Allosteric modulation by the fatty acid site in the glycosylated SARS-CoV-2 spike

A. Sofia F. Oliveira,[a,b*] Fiona L. Kearns,[c] Mia A. Rosenfeld,[c] Lorenzo Casalino,[c] Imre Berger,[b,e,f] Christiane Schaffitzel,[e] Andrew D. Davidson,[g] Rommie E. Amaro[d] and Adrian J. Mulholland[a,b*]

[a] Centre for Computational Chemistry, School of Chemistry, University of Bristol, Bristol BS8 1TS, UK

[b] School of Chemistry, University of Bristol, Bristol BS8 1TS, UK

[c] Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, CA, 92093 USA

[d] Department of Molecular Biology, University of California San Diego, La Jolla, CA, 92093 USA

[e] School of Biochemistry, University of Bristol, Bristol BS8 1TD, UK

[f] Max Planck Bristol Centre for Minimal Biology, School of Chemistry, Bristol BS8 1TS, UK.

[g] School of Cellular and Molecular Medicine, University of Bristol, University Walk, Bristol, BS8 1TD, UK

*Corresponding authors: adrian.mulholland@bristol.ac.uk and sofia.oliveira@bristol.ac.uk

Abstract

The trimeric spike protein plays an essential role in the SARS-CoV-2 virus lifecycle, facilitating virus entry through binding to the cellular receptor angiotensin-converting enzyme 2 (ACE2) and mediating viral and host membrane fusion. The SARS-CoV-2 spike contains an allosteric fatty acid (FA) binding site at the interface between two neighbouring receptor-binding domains. This site, also found in some other coronaviruses, binds free fatty acids such as linoleic and oleic acid, and other small molecules. Understanding allostery and how this site modulates the behaviour of different regions in this protein could potentiate the development of promising alternative strategies for new coronavirus therapies. Here, we apply dynamical nonequilibrium molecular dynamics (D-NEMD) simulations to investigate allosteric effects and identify the communication pathways in the fully glycosylated spike in the original SARS-CoV-2 ancestral variant. The results reveal the allosteric networks that connect the FA site to important functional regions of the protein, including some more than 40 Å away. These regions include the receptor binding motif, an antigenic supersite in the N-terminal domain, the furin cleavage site, the regions surrounding the fusion peptide and a second allosteric site known to bind heme and biliverdin. The networks identified here highlight the complexity of the allosteric modulation in this protein and reveal a striking and unexpected connection between different allosteric sites. Notably, 65% of amino acid substitutions, deletions and insertions in the Alpha, Beta, Delta, Gamma and Omicron variants map onto or close to the identified allosteric pathways.

Main paper

The SARS-CoV-2 virus, like other β-coronaviruses, uses the spike protein to mediate virus entry into host cells. The spike is a trimeric glycoprotein embedded in the virus envelope. During the initial stage of the SARS-CoV-2 infection process, the spike binds to host target cells, primarily *via* the receptor angiotensin-converting enzyme 2 (ACE2) (1, 2), but it can also bind to other targets, such as neuropilin-1 (3, 4), estrogen receptor α (5) and potentially to nicotinic acetylcholine receptors (6-8) and sugar receptors (9). Given its crucial role in the infection process and the fact that it is one of the main targets for virus neutralisation, the spike has emerged as one of the most important targets for developing COVID-19 therapies and vaccines (e.g. (10-14)).

Each monomer of the SARS-CoV-2 spike is composed of two subunits (S1 and S2), which contain the structural motifs that directly bind to the host receptors as well as those needed for the membrane fusion process (15-17). The spike also contains three fatty acid (FA) binding sites at the interfaces between neighbouring receptor binding domains (RBDs) (18) (Figure 1A). Each FA binding site is defined by a hydrophobic pocket formed by two RBDs, with one RBD providing the aromatic and hydrophobic residues to accommodate the FA hydrocarbon tail and the other providing the polar and positively charged residues that bind the FA carboxylate headgroup (Figure 1B). The essential FA linoleic acid (LA) binds (as linoleate) with high affinity in the FA pocket, stabilising the spike in a locked conformation, in which the RBDs are all 'down' with the receptor-binding motifs (RBMs) occluded inside the trimer, and thus inaccessible for binding to ACE2 (18). The discovery of this allosteric site inspired the development of new spike-based therapies based on FAs or other natural, repurposed, or specifically designed small molecules able to bind to the FA site (10, 19-21). Several cryo-EM structures of the SARS-CoV-2 spike in complex with small molecules, such as linoleic, oleic, and all-trans retinoic acid and SPC-14, bound to the FA site, are now available (10, 18-20). Following this discovery, equivalent FA sites have been identified in several closely related coronavirus spikes (18, 22, 23). Surface plasmon resonance experiments and cryo-EM structures show that the FA site is conserved in the spike proteins of highly pathogenic β-coronaviruses, such as SARS-CoV, MERS-CoV, SARS-CoV-2, but not in the spikes of common, mild disease-causing β-coronaviruses (24).

Simulations using the dynamical-nonequilibrium molecular dynamics (D-NEMD) approach showed that the FA site allosterically modulates the behaviour of functional motifs in both the ancestral (also known as wild type, 'early 2020', or original) spike and in several variants (25-27). These D-NEMD simulations showed that the FA site is allosterically connected to the RBM, N-terminal domain (NTD), furin cleavage site, and the region sourrounding the fusion peptide (FP) (25-27). These regions have significantly different allosteric behaviours when comparing the ancestral, Alpha, Delta, Delta plus, and Omicron BA.1 variants (25-27). They differ not only in the amplitude of the structural response of specific regions in the S1 and S2 subunits to LA removal but also in the rates at which the structural changes propagate (25-27). Some mutations (such as L18F,

T19R, G142D, E156Δ–F157Δ, T478K and P681H/R) were shown to enhance connections between important functional motifs and the FA site, but others (such as H69Δ–V70Δ, Y144Δ, and D614G) reduce them (25, 26). Nonetheless, all such previous D-NEMD simulations did not incorporate the spike's many *N*- and *O*-linked glycans. However, glycans are critical to the function of the spike, not only in protecting it from immune recognition, but also in modulating its dynamics (28-31). A crucial unresolved question, therefore, is whether and how glycosylation affects allosteric communication with the FA site within the spike protein.

The spike is heavily glycosylated with 22 *N*-linked glycosylation sites per monomer (15, 32), of which 18 are constantly modified (32-35). It also contains at least two *O*-glycosylation sites per monomer with low occupancy (34, 36, 37). This thick glycan coating plays a crucial role in shielding (30, 31) and infection (28-31). For example, the dense layer of weakly immunogenic complex glycans forms a shield around the spike which facilitates evasion of the host immune system (30). Glycans also impact the dynamics and stability of the essential regions of the protein, including the RBDs, and modulate binding to ACE2 (e.g. (28-30)). Here, we use D-NEMD simulations (38, 39) to characterise the response to LA removal of the fully glycosylated SARS-CoV-2 ancestral spike to investigate allosteric modulation by the FA site and the effects of glycans on allosteric behaviour. D-NEMD simulations (38, 39) have proved to be an effective approach to investigate responses of receptors (40-42), identify allosteric effects (25-27, 43, 44) and study effects of pH changes (45) in biomolecular systems, including SARS-CoV-2 targets (25-27, 43, 45).

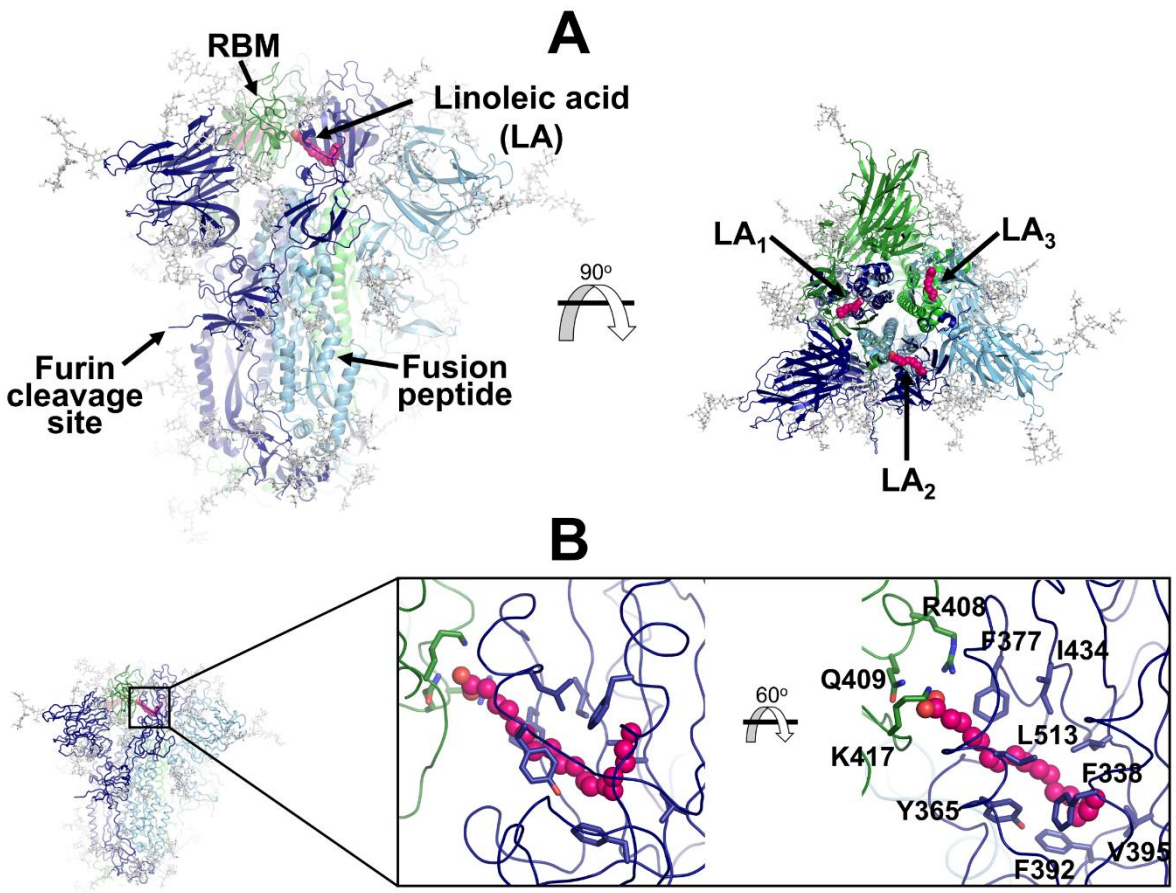

Figure 1. **Structure of the glycosylated head region of the ancestral SARS-CoV-2 spike with linoleate bound to the free fatty acid (FA) binding site.** (**A**) Model of the ectodomain of the glycosylated SARS-CoV-2 spike with linoleate (LA) bound. The spike-LA complex model was built using the cryo-EM structure 7JJI as a reference (23). Each monomer in the spike homotrimer is shown in a different colour: dark blue, light blue and green. The glycans are indicated with grey sticks, and LA molecules are highlighted with magenta spheres. Three FA binding sites exist in the trimer, each located at the interface between two neighbouring monomers. Note that in this model, all three receptor-binding motifs (RBMs) are in the 'down' conformation, and the protein is cleaved at the furin recognition site at the S1/S2 interface. (**B**) Detailed view of the FA binding site. This hydrophobic site is formed by two RBDs, with one providing the hydrophobic pocket for the FA hydrocarbon tail and the other providing polar (Q409) and positively charged (R408 and K417) residues to bind the negatively charged FA headgroup.

We performed extensive equilibrium MD simulations, followed by hundreds of D-NEMD simulations, to analyse the response of the fully glycosylated, cleaved (at the furin recognition site) spike to LA removal. The model of the locked state (with all NDBs down) with LA bound remained stable over the equilibrium simulations (Figure S1A), showing structural convergence after ~50 ns and minimal secondary structure loss after 750 ns (Figure S1B). Principal component analysis was performed to check the equilibration and sampling of the replicates (6-8) (Figure S1C). All LA molecules remained stably bound during the equilibrium simulations (Figures S2A and S2D), with the carboxylate head group of LA making consistent salt-bridge interactions with K417 and occasional interactions with R408 (Figures S2B and S2C).

The dynamic behaviour of the glycans over the equilibrium trajectories was also analysed (Figures S3-S4). As previously observed (30), the glycans are very flexible, exhibiting diverse levels of motion depending on their sequence, branching and solvent exposure (Figure S3). Generally, the *N*-glycans located in the NTD show higher fluctuations than those of the RBD (Figure S3). The glycan linked to N331 from chain C is an exception, with one of the largest RMSF values. The *O*-glycans connected to T323 and S325, close to the RBD, are less flexible than the *N*-glycans (Figure S3). The highly dynamic profile of the glycans (Figure S4) helps the spike to evade the host immune response by masking immunogenic epitopes, thus preventing them from being targeted by the host's neutralising antibodies. To quantify the shielding effect, the spike accessible surface area (ASA) covered by the glycans was determined for probe radii ranging from 0.14 (approximate radius of a water molecule) to 1.1 (approximate radius of a small antibody molecule) nm. As can be seen in Figures S4B and S4C, consistently with previous findings reported by Casalino *et al.* (30) for the closed state (with all NBDs in the 'down' conformation), the spike head has a thick glycan shield, which covers ~60% of the protein accessible area for a 1.0-nm-radius probe and restricts the binding of medium size molecules to the protein. However, small molecules (probes with a radius 0.14-0.3nm) can potentially penetrate the shield more easily as it only covers ~26% of the area of the protein accessible to smaller probes (Figure S4C).

An ensemble of 210 conformations (70 configurations per replicate) was extracted from the equilibrium MD simulations and used as starting points for the D-NEMD simulations, which investigated the effect of LA removal (Figure S5). The perturbation used here, the instantaneous removal of LA from the FA sites, is the

same as in our previous work (25-27). This perturbation is designed to force the system out of equilibrium, creating the driving force necessary for the conformational changes of interest to occur. LA removal from the FA sites triggers the structural response of the protein as it adapts to an empty FA site. Analysis of the evolution of the structural changes reveals the mechanical and dynamical coupling between the structural elements involved in response to LA removal and identifies the allosteric pathways connecting the FA site to the rest of the protein. The evolution of the structural response of the protein is extracted using the Kubo-Onsager relation (38, 39, 46, 47) from the difference between the equilibrium and nonequilibrium trajectories at equivalent points in time (Figure S5). The response obtained for each pair of equilibrium and nonequilibrium simulations is then averaged over all 210 trajectories, hence reducing noise (38, 39), and allowing the statistical significance of the responses to be assessed from the standard error of the mean (Figure S6-S8) (38).

LA removal initiates a complex chain of structural changes that are, over time, propagated within the protein. The deletion of the LA molecules immediately triggers a structural change in the FA site, which contracts as the sidechains of the residues lining it move closer to each other, filling the space once occupied by the LA molecule (Figure S9). The changes in the FA site are then swiftly transmitted to well defined regions of the protein, notably NTD, RBM and FP-surrounding regions (Figures 2 and S10-S11).

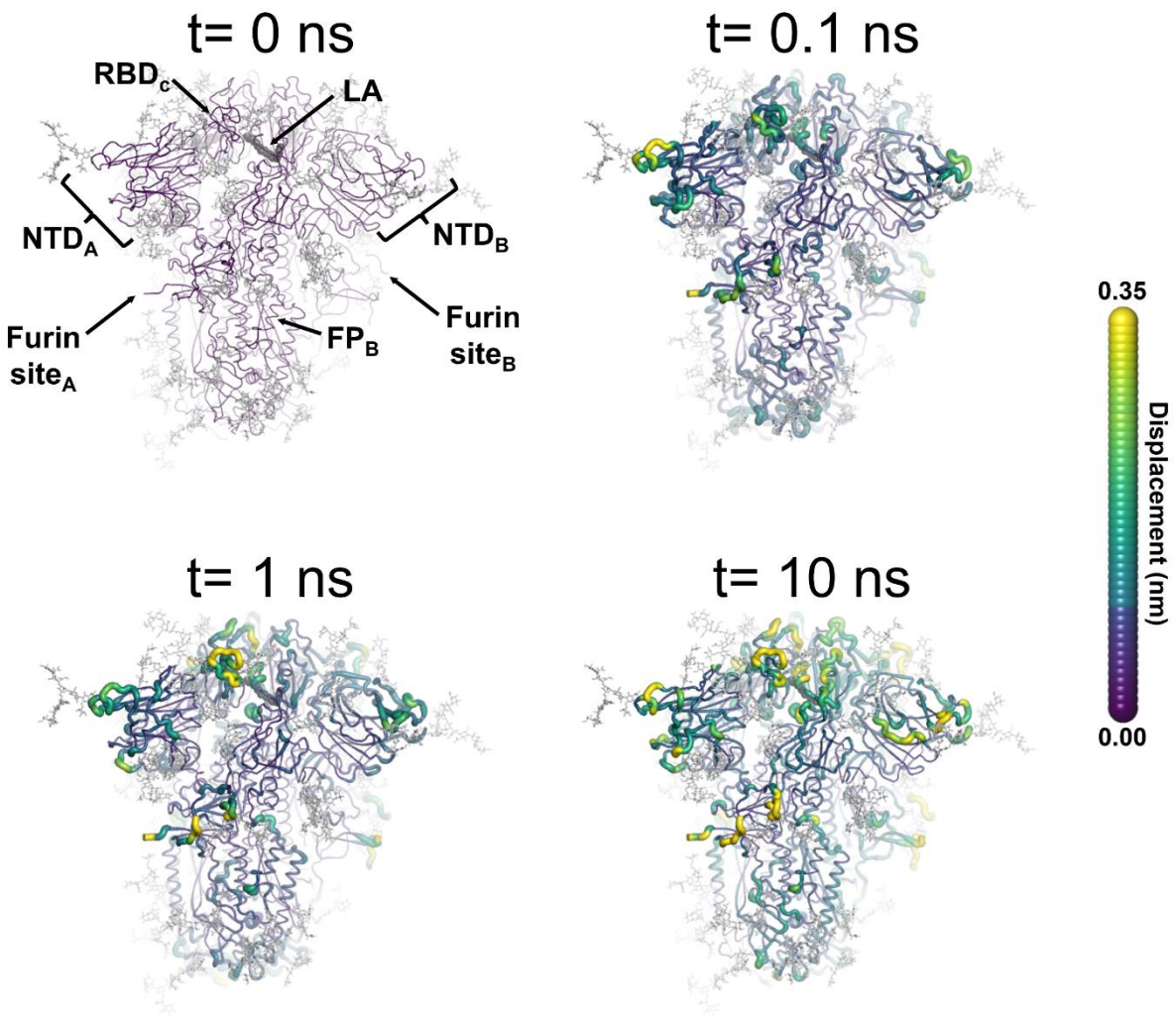

Figure 2. Structural response of the glycosylated spike to LA removal. The average Cα displacements 0.1, 1 and 10 ns after LA removal from the FA binding sites are shown, mapped onto the starting structure for the equilibrium simulations. The norm of the average Cα displacement vector between the D-NEMD apo and equilibrium LA-bound simulations was calculated for each residue using the Kubo-Onsager relation (38, 39, 46, 47). The final displacement values are the averages obtained over the 210 pairs of simulations (Figures S6-S8). The cartoon thickness and structure colours (scale on the right) indicate the average Cα-positional displacement. Each RBD, NTD, furin site and FP are subscripted with their chain ID (A, B or C). Glycans are shown as light grey sticks, whereas the dark grey spheres highlight the position of the LA molecule. The FA site shown in this figure corresponds to FA site 1, which is located at the interface between chains C and A (see Figures S10-S11 for the responses of the other two FA sites, which are generally similar).

Despite differences in the amplitude of the structural responses, the cascade of events observed here for the fully glycosylated ancestral spike is similar to that of the non-glycosylated protein (25-27) (Figure S12). Figure S12 shows a strong positive correlation between the responses obtained for the non-glycosylated and glycosylated spikes, with the largest differences in the responses between our current and previous D-NEMD simulations being around the furin recognition site. This is expected, because this site is cleaved in the glycosylated simulations and uncleaved in the non-glycosylated ones. The furin cleavage/recognition site is a

polybasic four-residue insertion located on a solvent-exposed loop at the S1/S2 junction (15, 16). This site is important for the activation of the spike (48), and its presence affects viral infectivity (e.g. (26, 48-51)).

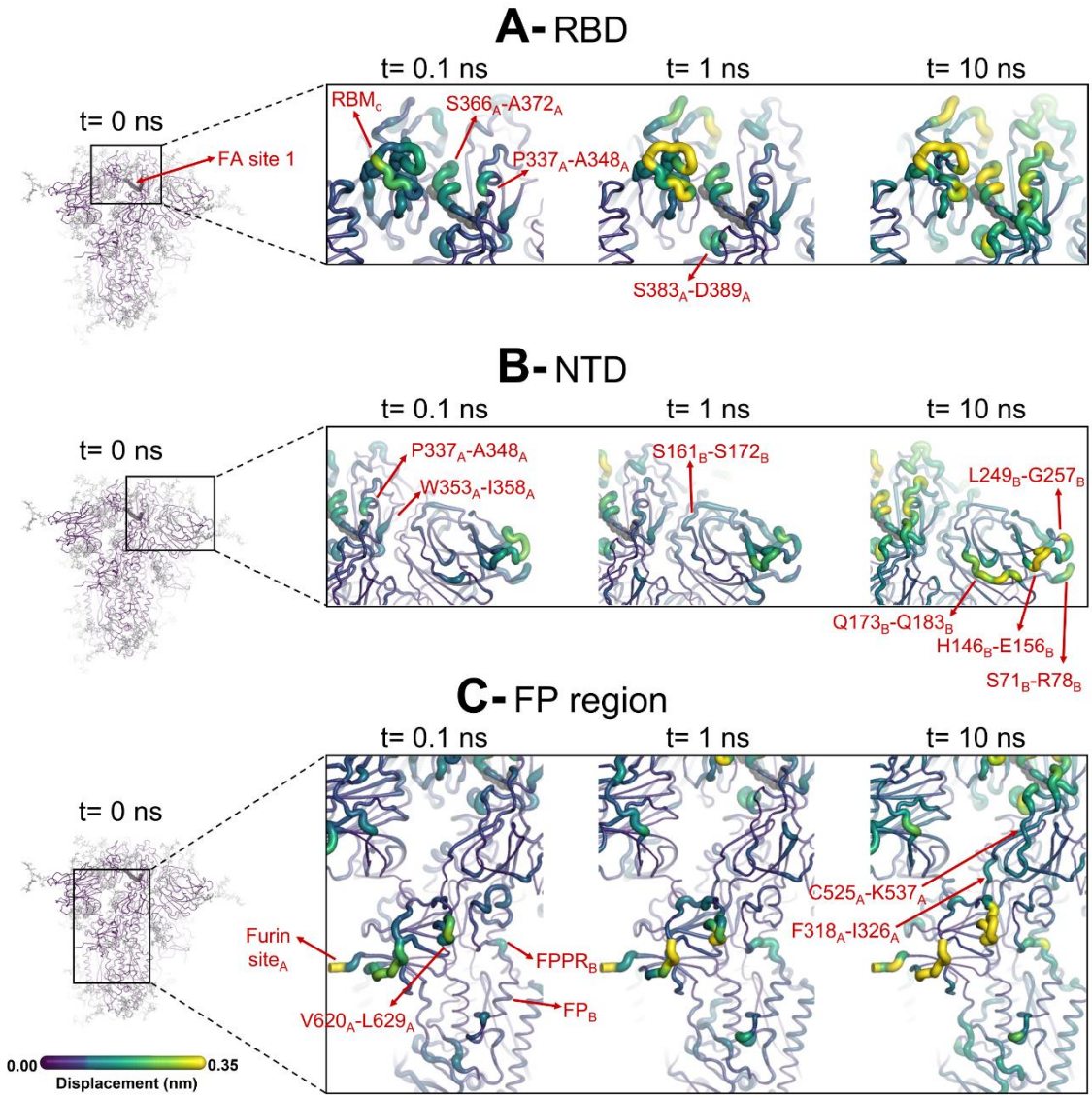

Figure 3. Structural responses of functional regions of the spike. Close-up view of the structural response of the RBD (**A**), NTD (**B**) and FP surrounding regions (**C**) to LA removal. The FA site shown here is FA site 1, located at the interface between chains C and A (see Supplementary Figures S13-S14 for the responses of the other two FA sites, which are similar). Structure colours and cartoon thickness indicate the average Cα displacement values. Each region is subscripted with its chain ID (A, B or C). The dark grey spheres highlight the FA binding site. In the images representing the spike at *t*=0 ns (left side images **A**, **B** and **C**), the glycans are shown as light grey sticks, whereas the dark grey spheres highlight the FA binding site. Glycans are omitted from the panels showing the responses at *t*=0.1, 1 and 10 ns, but are present in the simulations. For more detail, see the legend of Figure 2.

The evolution of the response of the spike to LA removal reveals the pathways through which structural changes propagate from the FA site to functional motifs (e.g. motifs involved in membrane fusion and antigenic epitopes (Figures 3 and S13-S14)). The structural changes induced by LA removal, which start in

the FA site (mainly in the P337-A348 and S366-A372 regions of one monomer and T415-K417 of the other one), are rapidly transmitted to the rest of the RBD. The R454-K458 region is particularly important because it mediates the transmission of the structural changes to the A475-C488 segment in the RBM.

In addition to the amplitude of the structural changes induced by LA removal, the average direction of the motion can also be computed by determining the average displacement vector of Cα atoms (43) between the equilibrium and nonequilibrium trajectories at equivalent time points (Figures 4 and S16-S19). Indeed, the amplitude of the structural changes in Figures 2-3, S10-S11 and S13-S14 corresponds to the norm of the average Cα displacement vector. Upon LA removal, the spike regions that form the FA site show responses with well-defined directions. These segments include P337-A348 and S366-A372, regions that contain residues whose side chains form the FA pocket (Figures 4 and S16-S19). Soon after LA deletion, P337-A348 and S366-A372 move inwards towards the FA site. These motions collectively reflect the contraction of the FA site upon LA removal. The directions of RBM motions are more diverse, with two of the RBMs (namely RBM_B and RBM_C) displaying an upward movement and the third showing an opposite downward one (RBM_A) at t=10 ns (Figures 4 and S16-S19).

SARS-CoV-2 spike has two glycans located on the RBD, notably *N*-glycans at position N331 and N343 (32). The N343 glycan is particularly interesting because it is situated immediately after one of the regions that transmits structural changes from the FA site, namely P337-F342. As well as being close to the FA site, this glycan can also directly bridge the two neighbouring RBDs (Figure S15), which may strengthen the connection between the FA site and these regions. N343 has been shown to play a role in the RBD opening mechanism by acting as a gate for the change from the ‘down’ to the ‘up’ conformation (29).

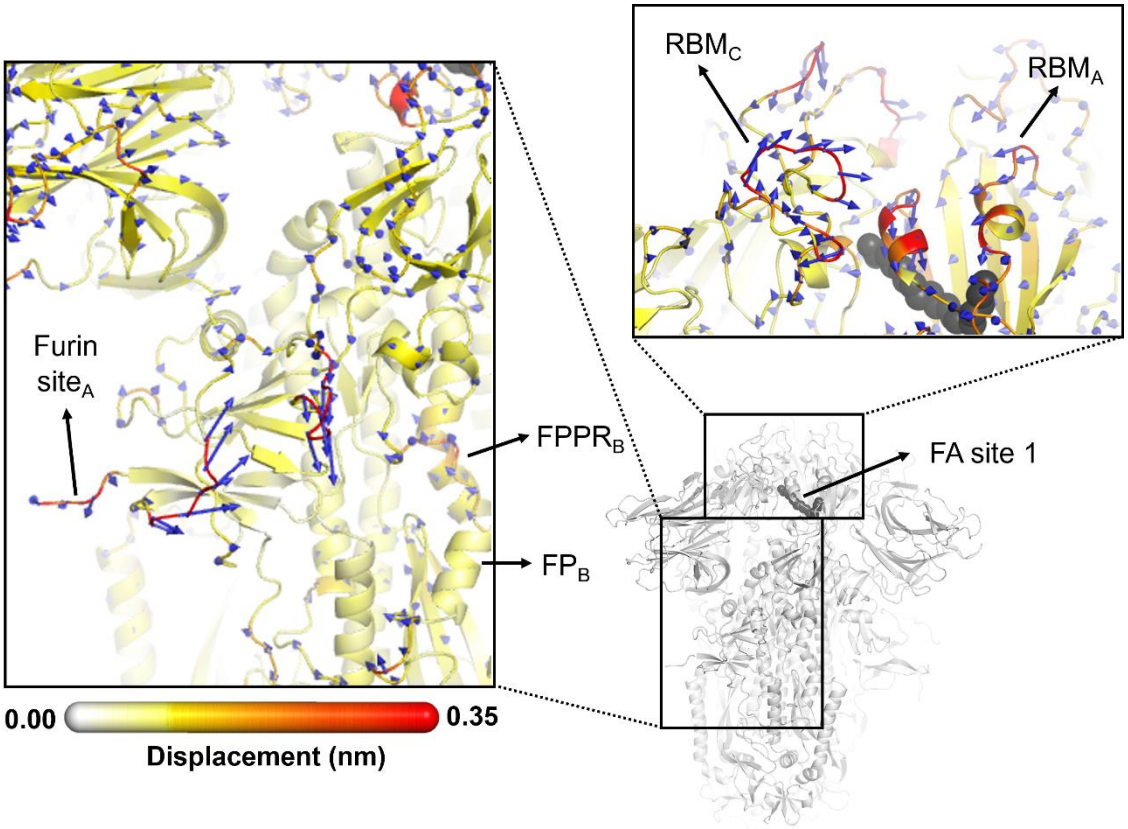

Figure 4. Direction of the structural responses of the RBD and FP-surrounding regions to LA removal. The average Cα displacement vectors at t=10 ns are shown. These vectors were determined by averaging Cα displacement vectors between the equilibrium and nonequilibrium trajectories over the 210 replicas. Vectors with a length ≥0.1 nm are displayed as blue arrows with a scale-up factor of 10. The average displacement magnitudes are represented on a white-yellow-orange-red scale. The dark grey spheres represent the FA site. This figure shows the direction of the responses around FA site 1, which is located at the interface between chains C and A (see Supplementary Figures S18-S19 for the direction of the motions in the other two FA sites).

The structural changes induced by LA removal are also swiftly propagated to the NTD via P337-F342, W353-I358, and C161-P172 (Figures 3 and S13-S14). The response of the protein, which starts in the P337-A348 segment in the FA site, is transmitted to W353-I358, C161-P172 and then to several antigenic epitopes located in the periphery of the NTD (52). The external regions showing high displacements, namely S71-R78, H146-E156, Q173-Q183 and L249-G257, are all part of an antigenic supersite in the distal-loop region of the NTD (53). In particular, the GTNGTKR motif in S71-R78, besides being an antigenic epitope, has also been suggested to be involved in binding to other receptors, such as sugar receptors (9).

The conformational response of the Q173-Q183 region is of particular interest. The structural response initiated in the FA site propagates through the NTD, reaching the Q173-Q183 segment (Figures 3 and 5). This region, which is located in the distal face of the NTD, forms the entrance of an allosteric site (Figure 5) that has been shown to bind heme (54) and its metabolite biliverdin (55). X-ray crystallography, cryo-EM and mutagenesis experiments together with modelling show that heme and biliverdin bind to a deep cleft in the NTD (54, 55) gated by the Q173-Q183 flexible loop (Figure 5B). Physiological concentrations of biliverdin

suppress binding of some neutralising antibodies to the spike (55). Such data suggested a new mode of immune evasion of the spike via the allosteric effect of biliverdin/heme binding (55). In our simulations, two of the three Q173-Q183 regions (in chains A and B) show well-defined outward motions in response to LA removal (Figures 5 and S20). Our results show a clear connection between the FA and biliverdin/heme allosteric sites via internal conformational motions. This connection captured by the D-NEMD approach is remarkable and illustrates the complexity of the potential allosteric modulation of the spike. These results suggest that the presence of heme or its metabolite in the NTD site will affect the internal networks and how dynamic and structural changes are transmitted to and from the FA site. The presence of heme/biliverdin may modulate the response of the spike to fatty acids and vice-versa, potentially affecting the rate and/or affinity of binding of the molecules to their respective allosteric sites. This apparent connection between allosteric sites may be worthy of experimental investigation.

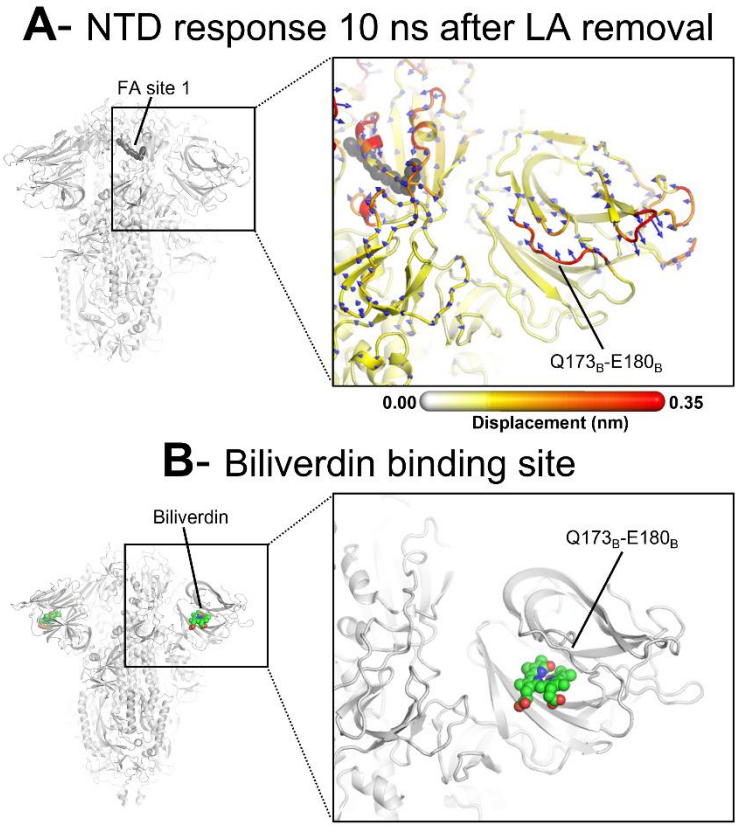

Figure 5. D-NEMD analysis of displacement vectors shows connection between the FA site and the heme/biliverdin binding site in the NTD. (**A**) View of the NTD 10 ns after LA removal, focusing on the heme/biliverdin binding site (54, 55) (which is not occupied in the simulations here). Note that the Q173-Q183 segment, which contains residues forming the heme/biliverdin binding site, shows an outward motion upon LA removal. The magnitudes of the displacements are represented on a white-yellow-orange-red colour scale. Vectors with a length ≥0.1 nm are displayed as blue arrows with a scale-up factor of 10. The dark grey spheres represent the FA site. This figure shows the direction of the structural responses around FA site 1 (see Figure S20 for the direction of the motions in the other two FA sites). (**B**) Cryo-EM structure showing the biliverdin binding site in the NTD (PDB code: 7NT9)(55). The protein is coloured in grey. The biliverdin molecules are shown with spheres.

There are eight *N*-glycans on the NTD, linked to N17, N61, N74, N122, N149, N165, N234 and N282 (32). From these, four (N74, N122, N149 and N165) are located in or close to the segments that respond to FA site occupancy (Figure S15). N74, N122 and N149 are involved in shielding the spike protein from the host immune system (30), whereas the role of N165 and N234 goes beyond shielding: it is involved in the stabilisation of the RBD 'up' conformation (30).

The structural changes induced by LA removal are not restricted to the RBD and NTD: they are transmitted to several regions far from the FA site, notably the furin and S2' cleavage sites, V620-L629 loop and FP-surrounding regions (Figures 3 and S13-S14). The structural responses starting in the FA site are quickly propagated downwards to the furin cleavage sites and V620-L629 loop via the F318-I326 and C525-K537 segments (Figures 4 and S17-S19). The furin cleavage site, which harbours a polybasic motif containing multiple arginine residues, is located at the boundary between the S1 and S2 subunits and more than 4 nm away from the FA site. This site is important for the protein's activation, and its removal reduces virus infectivity (26, 48, 50, 56). The addition of extra positively charged residues near the furin cleavage site, as observed in several variants, has been suggested to increase proteolytic processing (57), and has been shown to increase the rate of binding and the affinity of glycosaminoglycans such as heparin and heparan sulfate to this area (58). D-NEMD simulations also show that the addition of extra positive flanking charges, which is observed in some variants (such as P681R in Delta and N679K in Omicron), strengthens the allosteric connection between the FA and furin cleavage site (25). Overall, the allosteric effects are qualitatively similar to what was previously observed for the non-glycosylated spike (25-27): the same regions are connected to the FA site and are significantly affected by ligand removal from this site. This similarity indicates that the presence of glycans on the exterior of the protein does not substantially affect the internal allosteric communication pathways within the spike.

The structural changes starting in the FA site are transmitted to the furin cleavage site and V620-L629, and from there, over time, propagated laterally to the FP proximal region (FPPR) and S2' cleavage site (Figures 4 and S17-S19). Besides being a known epitope for neutralising antibodies (59, 60), the S2' cleavage site is also crucial for infection (51, 61). This proteolytic site is located immediately before the hydrophobic FP in the S2 subunit, and its cleavage is mediated by the transmembrane protease serine 2 (TMPRSS2) after binding to the host receptor (61). The FPPR is located after the FP, and it is thought to have a functional role in membrane fusion by mediating the transitions between pre and post-fusion structures of the protein (17). Upon the removal of LA, the residues of the FPPR in direct contact with the FP show a well-defined response upwards towards the C-terminal domain 1 (CTD1) in two of the chains, namely chains B and C, and an outwards motion in chain A. The CTD1 has been suggested (based on cryo-EM structures) to be a structural relay between RBD and FP, sensing the displacement on either side (17)) (Figures 4 and S16-S19). Interestingly, the chains displaying an FPPR motion towards CTD1, notably chains B and C, are the same ones that also

exhibit an RBM upward movement towards the solvent. The direction of the S2' motion observed in the D-NEMD simulations is diverse, with two of the sites ($S2'_A$ and $S2'_B$) displaying a motion towards the FP and one showing an opposite movement away from the FP ($S2'_C$) after t = 10 ns (Figures 4 and S16-S19).

The spike contains several complex *N*- and *O*-glycans in or close to the furin and S2' cleavage sites, FPPR, F318-I326, C525-K537 and V620-L629 (32). All three monomers contain one *O*- and two *N*-glycans (at positions and T323, N616 and N657) close to the pathway that connects that FA site to the furin cleavage site and FP surrounding regions (Figure S15). Monomer A also contains an additional *O*-glycan linked to S323 (Figure S15). Notably, interactions between the *O*-glycans S323 and T325 and the *N*-glycan at N234 create a direct connection between the NTD and the F318-I326 region of the same monomer (Figure S15). This glycan 'link' may facilitate and enhance the transmission of structural changes within an individual monomer. The glycan at position N234 has also been suggested to play a mechanical role in the spike working mechanism by helping to stabilise the RBD in the 'up' conformation (30).

To assess whether substitutions, deletions and insertions as seen in the variants lie on the allosteric communication pathways identified here, we overlapped, in the same 3D structure, the spatial positions of the variations with the allosteric networks (Figure S21). In Figure S21, substitutions, deletions and insertions within the allosteric pathways are shown with red spheres, and those close (<0.6 nm away) to the pathways are highlighted in dark blue. The results are interesting: overall, 22 out of the 77 amino acid positions per chain known to vary in the Alpha, Beta, Gamma, Delta and Omicron (BA.1, BA.2, BA.4, BA.5, BQ.1.1 and XBB.1.5) variants directly map onto the allosteric communication pathway identified using the D-NEMD approach. Additionally, a further 28 out of the 77 variations are in direct contact with, or very close proximity to, these networks. H655Y (present in Gamma and all Omicron sub-variants), T547K (in Omicron BA.1), D614G (in Alpha, Betta, Gamma, Delta and all Omicron sub-variants), W856 (in Omicron BA.1) and S982A (in Alpha) are all examples of mutations adjacent to the communication pathways: they modulate the connection to the FA site. These mutations are responsible for the differences in allosteric behaviour observed between SARS-CoV-2 variants. As future variants emerge, it will be of interest to establish if more mutations lie along the allosteric pathways described here. Differences in allosteric behaviour and regulation in the spike are likely to be of functional relevance and useful in understanding differences between SARS-CoV-2 variants.

Conclusions

D-NEMD simulations show important allosteric effects in the fully glycosylated ancestral SARS-CoV-2 spike. We identify the pathways that link the FA site with functional regions for membrane fusion, antigen antibody recognition and allosteric modulation in the fully glycosylated, cleaved spike. These simulations highlight the structural responses resulting from LA removal and demonstrate the complex connection between the FA site and the RBM, an antigenic supersite in the NTD, the allosteric heme/biliverdin binding site, the furin and S2'

cleavage sites and the FP-surrounding regions. The ordering of the conformational changes observed in response to LA removal from the FA sites identifies the initial steps of communication between the different secondary structure elements in the protein, thus constituting the pathways enabling effective signal propagation within the protein. Most tellingly, more than 65% of the substitutions, deletions and additions in the Alpha, Beta, Gamma, Delta, and Omicron variants are located either in or close to the allosteric pathways identified using the D-NEMD approach.

Our findings also show that the presence of glycans on the exterior of the protein, despite affecting the amplitude of the structural responses, does not qualitatively change the cascade of events connecting the FA site to the rest of the spike. Some glycans can influence the allosteric pathways, facilitating the transmission of the structural changes within and between monomers. For example, the interactions between the glycans linked to N234, T373 and S375 can create a direct connection between the NTD and the F318-I326 region of the same monomer, thus helping the propagation of the structural changes within the monomer.

Furthermore, our D-NEMD results reveal an unexpected connection between the FA site and a second allosteric site known to bind heme and its metabolite biliverdin. It will be of interest to understand how heme/biliverdin binding affects the dynamics and structural changes of the spike, and links to the FA site, and potentially other allosteric sites. While the effects of the apparent coupling between the heme/biliverdin site and the FA site remains to be investigated, this work has reinforced the ability of the D-NEMD approach to find allosteric sites and to map communication pathways between sites. The results here further point to complex allosteric effects in the SARS-CoV-2 spike, of potential functional relevance.

Acknowledgements

AJM and ASFO thank the Biotechnology and Biological Sciences Research Council (BBSRC grant number BB/W003449/1). ASFO was supported at the University of Bristol by Oracle for Research and the Biological and Biotechnological Sciences Research Council ([BB/X009831/1]). This work received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 101021207; project information: PREDACTED). MD simulations were carried out using the Oracle Public Cloud Infrastructure (https://cloud.oracle.com/en_US/iaas) under an award to AJM and ASFO from Oracle for Research for COVID-19 research. We thank EPSRC via HECBIOSIM (hecbiosim.ac.uk) for providing ARCHER/ARCHER2 time through a COVID-19 rapid response call. We also thank the Bristol UNCOVER Group and the University of Bristol for their support. IB and CS are investigators of the Wellcome Trust (202904/Z/16/Z, 206181/Z/17/Z, 221708/Z/20/Z). ADD is a member of the G2P-UK National Virology consortium funded by the Medical Research Council/UKRI (Grant MR/W005611/1)

Data availability statement

All D-NEMD simulation data (including input and trajectories files) are openly available from the MolSSI/BioExcel COVID-19 public data repository for biomolecular simulations of COVID proteins (https://covid.molssi.org/simulations/).

Competing interests

The authors declare competing interests. CS and IB report shareholding in Halo Therapeutics Ltd related to this Correspondence.

References

1. R. Yan *et al.*, Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* **367**, 1444-1448 (2020).
2. Q. Wang *et al.*, Structural and functional basis of SARS-CoV-2 entry by using human ACE2. *Cell* **181**, 894-904.e899 (2020).
3. J. L. Daly *et al.*, Neuropilin-1 is a host factor for SARS-CoV-2 infection. *Science* **370**, 861-865 (2020).
4. L. Cantuti-Castelvetri *et al.*, Neuropilin-1 facilitates SARS-CoV-2 cell entry and infectivity. *Science* **370**, 856-860 (2020).
5. O. Solis *et al.*, The SARS-CoV-2 spike protein binds and modulates estrogen receptors. *Sci Adv* **8**, eadd4150 (2022).
6. B. C. V. O'Brien, L. Weber, K. Hueffer, M. M. Weltzin, SARS-CoV-2 spike ectodomain targets alpha7 nicotinic acetylcholine receptors. *J Biol Chem* **299**, 104707 (2023).
7. J. Chrestia *et al.*, A Functional Interaction Between the SARS-CoV-2 Spike Protein and the Human α7 Nicotinic Receptor. *Mol Neurobiol* **59**, 6076-6090 (2022).
8. A. S. F. Oliveira *et al.*, A potential interaction between the SARS-CoV-2 spike protein and nicotinic acetylcholine receptors. *Biophys J* **120**, 983-993 (2021).
9. N. Behloul, S. Baha, R. Shi, J. Meng, Role of the GTNGTKR motif in the N-terminal receptor-binding domain of the SARS-CoV-2 spike protein. *Virus Res* **286**, 198058 (2020).
10. Q. Wang *et al.*, In Silico Discovery of Small Molecule Modulators Targeting the Achilles' Heel of SARS-CoV-2 Spike Protein. *ACS Cent Sci* **9**, 252-265 (2023).
11. D. Focosi *et al.*, Monoclonal antibody therapies against SARS-CoV-2. *Lancet Infect Dis* **22**, e311-e326 (2022).
12. L. Wang *et al.*, Discovery of potential small molecular SARS-CoV-2 entry blockers targeting the spike protein. *Acta pharmacol Sin* **43**, 788-796 (2022).

13. D. Bojadzic *et al.*, Small-Molecule Inhibitors of the Coronavirus Spike: ACE2 Protein-Protein Interaction as Blockers of Viral Attachment and Entry for SARS-CoV-2. *ACS Infect Dis* **7**, 1519-1534 (2021).
14. S. Pomplun, Targeting the SARS-CoV-2-spike protein: from antibodies to miniproteins and peptides. *RSC Med Chem* **12**, 197-202 (2020).
15. A. C. Walls *et al.*, Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* **181**, 281-292 (2020).
16. D. Wrapp *et al.*, Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260-1263 (2020).
17. Y. Cai *et al.*, Distinct conformational states of SARS-CoV-2 spike protein. *Science* **369**, 1586-1592 (2020).
18. C. Toelzer *et al.*, Free fatty acid binding pocket in the locked structure of SARS-CoV-2 spike protein. *Science* **370**, 725-730 (2020).
19. L. Tong *et al.*, A Retinol Derivative Inhibits SARS-CoV-2 Infection by Interrupting Spike-Mediated Cellular Entry. *Mbio* **13**, e0148522 (2022).
20. J. Ma *et al.*, Cryo-EM structure of S-Trimer, a subunit vaccine candidate for COVID-19. *J Virol* **95** (2021).
21. D. K. Shoemark *et al.*, Molecular simulations suggest vitamins, retinoids and steroids as ligands of the free fatty acid pocket of the SARS-CoV-2 spike protein. *Angew Chem Int Ed Engl* **60**, 7098-7110 (2021).
22. S. Zhang *et al.*, Bat and pangolin coronavirus spike glycoprotein structures provide insights into SARS-CoV-2 evolution. *Nat Commun* **12**, 1607 (2021).
23. S. Bangaru *et al.*, Structural analysis of full-length SARS-CoV-2 spike protein from an advanced vaccine candidate. *Science* **370**, 1089-1094 (2020).
24. C. Toelzer *et al.*, The free fatty acid-binding pocket is a conserved hallmark in pathogenic beta-coronavirus spike proteins from SARS-CoV to Omicron. *Sci Adv* **8**, eadc9179 (2022).
25. A. S. F. Oliveira *et al.*, SARS-CoV-2 spike variants of concern differ in their allosteric response to linoleic acid. *J Mol Cell Biol* **15,** mjad021 (2023).
26. K. Gupta *et al.*, Structural insights in cell-type specific evolution of intra-host diversity by SARS-CoV-2. *Nat Commun* **13**, 222 (2022).
27. A. S. F. Oliveira *et al.*, The fatty acid site is coupled to functional motifs in the SARS-CoV-2 spike protein and modulates spike allosteric behaviour. *Comput Struct Biotechnol J* **20**, 139–147 (2022).
28. A. M. Harbison *et al.*, Fine-tuning the spike: role of the nature and topology of the glycan shield in the structure and dynamics of the SARS-CoV-2 S. *Chem Sci* **13**, 386-395 (2022).
29. T. Sztain *et al.*, A glycan gate controls opening of the SARS-CoV-2 spike protein. *Nat Chem* **13**, 963-968 (2021).

30. L. Casalino *et al.*, Beyond shielding: the roles of glycans in the SARS-CoV-2 spike protein. *ACS Cent Sci* **6**, 1722-1734 (2020).
31. B. Turoňová *et al.*, In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. *Science* **370**, 203-208 (2020).
32. Y. Watanabe, J. D. Allen, D. Wrapp, J. S. McLellan, M. Crispin, Site-specific glycan analysis of the SARS-CoV-2 spike. *Science* **369**, 330-333 (2020).
33. Y. Watanabe *et al.*, Native-like SARS-CoV-2 Spike Glycoprotein Expressed by ChAdOx1 nCoV-19/AZD1222 Vaccine. *ACS Cent Sci* **7**, 594-602 (2021).
34. A. Shajahan, N. T. Supekar, A. S. Gleinich, P. Azadi, Deducing the N- and O-glycosylation profile of the spike protein of novel coronavirus SARS-CoV-2. *Glycobiology* **30**, 981-988 (2020).
35. H. Yao *et al.*, Molecular architecture of the SARS-CoV-2 virus. *Cell* **183**, 730-738.e713 (2020).
36. I. Bagdonaite *et al.*, Site-Specific O-Glycosylation Analysis of SARS-CoV-2 Spike Protein Produced in Insect and Human Cells. *Viruses* **13** (2021).
37. M. Sanda, L. Morrison, R. Goldman, N- and O-Glycosylation of the SARS-CoV-2 Spike Protein. *Anal Chem* **93**, 2003-2009 (2021).
38. A. S. F. Oliveira, G. Ciccotti, S. Haider, A. J. Mulholland, Dynamical nonequilibrium molecular dynamics reveals the structural basis for allostery and signal propagation in biomolecular systems. *Eur Phys J B* **94**, 144 (2021).
39. G. Ciccotti, M. Ferrario, Non-equilibrium by molecular dynamics: a dynamical approach. *Mol Simul* **42**, 1385-1400 (2016).
40. A. Oliveira *et al.*, Structural and temporal basis for agonism in the α4β2 nicotinic acetylcholine receptor. *bioRxiv* **DOI: 10.1101/2022.02.23.481608** (2022).
41. A. S. F. Oliveira *et al.*, Identification of the initial steps in signal transduction in the α4β2 nicotinic receptor: insights from equilibrium and nonequilibrium simulations. *Structure* **27**, 1171-1183 (2019).
42. A. S. F. Oliveira *et al.*, A general mechanism for signal propagation in the nicotinic acetylcholine receptor family. *J Am Chem Soc* **141**, 19953–19958 (2019).
43. H. T. H. Chan, A. S. F. Oliveira, C. J. Schofield, A. J. Mulholland, F. Duarte, Dynamical nonequilibrium molecular dynamics simulations identify allosteric sites and positions associated with drug resistance in the SARS-CoV-2 main protease. *J Am Chem Soc Au* **3,** 1767–1774 (2023).
44. I. Galdadas *et al.*, Allosteric communication in class A β-lactamases occurs via cooperative coupling of loop dynamics. *eLife* **10**, e66567 (2021).
45. A. Dommer *et al.*, #COVIDisAirborne: AI-Enabled Multiscale Computational Microscopy of Delta SARS-CoV-2 in a Respiratory Aerosol. *Int J High Perform C* **37**, 28-44 (2023).
46. G. Ciccotti, G. Jacucci, I. R. McDonald, Thought-experiments by molecular dynamics. *J Stat Phys* **21**, 1-21 (1979).

47. G. Ciccotti, Molecular dynamics simulation of non equilibrium phenomena and rare dynamical events in Computer simulation in material science, P. V. Meyer M, Ed. (Kluwer Academic Publishers, 1991), pp. 119–137.

48. A. D. Davidson *et al.*, Characterisation of the transcriptome and proteome of SARS-CoV-2 reveals a cell passage induced in-frame deletion of the furin-like cleavage site from the spike glycoprotein. *Genome Med* **12**, 68 (2020).

49. T. P. Peacock *et al.*, The furin cleavage site in the SARS-CoV-2 spike protein is required for transmission in ferrets. *Nat Microbiol* **6**, 899–909 (2021).

50. M. Hoffmann, H. Kleine-Weber, S. Pöhlmann, A multibasic cleavage site in the spike protein of SARS-CoV-2 is essential for infection of human lung cells. *Mol Cell* **78**, 779-784.e775 (2020).

51. M. Hoffmann *et al.*, SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* **181**, 271-280.e278 (2020).

52. X. Chi *et al.*, A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. *Science* **369**, 650-655 (2020).

53. G. Cerutti *et al.*, Potent SARS-CoV-2 neutralizing antibodies directed against spike N-terminal domain target a single supersite. *Cell Host Microbe* **29**, 819-833.e817 (2021).

54. S. Freeman *et al.*, Heme binding to the SARS-CoV-2 spike glycoprotein. *J Biol Chem* **299**, 105014 (2023).

55. A. Rosa *et al.*, SARS-CoV-2 can recruit a heme metabolite to evade antibody immunity. *Sci Adv* **7** (2021).

56. S. Matsuyama *et al.*, Enhanced isolation of SARS-CoV-2 by TMPRSS2-expressing cells. *Proc Natl Acad Sci U S A* **117**, 7001-7003 (2020).

57. G. R. Whittaker, SARS-CoV-2 spike and its adaptable furin cleavage site. *Lancet Microbe* **2**, e488-e489 (2021).

58. S. H. Kim *et al.*, SARS-CoV-2 evolved variants optimize binding to cellular glycocalyx. *Cell Rep Phys Sci* **4**, 101346 (2023).

59. L. Farrera-Soler *et al.*, Identification of immunodominant linear epitopes from SARS-CoV-2 patient plasma. *PloS one* **15**, e0238089 (2020).

60. C. M. Poh *et al.*, Two linear epitopes on the SARS-CoV-2 spike protein that elicit neutralising antibodies in COVID-19 patients. *Nat Commun* **11**, 2806 (2020).

61. M. Takeda, Proteolytic activation of SARS-CoV-2 spike protein. *Microbiol Immunol* **66**, 15-23 (2022).