ELSEVIER

Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

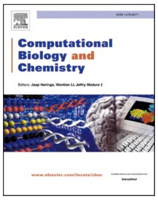
Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Computational Biology and Chemistry 98 (2022) 107668

Contents lists available at ScienceDirect

Computational Biology and Chemistry

journal homepage: www.elsevier.com/locate/cbac



Glycosylation is key for enhancing drug recognition into spike glycoprotein of SARS-CoV-2

Georcki Ropón-Palacios [a,*], Jhon Pérez-Silva [a], Ricardo Rojas-Humpire [a], Gustavo E. Olivos-Ramírez [a], Manuel Chenet-Zuta [b], Victor Cornejo-Villanueva [d], Sheyla Carmen-Sifuentes [a], Kewin Otazu [a], Yaritza L. Ramirez-Díaz [a], Karolyn Vega Chozo [a], Ihosvany Camps [a,c,**]

[a] *Laboratório de Modelagem Computacional, Instituto de Ciências Exatas, Universidade Federal de Alfenas, Brazil*
[b] *Escuela de Posgrado, Universidad San Ignacio de Loyola, Peru*
[c] *High Performance & Quantum Computing Labs, Waterloo, Canada*
[d] *Escuela de Genética y Biotecnología, Universidad Nacional Mayor de San Marcos, Peru*

ARTICLE INFO

Keywords:
COVID-19
Molecular dynamics
N- and O-glycosylation
Cryptic pocket
Induced-Fit binding
Roto-translation phenomenon
Free Energy Landscape

ABSTRACT

The emergence of COVID-19 caused by SARS-CoV-2 and its spread since 2019 represents the major public health problem worldwide nowadays, which has generated a high number of infections and deaths. The spike protein (S protein) is the most studied protein of SARS-CoV-2, and key to host-cell entry through ACE2 receptor. This protein presents a large pattern of glycosylations with important roles in immunity and infection mechanisms. Therefore, understanding key aspects of the molecular mechanisms of these structures, during drug recognition in SARS-CoV-2, may contribute to therapeutic alternatives. In this work, we explored the impact of glycosylations on the drug recognition on two domains of the S protein, the receptor-binding domain (RBD) and the N-terminal domain (NTD) through molecular dynamics simulations and computational biophysics analysis. Our results show that glycosylations in the S protein induce structural stability and changes in rigidity/flexibility related to the number of glycosylations in the structure. These structural changes are important for its biological activity as well as the correct interaction of ligands in the RBD and NTD regions. Additionally, we evidenced a roto-translation phenomenon in the interaction of the ligand with RBD in the absence of glycosylation, which disappears due to the influence of glycosylation and the convergence of metastable states in RBM. Similarly, glycosylations in NTD promote an induced fit phenomenon, which is not observed in the absence of glycosylations; this process is decisive for the activity of the ligand at the cryptic site. Altogether, these results provide an explanation of glycosylation relevance in biophysical properties and drug recognition to S protein of SARS-CoV-2, which must be considered in the rational drug development and virtual screening targeting S protein.

1. Introduction

Coronaviruses (CoVs) are a diverse group of enveloped, positive sense, single-stranded RNA viruses causing mild to severe respiratory infections in humans (Pal et al., 2020). In December 2019, a novel coronavirus, designated as Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), outbroke and rapidly spread causing an epidemic of unusual pneumonia in Wuhan, China (Woo et al., 2020). This novel coronavirus disease was declared as COVID-19 by the World Health Organization (WHO) at the beginning of 2020; since then, it has spread fast all over the world and has become the major public health problem. Such has been the impact of this pandemic that, until this submission, more than 353 million confirmed cases and more than 5 million deaths have been reported (Dong et al., 2020). For that reason, governments and the scientific community developed strategies to contain the impact of COVID-19, based on epidemiological surveillance, vaccination, and infection mechanisms research (Ibrahim, 2020). This has led to accelerated development of vaccines and the initiation of a worldwide vaccination campaign, which to date has seen more than 9 billion people vaccinated (Randall et al., 2021). However, many countries continue to face serious problems in obtaining vaccines and continue to record numerous cases. This situation highlights the need to

---

* Corresponding author.
** Corresponding author at: Laboratório de Modelagem Computacional, Instituto de Ciências Exatas, Universidade Federal de Alfenas, Brazil
E-mail addresses: groponp@gmail.com (G. Ropón-Palacios), icamps@unifal-mg.edu.br (I. Camps).

https://doi.org/10.1016/j.compbiolchem.2022.107668
Received 27 September 2021; Received in revised form 14 March 2022; Accepted 17 March 2022
Available online 23 March 2022
1476-9271/© 2022 Elsevier Ltd. All rights reserved.

discover a drug to help treat severe cases of COVID-19.

One of the most studied biological processes in SARS-CoV-2 is the recognition of the host cell-mediated by a membrane glycoprotein named Spike protein (S protein). The S protein is a trimeric class I fusion protein. Its extracellular domain contains the subunits S1 (residues 14–685) and S2 (residues 86–1273) (Li, 2016), associated with receptor binding and membrane fusion, respectively (Song et al., 2018). The S1 subunit contains the Receptor Binding Domain (RBD, residues 333–527) that interacts directly with the human Angiotensin-converting enzyme 2 (ACE2) (Monteil et al., 2020). This region also contains the Recognition Binding Motif (RBM, residues 438–506) that directly contacts ACE2 (Yang et al., 2020). In such interaction, the S protein undergoes conformational changes that allow an asymmetric rearrangement that drives one of its three RBDs to its "up" conformation. Only RBDs in this conformation can bind ACE2 (Roy et al., 2020). Since this is a vital process for the initiation of infection, a lot of research has pointed to the S protein as one of the most important therapeutic targets (Krumm et al., 2021).

The earliest studies of cryo-electron microscopy (cryo-EM) structure of the S protein in SARS-CoV-2 revealed that it is heavily glycosylated (Yao et al., 2020). Each promoter contains 22 canonical N-linked glycosylation and 3O-glycosylation sequons (Zhao et al., 2020). Glycosylation plays a pivotal role in the viral life cycle because it aids in evading the host immune system (Kasuga et al., 2021). Furthermore, glycosylations contribute to protein folding and the thermostability of the viral particle (Kawase et al., 2019). They also can be involved in viral entry by specific interactions with the host surface (Raman et al., 2016). Therefore, understanding key aspects of the molecular mechanisms of these structures, during drug recognition in SARS-CoV-2, may contribute to therapeutic alternatives.

Inhibition of the protein-protein interaction (PPI) of SARS-CoV-2-S-hACE2 has been widely studied mostly with antibodies (Guillon et al., 2008; Jiang et al., 2020). Here, the ability to disrupt the interaction is determined by the affinity of the antibodies to bind to the antigenic region of the protein (Fiedler et al., 2021). On the other hand, the use of molecules has not achieved the same success, due to the lack of a drugable or pocket cavity in the RBM domain. Moreover, conformational changes in the binding region can be difficult to target in a non-bonded state (Perez et al., 2021). Despite this, some small molecules have been reported to be potential inhibitors, due to they exhibit a high binding energy in the RBD domain (Bagdonaite and Wandall, 2018; Singh et al., 2021). This leads us to the compulsion of searching for molecules that possess a high affinity for the S protein; especially those new chemical entities that have not been explored so far, as is the case with The Pathogen Box library, a collection of drug-like compounds selected for their potential against several of the world's most important diseases (Veale, 2019).

Nowadays, computational methods have long been used in drug design and discovery (Lin et al., 2020). By applying these, molecular interactions can be described at an atomistic level (Zheng et al., 2018). These *in silico* methods are a fundamental part of the laboratory workflow, that have made it possible to transfer part of the screening process to computers. Among these, methods such as MD simulations and virtual screening and docking methods have been widely used to accelerate the drug discovery process (Zheng et al., 2018).

In addition, exploring new druggable sites in the S protein could potentially lead to inhibition of its function. The discovery of cryptic sites precisely focuses on that these forms a pocket in a holo-structure that can be identified transiently due to conformational changes (Sztain et al., 2021). These sites are generally not detected during the atomic resolution process (Beglov et al., 2018); however, new algorithms have been developed to help predict potential cryptic sites, for instance, CryptoSite (Cimermancic et al., 2016). Evaluating new druggable sites is a recent alternative being studied in SARS-CoV-2 with the aim of expanding its druggable proteome (Cavasotto et al., 2021).

Since there is a need to develop and find effective drugs to treat COVID-19, here, we evaluated the impact of glycosylation in the biophysical properties and drug recognition in RBD and N-terminal domain (NTD) of SARS-CoV-2. We have also estimated the implications of non-glycosylated systems in both. For this purpose, we have performed extensive molecular dynamics simulations and a thorough analysis of the structural changes upon drug-receptor interaction.

2. Material and methods

2.1. Data set

Information about the ligands and binding pockets in the S protein systems were taken from the protocol of Otazu et al. (2020), where it shows the importance of two small molecules, TCMDC-124223 and TCMDC-133766 (Veale, 2019). Both compounds belong to The Pathogen Box library which through molecular docking assays showed good binding energy with the S protein. About these, the first compound targeted the active site of the RBD region and the second the cryptic site of the NTD region. We build our systems from the docking modes described in this study.

2.2. MD simulation

To explore the biological role of glycosylations on small molecules, like drugs, we used all-atom molecular dynamics (MD) simulations. The systems were built on the CHARMM-GUI server (Jo et al., 2008) using a Spike glycosylated protein available on the COVID-19 Proteins archive (Woo et al., 2020). The solution builder module was used to generate the system topology on a cubic box with a padding of 1.5 nm. The TIP3P water was used to solvate the box, following ionization with sodium (Na+) and chlorine (Cl-) ions to neutralize the system at 154 mM. The CHARMM36m force field for the glycoprotein system and CHARMM General Force Field (CGenFF) for ligands were selected for the calculation of the interactions (Vanommeslaeghe et al., 2010).

MD simulations were performed in GROMACS v2021.4 (Abraham et al., 2015) in four steps. First, energy minimization using the steepest descent algorithm with 5000 steps or until reaching an energy < 10 kJ/mol/nm to eliminate bad contacts. Second, an NVT equilibrium phase at 310 K for 2 ns to equilibrate the system temperature. Third, an NPT equilibrium phase at 1 bar for 4 ns to equilibrate the system pressure. The Berendsen thermostat (Berendsen et al., 1998) and the Parrinello-Rahman barostat (Parrinello and Rahman, 1981) were used in the equilibrium phases. Fourth, a production simulation for 300 ns with integration steps of 2 fs, under constant pressure and temperature using leap-frog integration algorithm (Van Gunsteren and Berendsen, 1987). To generate the trajectories, the LINCS algorithm was used to constrain the interactions during equilibrium, while the Particle-Mesh Ewald algorithm was used to constrain the long-range ionic interactions.

2.3. Data analysis and biomolecular graphics

The trajectories were analyzed using geometric and structural properties to determine the influence of glycosylations on molecular recognition. RMSD, RMSF, FEL, β-factor, number and types of contacts were calculated with *in-house* python scripts using MDAnalysis library (Michaud-Agrawal et al., 2011). Geometric calculations of roto translation distances, angles and SASA were done using a pool of *in-house* scripts based on tcl embedded in VMD (Humphrey et al., 1996). The residues interacting with the ligand were analyzed with PLIP v2.1.6 software (Adasme et al., 2021), via a script that performs the analysis for each frame, iterates the interactions and plots the interaction percentages per residue. All images were rendered with VMD and all the animations with Molywood (Wieczór et al., 2020).

3. Results

This work explores the impact of glycosylation and ligand interactions on RBD and NTD of SARS-CoV-2. In Fig. 1, six systems built in the presence or absence of glycosylations or ligands are shown. Two systems without ligands were built to be used as controls (figura 1a). Two systems with the interaction of ligand TCMDC-124223 to RBD (Fig. 1b) and two others with the interaction of ligand TCMDC-133766 to NTD (Fig. 1c). In the figure, the systems in the top (bottom) row are in the absence (presence) of glycosylations.

3.1. Glycosylation induces important protein stability

Our RMSD analysis suggests that both RBD and NTD have a similar flexibility independent of glycosylation in the APO structure. However, the glycosylated protein presented a slightly higher RMSD, and especially in the RBD, a conformational change of ~2 Å after 250 ns was observed (Fig. 2a). In terms of the motifs of each region, they maintain a similar flexibility in both states, which can be distinguished in the heat map (Fig. 2b).

Simulations with the ligand indicated that glycosylations have a significant effect on the interaction. As can be seen in Fig. 1a (Holo), the glycosylated RBD presented an RMSD of up to ~4 Å, while in the absence of glycosylations the RMSD increases up to ~5 Å. In addition, we note that the glycosylated structure stabilizes early, at 50 ns, while the other stabilizes after 200 ns. Here, we also note that the ligand has a better fit in the RBM when glycosylations are present, being observed to integrate favorably in the loop region between residues T470-F490 (Fig. 2c up). This also affect the flexibility of the loop, tending to have less RMSD.

On the other hand, the analysis of NTD reveals that glycosylations induce modifications in the stability of the Holo structure. Here, we observed that the protein without glycosylations promotes major flexibility in the R246-G260 residues. In contrast, the presence of glycosylations tends to stabilize this region in the NTD. This effect is in agreement with the lower RMSD in this region (Fig. 1a and Fig. 2c down).

3.2. Rigid/flexibility changes in RBD and NTD by glycosylation

To evaluate changes in the rigid/flexibility of protein structure, RMSF analysis of the backbone was performed. The glycosylated and non-glycosylated NTD structures without ligand (Apo) did no presented differences about flexibility, with maximum RMSF values of ~50 Å. Interestingly, we observe that the motifs in the NTD are much rigid only when they are close to a glycosylation (Fig. 3b, down). In contrast, the RBD region tends to be much more flexible across the domain, even with a single glycosylation (Fig. 3b, up).

Besides, we have been able to corroborate modifications in the flexibility of each domain due to the interaction with each ligand. In the RBD domain, the Holo structure is much more rigid compared to the Apo structure, however the absence of glycosylations stabilizes the RBD domain more. On the other hand, our analysis of the NTD domain shows that the non-glycosylated Holo state modifies the flexibility of the whole protein compared to the Apo; whereas, the glycosylated Holo state causes an opposite effect, since we observed lower RMSF in the whole protein. This is partly due to an induced adjustment effect that compacts the whole domain (Fig. 3a, down and Fig. 3c, down).

3.3. Glycosylation changes roto-translation to the fix binding mode

The molecular dynamics of the RBD structures revealed that the ligand undergoes through different conformational states promoted by roto-translational phenomena in the RBM region in the absence of

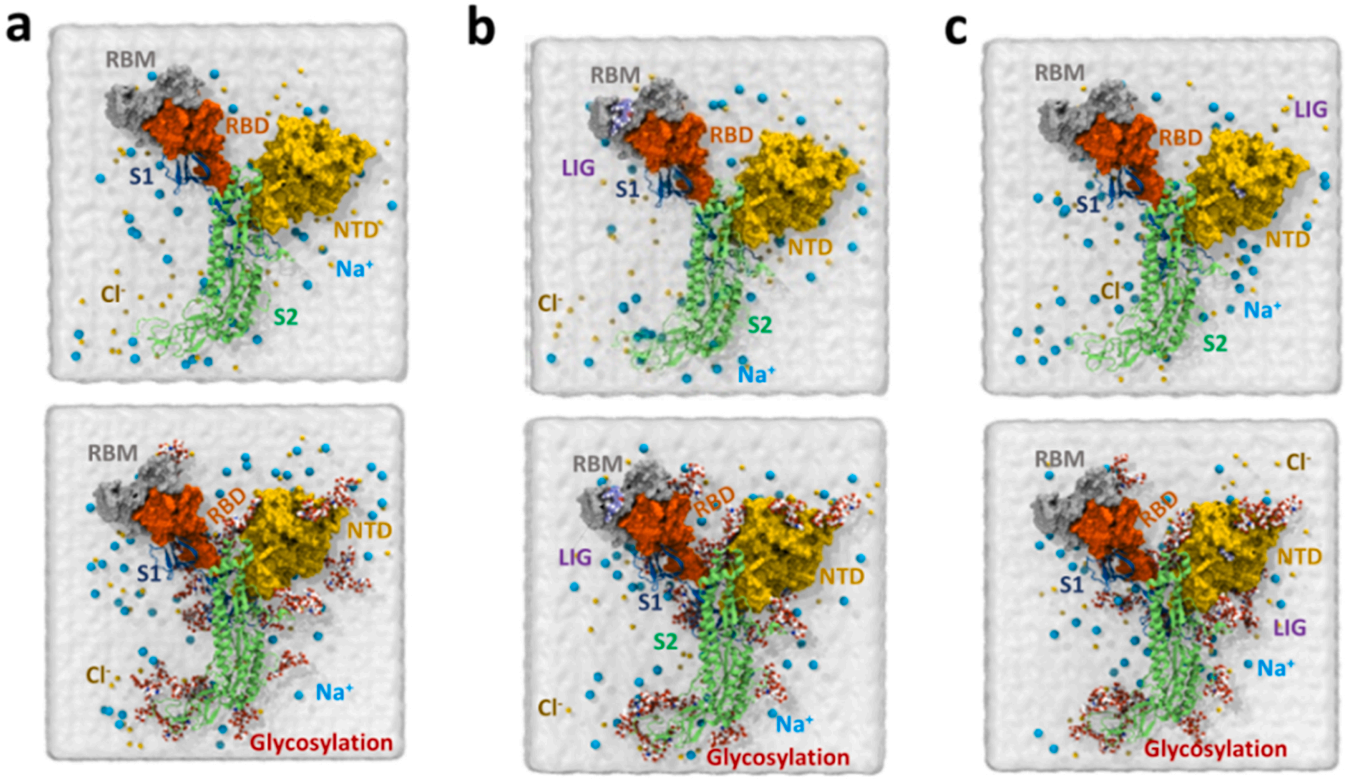


Fig. 1. System setup for molecular dynamics simulation. (a) Representation of S protein without interaction of ligands in RBD or NTD in absence (up) or presence (down) of glycosylations. All systems contain ions of Na^+ (skyblue) and Cl^- (yellow), while the water is represented as the surface in white color (b) Representation of S protein with interaction of TSMDC-124223 in RBD in absence (up) and presence (down) of glycosylations. (c) Representation of S protein with interaction of TCMDC-133766 in NTD in absence (up) or presence (down) of glycosylations.

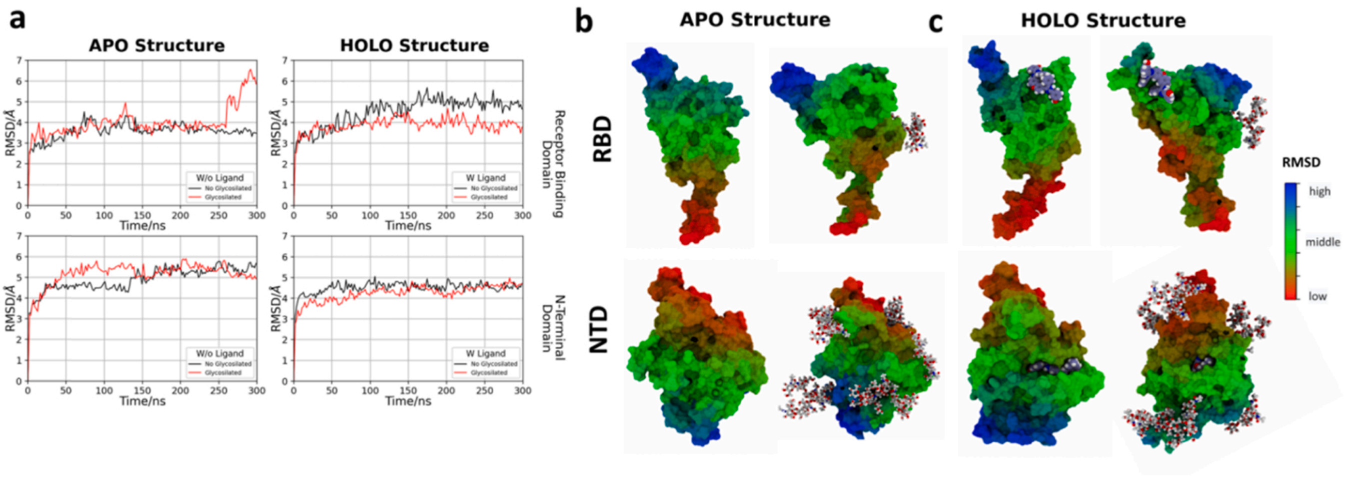


Fig. 2. Conformational stability induced by glycosylations. (a) Comparison of structural changes of RBD and NTD in the Apo and Holo structure, in presence and absence of glycosylations. (b) Local structural changes in RBD (up) and NTD (down) in Apo structure; left structure without glycosylation and right with glycosylation. The major conformational changes are observed in the loop region between residues T470-F490 of the RBD (blue site); while in the NTD domain, residues R246-G260 from the most unstable region (blue site). This dynamics was observed for both, with and without glycosylations, with a slight increase in instability in the glycosylated structure. (c) Changes in the stability of the loop region between residues T470-F490 of the RBD are observed in the ligand-bound and glycosylated structure (right), while the same region with ligand but without glycosylation shows no significant differences in stability compared to the APO structure. The same dynamics is observed in the NTD domain, where the R246-G260 region is much stable in the glycosylated state.

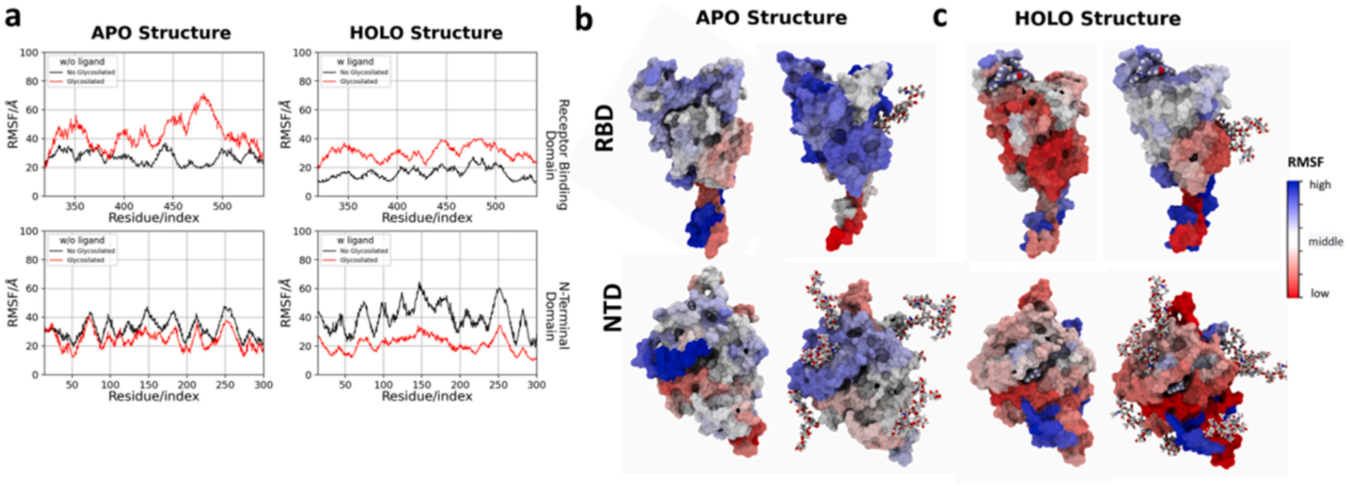


Fig. 3. Local changes in rigid/flexibility induced by glycosylations. (a) Comparison of rigid/flexibility on backbone of RBD and NTD with (Holo structure) or without (Apo structure) ligand interaction in presence or absence of glycosylations. (b) Rigid/flexibility changes in RBD with ligand interaction; left structure without glycosylation and right with glycosylation. (c) Rigid/flexibility changes in NTD with ligand interaction; left structure without glycosylation and right with glycosylation. (d) Rigid/flexibility changes in RBD and NTD without ligand interaction; left structure without glycosylation and right with glycosylation. Hue changes shown in the color bar indicate minimal (red), intermediate (white), or high (blue) flexibility.

glycosylations. These changes are observed from early stages of molecular dynamics, followed by a perpendicular rotation at 29.7 ns and translation until 60 ns, finally the ligand remains at the opposite end of RBM until the end of the simulation (Fig. 4a,d). In contrast, glycosylation induces a sustained ligand interaction in loop/β structure formed by residues T470-F490, which promote less conformational changes and rotations until the end of simulation (Fig. 4b,d).

To determine the characteristics of roto-translation phenomenon, the distance of central benzofuranic structure of the ligand to the RBM residue G446 was calculated, while for angle rotations the RBM residue Q493 was used as the origin to generate angles θ with respect to the central structure of the ligand. Then, the generation of metastable states was evaluated through free energy landscape (FEL) matrices (Fig. 4c,g, h). The displacement to 10 Å of the G446 residue of the ligand in the non-glycosylated structure at 20 ns which was maintained throughout the simulation was evidenced, this is related to the important rotational changes generated in the first ns of the simulation and the generation of metastable states from 18 Å to 10 Å at θ angles from 80° to 150° (figure e-g). In contrast, in the glycosylated structure the ligand remains approximately 22 Å from the G446 residue for a longer time, this is related to the generation of metastable states at 22 Å and θ angle between 100° and 140°, then a shift is observed before 30 ns and at the end of the simulation, this shift occurs in parallel to the rotational changes in time (figure e, f, h). The roto-translation phenomenon can be viewed in the supplementary videos (S1_movie_1_v2).

3.4. *Glycosylation promotes recognition by induced fit*

The molecular dynamics of the glycosylated and non-glycosylated NTD structures revealed that ligand remains inside the cryptic pocket,

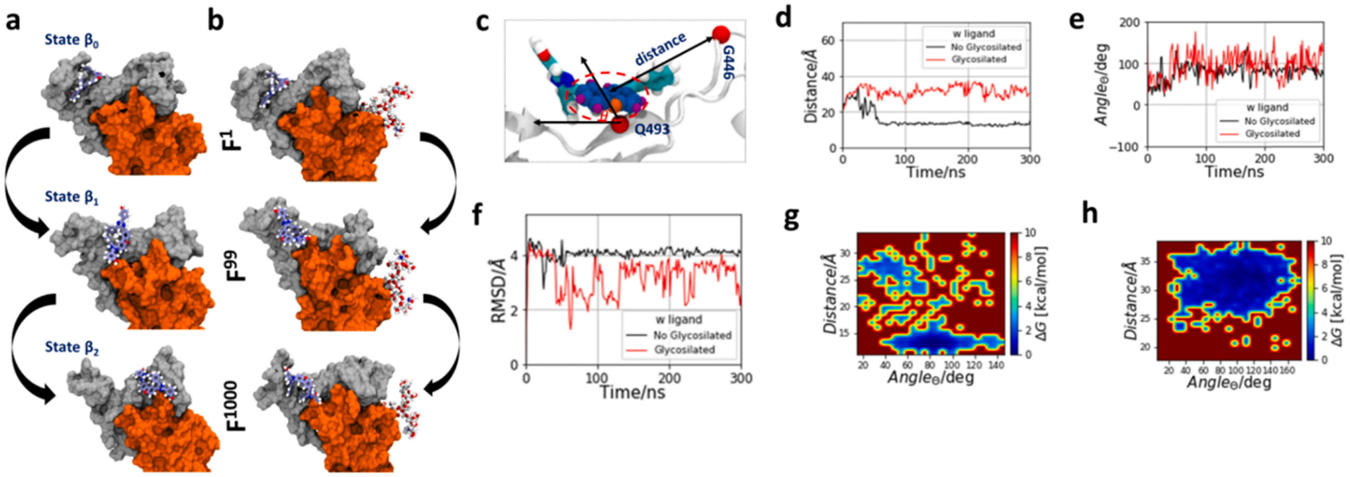


Fig. 4. Interaction of ligand TCMDC-124223 on RBM. Evolution of ligand conformational states along simulation. Gray structure represents RBM and complete gray and orange structure represents RBD. The β_0 state was the initial ligand conformation, the β_1 state was an intermediated conformation of ligand and β_2 state was the final conformation of ligand in the simulation. (a) Ligand conformation β_0 at frame 1 (0.3 ns), β_1 at frame 99 (29.7 ns), and β_2 at frame 1000 (300 ns). An effect of roto-translation of ligand is evident in absence of glycosylations. (b) Ligand conformation β_0 at frame 1 (0.3 ns), β_1 at frame 99 (29.7 ns), and β_2 at frame 1000 (300 ns), system in presence of glycosylations. The ligand is more stable in the presence of glycosylations. (c) Ligand distance to G466 and θ angle formed in the roto-translation. (d) Translation changes of ligand in presence and absence of glycosylations. (e) Rotational changes of ligand in presence and absence of glycosylations. (f) Conformational changes of ligand in presence and absence of glycosylations. (g) Free energy landscape (FEL) matrix of ligand with different θ angles and distance of G466 in absence of glycosylations. (h) FEL matrix of ligand with different θ angles and distance of G466 in presence of glycosylations. Hue changes shown in the color.

throughout the simulation. In this sense, we observed that the glycosylated structure promotes the induced fit, encapsulating the ligand at the end of the simulation. In contrast, the non-glycosylated structure does not present this phenomenon since it maintains the structure of the cryptic pocket without significant changes (Fig. 5a,b). To determine the characteristics of the ligand in the induced fitting process we evaluated the RMSD of the ligand, the solvent accessible surface areas (SASA), and the generation of metastable states through FEL matrices at a temperature of 310 K between the number of contacts and the SASA.

In the non-glycosylated structure, the ligand presented a progressive reduction of the RMSD up to 100 ns, after which it is more stable. On the other hand, the glycosylated structure presented marked fluctuations in the RMSD throughout the simulation, due to the effect of the adjustment induced in this cryptic site (Fig. 5c). Likewise, the SASA values were higher in the non-glycosylated structure from 140 ns of simulation, while in the glycosylated structure the SASA was lower due to the effect of the induced adjustment (Fig. 5d). This process of encapsulation by the induced fit significantly restricts the number of ligand contacts from 20 ns to the end of the simulation (Fig. 5e). On the contrary, the non-glycosylated structure maintains the ligand without significant variations until 40 ns where it reaches another stable conformation. Related to this, the induced fit phenomenon is not evidenced since the SASA

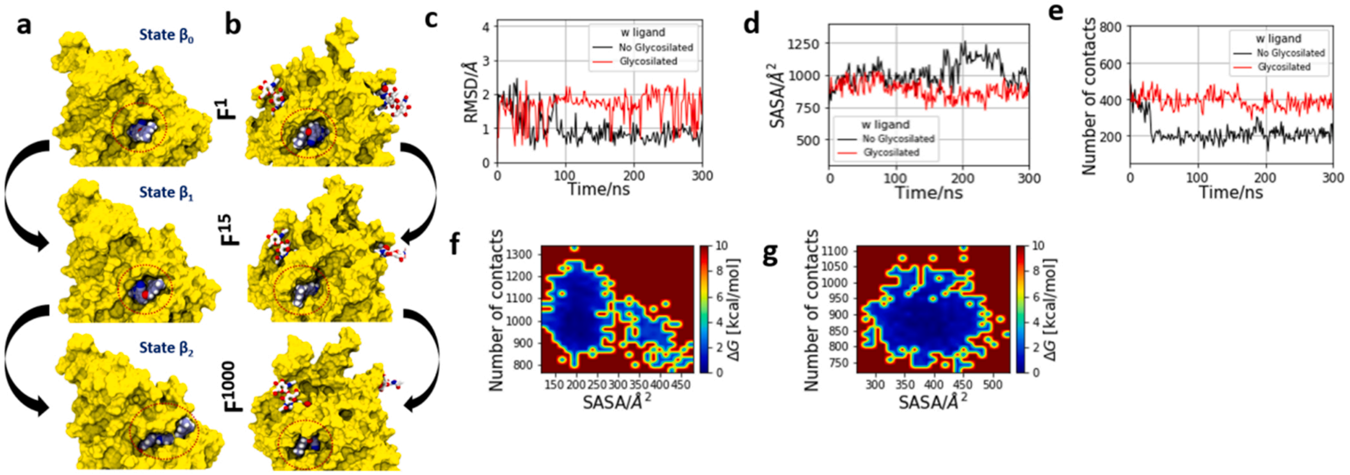


Fig. 5. Interaction of ligand TCMDC-133766 on the cryptic pocket of NTD. Evolution of ligand conformational states along simulation; yellow structure represents the NTD of S protein. The β_0 state was the initial ligand conformation, the β_1 state was an intermediated conformation of ligand and β_2 state was the final conformation of ligand in the simulation (a) Evolution of ligand interaction in the cryptic pocket at frames 1 (0.3 ns), 15 (4.5 ns), and 1000 (300 ns) in absence of glycosylations. (b) Evolution of ligand interaction in the cryptic pocket at frames 1 (0.3 ns), 15 (4.5 ns), and 1000 (300 ns). An induced fit of ligand is observed in presence of glycosylations. (c) Conformational changes of ligand in presence and absence of glycosylations. (d) Solvent accessible surface areas (SASA) changes of ligand in presence and absence of glycosylations. (e) Changes in the number of protein contacts of ligand in presence and absence of glycosylations. (f) Free energy landscape (FEL) matrix of ligand with different contacts of protein and SASA in absence of glycosylations. (g) FEL matrix of ligand with different contacts of protein and SASA in presence of glycosylations. Hue changes shown in the color bar indicate minimal (red), intermediate (green), or high (blue) probability of metastable states in FEL matrix.

presents slight variations increasing the solvent area that reaches high values until the end of the simulation. This process generates an increase in the degrees of freedom, evidenced by the increase in the number of contacts. Additionally, the glycosylated structure generates fewer metastable states than the non-glycosylated structure, due to the confinement by the induced fit, unlike the non-glycosylated structure that presents dispersed metastable states due to the higher SASA and freedom of contacts (Fig. 5g,f). The induced fit phenomenon can be reviewed in the supplementary videos (S2_movie_2_v2).

3.5. Glycosylation modify the pattern of molecular interactions

Clear modifications in the interaction patterns due to the presence of glycosylations have been observed (Fig. 6). In the RBD, TCMDC-124223 interacts mainly with 4 residues (R403, D405, S494, and G504) via hydrogen bonds, while 4 other residues form hydrophobic interactions (Y505, Y489, Y449, and Q493). In the presence of glycosylations, the ligand in RBD forms major hydrophobic contacts with 7 residues (Y489, Y473, F486, N487, L455, F456, and A475) and hydrogen bonds only with 2 residues (Y473 and N487). Besides, TCMDC-133766 presented hydrophobic interactions mostly in the NTD. In the non-glycosylated structure, only 9 interactions were observed (W104, F194, F192, F168, L229, I231, I128, I119, and F92); whereas, in the glycosylated NTD, 11 mostly hydrophobic interactions were observed (W104, T240, I101, F92, F192, F175, V126, Q173, L226, F194, and F106) which are related to the induced fit effect.

4. Discussion

In this work, we employed molecular dynamics simulations to study the effects of glycosylations on biophysical properties and drug recognition upon RBD and NTD of SARS-CoV-2. Our results confirmed the hypothesis that glycosylations introduce important structural stability, local changes in rigid/flexibility, and enhance interactions of drugs on RBD and NTD.

The analysis showed that, independently of the glycosylations, the APO state does not undergo significant structural modifications, both in the RBD and in the NTD of the spike protein. We emphasize that the most flexible regions of these domains are found between the T470-F490 loop/β of RBD. Such flexibility was to be expected since this site belongs to the RBM subdomain that directly contacts the human ACE2 receptor (Roy et al., 2020). In contrast, our observations suggest that glycosylations significantly modify the stability of the protein in a Holo state. The improved binding of the TCMDC-124223 ligand to the RBD may be due to a long-range effect exerted by glycosylation at N343. This glycosylation is of special interest since it has been reported to be responsible for the conformational change from the "down" to "up" state (Tian et al., 2021). In this sense, it is consistent to expect that the presence of small molecules will have far-reaching effects of glycosylations in this system. Likewise, the effects caused by glycosylation during drug-receptor interaction are reflected in the fluctuations of the RMSD. Thus, we can attribute the difference between the RMSD values observed in the RBD domain to this effect.

Furthermore, RMSF analysis showed an increase of superior flexibility in the drug binding region on the RBD glycosylated. Whereas, multiple glycosylations of NTD conferred greater structural rigidity. Studies on bovine alpha-chymotrypsin reveal that the increase of the molar mass and amount of glycans in the protein structure correlates with the degree of restriction of the structure dynamics (Solá and Griebenow, 2006). These changes can be explained by the local steric constraints that O- and N-glycans generate in different protein systems (Sola and Griebenow, 2009; Watanabe et al., 2004). For example, the study by Lee et al. showed that N-glycosylations do not induce significant global and local structural changes in different mammalian proteins, this was attributed to reduced structural rearrangement and thus increased stability (Lee et al., 2015). However, changes in rigidity/flexibility are apparently influenced by the structural complexity of the glycosylated proteins, since in some systems it is significant while in others it is not transcendent (Sarkar and Wintrode, 2011; Weiß et al., 2021). The latter may be the case for the RBD domain, which needs to possess characteristic flexibility to accomplish its biological function. Therefore, we consider it necessary to evaluate the effect of O- and N-glycans in each particular system, since there may be greater or lesser effects depending on the type of system.

In addition, our simulations suggest a rotation phenomenon on the RBM as a characteristic ligand interaction pattern in the presence of RBD glycosylations. The local structural flexibility of the pocket plays a key role in allowing or restricting this phenomenon by modifying the ligand entropy (Chang and Gilson, 2004). Therefore, the increased amount of hydrophobic interactions of the ligand with the flexible binding sites of RBD may limit its translational freedom, but allow its rotation, leading to a longer interaction time in the T470-F490 loop/β-loop of RBD. On

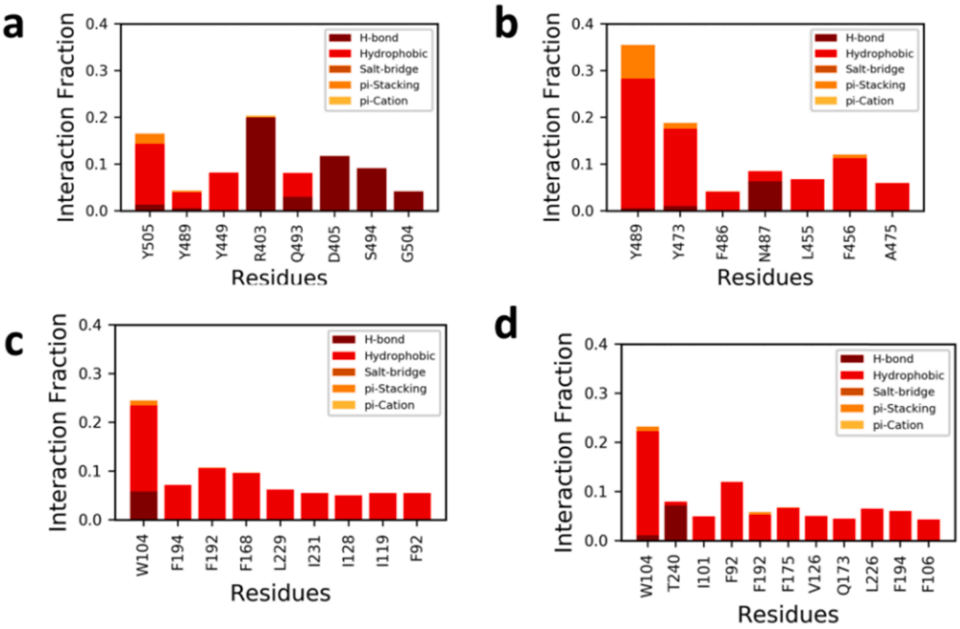


Fig. 6. Ligand-interaction analysis. Interactions appearing with more than 4% frequency are shown for both ligands in the RBD and NTD with and without glycosylations. (a) The TCMDC-124223 ligand interacts preferentially by hydrophobic interactions and hydrogen bridges in the absence of glycosylations. (b) In the presence of glycosylations, a completely different interaction pattern is observed where hydrophobic interactions predominate. (c) The ligand TCMDC-133766 interacts mainly by hydrophobic interactions with 9 NTD residues. (d) Hydrophobic interactions are observed with 11 residues corresponding to the induced fitting effect in the glycosylated structure.

the other hand, the number of glycosylations in RBD may explain the changes in local flexibility, since a higher degree of glycosylations reduces the free energy and structural fluctuations in various folded state proteins (Shental-Bechor and Levy, 2009). However, stable ligand interactions from a physicochemical point of view depend on other factors such as hydrogen potential, temperature of the system, number and types of bonds formed (Majewski et al., 2019).

The structural changes observed during the simulation confirm the presence of an induced fit phenomenon in NTD, generated by glycosylations in the presence of a ligand. This is due to the fact that the presence of sugars induces changes that trigger an induced adjustment of the ligand towards the inner part of the NTD domain. The adaptation of the ligand in the induced fit requires prior molecular matching, sufficient affinity and rapid kinetic equilibrium (Nussinov et al., 2014). In our simulations, the ligand TCMDC-133766 exhibited optimal affinity characteristics, and the FEL suggests a single metastable energy state with a high number of mostly hydrophobic contacts. We attribute this phenomenon to the changes in rigidity/flexibility when forming the glycosylated complex with ligand and confirmed by SASA and contacts analysis. However, the degree of glycosylations could also suggest the presence of a possible allosteric effect. This phenomenon has been reported in glycoproteins such as acid-β-glucosidase, a lysosomal hydrolase, where N-glycosylations in a region distant from the active site generated conformations with catalytic residues (E430 and E235) more than 10 Å from the substrate (Souffrant et al., 2020). On the other hand, there are systems, such as G protein, where the allosteric phenomenon can occur even in the absence of ligands (Renault et al., 2019). The results of our work do not allow us to affirm the presence of allosterism induced by glycosylations in S protein, for which further replicate simulations are required.

Glycosylations of viral envelopes serve a wide range of functions, including regulation of viral tropism, host immunity, and protein stability (Bagdonaite and Wandall, 2018). Our study shows that glycosylations are important components for the biophysical properties of SARS-CoV-2 protein S such as NTD dynamics promoting the up state of RBD in different types of coronaviruses including the current SARS-CoV-2 (Qing et al., 2021) or the flexibility of the T470-F490 loop/β-loop important for the affinity properties of SARS-CoV-2 RBD to human ACE2 (Reis et al., 2021). Moreover, glycosylations have implications for drug recognition in the RBD and NTD domain, so future virtual screening and molecular dynamics should consider N- and O-glycosylations for more accurate results on ligand interaction on drug targets to inhibit SARS-CoV-2 infection. The absence of these structural features may lead to misinterpretations.

5. Conclusion

We conclude that non-protein structural components such as glycosylations play an important role in the biophysical structural properties in S protein of SARS-CoV-2. These biophysical changes enhance drug recognition in RBD and NTD, key structures for the biological properties of SARS-CoV-2, therefore the development of new drugs.

Our main findings were the induction of structural stability and local changes in rigidity/flexibility related to the number of glycosylations in RBD and NTD. These structural changes are important for its biological activity and drug recognition, evidenced by roto-translation phenomenon in the interaction of the ligand with RBD in the absence of glycosylation and the induced fit phenomenon of ligand in NTD in presence of glycosylations. Therefore, glycosylations must be placed into account for rational drug development, virtual screening, and molecular dynamics targeting S protein.

CRediT authorship contribution statement

Georcki Ropón-Palacios: Conceptualization, Methodology, Software,Data analysis, Validation, Writing- Original draft preparation, Writing- Reviewing and Editing, Supervision. **Jhon Pérez-Silva:** Methodology, Software, Validation, Writing- Original draft preparation. **Ricardo Rojas-Humpire:** Methodology, Software, Validation, Writing- Original draft preparation, Writing- Reviewing and Editing. **Gustavo E. Olivos-Ramírez:** Methodology, Software, Validation, Writing- Original draft preparation, Writing- Reviewing and Editing. **Manuel Chenet-Zuta:** Data analysis. **Victor Cornejo-Villanueva:** Data curation. **Sheyla Carmen-Sifuentes:** Data curation. **Kewin Otazu:** Data curation. **Yaritza L. Ramirez-Díaz:** Data curation, Writing- Reviewing and Editing. **Karolyn Vega Chozo:** Data curation, Writing- Reviewing and Editing. **Ihosvany Camps:** Conceptualization, Supervision.

Supplementary material

Movies of molecular dynamics trajectories made with VMD can be found in https://doi.org/10.5281/zenodo.6326207.

Acknowledge

Part of the results presented here were developed with the help of CENAPAD-SP (Centro Nacional de Processamento de Alto Desempenho em São Paulo) grant UNICAMP/FINEP-MCT, CENAPAD-UFC (Centro Nacional de Processamento de Alto Desempenho, at Universidade Federal do Ceará) and Compute Canada Database (CCDB).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.compbiolchem.2022.107668.

References

Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B., Lindahl, E., 2015. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1–2, 19–25. https://doi.org/10.1016/J.SOFTX.2015.06.001.

Adasme, M.F., Linnemann, K.L., Bolz, S.N., Kaiser, F., Salentin, S., Haupt, V.J., Schroeder, M., 2021. PLIP 2021: expanding the scope of the protein–ligand interaction profiler to DNA and RNA. Nucleic Acids Res. 49 (W1), W530–W534. https://doi.org/10.1093/NAR/GKAB294.

Bagdonaite, I., Wandall, H.H., 2018. Global aspects of viral glycosylation. Glycobiology 28 (7), 443–467. https://doi.org/10.1093/GLYCOB/CWY021.

Beglov, D., Hall, D.R., Wakefield, A.E., Luo, L., Allen, K.N., Kozakov, D., Whitty, A., Vajda, S., 2018. Exploring the structural origins of cryptic sites on proteins. Proc. Natl. Acad. Sci. USA 115 (15), E3416–E3425. https://doi.org/10.1073/PNAS.1711490115.

Bitar, G., Badreddine, A., 2020. N-glycosylation and gaucher disease mutation allosterically alter active-site dynamics of acid-β-glucosidase. ACS Catal. 10 (3), 1810–1820. https://doi.org/10.1021/ACSCATAL.9B04404/SUPPL_FILE/CS9B04404_SI_001.PDF.

Cavasotto, C.N., Lamas, M.S., Maggini, J., 2021. Functional and druggability analysis of the SARS-CoV-2 proteome. Eur. J. Pharmacol. 890, 890. https://doi.org/10.1016/J.EJPHAR.2020.173705.

Chang, C.E., Gilson, M.K., 2004. Free energy, entropy, and induced fit in host-guest recognition: calculations with the second-generation mining minima algorithm. J. Am. Chem. Soc. 126 (40), 13156–13164. https://doi.org/10.1021/JA047115D/SUPPL_FILE/JA047115DSI20040524_062809.PDF.

Cimermancic, P., Weinkam, P., Rettenmaier, T.J., Bichmann, L., Keedy, D.A., Woldeyes, R.A., Schneidman-Duhovny, D., Demerdash, O.N., Mitchell, J.C., Wells, J.A., Fraser, J.S., Sali, A., 2016. CryptoSite: expanding the druggable proteome by characterization and prediction of cryptic binding sites. J. Mol. Biol. 428 (4), 709–719. https://doi.org/10.1016/J.JMB.2016.01.029.

Dong, E., Du, H., Gardner, L., 2020. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect. Dis. 20 (5), 533–534. https://doi.org/10.1016/S1473-3099(20)30120-1/ATTACHMENT/9DE52FFD-61D4-4C3F-826F-841D978D469D/MMC1.PDF.

Fiedler, S., Piziorska, M.A., Denninger, V., Morgunov, A.S., Ilsley, A., Malik, A.Y., Schneider, M.M., Devenish, S., Meisl, G., Kosmoliaptsis, V., Aguzzi, A., Fiegler, H., Knowles, T., 2021. Antibody affinity governs the inhibition of SARS-CoV-2 spike/ ACE2 binding in patient serum. ACS Infect. Dis. 7 (8), 2362–2369. https://doi.org/10.1021/ACSINFECDIS.1C00047/SUPPL_FILE/ID1C00047_SI_001.PDF.

Guillon, P., Clément, M., Sébille, V., Rivain, J.G., Chou, C.F., Ruvoën-Clouet, N., Le Pendu, J., 2008. Inhibition of the interaction between the SARS-CoV spike protein and its cellular receptor by anti-histo-blood group antibodies. Glycobiology 18 (12), 1085–1093. https://doi.org/10.1093/GLYCOB/CWN093.

Humphrey, W., Dalke, A., Schulten, K., 1996. VMD: visual molecular dynamics. J. Mol. Graph. 14 (1), 33–38. https://doi.org/10.1016/0263-7855(96)00018-5.

Ibrahim, N.K., 2020. Epidemiologic surveillance for controlling Covid-19 pandemic: types, challenges and implications. J. Infect. Public Health 13 (11), 1630–1638. https://doi.org/10.1016/J.JIPH.2020.07.019.

Jiang, S., Zhang, X. and Du, L. (2020) Therapeutic antibodies and fusion inhibitors targeting the spike protein of SARS-CoV-2, 'Therapeutic antibodies and fusion inhibitors targeting the spike protein of SARS-CoV-2′, https://doi.org/10.1080/14728222.2020.1820482, 25(6), pp. 415–421. doi: 10.1080/14728222.2020.1820482.

Jo, S., Kim, T., Iyer, V.G., Im, W., 2008. CHARMM-GUI: a web-based graphical user interface for CHARMM. J. Comput. Chem. 29 (11), 1859–1865. https://doi.org/10.1002/JCC.20945.

Kasuga, Y., Zhu, B., Jang, K.J., Yoo, J.S., 2021. Innate immune sensing of coronavirus and viral evasion strategies. Exp. Mol. Med. 53 (5), 723–736. https://doi.org/10.1038/s12276-021-00602-1.

Kawase, M., Kataoka, M., Shirato, K., Matsuyama, S., 2019. Biochemical analysis of coronavirus spike glycoprotein conformational intermediates during membrane fusion. J. Virol. 93 (19) https://doi.org/10.1128/JVI.00785-19/ASSET/43E66A52-6DB5-4D39-847E-D935E5988851/ASSETS/GRAPHIC/JVI.00785-19-F0011.JPEG.

Krumm, Z.A., Lloyd, G.M., Francis, C.P., Nasif, L.H., Mitchell, D.A., Golde, T.E., Giasson, B.I., Xia, Y., 2021. Precision therapeutic targets for COVID-19. Virol. J. 18 (1), 1–22. https://doi.org/10.1186/S12985-021-01526-Y.

Lee, H.S., Qi, Y., Im, W., 2015. Effects of N-glycosylation on protein conformation and dynamics: protein Data Bank analysis and molecular dynamics simulation study. Sci. Rep. 5 (1), 1–7. https://doi.org/10.1038/srep08926.

Li, F., 2016. Structure, Function, and Evolution of Coronavirus Spike Proteins, 'Structure, Function, and Evolution of Coronavirus Spike Proteins', https://doi.org/10.1146/annurev-virology-110615–042301, 3, pp. 237–261. doi: 10.1146/ANNUREV-VIROLOGY-110615–042301.

Lin, Xiaoqian, Li, X., Lin, Xubo, 2020. A review on applications of computational methods in drug screening and design. Molecules. https://doi.org/10.3390/MOLECULES25061375.

Majewski, M., Ruiz-Carmona, S., Barril, X., 2019. An investigation of structural stability in protein-ligand complexes reveals the balance between order and disorder. Commun. Chem. 2 (1), 1–8. https://doi.org/10.1038/s42004-019-0205-5.

Michaud-Agrawal, N., Denning, E.J., Woolf, T.B., Beckstein, O., 2011. MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. J. Comput. Chem. 32 (10), 2319–2327. https://doi.org/10.1002/JCC.21787.

Miller, A., Carchman, R., Long, R., Denslow, S.A., 1998. Molecular dynamics with coupling to an external bath. J. Chem. Phys. 81 (8) https://doi.org/10.1063/1.448118, 3684-42.

Monteil, V., Kwon, H., Prado, P., Hagelkrüys, A., Wimmer, R.A., Stahl, M., Leopoldi, A., Garreta, E., Hurtado Del Pozo, C., Prosper, F., Romero, J.P., Wirnsberger, G., Zhang, H., Slutsky, A.S., Conder, R., Montserrat, N., Mirazimi, A., Penninger, J.M., 2020. Inhibition of SARS-CoV-2 infections in engineered human tissues using clinical-grade soluble human ACE2. Cell 181 (4), 905–913.e7. https://doi.org/10.1016/J.CELL.2020.04.004.

Nussinov, R., Ma, B., Tsai, C.J., 2014. Multiple conformational selection and induced fit events take place in allosteric propagation. Biophys. Chem. 186, 22–30. https://doi.org/10.1016/J.BPC.2013.10.002.

Otazu, K. et al., 2020. 'Targeting Receptor Binding Domain and Cryptic Pocket of Spike glycoprotein from SARS-CoV-2 by biomolecular modeling'. doi: 10.48550/arxiv.2006.06452.

Pal, M., Berhanu, G., Desalegn, C., Kandi, V., 2020. Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2): an update. Cureus 12 (3). https://doi.org/10.7759/CUREUS.7423.

Parrinello, M., Rahman, A., 1981. Polymorphic transitions in single crystals: a new molecular dynamics method. J. Appl. Phys. 52 (12) https://doi.org/10.1063/1.328693, 7182-42.

Perez, J.J., Perez, R.A., Perez, A., 2021. Computational modeling as a tool to investigate PPI: from drug design to tissue engineering. Front. Mol. Biosci. 8, 451. https://doi.org/10.3389/FMOLB.2021.681617/BIBTEX.

Qing, E., Kicmal, T., Kumar, B., Hawkins, G.M., Timm, E., Perlman, S., Gallagher, T., 2021. Dynamics of sars-cov-2 spike proteins in cell entry: Control elements in the amino-terminal domains. mBio 12 (4), 0159021. https://doi.org/10.1128/MBIO.01590-21/SUPPL_FILE/MBIO.01590-21-SF005.EPS.

Raman, R., Tharakaraman, K., Sasisekharan, V., Sasisekharan, R., 2016. Glycan-protein interactions in viral pathogenesis. Curr. Opin. Struct. Biol. 40, 153–162. https://doi.org/10.1016/J.SBI.2016.10.003.

Randall, T. et al., 2021. More Than 10.8 Billion Shots Given: Covid-19 Vaccine Tracker. Available at: ⟨https://www.bloomberg.com/graphics/covid-vaccine-tracker-global-distribution/⟩ (Accessed: 3 March 2022).

Reis, C.A., Tauber, R., Blanchard, V., 2021. Glycosylation is a key in SARS-CoV-2 infection. J. Mol. Med. 99 (8), 1023–1031. https://doi.org/10.1007/S00109-021-02092-0.

Renault, P., Louet, M., Marie, J., Labesse, G., Floquet, N., 2019. Molecular dynamics simulations of the allosteric modulation of the adenosine A2a receptor by a mini-G protein. Sci. Rep. 9 (1), 1–12. https://doi.org/10.1038/s41598-019-41980-x.

Roy, S., Jaiswar, A., Sarkar, R., 2020. Dynamic asymmetry exposes 2019-nCoV prefusion spike. J. Phys. Chem. Lett. 11 (17), 7021–7027. https://doi.org/10.1021/ACS.JPCLETT.0C01431/SUPPL_FILE/JZ0C01431_SI_003.MP4.

Sarkar, A., Wintrode, P.L., 2011. Effects of glycosylation on the stability and flexibility of a metastable protein: the human serpin α1-antitrypsin. Int. J. Mass Spectrom. 302 (1–3), 69–75. https://doi.org/10.1016/J.IJMS.2010.08.003.

Shental-Bechor, D., Levy, Y., 2009. Folding of glycoproteins: toward understanding the biophysics of the glycosylation code. Curr. Opin. Struct. Biol. 19 (5), 524–533. https://doi.org/10.1016/J.SBI.2009.07.002.

Singh, R., Bhardwaj, V.K., Sharma, J., Kumar, D., Purohit, R., 2021. Identification of potential plant bioactive as SARS-CoV-2 Spike protein and human ACE2 fusion inhibitors. Comput. Biol. Med. 136, 136. https://doi.org/10.1016/J.COMPBIOMED.2021.104631.

Sola, R.J., Griebenow, K., 2009. Effects of glycosylation on the stability of protein pharmaceuticals. J. Pharm. Sci. 98 (4), 1223–1245. https://doi.org/10.1002/JPS.21504.

Solá, R.J., Griebenow, K., 2006. Chemical glycosylation: new insights on the interrelation between protein structural mobility, thermodynamic stability, and catalysis. FEBS Lett. 580 (6), 1685–1690. https://doi.org/10.1016/J.FEBSLET.2006.02.001.

Song, W., Gui, M., Wang, X., Xiang, Y., 2018. Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. PLoS Pathog. 14 (8), e1007236 https://doi.org/10.1371/JOURNAL.PPAT.1007236.

Sztain, T., Amaro, R., McCammon, J.A., 2021. Elucidation of cryptic and allosteric pockets within the SARS-CoV-2 main protease. J. Chem. Inf. Model. 61 (7), 3495–3501. https://doi.org/10.1021/ACS.JCIM.1C00140.

Tian, Y., Parsons, L.M., Jankowska, E., Cipollo, J.F., 2021. Site-specific glycosylation patterns of the SARS-CoV-2 spike protein derived from recombinant protein and viral WA1 and D614G strains. Front. Chem. 9, 943. https://doi.org/10.3389/FCHEM.2021.767448/BIBTEX.

Van Gunsteren, W.F., Berendsen, H.J.C., 1987. A leap-frog algorithm for stochastic dynamics, 1 (3), 173–185. https://doi.org/10.1080/08927028808080941.

Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I., Mackerell Jr., A.D., 2010. CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J. Comput. Chem. 31 (4), 671–690. https://doi.org/10.1002/JCC.21367.

Veale, C.G.L., 2019. Unpacking the pathogen box—an open source tool for fighting neglected tropical disease. ChemMedChem 14 (4), 386–453. https://doi.org/10.1002/CMDC.201800755.

Watanabe, I., Zhu, J., Recio-Pinto, E., Thornhill, W.B., 2004. Glycosylation affects the protein stability and cell surface expression of Kv1.4 but Not Kv1.1 potassium channels: a pore region determinant dictates the effect of glycosylation on trafficking. J. Biol. Chem. 279 (10), 8879–8885. https://doi.org/10.1074/JBC.M309802200.

Weiß, R.G., Losfeld, M.E., Aebi, M., Riniker, S., 2021. N-glycosylation enhances conformational flexibility of protein disulfide isomerase revealed by microsecond molecular dynamics and Markov state modeling. J. Phys. Chem. B 125 (33), 9467–9479. https://doi.org/10.1021/ACS.JPCB.1C04279.

Wieczór, M., Hospital, A., Bayarri, G., Czub, J., Orozco, M., 2020. Molywood: streamlining the design and rendering of molecular movies. Bioinformatics 36 (17), 4660–4661. https://doi.org/10.1093/BIOINFORMATICS/BTAA584.

Woo, H., Park, S.J., Choi, Y.K., Park, T., Tanveer, M., Cao, Y., Kern, N.R., Lee, J., Yeom, M.S., Croll, T.I., Seok, C., Im, W., 2020. Developing a fully glycosylated full-length SARS-COV-2 spike protein model in a viral membrane. J. Phys. Chem. B 124 (33), 7128–7137. https://doi.org/10.1021/ACS.JPCB.0C04553/SUPPL_FILE/JP0C04553_SI_001.MP4.

Yang, M., Shang, Y.X., Tian, Z.Y., Xiong, M., Lu, C.L., Jiang, Y., Zhang, Y., Zhang, Y.Y., Jin, X.Y., Jin, Q.B., Zhang, Y., Willcox, M.L., Liu, J.P., 2020. Characteristics of registered studies for Coronavirus disease 2019 (COVID-19): a systematic review. Integr. Med. Res. 9 (3), 100426 https://doi.org/10.1016/J.IMR.2020.100426.

Yao, H., Song, Y., Chen, Y., Wu, N., Xu, J., Sun, C., Zhang, J., Weng, T., Zhang, Z., Wu, Z., Cheng, L., Shi, D., Lu, X., Lei, J., Crispin, M., Shi, Y., Li, L., Li, S., 2020. Molecular architecture of the SARS-CoV-2 virus. Cell 183 (3), 730–738.e13. https://doi.org/10.1016/J.CELL.2020.09.018.

Zhao, P., Praissman, J.L., Grant, O.C., Cai, Y., Xiao, T., Rosenbalm, K.E., Aoki, K., Kellman, B.P., Bridger, R., Barouch, D.H., Brindley, M.A., Lewis, N.E., Tiemeyer, M., Chen, B., Woods, R.J., Wells, L., 2020. Virus-receptor interactions of glycosylated SARS-CoV-2 spike and human ACE2 receptor. Cell Host Microbe 28 (4), 586–601.e6. https://doi.org/10.1016/J.CHOM.2020.08.004.

Zheng, M., Zhao, J., Cui, C., Fu, Z., Li, X., Liu, X., Ding, X., Tan, X., Li, F., Luo, X., Chen, K., Jiang, H., 2018. Computational chemical biology and drug design: facilitating protein structure, function, and modulation studies. Med. Res. Rev. 38 (3), 914–950. https://doi.org/10.1002/MED.21483.