scientific reports

Check for updates

OPEN

Exploring the effects of N234 and N343 linked glycans to SARS CoV 2 spike protein pocket accessibility using Gaussian accelerated molecular dynamics simulations

Ronny L. Cheng, James Peter L. Lim, Myrnel A. Fortuna, Donnifer V. Reyes Jr., Earl Adrian D. R. Hans & Ricky B. Nellas✉

The N234 and N343-linked glycans of the SARS-CoV 2 spike protein are known to stabilize the up-conformation of its receptor-binding domains (RBDs), enabling human angiotensin enzyme 2 (hACE2) receptor binding. However, the effect of spike-hACE2 binding on these important glycans remains poorly understood, and these changes could have implications in the development of drugs that inhibit viral entry. In this study, Gaussian accelerated molecular dynamics (GaMD) simulations of the hACE2-free and hACE2-bound spike protein are performed. Biophysical analyses were focused on the accessibility of three previously suggested druggable pockets underneath the three RBD subunits. A shielding effect by N234-linked glycans on the components of their adjacent pockets was observed. Although deshielding of central scaffold residues was observed in the hACE2-bound state, pocket A's accessibility was reduced due to an increase in NTD_B–RBD_B contacts, restricting entry into the pocket. For pocket B, changes in $N234_C$ and $N343_C$ expose the central scaffold residues in the bound state, increasing accessibility. In Pocket C, increased shielding due to $N234_A$ was found in the bound state, reducing accessibility. Despite these changes, the pockets remain accessible to ligands in both states and are still valid targets for drug development studies.

Keywords COVID-19, SARS-CoV-2 spike protein, Gaussian accelerated molecular dynamics, Glycosylation, Biophysical analysis

The interaction between the SARS-CoV-2 spike protein and the human angiotensin-converting enzyme 2 (hACE2) receptor facilitates viral entry into cells and is a crucial determinant in the infectivity of COVID-19[1]. Glycosylations within both the spike protein and the hACE2 receptor are known to play a key role in the spike-hACE2 interaction, modulating viral binding and cellular entry[2–4]. By understanding glycan-mediated interactions, future therapeutic interventions that disrupt them can be developed.

The spike protein is a homotrimeric glycoprotein embedded in its viral membrane[5,6]. It has three receptor binding domains (RBDs) that can assume a combination of up, down, or transient states[3,7]. In the absence of hACE2, the RBDs fluctuate between these, with the stochastic up-down transitions occurring at the millisecond timescale[8,9]. When an RBD binds to hACE2, its up state is stabilized, and either of the other two RBDs adopts an induced up state[2,10]. When at least one of the RBDs is in the up state, the spike protein is said to be in the open conformation, with an RBD available for hACE2 binding.

The N-glycans linked to N234, N343, and N165 play a role in the binding of the spike protein to hACE2. Studies suggest that the N234 N-glycans stabilize the up conformation of the RBDs[2]. Meanwhile, the $N343_B$ N-glycan serves a role in controlling the RBD_A conformational states[3]. Finally, the $N165_A$-linked glycan is found to interact with the hACE2 receptor or its associated N-glycans directly[4,11]. Mechanistic investigations have elucidated that glycans facilitate long-range interactions between the spike protein and hACE2 through a catch-slip mechanism[12]. Characterized by hydrogen bond relaxation and reformation, this mechanism is driven by the increased flexibility imparted by glycans.

Institute of Chemistry, University of the Philippines Diliman, Quezon City 1101, Philippines. ✉email: rbnellas@up.edu.ph

As N-glycans can affect the stability and structure of the spike protein, it is essential to understand their behavior and interactions in the open conformation with or without hACE2, especially for those that do not directly interact with hACE2 or RBD_A. One aspect of the SARS-CoV-2 spike protein that has yet to be explored is the impact of hACE2 binding on its glycans. A structural and molecular approach can shed light on potential conformational changes in spike protein N-glycans in the presence and absence of hACE2.

Aside from stabilizing the spike-hACE2 interaction[13,14], glycosylations serve as physical barriers or shields against neutralizing antibodies[4,15]. Glycan shielding blocks up to 40% of the protein surface from antibody access. Previously, the authors identified three druggable pockets beneath the RBDs[16]. To determine whether or not shielding occurs in the pockets, this paper focuses on discerning how N-glycans influence the accessibility of drug candidates to these sites.

In this study, biophysical analysis from Gaussian accelerated molecular dynamics (GaMD) simulations of the spike protein head under these two conditions (hACE2-free and hACE2-bound) is performed to observe potential N-glycan contact differences and how hACE2 binding can affect the accessibility of druggable pockets within the spike protein. These findings can be applied in structural studies of the spike protein and future drug discovery studies against future coronaviruses.

Results

Global contacts and dynamics differ between the hACE2-free and hACE2-bound systems

Gaussian accelerated molecular dynamics simulations were used to sample the conformational space of hACE2-free protein and hACE2-bound spike protein systems[17]. The simulations utilized a boost potential, enabling the exploration of conformational space that would typically require longer time scales or be inaccessible with classical simulations[17]. In this study, 200 ns GaMD simulations were deemed sufficient to investigate the accessibility of highly localized pockets in an unbiased manner within the conformational landscape of hACE2-free and hACE2-bound spike proteins. This approach focuses on capturing the general protein dynamics without the constraints of biased techniques, such as steered MD or umbrella sampling, which are typically required to observe rare conformational changes occurring on much longer time scales.

The receptor binding domain (RBD) of the spike protein can exhibit multiple conformations, i.e., up, transient, and down[7,10]. Binding requires the RBD to adopt an up conformation, which is stabilized in the presence of hACE2[18,19]. hACE2 binding also contributes to the conformational change of the neighboring RBDs to an up state. Simulations for both hACE2-free and hACE2-bound spike protein systems depict the starting RBD conformations as up for RBD_A, transient for RBD_B, and down for RBD_C. Throughout the simulations, the conformations for the RBDs remain mostly stable and unchanged for both hACE2-free and hACE2-bound systems. This is characterized by the distance distribution histograms between the RBD center of mass and the central scaffold (Supplementary Fig. S1)[2,3]. Additionally, RMSD analyses of the RBDs in Supplementary Fig. S2a-S2c show a ~2Åadeviation between the hACE2-free and hACE2-bound systems, indicating minimal change between the systems. Conformational transitions in the RBD occur on millisecond-to-second time scales and are beyond the reach of accelerated simulations at the nanosecond level for large proteins[8]. However, the focus of this study is not on the conformational transitions of the spike RBD, but rather on the effect of spike protein glycosylations to pocket accessibility in the presence and absence of hACE2. Such interaction changes can be correlated with conformational changes and domain accessibility[20,21].

Spike protein-glycan interactions are meaningful interactions to consider because they are found to influence shielding and conformational transitions of the RBD. Specifically, the N234 and N343-linked glycans surround the RBD domains within the spike protein[2,3,9,12]. The $N234_B$ and $N343_B$ glycans are situated along the RBD_A, while the $N343_C$ glycan is situated at the upper portion of the interface between the RBD_B and RBD_Cdomains in the hACE2-free state. The locations of the N234 glycans are also situated adjacent to the three proposed allosteric pockets below the RBDs[16]. These N-glycans are depicted in Fig. 1a.

The flexibility changes of the N234 and N343-linked glycan cores (illustrated in Fig. 1b) were assessed through root-mean-square fluctuation (RMSF) analysis. In Fig. 1c, the $N234_B$ glycan radial plot indicates a two-fold difference in RMSF of all glycan core carbohydrate units between the two spike protein systems, with lower RMSF observed at hACE2-bound state. Meanwhile, the $N343_C$ N-glycan exhibits higher RMSF for all glycan core carbohydrate units at hACE2-bound state, as shown in Fig. 1d. This suggests that the glycan has greater stability in the absence of hACE2. The change in RMSF across the remaining N234 and N343-linked glycans was also calculated, as shown in Supplementary Fig. S3. Given the similarity in RMSF values of these glycan cores, the changes observed in the $N234_B$ and $N343_C$ glycans can be interpreted as asymmetric glycan movement.

Changes in accessibility

Previously, the authors identified three potential druggable pockets beneath each RBD chain and adjacent to the central scaffold of the spike protein[16]. The residues forming the pockets are listed in Table S1. These pockets were evaluated for druggability in both hACE2-bound and hACE2-free systems, with the primary descriptor being the mean local hydrophobic density (ρ). The calculation of ρ depends on pocket compactness. The latter is characterized by parameters including solvent-accessible surface area (SASA) and the number of native contacts[23]. Here, pocket accessibility is further investigated by performing biophysical analyses, including SASA and contacts analysis. These analyses target the three N234-linked glycans near the pockets, as depicted in Fig. 1a, and aim to understand the potential druggability of each pocket further.

The Shrake-Rupley method, implemented in VMD[24], utilizes the Voronoi algorithm[25,26] to calculate the SASA. This method partitions space into atom-associated regions using contact planes and a hypothetical sphere. Each atom is surrounded by a sphere of uniformly spaced dots, simplifying calculations by removing overlaps, and SASA is computed as the total volume of the remaining dots. Understanding pocket accessibility is of paramount importance in drug development. It offers crucial insights into the potential binding affinities of

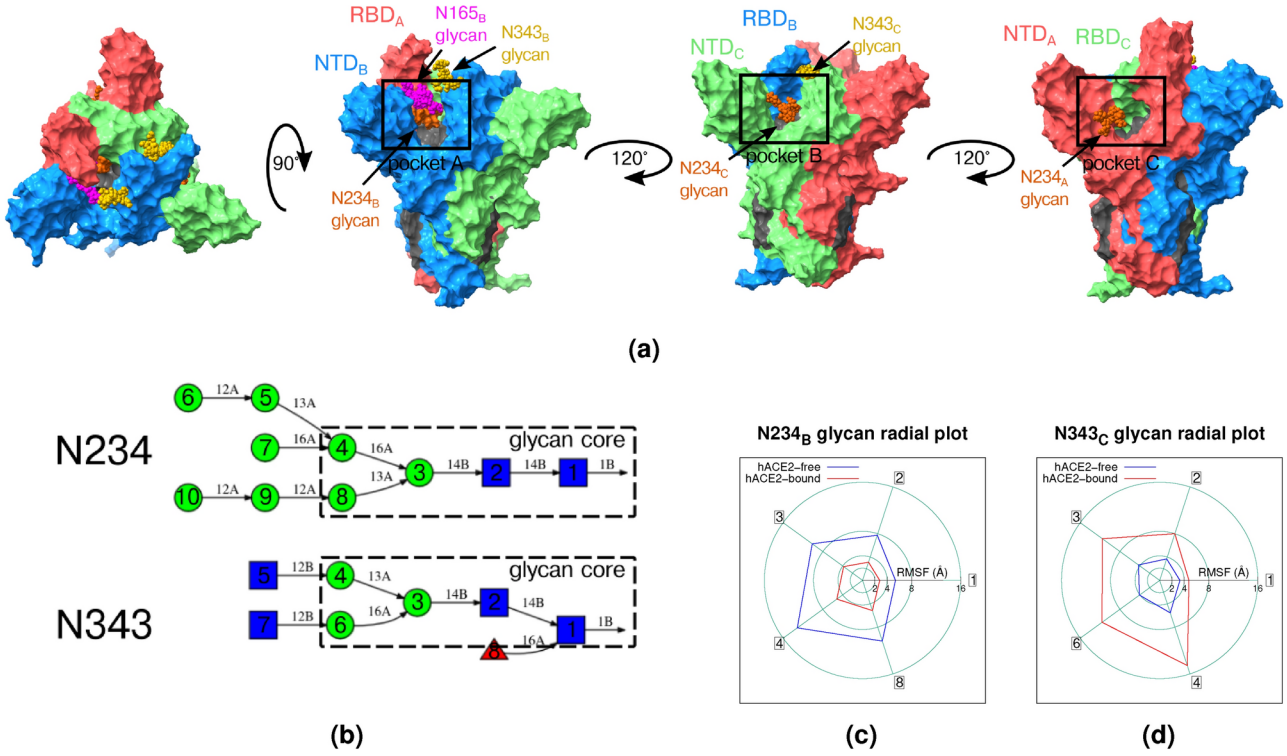

Fig. 1. (**a**) Depiction of the locations of $N234_A$, $N234_B$, $N234_C$ and $N343_C$relative to the hACE2-free spike protein. (**b**) The glycan sequences of N234 and N343 glycans are depicted using the format of Glycan Sequence Reader (GRS)[22]. The glycan radial plots for the glycan cores of (**c**) $N234_B$, and (d) $N343_C$.

drug-like compounds with these specific protein regions. According to Lipinski's rule of five, oral drugs typically possess a molecular weight below 500 Da, corresponding to an approximate protein surface area of 300 $Å^2$[27].

This study used different probe sizes to assess SASA across 2000 frames for the three pockets and the RBDs. Probe sizes range from the default size of 1.4 Å to larger sizes representing small drug-like molecules (2 to 5 Å). The calculation also considered the presence of glycans on accessibility, and the results are presented in Supplementary Tables S2 and S3. Additionally, a per-hydrophobic residue SASA calculation was conducted to evaluate the exposure of spike domains in each pocket. The SASA analysis reveals that Pocket C demonstrates the highest accessibility among all probe sizes (refer to Supplementary Tables S2, S3 and Supplementary Fig. S4a). This increased accessibility implies a heightened ligand or small molecule binding interaction potential. Conversely, despite its larger dimensions, Pocket A exhibits comparable accessibility to Pocket B, indicating similar binding interaction capacities. Consequently, the observed SASA range suggests that drug-like molecules within Lipinski's defined size range may effectively engage with the identified pockets, notably Pocket C, given its pronounced accessibility.

Accessibility of the RBDs (Supplementary Fig. S4b) reveals the most accessible to be the hACE2-free RBD_A, followed by hACE2-free RBD_B and hACE2-bound RBD_B. This trend is expected as the most accessible RBD should be the more exposed up conformer (RBD_A) ready for hACE2 binding. Upon binding of the hACE2 on the RBD_A interface, an intuitive drop in accessibility can be observed. The second most accessible RBD is RBD_B, both in the bound and free state. This alludes to a more solvent-exposed placement of the RBD_B, characteristic of an RBD in the transient conformation. On the other hand, RBD_C, both in the bound and free state, has the lowest accessibility, which is expected for an RBD in the down conformation.

The calculated SASAs of hydrophobic residues per pocket were then grouped into respective domains to compare the SASA in terms of the change in accessibility within components of the pockets, as shown in Fig. 2a. Generally, the SASA contributions of the RBD residues located above their respective pockets are similar, except for RBD_B in hACE2-free, which exhibited slightly higher SASA in hACE2-bound. Instead, differences can be observed in the contribution of the central scaffold residues adjacent to their respective pockets. The lower SASA of the hACE2-bound state for Pocket A and Pocket C can be attributed to the lower SASA in the central scaffold residues. The higher SASA in central scaffold residues in Pocket B could also rationalize the higher SASA observed in the hACE2-bound spike protein.

The RMSF bound-free ratios (Fig. 2b) reveal that Pocket C is substantially more stabilized in the spike-hACE2-free form, corroborating the SASA information on Pocket C. As visualized in Fig. 2c, the RMSF ratio differences across pocket residues are not substantial across the three pockets. Evidence from the slopes of the SASA-probe radii of the pockets (Supplementary Fig. S4a) reveals a potentially marked difference in the topographies of the pockets. Pocket C SASA-probe radii slopes are substantially steeper than Pockets A and B, suggestive of a rough surface pocket topography, potentially indicative of more residues buried in concavities

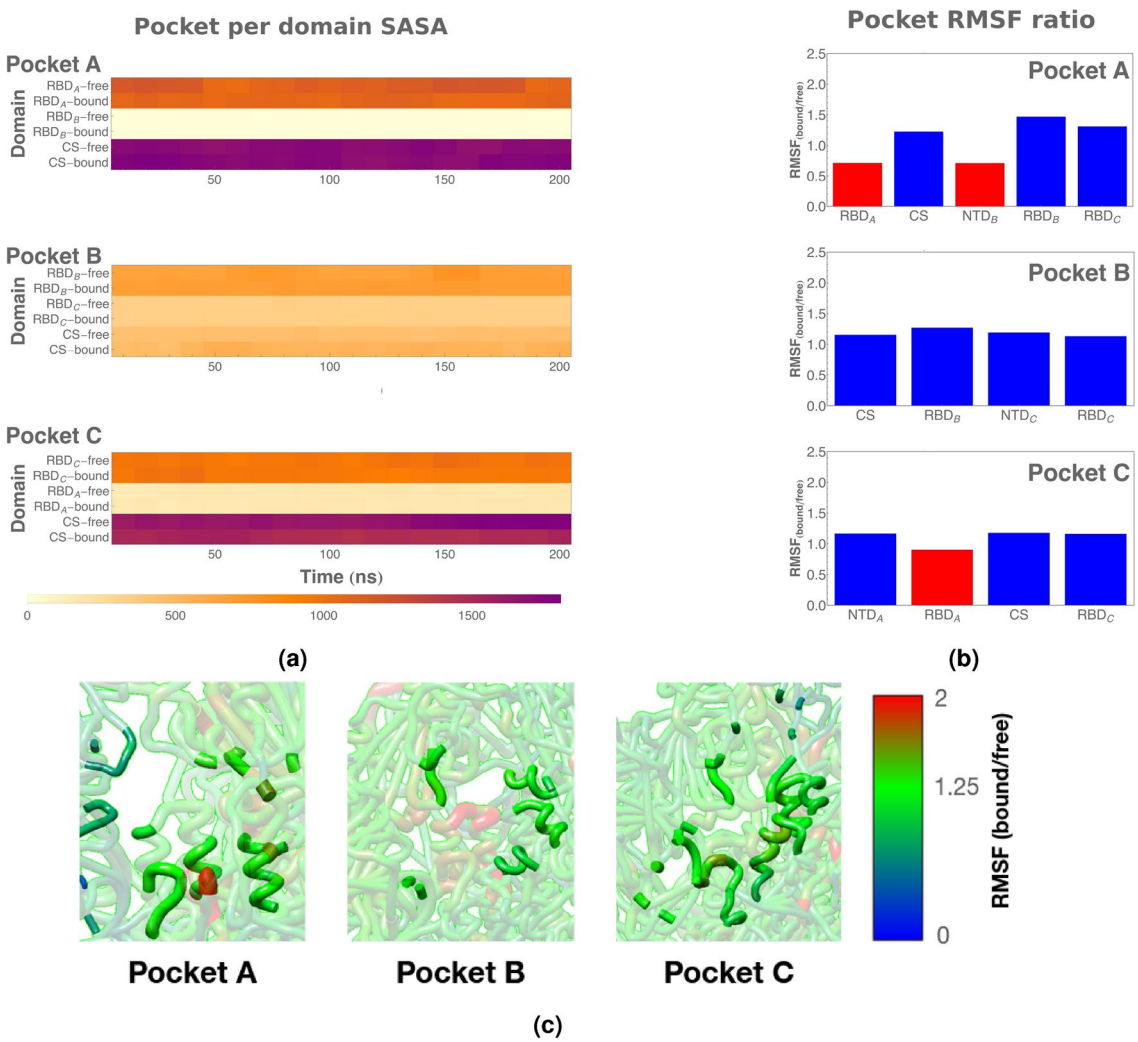

Fig. 2. The SASA per domain and RMSF ratio of the three pockets. The (**a**) total per domain of the mean SASA (in Å^2) of pocket hydrophobic residues. Each bin represents the average of 100 frames. The (**b**) RMSF ratios between the hACE2-bound and hACE2-free spike protein, with color coding indicating ratios < 1 in red (higher stability in the hACE2-bound state) and ratios > 1 in blue (higher flexibility in the hACE2-bound state). The (**c**) changes in RMSF along Pockets A, B and C relative to the hACE2-free system are visualized as an RMSF worm plot. Figure 2c is visualized using Chimera[28].

in the pocket. In contrast, Pocket B residues in the spike-hACE2 bound state are collectively less rigid (Fig. 2c) compared to the unbound form, further supporting the idea of a distinguishable increase in pocket accessibility.

Interplay of glycans and RBD state affect the accessibility of pocket A

Informed by the SASA results and the locations of glycans along the pockets, contact analysis was performed for extensive assessment of protein-glycan and glycan-glycan interactions, which are mapped for both the hACE2-free and hACE2-bound systems as shown in Supplementary Fig. S5.

In a previous study, the contacts of N234$_B$ glycan were deemed vital in the stability of the RBD$_A$up conformation[2]. The glycan is also within Pocket A, which lies below the RBD$_A$. Identified residues in the Pocket A are comprised of residues in RBD$_A$, NTD$_B$, RBD$_C$, and the central scaffold. Understanding where the contacts can occur is vital, as glycan interactions can effectively shield pocket residues from potential inhibitors. Interestingly, comparison of N234$_B$-Pocket A interactions between the hACE2-free and hACE2-bound systems in Figs. 3a and 3b shows a drastic reduction in contacts at the hACE2-bound spike protein.

A closer look at Pocket A reveal partial deshielding of the central scaffold to be possible at the hACE2-free state (Figs. 3c and 3d), while a lack of glycan interaction near the central scaffold is observed at the hACE2-bound state (Fig. 3e). Visualization of N234$_B$ glycans are displayed in Figs. 3c, 3d, and 3e. A closer look at Pocket A reveals partial deshielding of the central scaffold to be possible at the hACE2-free state (Figs. 3c and 3d), while a lack of glycan interaction near the central scaffold is observed at the hACE2-bound state (Fig. 3e). Overall, the contacts towards nearby domains are more frequent in the hACE2-free state compared to the hACE2-bound, as shown in Supplementary Fig. S6a and S6b. In previous studies, the deshielding of the central scaffold region

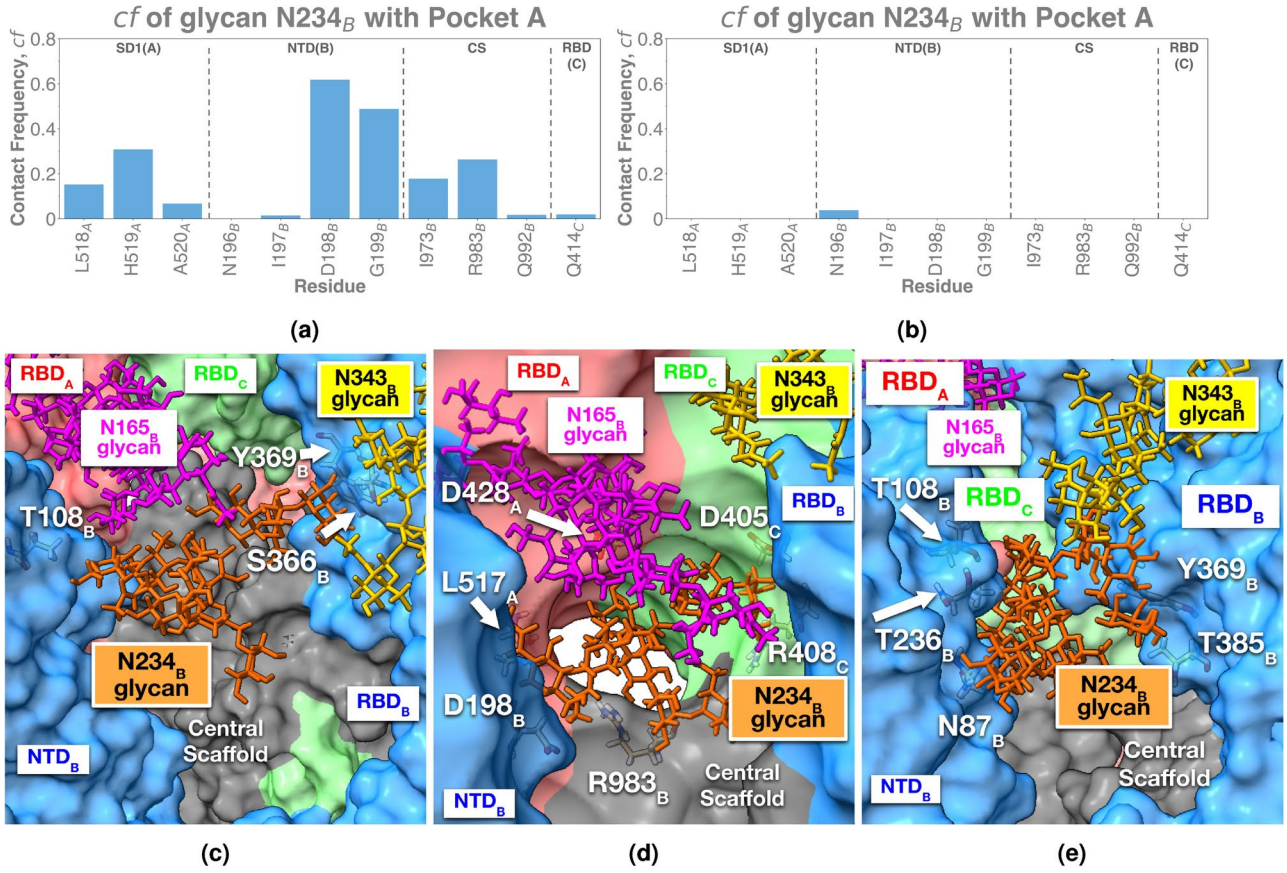

Fig. 3. Contact frequencies for the $N234_B$ glycan towards Pocket A residues exhibit a big difference between (**a**) the hACE2-free and (**b**) the hACE2-bound spike protein. Residues are grouped by dashed lines according to domain: N-terminal domain (NTD), receptor binding domain (RBD), and central scaffold (CS). Visual representation of the $N234_B$ glycan contacts. (**c**) The initial state of the hACE2-free state showing $N234_B$ glycan contacts towards the RBD_B domain. (**d**) The latter part of the simulations show the hACE2-free state forming $N234_B$ glycan contacts towards RBD_A and RBD_C domains. (**e**) The hACE2-bound state maintains contacts between $N234_B$ glycan and RBD_Bdomain throughout the simulation. Figures are depicted using ChimeraX[29].

was observed on the domain level with RBD[10,30]. RMSD analyses with the central scaffold (CS) region in Figs. S2d-S2f have shown higher values for the hACE2-bound system, which could indicate higher deviation from the starting structure. This deviation could probably be due to internal contacts within the CS that could lead to the structural rearrangement of this domain when the spike is fully open[10]. The glycan deshielding might be an important step prior to the opening of the spike protein and may be attributed to the shift in glycan-glycan interactions of $N234_B$ glycan with the $N165_B$ and $N343_B$ glycans.

Simulations show the $N234_B$ glycan to form contacts with the $N165_B$ glycan in the hACE2-free system. In the hACE2-bound system, the $N234_B$ glycan lacks this contact with the $N165_B$ glycan and instead forms contacts with the $N343_B$ glycan. Contacts are also formed with the $N165_B$ glycan in the hACE2-bound system. Fig. 3c-e shows representative snapshots of these glycan-glycan interactions. The lack of $N165_B$-$N234_B$ glycan-glycan contacts suggests that the $N165_B$ and $N234_B$ glycans could shift away from each other due to conformational changes within their parent domain NTD_B, and are loosely bridged by the $N343_B$ glycan. The changes in protein-glycan and glycan-glycan contacts within the $N234_B$ glycan could also help explain its increased flexibility in the absence of hACE2, as revealed in the RMSF glycan radial plot.

Comparing the glycan contacts between both systems in Supplementary Fig. S6a, S6b, S7a, and S7b also reveals a shift in contacts of $N234_B$ on the domain level: the glycan shifts from RBD_A, RBD_C, and central scaffold contacts in the hACE2-free, to the RBD_B in hACE2-bound. The $N165_B$ glycan also changes its contacts from mostly RBDs in hACE2-free to RBD_A and NTD_B in hACE2-bound, as seen in Supplementary Fig. S6c and S6d. Likewise, the $N343_B$ glycan changes from only RBD_B contacts in the hACE2-free to the NTD_B contacts in the hACE2-bound system (Supplementary Fig. S6e). These shifts in interactions could be due to an internal conformational change in NTD_B, which affected the interactions between the glycans $N165_B$ and $N234_B$.

Another potential explanation is the reduced distance between NTD_B and the RBD_B in the hACE2-bound system, as seen in Supplementary Fig. S7e. This interaction could result from contacts between the bound state of RBD_A and the $N343_B$ glycan. The $N343_B$ glycan serves as a gate that facilitates the transition of $RBD_A$from down to up state[3]. No contacts were observed between the $N343_B$-linked glycan and $RBD_A$for the hACE2-free spike protein in this study. This is in agreement with microsecond-length weighted-ensemble simulations studying

the down-to-up transition of the spike protein[3]. Since the up state of RBD_A is stabilized in the hACE2-bound spike protein, the glycan could instead engage with the neighboring NTD_B and glycan residues. Such RBD-NTD interactions could be an initial step to the up state transition of RBD_B according to the wedge model[31]. However, the presence of these contacts limits the entrance to Pocket A, which negates the effects of glycan deshielding. These could explain the similar accessibility of Pocket A in Fig. 2a in the hACE2-free and hACE2-bound spike proteins.

Despite the negligible changes to the accessibility of Pocket A residues between the hACE2-free and hACE2-bound states, the central scaffold remains highly accessible. Coupled with the deshielding of glycans, this opens the possibility of exploring potential inhibitors for the hACE2-bound spike protein (1-up bound).

$N234_C$ and $N343_C$ interactions influence pocket B accessibility

As the SASA reveals that Pocket B has higher SASA in the hACE2-bound state than in the hACE2-free state, the question arises as to why this phenomenon occurs. Contacts are formed with the RBD_B, RBD_C, and central scaffold residues assigned to Pocket B (Figs. 4a and 4b). Notable contact differences are observed between the hACE2-free (Supplementary Fig. S8a) and hACE2-bound (Supplementary Fig. S8b) states. In the hACE2-free state, $N234_C$ glycans form additional interactions with the chain C subdomain regions of Pocket B at the hACE2-free state, while additional contacts to RBD_B and RBD_C are formed in the bound state. These subdomain regions can serve as passageways to Pocket B, which affects access towards the central scaffold residues of the pocket. Likewise, in the hACE2-bound state, the $N234_C$ glycan is buried underneath the RBD_B, which could allow unimpeded accessibility of central scaffold residues in Pocket B.

Time evolution series of Pocket B residues and $N234_C$ glycan contacts in the hACE2-free state (Fig. 5a) shows the stable interaction of the $N234_C$ glycans with the RBD_B residues of Pocket B. Probing the time evolution of additional contact events in the hACE2-bound state (Fig. 5a), notable movements of the $N234_C$ glycans in Pocket B can be observed. Noteworthy is the shifting interactions of the $N234_C$ glycans between RBD_B-RBD_C and NTD_C domains, which explains higher RMSF of the Pocket B subdomains (Fig. 2c). Visualization of these contacts can be found in Fig. 4e-g.

Given the position of the $N343_C$ glycan at the top of the RBD_B and RBD_C interface (Fig. 4e) and its associated high RMSF, the potential effects of the $N343_C$ glycans to Pocket B accessibility was also considered. The differences in the contact frequency events of the $N343_C$ glycans with RBD_B between the hACE2-free (Fig. 4c and Supplementary Fig. S7c) and hACE2-bound (Fig. 4d and Supplementary Fig. S7d) states show significant loss of the $N343_C$ glycan contacts to the RBD_B in the hACE2 bound state. Further investigation using the time evolution of the contacts shows that in the hACE2-free state (Fig. 5b), $N343_C$ glycans remain in contact with the RBD_B domain interface throughout the simulation. In contrast, in the hACE2-bound state (Fig. 5b), it can be seen that in the latter half of the GaMD simulation, the minimum distance of the $N343_C$ glycan to RBD_B increases to around 10 Å. This shows dissociation of the glycan from the interface (Fig. 4f) outwards into the solvent (Fig. 4g), breaking its interaction with the RBD_B domain.

This observation is interesting as the $N343_B$ glycan was suggested to follow a gating mechanism to shift the RBD_A from a down-to-up conformation by moving away from the RBD_A-RBD_Binterface[3]. Current literature supports the possible formation of the 2-up from the 1-up spike-hACE2 state accompanied by the down-to-up transition of the RBD domain[10]. In this GaMD simulation, a similar glycan gating mechanism can be deduced as the $N343_C$ glycans move away from the RBD_B-RBD_C interface, gearing the RBD_B from the initial transient state into a potentially more open form ready for hACE2 binding. The RMSF of the RBD_B region of Pocket B is substantially higher in the bound state ($\sim 30\%$), which further corroborates the increased mobility of the RBD_B that could potentially be a consequence of the transient-to-up conformational change. The possible shifting of RBD_B conformation from $N343_C$ activity could provide an opportunity for $N234_C$ to move inwards towards the lower part of RBD_B, allowing the central scaffold to be more accessible at the hACE2-bound state. Elucidation of the exact mechanism of the conformational transitions of these glycans warrants further investigation, which can be realized in more extended time regimes.

Dynamics of pocket C

The changes in Pocket C dynamics were explored from the perspective of $N234_A$ glycan motions. Contact frequency maps in Figs. 6a and 6b show that glycan shielding via contacts in RBD_C, subdomain region 1 of chain A, and the central scaffold adjacent to RBD_C is possible. However, contacts towards these regions appear more frequent in hACE2-bound rather than hACE2-free. Similar patterns are evident in the global frequency maps of the $N234_A$ glycan, showing increased contact frequency at the subdomain regions 1 and 2 of chain A, as well as with RBD_C in the hACE2-bound state (Supplementary Fig. S9). Visualization of these contacts in Fig. 6c shows the partial shielding of the pocket passageway along the three aforementioned regions. While this shielding can still exist in the hACE2-free simulations, it occurs less frequently than the hACE2-bound state, which could explain the lower SASA of the central scaffold surrounding Pocket C.

Discussion

The SARS-CoV-2 spike protein has been established as an integral component in viral pathogenicity[2,32]. Interactions of its glycans are essential for hACE2 binding, shielding, and domain conformational changes[2,3,33–36]. Spike glycans have also been known to mediate conformational states of the RBD and stabilize the core of the spike protein[2,3]. The N234-linked glycans interact with the three potential pockets of the spike protein. These pockets have been identified in a previous study through coarse-grained simulations of non-glycosylated spike protein[16]. Glycan conformations and interactions could potentially impact the shielding and accessibility of pockets to bind drug candidates.

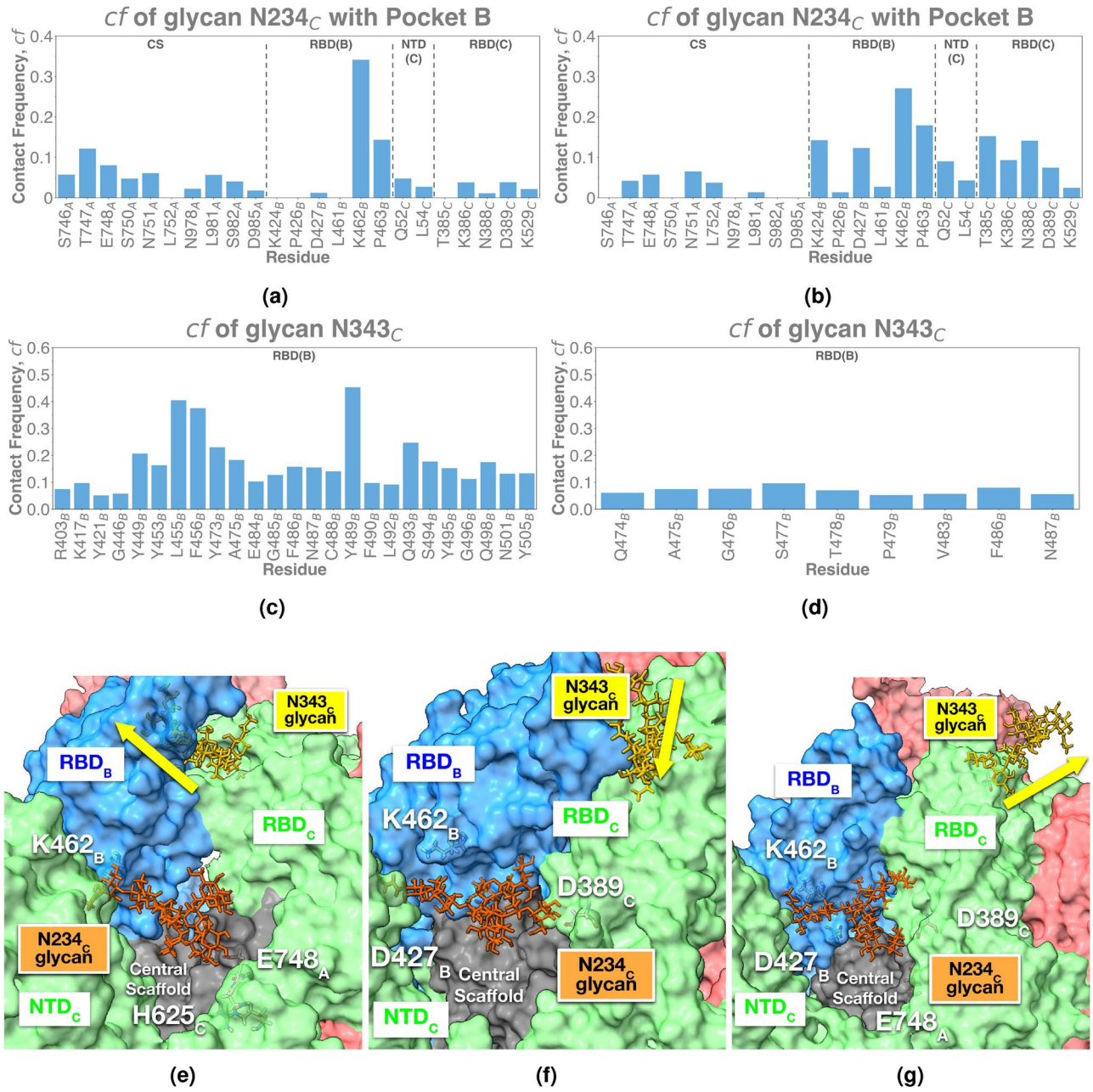

Fig. 4. Contact frequencies of glycans $N234_C$ with Pocket B residues show only a slight difference between the (**a**) hACE2-free and (**b**) hACE2-bound systems. Residues are grouped by dashed lines according to domain: N-terminal domain (NTD), receptor binding domain (RBD), and central scaffold (CS). A substantial decrease in contact frequencies between the glycan linked to $N343_C$ and the RBD_B in the (**c**) hACE2-free and the (**d**) hACE2-bound system were observed. (**d**) Simulations show the $N343_C$ glycan at hACE2-free state to be maintained in the interface of RBD_B and RBD_C domains. (**e**) The hACE2-bound state initially is also located at the interface of RBD_B and RBD_C, with the direction of the $N343_C$ glycan drawn as a yellow arrow. (**f**) which then shifts to favor outwards towards solvent. (**g**) The $N343_C$ moves outwards in hACE2-bound state, with the direction of the $N343_C$glycan drawn as a yellow arrow. Figures are depicted using ChimeraX[29].

This study explores the interactions and behavior of the N234 and N343 glycans in the presence and absence of hACE2. The spike RBDs can exist in up, transient, or down states, where hACE2 binds to the former[19,37]. RBD core-central scaffold distance and RMSD analysis of the RBDs show that they each assumed a unique state at which they remained for the duration of the simulations[3]. However, RBD-hACE2 binding could have subtle effects, i.e., changes in interactions, which could affect glycan shielding and conformations of the neighboring RBDs. The study measured these changes using SASA analysis, glycan RMSF, and contact analyses.

Previously, three pockets between the RBDs and the central scaffold were characterized as potential sites for drug binding[16]. Analyses of the pocket SASA revealed that central scaffold accessibility can correlate with pocket

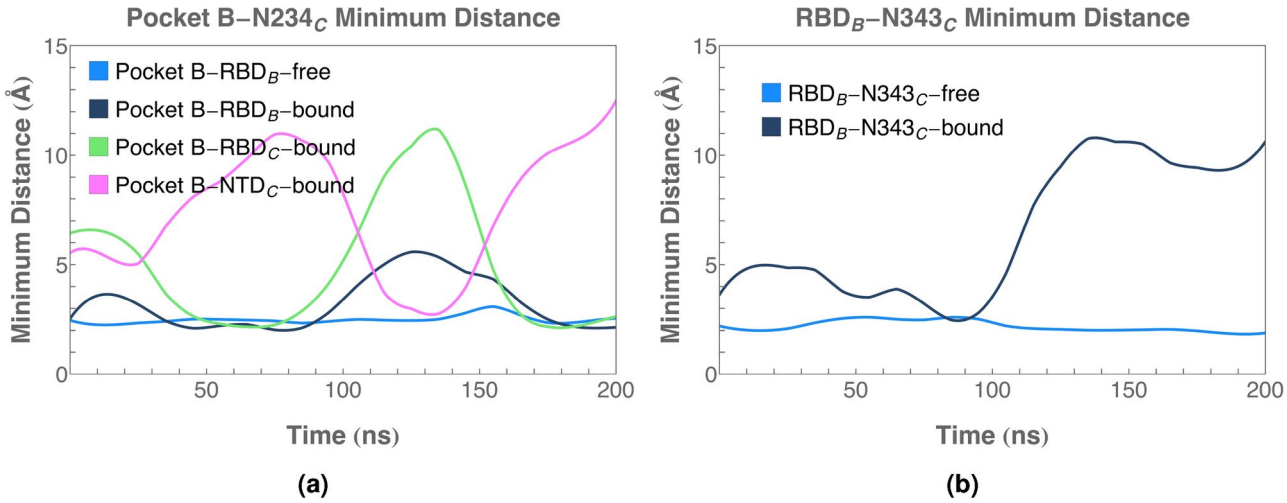

Fig. 5. Time evolution of the minimum distance between (**a**) the Pocket B and the $N234_C$ glycan, and (**b**) the RBD_B and the $N343_C$ glycan.

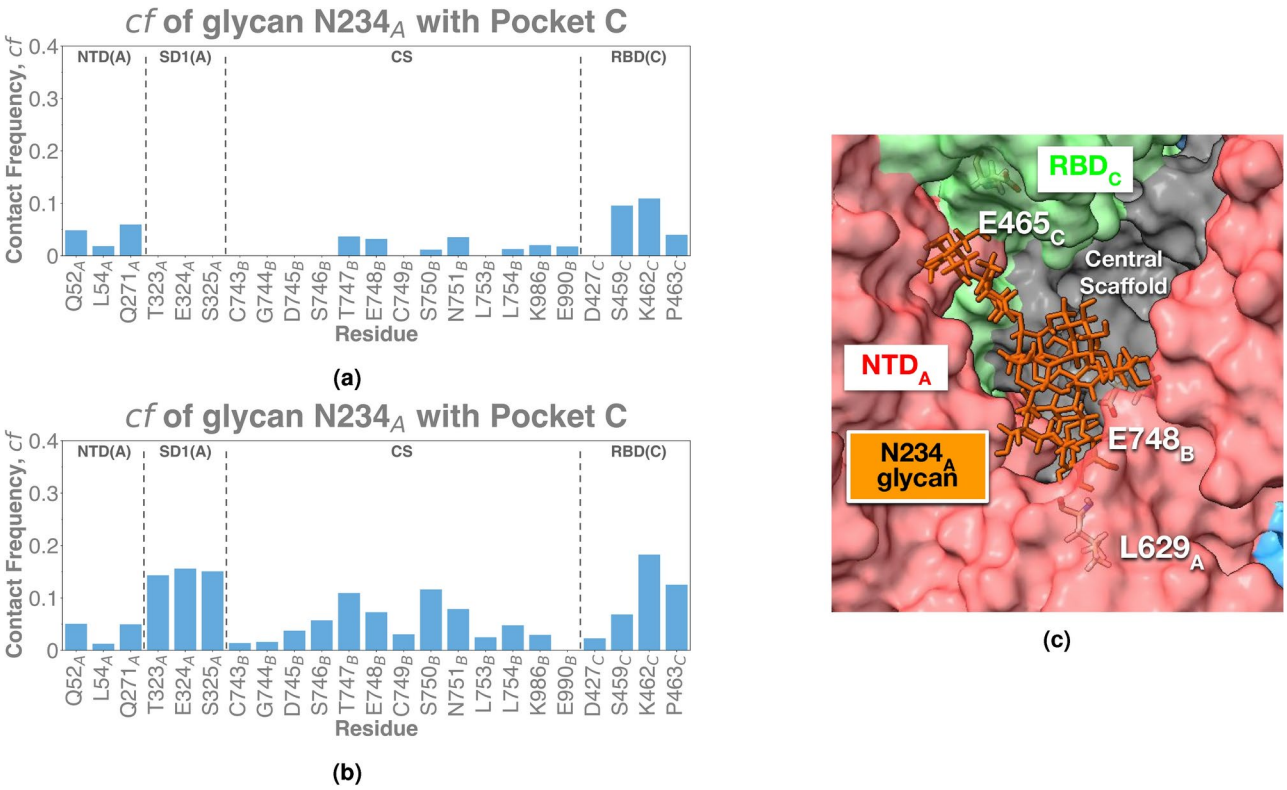

Fig. 6. Contact frequencies for glycans $N234_A$ with pocket C residues exhibit a slight increase between the (**a**) hACE2-free and (**b**) hACE2-bound systems. Residues are grouped by dashed lines according to domain: N-terminal domain (NTD), subdomain 1 (SD1), receptor binding domain (RBD), subdomain 2 (SD2), and central scaffold (CS). (**c**) Visual representation of contacts for $N234_A$ glycan. (c) The initial state of the hACE2-free state showing contacts of $N234_A$ glycan towards the RBD_Bdomain. Figures are depicted using ChimeraX[29].

druggability. However, nearby glycans, specifically glycan N234, could hinder access to these pockets. Contact analyses and RMSF radial plots revealed differences in pocket shielding and glycan conformations between the free spike and hACE2-bound systems, respectively. In Pocket A, a glycan deshielding of the pocket is observed in the hACE2-bound state, as supported by the lack of $N234_B$ contacts. The mechanism of deshielding could be due to the interplay of $N165_B$, $N234_B$, and $N343_B$ glycans in the hACE2-bound system. The interplay could be due

to a shift in interactions of glycan $N343_B$ from the RBD_B towards the nearby NTD_B and its glycans. This shift in interactions could be due to a decrease in distance between NTD_B and RBD_B, which could constrict the entrance of Pocket A. While deshielding of the central scaffold occurs as a result of viral fusion during hACE2 binding, the constriction of NTD_B and RBD_B could negate the CS accessibility and explain the similar SASA between the free and bound systems. Further studies are needed to probe and understand the deshielding mechanism in the spike pockets.

In Pocket B, SASA, RMSF, and time evolution contact analyses of the dynamics of the $N234_C$ and $N343_C$ glycans similarly suggest the change in pocket accessibility to be affected by the accessibility of the central scaffold in the spike protein. This is based on contact analysis data that reveals increased shielding of the central scaffold at hACE2-free, while the $N234_C$ is buried underneath the RBD_B in the hACE2-bound state. Meanwhile, hACE2-bound spike proteins simulations show the $N343_C$ glycan to move away from the RBD_B-RBD_C interface, showing similarities to the glycan gating mechanism revealed in $N343_B$ glycan. Considering that RBD_B is in transient form, it is possible that this change could favor the RBD_B towards the up conformation for hACE2 binding. Thus, $N343_C$ glycan dynamics could be an indirect effect of a mechanism where polar interactions could be responsible for burying the $N234_C$ glycan underneath the RBD_B. Meanwhile, shielding was also observed by $N234_A$ glycan along Pocket C in the hACE2-free simulations, which could affect access to the central scaffold.

Based on the biophysical analyses, important contact contributions of the N234 and N343-linked glycans and spike protein dynamics reveal potential insights on druggability. In Pocket A, the differences in the mobilities of the RBD_A and NTD_B domains in the free and bound states imply that their stabilities are an important consequence of the hACE2 binding. Using contact information of the $N234_B$ glycans with Pocket A, the NTD_B domain in the hACE2-free state could be a crucial point of interaction that can be explored further for potential inhibitors for the hACE2-free spike protein. Potential ligands that can disrupt these $N234_B$ glycan-NTD_B interactions could be relevant. In Pocket B, culminating the SASA information and $N234_C$ interactions, the RBD_B region of the pocket is identified to be the most promising region for development of potential inhibitors that could disrupt the $N234_C$-RBD_B interactions both in the hACE2-free and the hACE2-bound (1-up) state. In Pocket C, SASA and $N234_A$ glycan contacts reveal the central scaffold and RBD_C regions to be the relevant targets for potential inhibitors.

Glycan-protein contact shifts between the free and bound states were observed in the simulations, which warrants further validation of the switching phenomenon. Since this study uses trajectories with different starting points, elucidation can be done using biased sampling methods such as metadynamics to simulate the spike protein-hACE2 binding event. Altogether, it is observed that the glycan dynamics can affect pocket accessibility by shielding the regions surrounding the pockets, especially along the central scaffold residues adjacent to them. While pocket accessibility can be affected, the SASA reveals that these pockets are still accessible in the presence of glycans. Caution is necessary to interpret virtual screening studies that utilize unglycosylated spike protein, as some hydrophobic drug interactions are affected by glycosylations. These results could offer insights for future drug discovery studies targeting the spike protein in free and hACE2-bound states.

Conclusion

In this study, computational biophysical analysis was employed to investigate the impact of the N234 and N343-linked glycosylations on the accessibility of the proposed pockets in the presence and absence of hACE2. SASA calculations of the three pockets show that their hydrophobic components remain accessible in both free and hACE2-bound states. Pockets A and C show slightly lower SASA when bound to hACE2, while Pocket B has higher SASA in the hACE2-bound state. This can be attributed to the difference in accessibility of central scaffold residues adjacent to the pockets.

Contact analysis shows that the N234-linked glycans affect pocket accessibility by deshielding pocket components or further exposing the central scaffold residues adjacent to the pocket. For Pocket A, $N234_B$ glycan contacts between RBD_B and RBD_C can potentially shield the central scaffold. While the central scaffold is deshielded in the hACE2-bound state, the increased interactions between the NTD_B and RBD_B constrict the pocket entrance and limit the central scaffold accessibility.

For Pocket B, central scaffold accessibility is affected by the location and formation of contacts of the $N234_C$ glycan. In the hACE2-free state, contacts between the $N234_C$ glycan and the adjacent RBD_C can contribute to the shielding of the central scaffold. On the other hand, in the hACE2-bound state, the $N234_C$ glycan is buried underneath RBD_B, exposing the central scaffold. Changes are also observed in the $N343_C$ glycan, where the glycan can be pushed outwards from the RBD_B-RBD_C interface at the hACE2-bound state.

Identified glycan contacts will be valuable considerations for drug design of anti-COVID-19 drugs. These results indicate that the pockets are still potentially druggable, even after the changes in glycan shielding effects due to hACE2 binding.

Methods

Systems preparation

The initial systems for the study are based on the atomistic structure of the wild-type trimeric spike protein (hACE2-free) from the CHARMM-GUI COVID-19 database[38]. This structure is based on the existing cryo-EM structure of the trimeric spike protein (PDB ID: 6VSB, GenBank Accession ID: YP_009724390.1)[39,40]. In this structure, the RBD_A, RBD_B, and RBD_C domains are at an up, transient, and down conformations, respectively[3]. For the hACE2-bound spike protein complex, its structure was derived from a combination of the previously described structure of the spike protein and the cryo-EM structure of the wild-type receptor-binding domain (RBD)-hACE2 complex (PDB ID: 6VW1)[41].

System	Water box dimensions	Number of atoms
hACE2-bound	252 Å x 203 Å x 201 Å	773,227 atoms (including 235,708 water molecules, 680 K^+ and 665 Cl^- atoms)
hACE2-free	202 Å x 202 Å x 202 Å	940,273 atoms (including 287,742 water molecules, 851 K^+ and 811 Cl^- atoms)

Table 1. Summary of systems run for GaMD simulation.

Loops absent in the crystal structure were added to the models using template-based modeling with the GalaxyTBM and FALC programs[42–44]. The spike protein tail and membrane were removed from the initial model, leaving only the residues that comprise the spike protein head (residues 1–1146) and its corresponding glycosylations, which include 19 N-glycans and 1 O-glycan[42]. The glycan composition of each site with the highest abundance was identified based on existing mass spectrometry data[45,46]. The same glycosylation profile was employed for the hACE2-bound model, which also includes 5 N-glycans at the hACE2 receptor[45,46]. The glycosylation sites alongside their corresponding added glycan sequence to the models are available at the COVID-19 Archive of the CHARMM-GUI Library[38,42].

The Solution Builder feature of the CHARMM-GUI server was used to prepare the proteins[47]. The systems were kept neutrally charged and solvated in a 0.15M KCl solution, where the positions of the ions were determined using the distance ion placing method. Protein, glycan, water, and ion atoms were treated with the CHARMM36m force field with hydrogen mass partitioning[48,49]. A summary of the systems is presented in Table 1.

Gaussian accelerated molecular dynamics simulations

All simulations were performed in Amber20[50]. Energy minimization of the systems was first performed using the steepest descent algorithm for 2500 steps with the protein and glycans constrained at 1.0 kcal $Å^2$. This was followed by a gradual NVT heating of the systems to 310 K for 500 ps, with temperature regulated by the Berendsen thermostat with a thermal coupling parameter of 0.5 ps[51]. NPT equilibration for 4 ns was then applied to the systems, with the first 2 ns still including the restraints and the final 2 ns removing the restraints. Isotropic position scaling was applied with a 1.0 ps pressure relaxation time[52].

A dual-boost GaMD simulation was performed for every system[17,53]. A 2 ns classical MD simulation was first performed, followed by a 10 ns GaMD equilibration run. The threshold energy was set as the minimum potential energy of each respective system, while the dihedral boost and potential energy boost were set to 6.0 kcal/mol. The average mean and standard deviations of dihedral and potential energies were updated every 0.2 ns. An NPT production GaMD run of 200 ns for every system was then performed using the dihedral and potential energy boosts from GaMD equilibration to maximize conformational sampling. Pressure and temperature are regulated by the Langevin piston and thermostat, respectively[54]. The coordinates of the GaMD production run were written into its trajectory every 1 ps.

All simulations were conducted under periodic boundary conditions. Long-range electrostatics were calculated using the Particle Mesh Ewald (PME) method with a cutoff distance of 8 Å[55]. The SHAKE algorithm[56] was applied to handle constraints for the bonded hydrogen.

Biophysical analysis

Trajectory analysis was performed using the cpptraj module of AmberTools20[50]. Flexibility within the glycoprotein is addressed by calculating the root-mean-square fluctuation (RMSF) of the Cα atoms of the amino acid residues. The carbon atoms of the two N-acetylglucosamines (NAG) and three mannose (MAN) units that form the N-glycan core were also calculated[45,46]. The total solvent-accessible surface area (SASA) was calculated using the *measure sasa* function of VMD 14.5[57]. Five probe sizes (1.4 to 5.0 Å) were used to measure the overall accessible surface areas of the three pockets, the RBDs, and the central scaffolds of the three spike units. To visualize the accessibility of each residue of the pockets during the simulation, a per-residue SASA analysis at 1.4 Å probe size was also performed. The Mathematica 14.0 Student Edition was used to visualize the results[58].

Contact analysis of the systems was performed with an initial script from the CAMERRA toolkit[20]. In a single frame, a contact matrix is constructed from all possible residue pairs, where a binary value (1/0) is assigned depending on the occurrence of a contact. In this study, a contact is defined when the distance between the nearest heavy atoms of a residue pair is less than 5 Å. Calculation of the contact frequencies was performed by taking the mean of the contact values of each residue pair for all extracted frames from the trajectories.

Data availability

The data underlying this study are available in the published article and its Supporting Information.

Received: 5 April 2024; Accepted: 1 January 2025
Published online: 27 February 2025

References

1. Borkotoky, S., Dey, D. & Hazarika, Z. Interactions of angiotensin-converting enzyme-2 (ace2) and sars-cov-2 spike receptor-binding domain (rbd): A structural perspective. *Mol. Biol. Rep.* **50**, 2713–2721 (2023).
2. Casalino, L. et al. Beyond shielding: the roles of glycans in the sars-cov-2 spike protein. *ACS Cent. Sci.* **6**, 1722–1734 (2020).
3. Sztain, T. et al. A glycan gate controls opening of the sars-cov-2 spike protein. *Nat. Chem.* **13**, 963–968 (2021).

4. Mehdipour, A. R. & Hummer, G. Dual nature of human ace2 glycosylation in binding to sars-cov-2 spike. *Proc. Natl. Acad. Sci.* **118**, e2100425118 (2021).
5. Huang, Y., Yang, C., Xu, X.-F., Xu, W. & Liu, S.-W. Structural and functional properties of sars-cov-2 spike protein: potential antivirus drug development for covid-19. *Acta Pharmacol. Sin.* **41**, 1141–1149 (2020).
6. Lan, J. et al. Structure of the sars-cov-2 spike receptor-binding domain bound to the ace2 receptor. *Nature* **581**, 215–220 (2020).
7. Egri, S. B. et al. Detergent modulates the conformational equilibrium of sars-cov-2 spike during cryo-em structural determination. *Nat. Commun.* **14**, 2527 (2023).
8. Lu, M. et al. Real-time conformational dynamics of sars-cov-2 spikes on virus particles. *Cell Host Microbe* **28**, 880–891 (2020).
9. Pang, Y. T., Acharya, A., Lynch, D. L., Pavlova, A. & Gumbart, J. C. Sars-cov-2 spike opening dynamics and energetics reveal the individual roles of glycans and their collective impact. *Commun. Biol.* **5**, 1170 (2022).
10. Benton, D. J. et al. Receptor binding and priming of the spike protein of sars-cov-2 for membrane fusion. *Nature* **588**, 327–330 (2020).
11. Isobe, A. et al. Ace2 n-glycosylation modulates interactions with sars-cov-2 spike protein in a site-specific manner. *Commun. Biol.* **5**, 1188 (2022).
12. Huang, Y. et al. Sars-cov-2 spike binding to ace2 is stronger and longer ranged due to glycan interaction. *Biophys. J.* **121**, 79–90 (2022).
13. Wang, L., Wang, L. & Zhuang, H. Profiling and characterization of sars-cov-2 mutants' infectivity and antigenicity. *Signal Transduct. Target. Ther.* **5**, 185 (2020).
14. Gong, Y., Qin, S., Dai, L. & Tian, Z. The glycosylation in sars-cov-2 and its receptor ace2. *Signal Transduct. Target. Ther.* **6**, 396 (2021).
15. Grant, O. C., Montgomery, D., Ito, K. & Woods, R. J. Analysis of the sars-cov-2 spike protein glycan shield reveals implications for immune recognition. *Sci. Rep.* **10**, 14991 (2020).
16. Cheng, R. L., Quirante, J. C., Vargas, L. E. Z., Gatchalian, A. F. & Nellas, R. B. Complementary pocket and network-based approach to search for spike protein allosteric pocket sites. *ACS Omega* **8**, 45313–45325 (2023).
17. Miao, Y., Feher, V. A. & McCammon, J. A. Gaussian accelerated molecular dynamics: Unconstrained enhanced sampling and free energy calculation. *J. Chem. Theory Comput.* **11**, 3584–3595 (2015).
18. Henderson, R. et al. Controlling the sars-cov-2 spike glycoprotein conformation. *Nat. Struct. Mol. Biol.* **27**, 925–933 (2020).
19. Yan, R. et al. Structural basis for the different states of the spike protein of sars-cov-2 in complex with ace2. *Cell Res.* **31**, 717–719 (2021).
20. Johnson, Q. R., Lindsay, R. J. & Shen, T. Camerra: an analysis tool for the computation of conformational dynamics by evaluating residue–residue associations (2018).
21. Gur, M. *et al.* Conformational transition of sars-cov-2 spike glycoprotein between its closed and open states. *J. Chem. Phys.* **153** (2020).
22. Park, S.-J. et al. Charmm-gui glycan modeler for modeling and simulation of carbohydrates and glycoconjugates. *Glycobiology* **29**, 320–331 (2019).
23. Bogatyreva, N. & Ivankov, D. The relationship between the solvent-accessible surface area of a protein and the number of native contacts in its structure. *Mol. Biol.* **42**, 932–938 (2008).
24. Shrake, A. & Rupley, J. A. Environment and exposure to solvent of protein atoms. lysozyme and insulin. *J. Mol. Bio.* **79**, 351–371 (1973).
25. Voronoi, G. Nouvelles applications des paramètres continus à la théorie des formes quadratiques. premier mémoire. sur quelques propriétés des formes quadratiques positives parfaites. *J. Reine Angew. Math.* **1908**, 97–102 (1908).
26. McConkey, B. J., Sobolev, V. & Edelman, M. Quantification of protein surfaces, volumes and atom-atom contacts using a constrained voronoi procedure. *Bioinformatics* **18**, 1365–1373 (2002).
27. Cheng, A. C. et al. Structure-based maximal affinity model predicts small-molecule druggability. *Nat. Biotechnol.* **25**, 71–75 (2007).
28. Pettersen, E. F. et al. Ucsf chimera-a visualization system for exploratory research and analysis. *J. Comp. Chem.* **25**, 1605–1612 (2004).
29. Meng, E. C. et al. Ucsf chimerax: Tools for structure building and analysis. *Protein Sci.* **32**, e4792 (2023).
30. Zhang, J., Xiao, T., Cai, Y. & Chen, B. Structure of sars-cov-2 spike protein. *Curr. Opin. Virol.* **50**, 173–182 (2021).
31. Li, Y. et al. Exploring the regulatory function of the n-terminal domain of sars-cov-2 spike protein through molecular dynamics simulation. *Advanced theory and simulations* **4**, 2100152 (2021).
32. Yang, Q. *et al.* Role for n-glycans and calnexin-calreticulin chaperones in sars-cov-2 spike maturation and viral infectivity. *Sci. Adv.* **8**, eabq8678 (2022).
33. Lenza, M. P. et al. Structural characterization of n-linked glycans in the receptor binding domain of the sars-cov-2 spike protein and their interactions with human lectins. *Angew. Chem* **132**, 23971–23979 (2020).
34. Li, Y. et al. The importance of glycans of viral and host proteins in enveloped virus infection. *Front. immunol.* **12**, 638573 (2021).
35. Raman, R., Tharakaraman, K., Sasisekharan, V. & Sasisekharan, R. Glycan-protein interactions in viral pathogenesis. *Curr. Opin. Struct. Biol.* **40**, 153–162 (2016).
36. Miller, N. L., Clark, T., Raman, R. & Sasisekharan, R. Glycans in virus-host interactions: A structural perspective. *Front. Mol. Biosci.* **8**, 666756 (2021).
37. Moreira, R. A., Guzman, H. V., Boopathi, S., Baker, J. L. & Poma, A. B. Characterization of structural and energetic differences between conformations of the sars-cov-2 spike protein. *Materials* **13**, 5362 (2020).
38. Jo, S., Kim, T., Iyer, V. G. & Im, W. Charmm-gui: a web-based graphical user interface for charmm. *J. Comput. Chem.* **29**, 1859–1865 (2008).
39. Wrapp, D. et al. Cryo-em structure of the 2019-ncov spike in the prefusion conformation. *Science* **367**, 1260–1263 (2020).
40. Wu, F. et al. A new coronavirus associated with human respiratory disease in china. *Nature* **579**, 265–269 (2020).
41. Shang, J. et al. Cell entry mechanisms of sars-cov-2. *Proc. Natl. Acad. Sci. USA* **117**, 11727–11734 (2020).
42. Woo, H. et al. Developing a fully glycosylated full-length sars-cov-2 spike protein model in a viral membrane. *J. Phys. Chem. B* **124**, 7128–7137 (2020).
43. Ko, J., Park, H. & Seok, C. Galaxytbm: template-based modeling by building a reliable core and refining unreliable local regions. *BMC Bioinform.* **13**, 1–8 (2012).
44. Ko, J. et al. The falc-loop web server for protein loop modeling. *Nucleic Acids Res.* **39**, W210–W214 (2011).
45. Watanabe, Y., Allen, J. D., Wrapp, D., McLellan, J. S. & Crispin, M. Site-specific glycan analysis of the sars-cov-2 spike. *Science* **369**, 330–333 (2020).
46. Shajahan, A., Supekar, N. T., Gleinich, A. S. & Azadi, P. Deducing the n-and o-glycosylation profile of the spike protein of novel coronavirus sars-cov-2. *Glycobiology* **30**, 981–988 (2020).
47. Lee, J. et al. Charmm-gui input generator for namd, gromacs, amber, openmm, and charmm/openmm simulations using the charmm36 additive force field. *Biophys. J.* **110**, 641a (2016).
48. Huang, J. et al. Charmm36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **14**, 71–73 (2017).
49. Hopkins, C. W., Le Grand, S., Walker, R. C. & Roitberg, A. E. Long-time-step molecular dynamics through hydrogen mass repartitioning. *J. Chem. Theory Comput.* **11**, 1864–1874 (2015).
50. Case, D. A. *et al.* Amber 2021 (2021).

51. Berendsen, H. J., Postma, J. v., Van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).
52. Martyna, G. J., Tobias, D. J. & Klein, M. L. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **101**, 4177–4189 (1994).
53. Wang, J. et al. Gaussian accelerated molecular dynamics: Principles and applications. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **11**, e1521 (2021).
54. Feller, S. E., Zhang, Y., Pastor, R. W. & Brooks, B. R. Constant pressure molecular dynamics simulation: The langevin piston method. *J. Chem. Phys.* **103**, 4613–4621 (1995).
55. Essmann, U. et al. A smooth particle mesh ewald method. *J. Chem. Phys.* **103**, 8577–8593 (1995).
56. Hess, B., Bekker, H., Berendsen, H. J. & Fraaije, J. G. Lincs: a linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).
57. Humphrey, W., Dalke, A. & Schulten, K. Vmd: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
58. Wolfram Research, Inc. Mathematica, Version 14.0. Champaign, IL, 2024.

Acknowledgements

The authors thank the Department of Science and Technology -Philippines for funding the project "Technical Battle Against COVID: Project 1. Design and Synthesis of SARS-COV2, ACE2, SARS- COV2:ACE2 Destabilizing Compounds" (DOST-PCHRD-UPD-8432) and to the DOST-Advanced Science and Technology Institute (DOST-ASTI) as well as the University of the Philippines Diliman Computational Science Research Center (UP-CSRC) for providing the computational resources used in this study. The authors also thank Sean Julian C. Fortuna of the National Institute of Physics, UP Diliman for his valuable assistance in the visuals of this study.

Author contributions

R.L.C. performed the GaMD simulations, analysis, and figure generation. J.P.L. helped with the analysis and graphical abstract. M.A.F. helped with the analysis and figure generation. D.V.R. helped with the analysis. E.d.R.H. helped with the analysis. R.B.N. conceived the original idea of this work and supervised the project. All authors discussed the results, contributed to the final manuscript, and approved its contents for submission.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-85153-5.

Correspondence and requests for materials should be addressed to R.B.N.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.