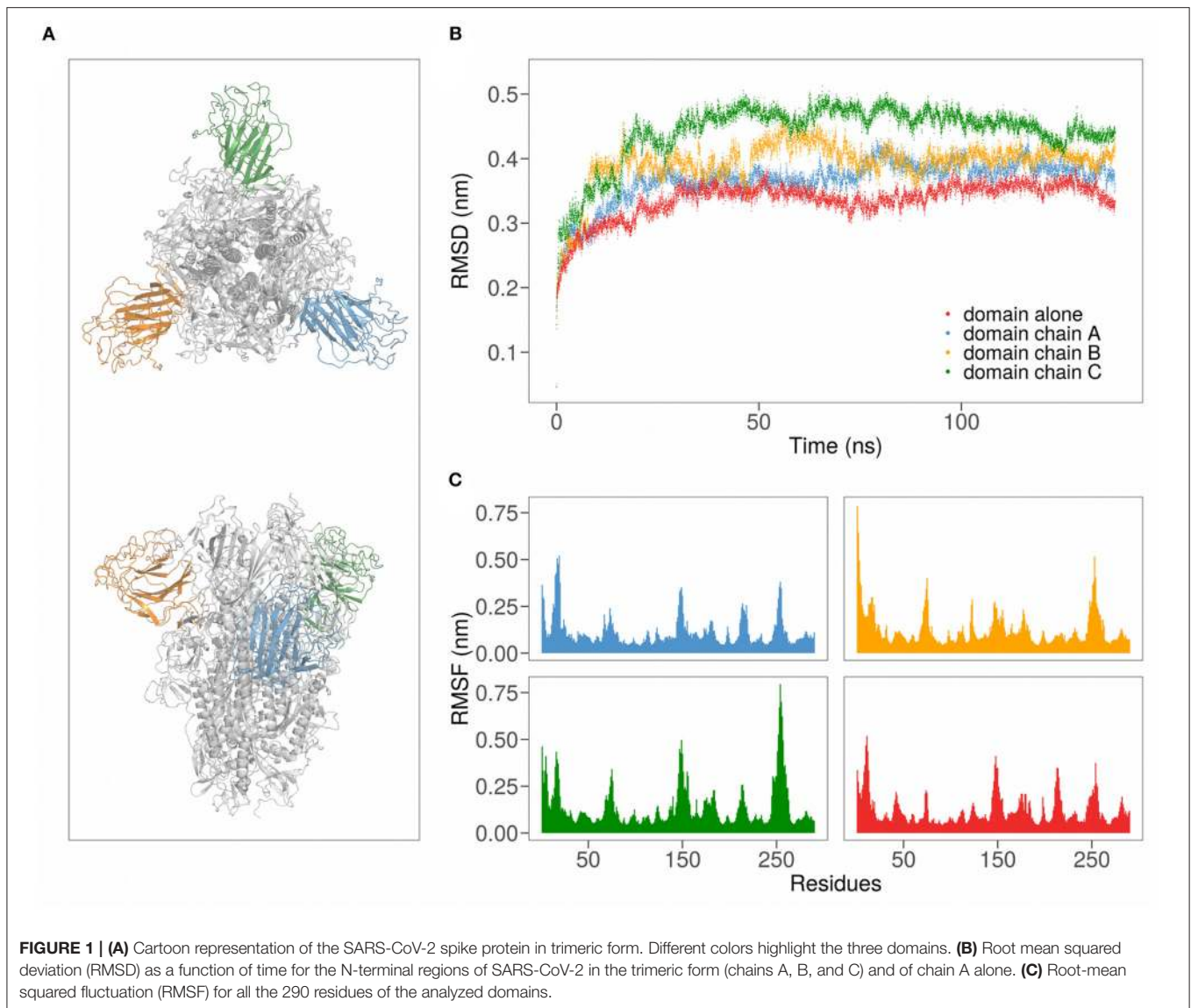
These results shed light on this important molecular interaction, that could provide an additional cell invasion mechanism to the virus possibly explaining the impressive rate of infection showed by SARS-CoV-2.

## 2. RESULTS AND DISCUSSION

To investigate protein-ligand interactions at the atomic level, molecular dynamics simulation (MD) with explicit solvent is a powerful computational approach because it allows us to take into account the exploration of the conformational space, as well as the solvent and entropic effects in a physically consistent manner (29), which are weakly considered in standard docking approaches. A crucial limitation of standard MD approaches is the exhaustiveness of the phase space sampling: in particular, when the energy barrier between the two states is too high, the study of the transition process is infeasible because the timescales of such transitions exceed those accessible to MD (30).

However, virus-glycan associations are typically characterized by low affinity and so they are inherently dynamic (31). Moreover, they are characterized by a high diffusion capacity of the molecule in water and by a few degrees of freedom. For these reasons, we studied the binding between spike protein of the SARS-CoV-2 and the sialic acid molecule through an extensive and completely unbiased MD simulation, intending to observe a spontaneous binding between the ligand and the protein.

To keep the computational costs low, increasing at the same time the simulation time to have a greater probability of observing the interaction between protein and ligand, we



used only the N-terminal domain (NTD) of the spike protein. Indeed, since both theoretical prediction and early experimental validations indicate that the sialic acid-binding region is localized in the NTD of the spike protein, we firstly compare the free motion of the domain (residues 1–290) when inserted in the trimeric spike complex and alone in water using 140 ns long MD trajectories (see section 4 for details and **Figure 1A** for a graphical representation).

## 2.1. Analysis of the Motion of SARS-CoV-2 N-Terminal Domain

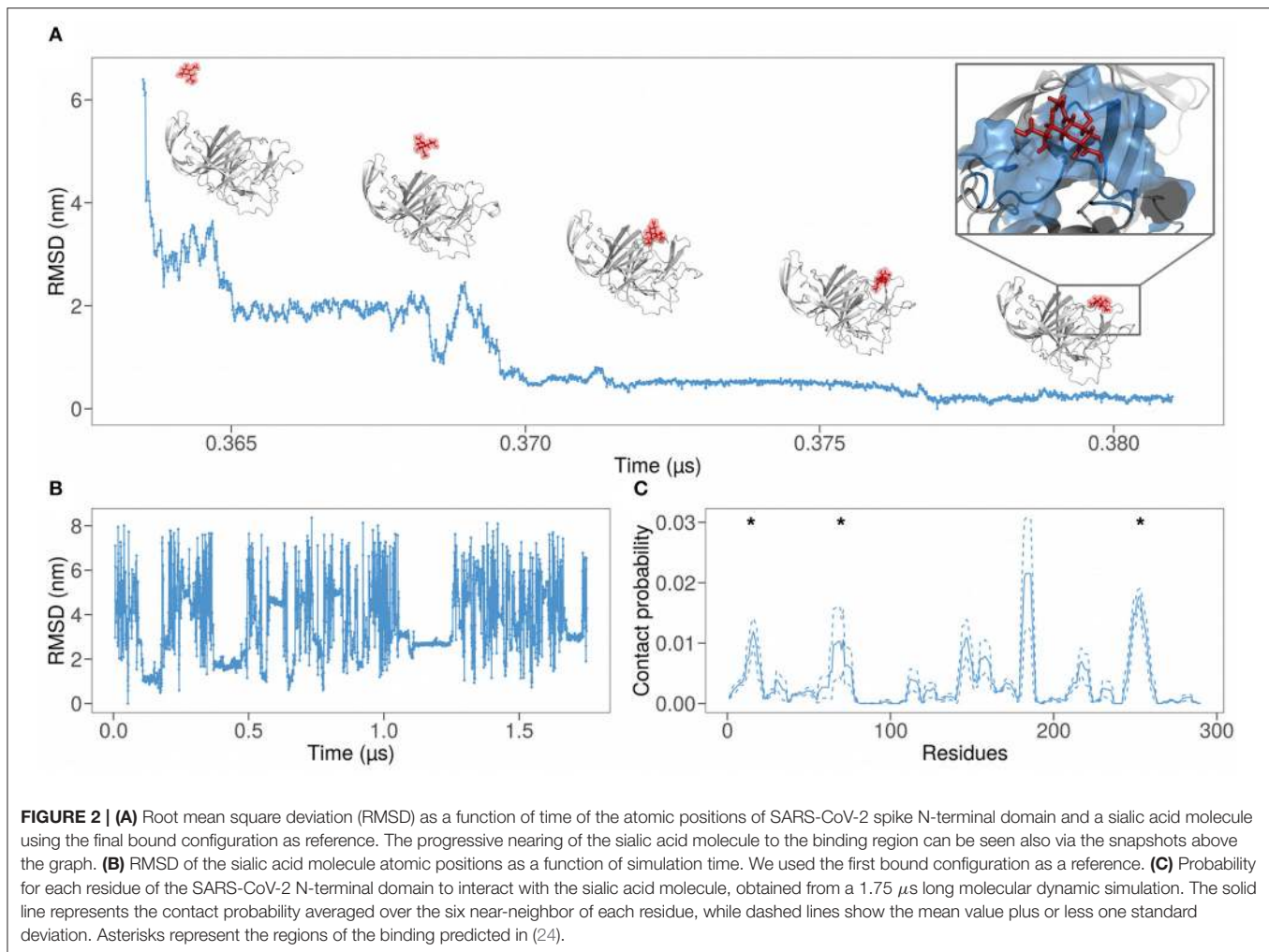
To describe the motion of the N-terminal Spike domain, we performed two MD simulations, one with the domain alone in the water and one with the trimeric form of Spike.

We thus evaluated both the root mean square deviation (RMSD) and the root mean square fluctuation (RMSF) of the backbone carbons of the Spike N-terminal domain (see

**TABLE 1 |** SARS-CoV-2 area under the curve values for the identified regions of high values of contact probability (peaks in **Figure 2C**).

Residues range	Area	Peak
1–23	0.117	Peak-1S
56–78	0.132	Peak-2S
140–165	0.153	Peak-3S
178–191	0.153	Peak-4S
241–265	0.194	Peak-5S

**Figures 1B,C**). Interestingly, in both the simulations the regions of the domain that present higher mobility (high RMSF) are formed by residues forming structural loops (see **Table 1**). Notably, these high motile regions are located on the Spike SARS-CoV-2 insertions with respect to SARS-CoV homolog.



Remarkably, both descriptors behave similarly when the domain is alone in water (red in the Figure) or is inserted in the trimeric complex (different chain domains are represented in blue, orange, and green, respectively). In particular, the RMSD curves show the same trend, with about the same time to reach equilibrium, when both the single domain and the trimeric complex are considered. More specifically, the average RMSD value at equilibrium (after 30 ns of the simulation) of the N-terminal domain alone in solution is very similar to that of the same domain belonging to the chain A of the trimer, meaning that the interdomain interactions are not essential for the overall fold maintenance.

On the other hand, the comparison between the residues RMSF obtained when the domain is simulated alone or in trimer shows that the most fluctuating and least fluctuating regions are conserved. Indeed, the average Pearson correlation coefficient between the RMSF values of the three domains of the trimer and the RMSF values of the domain alone is 0.63. Looking at the correlation of the motions between residues, we underlined how even the residues combined motions are conserved when the domain is simulated alone (see **Supplementary Material**).

Taken together, these results indicate that the inter-chain interactions do not notably modify the motion of the Spike N-terminal domain. This allows us to simulate only the N-terminal domain, thus significantly reducing the computational cost of our simulations.

## 2.2. Molecular Simulation of the Binding Between SARS-CoV-2 and a Sialic Acid Molecule

We then performed a long molecular dynamics simulation of the N-terminal domain of SARS-CoV-2 in the presence of a sialic acid molecule (see section 4 for details), in order to observe a blind binding. To keep the computational cost of the simulation low and to avoid ligand-ligand interactions, we setup the system with a Spike N-terminal domain and a single sialic acid molecule.

During the 1.75  $\mu\text{s}$  of the evolution, we observed various binding (and unbinding) events. In particular, each event consists of a progressive nearing of the sialic acid to the protein until a binding takes place. **Figure 2A** displays some snapshots of one of the observed binding events, where the positions of the sialic



acid molecule are measured in terms of RMSD, using the final bound configuration as a reference. As intuitively, we observe a progressive decrease of the RMSD, until the sialic acid binds to the protein.

As testified by the various constant stretch in **Figure 2B**, the binding is not unique, probably due to the aviditic nature of the interaction (31). Therefore, we defined a measure of local binding propensity, based on the number of contacts each residue does with sialic acid during the time evolution of the simulation. Indeed, for each residue, we calculate the fraction of frames in which the sialic acid center of mass is located at a distance lower than 6 Å to any heavy atoms of the residue.

In **Figure 2C**, the dynamics binding propensities for each residue are shown. We performed a smoothing procedure for each residue, averaging the contact probability of the examined residue (solid lines) on a sequence window of 7 residues centered on it. Dashed lines represent the fluctuations around the mean.

The contact probability as a function of the residues shows well-defined peaks. This means that the binding with sialic acid involves specific regions. The sum of all the areas under the most pronounced peaks, whose residue ranges are reported in **Table 1** and represent the regions with the highest contact frequency, consists of the 75% of the total contacts occurring between spike and sialic acid.

Asterisks in **Figure 2C** mark the regions of SARS-CoV-2 that show a high shape similarity with the MERS-CoV binding region, as explained in (24). Notably, these three out of five peaks coincide with the previously identified region, while the peak in the range 140–165 is in close proximity to the predicted region.

### 2.3. Possible Binding Conformations

To characterize the different observed binding events, we investigated the structural conformations of the formed complexes. To do so, we performed a principal component analysis (PCA) on the spatial coordinate of the heavy atoms of both the 111 interacting residues of the spike domain and the sialic acid for uniform sampling of the molecular dynamics trajectory. Since we found not an unique binding region, we looked for a compact way to study and isolate the different binding regions on the spike protein. The PCA allows us to compactly represent each molecular dynamics frame as a point in essential plane, where the difference with the other possible bound conformation is maximum. In this way we are sure that the clustering procedure groups conformations that are similar focusing only on the position of the interacting atoms, considering both the protein and the ligand.

**Figure 3A** shows the projection of the sampled bound configurations in the plane identified by the first two principal components, which represent 69% of the total variance. Since each binding pose is characterized by a different set of atomic coordinates, this protocol allows us to group similar structures in the essential space of the first two principal components. The different binding configurations form clusters in the plane, which can be isolated upon performing a clustering analysis with a density filter (see details in section 4). We end up with five clusters, highlighted in different colors in **Figure 3A**.

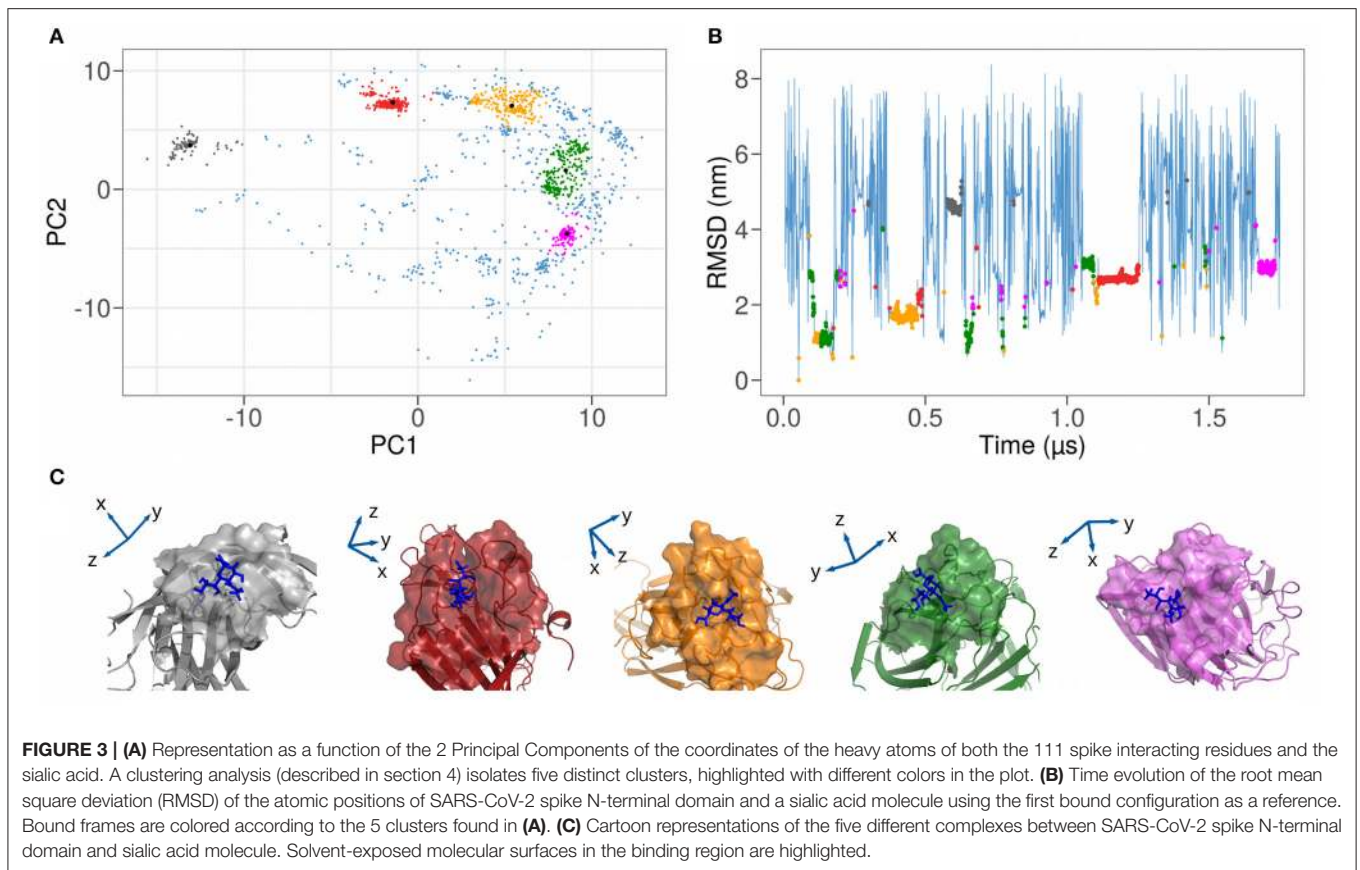
In **Table 2**, we report, for each identified cluster, the percentage of its population with regard to the total bound frames of the trajectory. For example, the most frequent binding region is cluster 2, with a percentage of all other binding conformations of 32%.

The groups obtained via the PCA analysis can be investigated following the time evolution of the simulation. Indeed, we show in **Figure 3B** the RMSD plot where each bound state was colored according to the coloring scheme defined with the PCA clustering and in **Figure 3C** an illustration of the 5 found binding modes. This RMSD representation allows the investigation of the duration time of each binding event. Indeed, even if the population of 2 clusters can be comparable, it is possible to have many unstable or few stable events. For example, cluster 2 seems to be more stable than cluster 4, since it is characterized by a unique and long binding event, even if the simulation time should be considerably increased to have a reliable conclusion.

### 2.4. Protein-Ligand Binding Characterization

The principal component analysis allows us to identify binding conformations between the N-terminal domain of the SARS-CoV-2 spike protein and the sialic acid molecule. Furthermore, selecting the most representative structure of each cluster and analyzing the protein-ligand inter-molecular contacts, we define 5 possible binding sites. Identified the interacting residues involved in the binding sites, we investigated their dynamic and structural properties, analyzing, on one hand, the correlation of motion between the residues during the unbound simulation time and on the other hand the effect of the binding in terms of structural stabilization.

The correlation analysis between the residues of the N-terminal domain, shown in **Figure 4A**, allows the identification of the sub-regions characterized by a high correlated motion in the unbound state, meaning that there is a dependence between the motion of these residues implying a possible functional role. More interestingly, in order to investigate the role of residues identified in interaction, we compare their distribution of Pearson correlations with those regarding set of patches centered on the residues of the loops (shown in **Table 1**). In particular, in order to compare the average correlation between the residues of each proposed binding site with a set of random non-interacting surface regions, we define 111 random patches, centered each time on one of the loops residues and composed by the residues closer than 8 Å to it. The mean correlation of the residues belonging to the randomly selected patches is 0.27 (**Figure 4B**, blue distributions), while for each binding site identified by the PCA the correlations are 0.33, 0.29, 0.26, 0.27, and 0.29 for the cluster 1, cluster 2, cluster 3, cluster 4, and cluster 5, respectively (**Figure 4B**, colored lines). Interestingly, the correlations of the motion of the residues of both cluster 3 and cluster 4 (which are indicated in orange and green in the figures) are less than or equal to the average correlation of residues belonging to non-interacting patches. On the other hand, clusters 1, 2, and 5 are characterized by correlations higher than the mean of the correlations of the random patches.



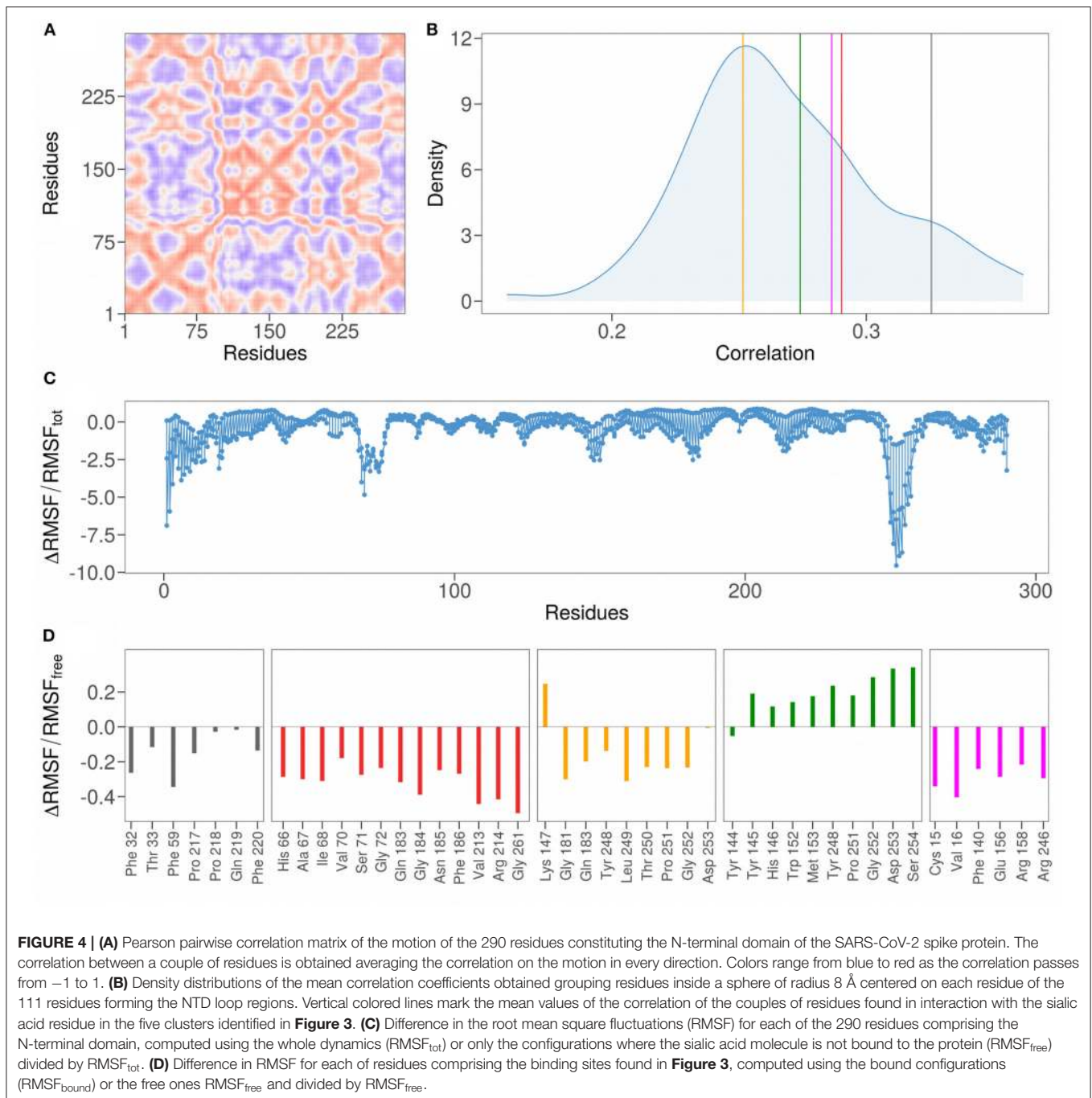
**TABLE 2 |** Summary of the relevant features for the five clusters identified in **Figure 3A**: percentage of bound conformations assigned to each cluster, list of residues in interaction with the sialic acid molecule, mean RMSF in free and bound state of the residues associated to each cluster.

Cluster population (%)	Residues	RMSF <sub>free</sub>	RMSF <sub>bound</sub>	
Cluster 1	8.91	Phe 32, Thr 33, Phe 59, Pro 217, Pro 218, Gln 219, Phe 220	0.11 ± 0.01	0.10 ± 0.01
Cluster 2	31.93	His 66, Ala 67, Ile 68, Val 70, Ser 71, Gly 72, Gln 183, Gly 184, Asn 185, Phe 186, Val 213, Arg 214, Gly 261	0.25 ± 0.03	0.17 ± 0.02
Cluster 3	22.52	Lys 147, Gly 181, Gln 183, Tyr 248, Leu 249, Thr 250, Pro 251, Gly 252, Asp 253	0.44 ± 0.07	0.37 ± 0.06
Cluster 4	22.42	Tyr 144, Tyr 145, His 146, Trp 152, Met 153, Tyr 248, Pro 251, Gly 252, Asp 253, Ser 254	0.38 ± 0.08	0.47 ± 0.11
Cluster 5	14.21	Cys 15, Val 16, Phe 140, Glu 156, Arg 158, Arg 246	0.13 ± 0.02	0.09 ± 0.01

We also analyzed the variation of the mobility of the interacting residues before and after the binding with the ligand. For this purpose, we calculated the percentage difference of the atomic fluctuation between the bound and unbound state for each N-terminal domain residue (**Figure 4C**). Notably, the region with the highest values, containing the Tryptophan 258, is characterized by high structural stabilization upon binding with the ligand, and the same region was predicted as a possible interacting region by (24) and (27). Similarly, two other regions show a high level of stabilization after binding: the N-terminal loop region and the region centered on Lysine 77. Note that the N-terminal loop is characterized by high mobility (high RMSF values) in the simulation where the NTD is considered free in solution, as can be seen in **Figure 1C**. This analysis shows that the N-terminal loop

could play a role in binding given its evident stabilization upon binding.

More specifically, we here also analyze the stabilizing effect of binding with the ligand by considering only the residues of the 5 binding sites identified. Similarly to what was found with the correlation analysis, clusters 1, 2, and 5 also show a highly stabilizing effect upon binding with the sialic acid molecule (**Figure 4D**). Interestingly, the residues belonging to cluster 4 (indicated in green) are characterized by a higher fluctuation of atomic mobility in their bound state with the ligand. The binding region with the highest degree of stabilization is that of cluster 2 (in red), having the average percentage difference in RMSF before and after binding of  $-0.32$ , in comparison with  $-0.15$ ,  $-0.16$ ,  $0.2$ , and  $-0.30$  for cluster 1, cluster 3, cluster 4, and cluster 5, respectively.



Thus, the analysis of the RMSD (**Figure 2B**) shows that the kinetics of the binding between a ligand and the binding site of cluster 4 is much faster than that between sialic acid and cluster 2 (the duration of the cluster 4 binding events is lower than that with cluster 2). This behavior could be explained by the after binding high atomic fluctuations of the cluster 4 residues and suggests the exclusion of this region as a possible region of interaction with sialic acid. More detailed data about the lifetime of binding events can be found in Supporting Information.

A more accurate study would be needed to analyze the binding kinetics, thus performing much longer simulations.

### 3. CONCLUSIONS

The N-terminal domain of SARS-CoV-2 spike protein is characterized by some insertion with respect to the SARS-CoV homologous one, and the surface regions generated by

these residues have acquired the capability to bind sialic acid molecule (24, 26, 27). However, at present, the atomic details of the molecular complex is still missing.

Here, we address this issue through an extensive molecular dynamics simulation, where we were able to observe spontaneous binding events occurring between the spike protein of SARS-CoV-2 and the sialic acid molecule. In particular, we simulated the free motion of both the sialic acid molecule and the spike protein for times long enough to detect several binding between the two molecules. The main advantage of this approach is to consider exhaustively the physics of the system, taking into consideration electrostatic interactions as well as solvent and entropic effects.

We identify 5 sialic acid-interacting regions belonging to the NTD of the spike protein.

It has to be noted that physiologically the Spike protein is extensively glycosylated, a property that often escape the experimental characterization. Indeed, each Spike protomer has 22 glycosylation sites, which are supposed to affect which cells the virus can infect (32–35). This glycan shield, beyond being structurally necessary to bind ACE2, allows the virus to elude the host immune response (36). Importantly, none of the residues we found to be involved in sialic acid contact are in close proximity to these glycosylation sites, and therefore the binding region we identified could be considered free and capable to interact with sialic acid.

Moreover, since sialic acid is negatively charged, it is remarkable that the selected interacting regions are never characterized by negative net charge. The binding mode 2 is characterized by the highest positive charge that make it a suitable region to recognize the sialic acid molecule, and it is probably because of this that it is the most populated cluster among the five we proposed. Indeed, even looking at the sialic acid experimental binding region observed in MERS-CoV Spike protein, it emerges that the residues involved generate an overall charge of +1 (in unit of electron charge), similar to those characterizing our proposed binding modes.

Dynamic and structural analyses on these binding sites allow us to characterize the regions most prone to binding with the molecule. In particular, we consider the binding region 4 unsuitable for binding to sialic acid since binding occurs several times but of short duration. In addition, the residues corresponding to this region have a lower degree of correlation of their motion and the binding of the sialic acid increases the atomic mobility. On the other hand, three regions (number 1, 2, and 5) are particularly interesting because they have a high average correlation in terms of the motion of the residues that compose them and undergo an evident stabilizing effect after binding with sialic acid.

## 4. MATERIALS AND METHODS

### 4.1. Datasets

- Unbound SARS-CoV-2 spike protein: modeled by I-TASSER server (37). The Top 10 structural analogs in PDB [as identified by TM-align (38)] are: 5x58, 6nzk, 6nb3, 3jcl, 5i08, 6cv0, 5szs, 5wrg, 6utk, 6b7n.

- *N*-Acetylneuraminic acid molecule: structure from PubChem database (id: 439197) (39). The parameters to perform the molecular dynamics simulation were obtained from the SwissParam server (40).

### 4.2. Molecular Dynamics Simulations

All simulations were performed using Gromacs 2019.3 (41). Topologies of the system were built using the CHARMM-27 force field (42). The protein was placed in a dodecahedral simulative box, with periodic boundary conditions, filled with TIP3P water molecules (43). For all simulated systems, we checked that each atom of the proteins was at least at a distance of 1.1 nm from the box borders. Each system was then minimized with the steepest descent algorithm. Next, a relaxation of water molecules and thermalization of the system was run in NVT and NPT environments each for 0.1 ns at 2 fs time-step. The temperature was kept constant at 300 K with v-rescale thermostat (44); the final pressure was fixed at 1 bar with the Parrinello-Rahman barostat (45).

LINCS algorithm (46) was used to constraint bonds involving hydrogen atoms. A cut-off of 12 Å was imposed for the evaluation of short-range non-bonded interactions and the Particle Mesh Ewald method (47) for the long-range electrostatic interactions. The described procedure was used for all the performed simulations. In the following, we provide further details, specific of each simulation.

#### 4.2.1. SARS-CoV-2 Spike Trimer

Simulation of the SARS-CoV-2 spike trimeric complex was performed starting from the model structure proposed by the I-Tasser server (48). The addition of 3 sodium counterions rendered the systems electroneutral. The water density was of 1,004 kg/m<sup>3</sup>, close to the experimental value. The system was simulated with a 2 fs time-step for 140 ns in periodic boundary conditions.

#### 4.2.2. SARS-CoV Spike-2 N-Terminal Domain (Residue Range: 1–290)

We simulate also the N-terminal region of the SARS-CoV-2 spike in the presence of one molecule of sialic acid (neu5ac) in solution. We selected the domain ranging from residue 1 to 290 of the chain A of the trimeric complex. We added 2 Cl molecules to neutralize the system, while water density was 1,008 kg/m<sup>3</sup>. The simulation was carried out for 1.75 μs with time steps of 2 fs.

### 4.3. Principal Component and Clustering Analysis

We performed a PCA on the covariances of the  $C_{\alpha}$  of the residues belonging to the peaks reported in **Table 1** (111 atoms) and on the heavy atoms of the sialic acid molecule (21 atoms) using Gromacs 2019.3 internal packages. The covariances were calculated sampling the frames of the molecular dynamics trajectory each 500 ps, and selecting only those frames where the sialic acid contacts at least one of the residues belonging to the peaks. A contact is defined if the center of mass of the sialic acid is at a distance lower than 6 Å from any heavy atom of the residue. We ended up with 1574 contact frames. In order to obtain the



clusters showed in **Figure 3A**, we first evaluated the density of the points in the PC1-PC2 plane; then we filtered the points according to a density threshold and performed hierarchical clustering. By varying the threshold, we ended up with 5 groups on the density grid. We then reassigned the points of the original essential plane according to the position they had on the density grid (see **Table 2**). Analyses were performed using R standard libraries (49).

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

## REFERENCES

- Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* (2020) 395:497–506. doi: 10.1016/S0140-6736(20)30183-5
- Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med.* (2020) 382:727–33. doi: 10.1056/NEJMoa2001017
- Drosten C, Günther S, Preiser W, Van Der Werf S, Brodt H-R, Becker S, et al. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N Engl J Med.* (2003) 348:1967–76. doi: 10.1056/NEJMoa030747
- Ksiazek TG, Erdman D, Goldsmith CS, Zaki SR, Peret T, Emery S, et al. A novel coronavirus associated with severe acute respiratory syndrome. *N Engl J Med.* (2003) 348:1953–66. doi: 10.1056/NEJMoa030781
- Zaki AM, Van Boheemen S, Bestebroer TM, Osterhaus ADME, Fouchier RAM. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med.* (2012) 367:1814–20. doi: 10.1056/NEJMoa1211721
- Liu DX, Liang JQ, Fung TS. Human coronavirus-229e-OC43-NL63, and-HKU1. *Refer Module Life Sci.* (2020). doi: 10.1016/b978-0-12-809633-8.21501-x
- Promptchara E, Ketloy C, Palaga T. Immune responses in COVID-19 and potential vaccines: lessons learned from sars and mers epidemic. *Asian Pac J Allergy Immunol.* (2020) 38:1–9. doi: 10.12932/AP-200220-0772
- Su S, Wong G, Shi W, Liu J, Lai ACK, Zhou J, et al. Epidemiology, genetic recombination, and pathogenesis of coronaviruses. *Trends Microbiol.* (2016) 24:490–502. doi: 10.1016/j.tim.2016.03.003
- Sulzer D, Antonini A, Leta V, Nordvig A, Smeyne RJ, Goldman JE, et al. COVID-19 and possible links with Parkinson's disease and parkinsonism: from bench to bedside. *NPJ Parkinson Dis.* (2020) 6:1–10. doi: 10.1038/s41531-020-00123-0
- Riva L, Yuan S, Yin X, Martin-Sancho L, Matsunaga N, Pache L, et al. Discovery of SARS-CoV-2 antiviral drugs through large-scale compound repurposing. *Nature* (2020) 586:113–9. doi: 10.1038/s41586-020-2577-1
- Serrano G, Kochergina I, Albors A, Diaz E, Oroval M, Hueso G, et al. Liposomal lactoferrin as potential preventative and cure for covid-19. *Int J Res Health Sci.* (2020) 8:8–15. doi: 10.5530/ijrhs.8.1.3
- Miotto M, Rienzo LD, Leonardo B, Boffi A, Ruocco G, Milanetti E. Molecular mechanisms behind anti SARS-CoV-2 action of lactoferrin. *arXiv [Preprint] arXiv:2007.07341* (2020).
- Grein J, Ohmagari N, Shin D, Diaz G, Asperges E, Castagna A, et al. Compassionate use of remdesivir for patients with severe covid-19. *N Engl J Med.* (2020) 382:2327–36. doi: 10.1056/NEJMoa2007016
- Schoeman D, Fielding BC. Coronavirus envelope protein: current knowledge. *Virology* (2019) 16:1–22. doi: 10.1186/s12985-019-1182-0
- Zhu Z, Lian X, Su X, Wu W, Marraro GA, Zeng Y. From SARS and MERS to COVID-19: a brief summary and comparison of severe acute respiratory infections caused by three highly pathogenic human coronaviruses. *Respir Res.* (2020) 21:1–14. doi: 10.1186/s12931-020-01479-w
- Ou X, Liu Y, Lei X, Li P, Mi D, Ren L, et al. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nat Commun.* (2020) 11:1–12. doi: 10.1038/s41467-020-15562-9
- Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* (2020) 579:270–3. doi: 10.1038/s41586-020-2012-7
- Li W, Zhang C, Sui J, Kuhn JH, Moore MJ, Luo S, et al. Receptor and viral determinants of sars-coronavirus adaptation to human ACE2. *EMBO J.* (2005) 24:1634–43. doi: 10.1038/sj.emboj.7600640
- Li F. Structural analysis of major species barriers between humans and palm civets for severe acute respiratory syndrome coronavirus infections. *J Virol.* (2008) 82:6984–91. doi: 10.1128/JVI.00442-08
- Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q. Structural basis for the recognition of the SARS-CoV-2 by full-length human ACE2. *Science.* (2020) 367:1444–8. doi: 10.1126/science.abb2762
- Hulswit RDG, Lang Y, Bakkers MJG, Li W, Li Z, Schouten A, et al. Human coronaviruses OC43 and HKU1 bind to 9-o-acetylated sialic acids via a conserved receptor-binding site in spike protein domain A. *Proc Natl Acad Sci USA.* (2019) 116:2681–90. doi: 10.1073/pnas.1809667116
- Schwegmann-Wefßels C, Herrler G. Sialic acids as receptor determinants for coronaviruses. *Glycoconj J.* (2006) 23:51–8. doi: 10.1007/s10719-006-5437-9
- Tortorici MA, Walls AC, Lang Y, Wang C, Li Z, Koerhuis D, et al. Structural basis for human coronavirus attachment to sialic acid receptors. *Nat Struct Mol Biol.* (2019) 26:481–9. doi: 10.1038/s41594-019-0233-y
- Milanetti E, Miotto M, Di Rienzo L, Monti M, Gosti G, Ruocco G. *In-silico* evidence for two receptors based strategy of SARS-CoV-2. *arXiv preprint arXiv:2003.11107* (2020). doi: 10.1101/2020.03.24.006197
- Wölfel R, Corman VM, Guggemos W, Seilmaier M, Zange S, Müller MA, et al. Virological assessment of hospitalized patients with COVID-2019. *Nature* (2020) 581:1–10. doi: 10.1038/s41586-020-2196-x
- Baker AN, Richards S-J, Guy CS, Congdon TR, Hasan S, Zwetsloot AJ, et al. The SARS-CoV-2 spike protein binds sialic acids and enables rapid detection in a lateral flow point of care diagnostic device. *ACS Cent Sci.* (2020) 11:2046–52. doi: 10.1021/acscentsci.0c00855
- Robson B. Bioinformatics studies on a function of the SARS-CoV-2 spike glycoprotein as the binding of host sialic acid glycans. *Comput Biol Med.* (2020) 122:103849. doi: 10.1016/j.combiomed.2020.103849
- Awasthi M, Gulati S, Sarkar DP, Tiwari S, Kateriya S, Ranjan P, et al. The sialoside-binding pocket of SARS-CoV-2 spike glycoprotein structurally resembles MERS-CoV. *Viruses.* (2020) 12:909. doi: 10.3390/v12090909
- Söderhjelm P, Tribello GA, Parrinello M. Locating binding poses in protein-ligand systems using reconnaissance metadynamics. *Proc Natl Acad Sci USA.* (2012) 109:5170–5. doi: 10.1073/pnas.1201940109

## AUTHOR CONTRIBUTIONS

EM and GR conceived the research. LB performed molecular dynamics simulations. LB, MM, LD, and EM performed calculations and computational analysis. All authors analyzed the results, wrote, and revised the paper.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmedt.2020.614652/full#supplementary-material>

30. Dror RO, Arlow DH, Maragakis P, Mildorf TJ, Pan AC, Xu H, et al. Activation mechanism of the  $\beta$ 2-adrenergic receptor. *Proc Natl Acad Sci USA*. (2011) 108:18684–9. doi: 10.1073/pnas.1110499108
31. Li W, Hulswit RJG, Widjaja I, Raj VS, McBride R, Wenjie P, et al. Identification of sialic acid-binding function for the middle east respiratory syndrome coronavirus spike glycoprotein. *Proc Natl Acad Sci USA*. (2017) 114:E8508–17. doi: 10.1073/pnas.1712592114
32. Watanabe Y, Berndsen ZT, Raghvani J, Seabright GE, Allen JD, Pybus OG, et al. Vulnerabilities in coronavirus glycan shields despite extensive glycosylation. *Nat Commun*. (2020) 11:1–10. doi: 10.1038/s41467-020-16567-0
33. Breiman A, Ruvèn-Cloutier N, Le Pendu J. Harnessing the natural anti-glycan immune response to limit the transmission of enveloped viruses such as SARS-CoV-2. *PLoS Pathog*. (2020) 16:e1008556. doi: 10.1371/journal.ppat.1008556
34. Miotto M, Di Rienzo L, Gosti G, Milanetti E, Ruocco G. Does blood type affect the covid-19 infection pattern? *arXiv [Preprint] arXiv:2007.06296* (2020).
35. Zhao J, Yang Y, Huang H-P, Li D, Gu D-F, Lu X-F, et al. Relationship between the abo blood group and the COVID-19 susceptibility. *medRxiv [Preprint]*. (2020). doi: 10.1101/2020.03.11.20031096
36. Casalino L, Gaieb Z, Goldsmith JA, Hjorth CK, Dommer AC, Harbison AM, et al. Beyond shielding: the roles of glycans in the SARS-CoV-2 spike protein. *ACS Cent Sci*. (2020) 6:1722–34. doi: 10.1021/acscentsci.0c01056
37. Yang J, Zhang Y. I-tasser server: new development for protein structure and function predictions. *Nucleic Acids Res*. (2015) 43:W174–81. doi: 10.1093/nar/gkv342
38. Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res*. (2005) 33:2302–9. doi: 10.1093/nar/gki524
39. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, et al. PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res*. (2018) 47:D1102–9. doi: 10.1093/nar/gky1033
40. Zoete V, Cuendet MA, Grosdidier A, Michielin O. SwissParam: a fast force field generation tool for small organic molecules. *J Comput Chem*. (2011) 32:2359–68. doi: 10.1002/jcc.21816
41. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. GROMACS: fast, flexible, and free. *J Comput Chem*. (2005) 26:1701–18. doi: 10.1002/jcc.20291
42. Brooks BR, Brooks CL, Mackerell AD, Nilsson L, Petrella RJ, Roux B, et al. CHARMM: The biomolecular simulation program. *J Comput Chem*. (2009) 30:1545–614. doi: 10.1002/jcc.21287
43. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys*. (1983) 79:926–35. doi: 10.1063/1.445869
44. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. *J Chem Phys*. (2007) 126:014101. doi: 10.1063/1.2408420
45. Parrinello M, Rahman A. Crystal structure and pair potentials: a molecular-dynamics study. *Phys Rev Lett*. (1980) 45:1196–9. doi: 10.1103/PhysRevLett.45.1196
46. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: a linear constraint solver for molecular simulations. *J Comput Chem*. (1997) 18:1463–72. doi: 10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H
47. Tcheatham TE III, Miller JL, Fox T, Darden TA, Kollman PA. Molecular dynamics simulations on solvated biomolecular systems: the particle mesh ewald method leads to stable trajectories of DNA, RNA, and proteins. *J Am Chem Soc*. (1995) 117:4193–4.
48. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc*. (2010) 5:725–38. doi: 10.1038/nprot.2010.5
49. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing (2020).

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Bò, Miotto, Di Rienzo, Milanetti and Ruocco. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.