



Communications

The fatty acid site is coupled to functional motifs in the SARS-CoV-2 spike protein and modulates spike allosteric behaviour

A. Sofia F. Oliveira^{a,b}, Deborah K. Shoemark^{b,c}, Amaury Avila Ibarra^d, Andrew D. Davidson^e, Imre Berger^{b,c,f}, Christiane Schaffitzel^{b,c}, Adrian J. Mulholland^{a,*}

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

^b BrisSynBio, Life Sciences Building, Tyndall Avenue, Bristol BS8 1TQ, UK

^c School of Biochemistry, University of Bristol, Bristol BS8 1TD, UK

^d Research Software Engineering, Advanced Computing Research Centre, University of Bristol, Bristol BS1 5QD, UK

^e School of Cellular and Molecular Medicine, University of Bristol, University Walk, Bristol BS8 1TD, UK

^f Max Planck Bristol Centre for Minimal Biology, Cantock's Close, Bristol BS8 1TS, UK



ARTICLE INFO

Article history:

Received 17 September 2021

Received in revised form 8 December 2021

Accepted 8 December 2021

Available online 11 December 2021

ABSTRACT

The SARS-CoV-2 spike protein is the first contact point between the SARS-CoV-2 virus and host cells and mediates membrane fusion. Recently, a fatty acid binding site was identified in the spike (Toelzer *et al.* Science 2020). The presence of linoleic acid at this site modulates binding of the spike to the human ACE2 receptor, stabilizing a locked conformation of the protein. Here, dynamical-nonequilibrium molecular dynamics simulations reveal that this fatty acid site is coupled to functionally relevant regions of the spike, some of them far from the fatty acid binding pocket. Removal of a ligand from the fatty acid binding site significantly affects the dynamics of distant, functionally important regions of the spike, including the receptor-binding motif, furin cleavage site and fusion-peptide-adjacent regions. Simulations of the D614G mutant show differences in behaviour between these clinical variants of the spike: the D614G mutant shows a significantly different conformational response for some structural motifs relevant for binding and fusion. The simulations identify structural networks through which changes at the fatty acid binding site are transmitted within the protein. These communication networks significantly involve positions that are prone to mutation, indicating that observed genetic variation in the spike may alter its response to linoleate binding and associated allosteric communication.

© 2021 Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

The COVID-19 pandemic, which is having a devastating social and economic impact worldwide, is caused by the severe acute respiratory syndrome 2 (SARS-CoV-2) coronavirus. Since the initial outbreak in late 2019, SARS-CoV-2 has caused >261 million confirmed cases of COVID-19 disease and >5.2 million deaths [1] (perhaps as many as 7–13 million [2] worldwide as of 29th November 2021). SARS-CoV-2 is an enveloped, single-stranded RNA virus that

belongs to the *Betacoronavirus* genus of the *Coronaviridae* family which includes pathogenic human coronaviruses that cause SARS severe acute respiratory syndrome) and MERS (Middle East respiratory syndrome) [3,4]. It initially infects respiratory epithelial cells by binding to the angiotensin-converting 2 enzyme (ACE2) [5,6]. When it first emerged as a human pathogen, SARS-CoV-2 was thought to cause predominantly respiratory disease, particularly pneumonia and severe acute respiratory distress syndrome [7,8]. However, it is now known that its effects are not limited to the respiratory tract: COVID-19 can cause severe inflammation and damage in other organs [9–11], including the heart, kidneys, liver and intestines, and can lead to neurological problems [12]. In common with other enveloped viruses, SARS-CoV-2 fuses its viral envelope with a host cell membrane to infect cells. Membrane attachment and fusion with the host cell is mediated by the SARS-CoV-2 spike protein, which primarily binds to the host ACE2 receptor [5,6]; however, the spike can also interact with

MD, Molecular dynamics; LA, Linoleic acid; RBM, receptor-binding motif; FA, fatty acid; SARS-CoV-2, severe acute respiratory syndrome 2; RBD, receptor binding domain; RMB, receptor binding motif; NTD, N-terminal domain; FP, fusion peptide; FPPR, fusion-peptide proximal region; HR1, heptad repeat 1; CH, central helix; CD, connector domain; ACE2, angiotensin-converting 2 enzyme; SARS, severe acute respiratory syndrome; MERS, middle east respiratory syndrome.

* Corresponding author.

E-mail address: adrian.mulholland@bristol.ac.uk (A.J. Mulholland).

neuropilin-1 [13,14] and potentially with other receptors [15,16]. The spike is a glycoprotein [17,18] and is found on the surface of the virion, in a trimeric form Fig. 1A).

Each monomer is formed of three regions: a large ectodomain, a transmembrane anchor and a short cytoplasmic tail [18]. The ectodomain comprises two subunits: S1 is responsible for binding to ACE2 [6,18], and S2 for viral-host membrane fusion [18,20]. The SARS-CoV-2 spike contains two proteolytic cleavage sites [18]: one 'furin protease recognition' site at the S1/S2 boundary, thought to activate the protein [21], and a second in the S2 subunit S₂' that releases the fusion peptide [18,20]. The SARS-CoV-2 spike contains three free fatty acid (FA) binding sites, each located at the interface between every two neighbouring receptor-binding domains (Fig. 1A and S3) [19]. The FA binding sites are lined by aromatic and hydrophobic residues (Fig. 1B) and a positively charged residue from a neighbouring monomer, namely R408, which acts as an anchor for the FA carboxylate headgroup [19]. The open spike conformation, with at least one RBD pointing upwards, is needed to interact with ACE2 receptors on the human host cell. It was shown by surface plasmon resonance that the presence of the FA linoleic acid (LA) reduces binding of the spike to ACE2 [19]. LA stabilizes the locked spike conformation, in which the receptor-binding motif (RBM) is occluded and cannot bind to the human ACE2 receptor [19], but there is no obvious connection between the FA sites and other structural motifs relevant for membrane fusion, or with antigenic epitopes. MD simulations showed persistent and stable interactions between LA and the spike trimer [19,22]. These simulations also revealed that LA rigidifies the FA binding site, and these effects extend to the N-terminus domain [22]. The cryo-EM structure of the spike from pangolin coronavirus (which is closely related to SARS-CoV-2) shows that the spike also binds LA in an equivalent FA pocket [23]. An equivalent FA binding

site has also been found on the Novavax SARS-CoV-2 construct expressed and purified from insect cells [24].

Simulations, particularly atomistic molecular dynamics (MD) simulations, have provided crucial atomic-level insight into the structure, dynamics and interactions of the SARS-CoV-2 spike [13,19,22,25–33]. Here, we apply dynamical-nonequilibrium MD simulations [34–37] to investigate the response of the SARS-CoV-2 spike to LA removal. We have shown this approach to be effective in identifying structural communication pathways in a variety of proteins, e.g. in identifying a general mechanism of interdomain signal propagation in nicotinic acetylcholine receptors [38,39] and mapping the networks connecting the allosteric and catalytic sites in two clinically relevant β -lactamase enzymes [40]. This approach is based on equilibrium simulations of the system in question, which generate configurations for multiple dynamical-nonequilibrium simulations, through which the effect of a perturbation can be studied. Running a large number of nonequilibrium simulations allows for the determination of the statistical significance of the structural response observed [37].

2. Dynamic response of the wild-type spike

A model of the locked wild-type spike was created from the cryo-EM structure (PDB code: 6ZB5) of the SARS-CoV-2 spike protein bound to three linoleate molecules [19]. Missing loops were built to generate the wild-type sequence according to the Uniprot accession number P0DTC2 for the unglycosylated ectodomain of the spike bound with LA (for details, see [Supplementary Material](#)). The locked structure had 42 disulphides per trimer. It remained intact and faithfully retained the structure and overall fold of the cryo-EM structure over the equilibrium simulation time [19,22]. Three equilibrium MD simulations (Fig. S1), 200 ns each, were per-

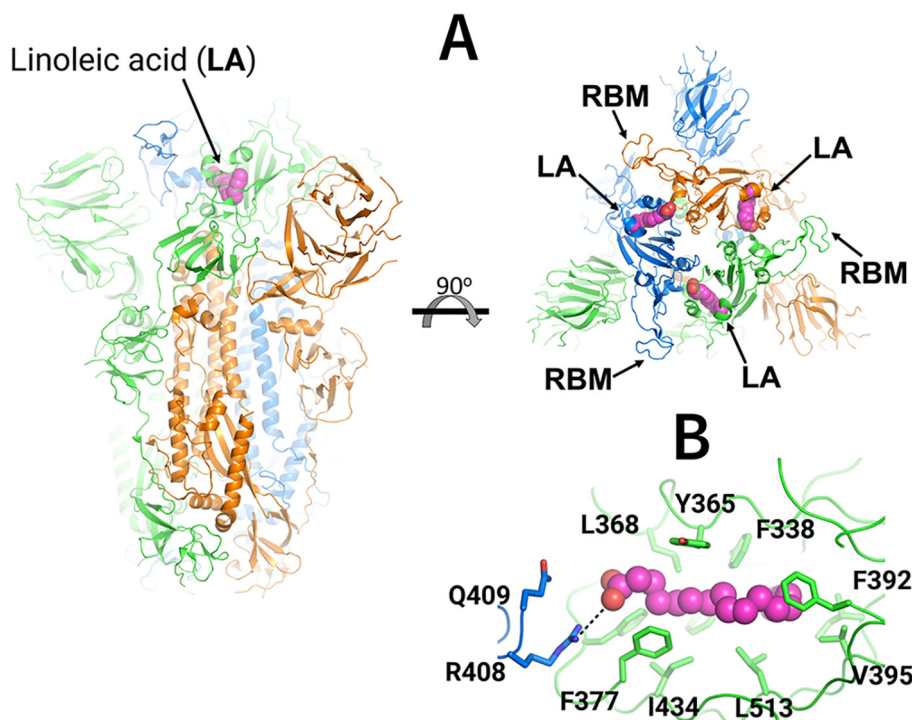


Fig. 1. Cryo-EM structure of the ectodomain of the SARS-CoV-2 spike trimer with linoleic acid (LA) bound to the fatty acid-binding sites [19]. (A) Three-dimensional structure of the complex of the locked (in which all receptor-binding motifs (RBMs) are occluded) ectodomain of the SARS-CoV-2 spike trimer with linoleic acid (PDB code: 6ZB5) [19]. The spike protein is a homotrimer [18]: each monomer is shown in a different colour, namely green, orange and blue. LA molecules are highlighted with spheres. Each fatty acid (FA) binding site is located at the interface between two neighbouring monomers, and is formed by residues from two adjacent receptor-binding domains. (B) Detailed view of the FA binding site: this pocket is lined by hydrophobic and aromatic residues, and the LA acidic headgroup is close to R408 and Q409. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

formed for the locked form of the unglycosylated and uncleaved (no cleavage at the S1/S2 interface) ectodomain of the spike bound with LA and used as starting points for 90 dynamical-nonequilibrium simulations (Fig. S2). Here, we used models of the uncleaved spike ectodomains in order to detect any potential effects on structurally distant sites influenced by ligand in the FA sites in the intact spike. In the nonequilibrium simulations, all LA molecules were (instantaneously) annihilated. This triggers a response of the protein, as it adapts to LA removal. This annihilation is carried out for multiple configurations sampled from equilibrium MD (top panel in Figure S2), and comparison between the equilibrium and short dynamical-nonequilibrium MD trajectories identifies the structural response of the protein. Running multiple (in this case, 90) dynamical-nonequilibrium simulations reduces the noise associated with the structural response of the protein and allows for the determination of the statistical significance of the observed response. Nonequilibrium simulations of this type are emerging as an effective tool to study signal transmission and identify communication networks within proteins [38–42]. Here, the direct comparison between the equilibrium LA-bound and nonequilibrium apo spike simulations using the Kubo-Onsager approach [34–37] (bottom panel in Figure S2), and the average of the results over all the 90 replicates, allows for identification of the temporal sequence of conformational changes associated with the response of the spike to LA removal (Figs. 2 and S4), and also the determination of their statistical significance (Figure S5). The structure that we simulate here corresponds to the unglycosylated wild-type spike (Uniprot accession number PODTC2), not cleaved at the ‘furin recognition/cleavage’ site. It was built based on the cryo-EM structure that originally revealed the FA binding site (PDB code: 6ZB5) [19]. Although a few glycans (e.g. at positions N165, N234, and N343) have been shown to be involved in the spike infection mechanism by altering the dynamics of receptor binding domain opening [25,43], and the glycan shield plays a vital role in the biological function of the spike, the internal networks and response of the protein scaffold identified here are not likely to be qualitatively altered by the glycans, which predominantly cover the exterior of the spike. The spike structure used for these simulations, namely the tightly packed closed structure [19], is only glycosylated on the outside, and as such, glycans are unlikely to affect the protein’s internal allosteric communication networks. As Casalino et al. have shown, glycan dynamics are fast relative to the dynamics of the protein [25]. Note that the perturbation introduced here (LA annihilation) is not intended

to mimic the physical process of LA (un)binding, but rather to promote a rapid response and force signal transmission within the protein, thus mapping the mechanical and dynamic coupling between the structural elements involved in this response. Note also that, due to the non-physical nature of the perturbation, the timescales observed for the protein’s response do not represent the physical timescales of conformational change, however, the responses of similar systems (e.g. wild-type and D614G spike) can be meaningfully compared.

Despite small variations in amplitude, the structural response to LA removal is similar for the three monomers (Figs. 2 and S4). Figs. 2 and S4 show the time evolution of the average C_{α} -positional deviation of each individual monomer in the 5 ns following LA. Comparing these figures shows that all the monomers respond similarly, with the same motifs and order of events associated with signal propagation observed for each.

The simulations show that the structural response to LA removal occurs in very specific and well-defined regions of the spike. It is striking that some functional motifs, including regions distant from the FA site, are particularly affected by LA removal (Figs. 2 and S4). Some of these conformational changes occur in solvent-exposed, flexible regions of the protein (Figs. S6–S8). This might initially be surprising, but it should be noted that flexibility means that little energy is required to induce the conformational rearrangement. This is one reason why loops often play essential roles in protein function and allostery (e.g. [44]), and are frequently involved in signal propagation in proteins (e.g. [38–41]). Upon LA removal, the response within the spike starts in the FA binding pocket region, and it rapidly propagates through the receptor-binding domain (RBD) to the N-terminal domain (NTD), furin cleavage/recognition site and residues surrounding the fusion peptide (FP). All of the monomers are closely intertwined, and therefore signal propagation does not occur simply within an individual monomer, but rather involves a complex network of conformational changes spanning all three chains (Figs. S6–S8). For example, LA removal from the FA pocket formed by monomers A and C induces structural responses in the NTD of monomer B (NTD_B), the RBD of monomer C (RBD_C) and the furin cleavage/recognition site of monomer A, as shown by the propagation of the signal from the first site to these regions (Figs. S6–S8).

Upon LA removal, the hydrophobic and aromatic residues lining the FA site reorient their side chains towards the inside of the FA site (Fig. S9), as part of the contraction and collapse of this site (Figs. S9–S10). Furthermore, in the absence of the partnering

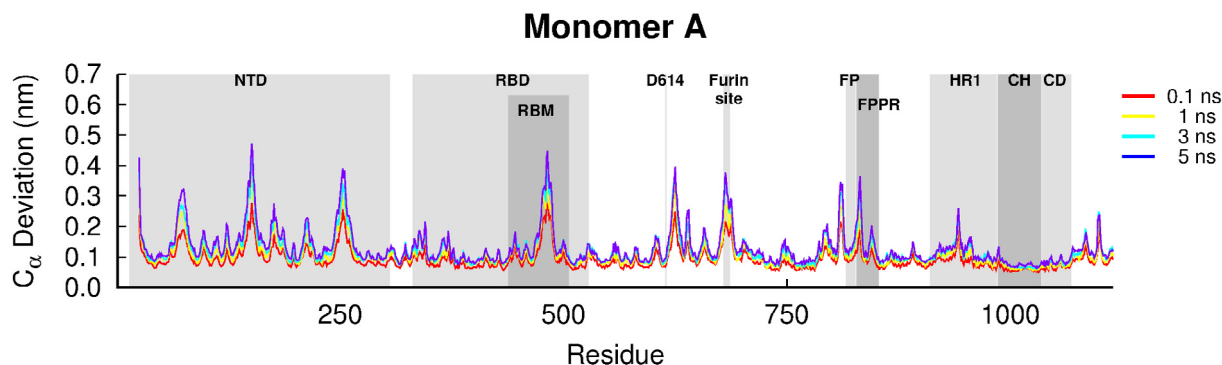


Fig. 2. Average C_{α} -positional deviation for the first monomer in the five nanoseconds after LA removal from the FA sites in the SARS-CoV-2 spike. The structural deviations were calculated using the Kubo-Onsager approach [34–37] for the pairwise comparison between the equilibrium LA-bound and dynamical-nonequilibrium apo spike simulations and averaged over all 90 replicates. A similar response to LA removal is observed for the other two monomers (Figure S4). Some relevant motifs are highlighted in grey, namely N-terminal domain (NTD), receptor-binding domain (RBD), receptor-binding motif (RBM), fusion peptide (FP), fusion-peptide proximal region (FPPR), heptad repeat 1 (HR1), central helix (CH) and connector domain (CD).

carboxylate headgroup of LA, R408 establishes alternative hydrogen bond interactions with nearby polar residues, such as Q404 and S375 (Fig. S11).

FA sites modulate key motifs for membrane fusion or antigenic epitopes in the spike

The RBMs respond rapidly to LA removal (Fig. 3 and S12–S13), due to their close proximity to the FA sites. From the time evolution of the protein response, signal transmission from the FA sites to the RBMs is apparently mediated by S366–A372 and R454–K458 (Fig. S14). The structural changes induced by LA removal in S366–A372 (a segment containing some of the residues lining the FA sites) are directly transmitted to R454–K458, and from there to the rest of the RBMs. 0.1 ns after LA removal, significant structural rearrangements are already apparent in the RBMs, mainly in the A475–C488 segment (Fig. 3 and S12–S13). Subsequently, a gradual increase in deviations is observed for A475–C488. The RBM lies between the β 4 and β 7 strands of the RBD and contains most of the residues that directly interact with ACE2 [45,46]. This motif is one of the most variable regions of SARS-CoV-2 spike [47] and a major target for neutralising antibodies [48–50].

The NTDs also show a fast and significant response to LA removal, in particular, H146–E156 and L249–G257 (Fig. 4 and S15–S16). The communication pathway connecting the FA sites to the NTDs apparently involve P337–A348 (a segment comprising some of the residues that directly interact with LA), W353–I358 and C166–P174 (Figure S17). The rearrangements induced by LA removal in the P337–A348 region promptly spread to W353–I358 and from there to the NTDs via the C166–P174 region (Figure S17). The NTD of the spike is a surface-exposed domain structurally linked to the RBD of a neighbouring monomer [19,45]. Although directly coupled to the RBD, the NTD does not bind to ACE2 [5,6] and its function in SARS-CoV-2 infection remains unclear. The spike NTDs of other related coronaviruses have been suggested to play a role in infection [51–53] and are known epitopes for neutralising antibodies [16,53]. Human antibodies targeting the NTD

of the SARS-CoV-2 spike have been isolated from convalescent COVID-19 patients e.g. [16,54,55]) and this region was shown to be a super-antigenic site [56]. A cryo-EM structure of the complex between the spike and the 4A8 monoclonal antibody shows that the NTD loops L141–E156 and R246–A260 (two of the regions that show the largest responses to LA removal in Fig. 4 and S15–S16) directly mediate the interaction between the proteins [54]. Both of these loops are candidates for vaccine and therapeutic developments [54]. The conformational changes in the H146–E156 and L249–G257 segments are further transmitted, over the following 5 ns, to other parts of the NTD, namely S71–R78, N122–N125 and F175–F186. The N122–N125 segment is a conserved NxxN sequence motif present in the NTD of spikes from several coronaviruses, and its function remains unknown [57]. The F175–F186 region is located immediately before a recently identified epitope for human antibodies [55]. The S71–R78 segment is part of the GTNGTKR insertion shared by the SARS-CoV-2 and bat-CoV RaTG13 spikes but not the SARS-CoV spike [57]. This motif, which is also found in structural proteins of several other viruses, and proteins from other organisms, has been suggested to allow the SARS-CoV-2 spike to bind to other receptors besides ACE2 [57]. The coupling identified here between the FA site and specific regions of the NTD is remarkable and highlights the complex allosteric connections within the spike, with distant sites apparently able to modulate the response of the NTDs.

Both the furin cleavage/recognition site and V622–L629 region, which are >40 Å away from the FA site, respond notably to the removal of LA (Fig. 5 and S18–S19). Both regions respond rapidly, with a significant conformational response observed almost immediately after LA removal. The furin cleavage/recognition site is located at the boundary between the S1 and S2 subunits [18,19] and furin cleavage is thought to be important for the activation of the spike [21]. This site contains a polybasic PRRA insertion not found in other SARS-CoV-related coronaviruses [58]. Cell-based assays show that deletion of the PRRA motif affects virus infectivity [21,58–62]. Note that in the simulations presented here, the furin cleavage/recognition site (located between R685 and S686) is not cleaved (see Supplementary Material). From the time evolution of the protein response to LA removal, signal propagation from the FA site to the furin cleavage/recognition site appears to occur via the C525–K537, F318–I326, and L629–Q644 regions (Figure S20). Upon LA removal, the structural changes in the FA site are first propagated to C525–K537 and then sequentially transmitted to F318–I326 and L629–Q644, ultimately reaching V622–L629 and the furin site. The furin cleavage/recognition site and V622–L629 region are among the spike regions most affected by LA removal and show increasingly large deviations (larger than most other loop regions of the protein) over the simulations. The conformational changes in these regions propagate to segments immediately adjacent to the fusion peptide, namely the downstream fusion peptide proximal region (FPPR) and the upstream D808–S813. The FPPR is a ~25-residue segment located in S2 immediately downstream of the fusion peptide, which has been suggested to play an essential role in the structural transitions between pre- and post-fusion conformations of the spike [20]. The D808–S813 region is located upstream of the fusion peptide (FP), immediately preceding the S2' protease recognition and cleavage site R815) [60]. Both proteolytic sites in the SARS-CoV-2 spike are known epitopes for neutralising antibodies [63,64].

The close connection between the furin cleavage/recognition site, V622–L629, and the regions adjacent to the FP, identified here for the intact, wild-type spike is remarkable. Due to this crosstalk, mutations in or close to the furin cleavage/recognition site or V622–L629 are likely to affect signal transmission to the FP-surrounding regions, i.e. the FPPR and the S2' cleavage site. This is worthy of experimental investigation.

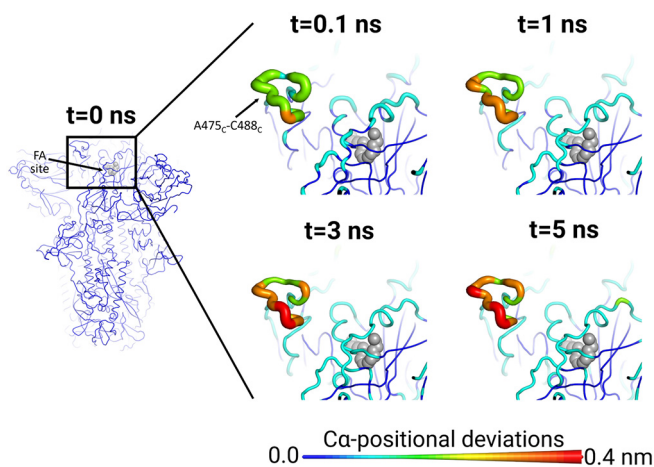


Fig. 3. The FA site allosterically affects the RBM in the SARS-CoV-2 spike. Average C α -positional deviation around the first FA site at times 0, 0.1, 1, 3 and 5 ns following LA removal from the FA binding pockets. The C α deviations between the simulations with and without LA were determined for each residue and averaged over the 90 pairs of simulations (Figs. 2 and S4). The C α average deviations are mapped onto the structure used as starting point for the LA-bound equilibrium simulations. Both the structure colours, and the cartoon thickness, relate to the average C α -positional deviation values. The location of the LA binding site is highlighted with grey spheres. The subscript letters in the labels correspond to the monomer ID. This figure shows the response around the first FA binding site, which is formed by monomers A and C. Results for the other sites show similar connections between the FA sites and the RBMs (see Figs. S12–S13 for the responses observed for the other two FA sites).

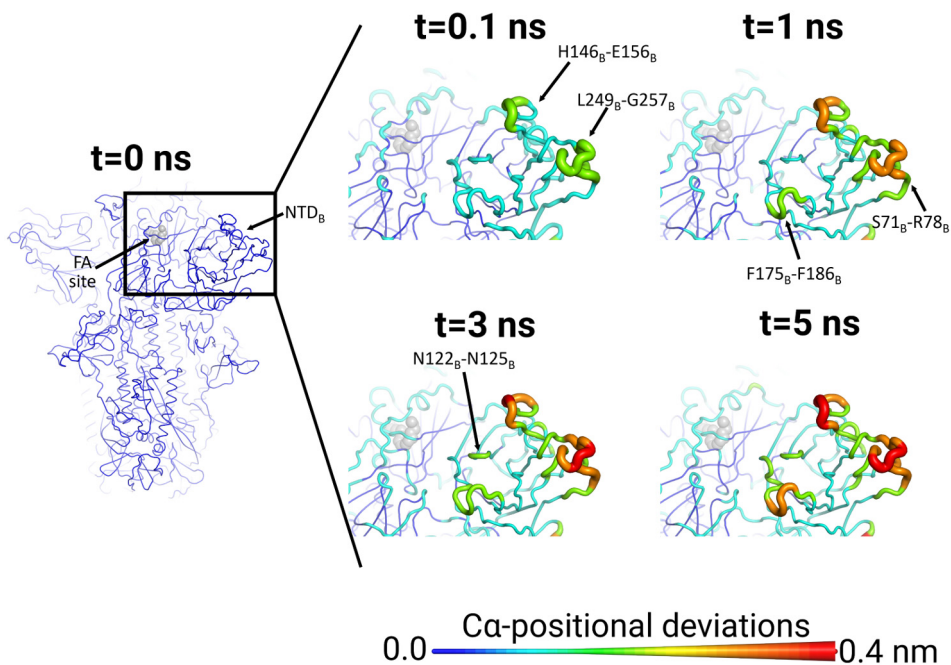


Fig. 4. The FA site allosterically affects the NTD in the SARS-CoV-2 spike. Average Ca -positional deviation around the first FA site at times 0, 0.1, 1, 3 and 5 ns following LA removal from the FA binding pockets (see Figs. S15–S16 for the responses observed for the other two FA sites). This figure shows the response around the first FA binding site, which is formed by monomer A and C. Similar results are observed for the other two FA sites and the NTD (Figs. S15–S16). For more details, see the legend of Fig. 3.

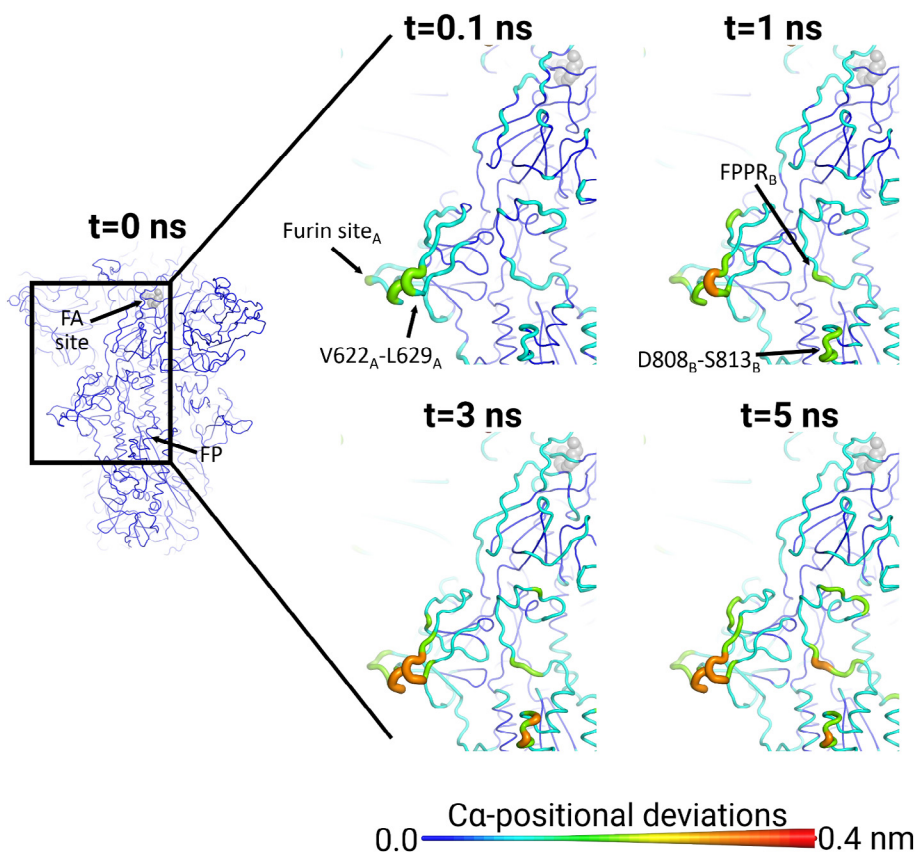


Fig. 5. The FA site allosterically affects the furin cleavage/recognition site, the fusion peptide proximal region (FPPR), and the residues immediately preceding the S2' cleavage site in the SARS-CoV-2 spike. Average Ca -positional deviations around the first FA site at times 0, 0.1, 1, 3 and 5 ns following LA removal from the FA binding pockets. This figure shows the response around the first FA binding site, which is formed by monomers A and C. A similar response is observed for all three monomers (see Figs. S18–S19 for the responses observed for the other two FA sites). Analogous connections between the other two FA sites and the furin cleavage/recognition sites, FPPR and S2' cleavage sites are also observed (Figs. S17–S18). For more details, see the legend of Fig. 3.

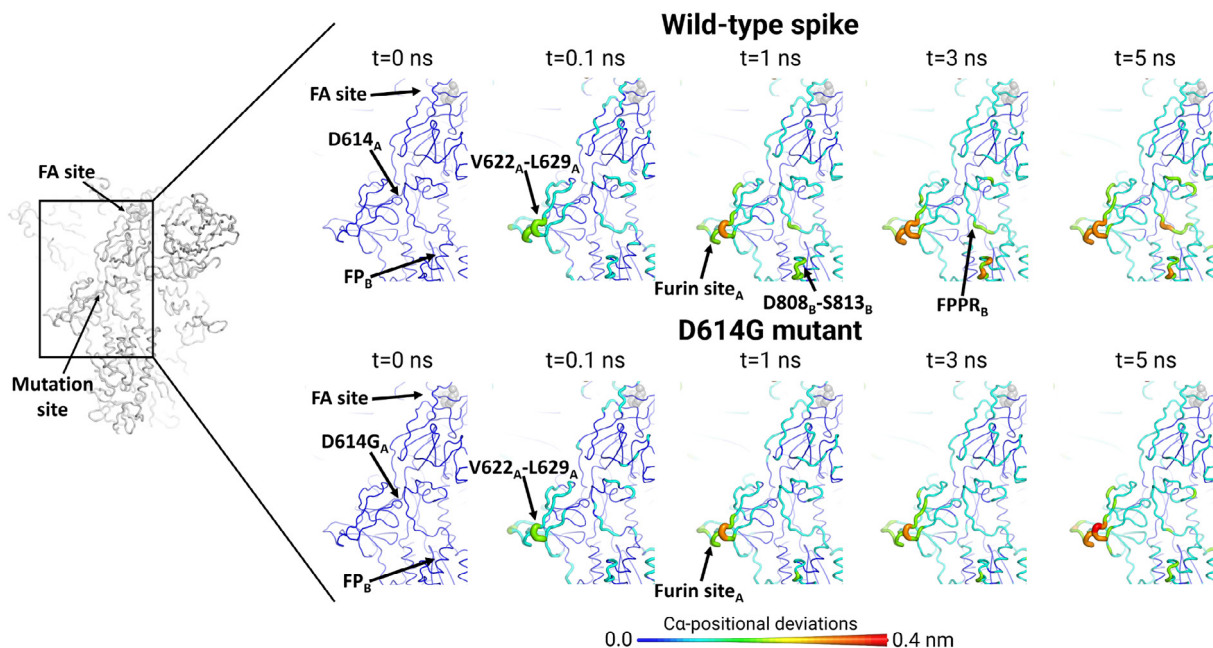


Fig. 6. Response of the wild-type and D614G spike to LA removal. Average C α -positional deviations around the first FA site at times 0, 0.1, 1, 3 and 5 ns following LA removal from the FA binding pockets (see Figs. S33–S34 for the responses observed for the other two FA sites). The C α deviations between the simulations with and without LA were determined for each residue, and the values averaged over the 90 pairs of simulations (Figs. 2, S4 and S25). The structure colours, and cartoon thickness, relate to the average C α -positional deviations. The location of the LA binding site is highlighted with grey spheres. The subscript letters in the labels correspond to the monomer ID.

Simulations of the D614G spike

We also performed equilibrium and dynamical-nonequilibrium simulations of the D614G mutant spike. The D614G mutation is now dominant in SARS-CoV-2 lineages circulating worldwide [65] and confers increased transmissibility. Other variants are emerging as increasing numbers of infections provide further opportunities for mutations to arise. The B.1.1.7 (also known as Alpha) variant, largely responsible for the surge in cases in the UK in the winter of 2020/21, has increased infectivity without the D614G mutation [66,67]. However, three of the four variants, involved in the April 2021 surge of cases in India as well as the B.1.617.2 (Delta) and B.1.1.529 (Omicron) variants do include D614G among the mutations (COVID-19 Genomics UK Consortium, <https://www.cogconsortium.uk> and <https://www.ecdc.europa.eu/>). That a single amino acid replacement of the aspartate residue in position 614 by a glycine leads to more efficient viral transduction into host cells and greater infectivity [68–71] is worthy of mechanistic exploration. Here, three equilibrium 200 ns MD simulations were performed for the locked form of the unglycosylated and uncleaved ectodomain region of the D614G spike modelled from a wild-type model [19,22] based on the cryo-EM structure 6ZB5 [19] with and without LA bound (see [Supplementary Material](#)). Note that D614G has the same sequence as the wild-type except for the residue in position 614, which was mutated from an aspartate to a glycine. In the original wild-type spike, D614 is located at the interface between monomers, with its sidechain directly interacting with residues across the subunit interface [61]. Root mean square fluctuation (RMSF) profiles from equilibrium MD for the wild-type and D614G apo spikes are similar (Figure S21). However, unlike the wild-type with or without LA, a single chain in one replicate of the D614G with LA bound exhibits greater fluctuation in the middle of the RBM corresponding to exposed loop residues Q474–N487 (Figure S21). The RBM loop flipped orientation coincident with a transient loss of a salt bridge on the stem of this loop between K458 and E471 of the same chain.

The Q474–N487 loop is interesting because, although the RBM in the closed conformation remains inaccessible for binding to ACE2, residues Q474–N487 of the RBM (shown in magenta in the insert in Figure S21F) it includes the epitope Y473–P479 (YQAGSTP) [72], which may still provide a target for neutralizing antibodies in the closed conformation, depending on the degree of glycan shielding [73] (being close to the S349 O-glycosylation site [21]). This behaviour, albeit in a single chain of our equilibrium MD simulations of the unglycosylated, uncleaved wild-type and D614G LA-bound spikes, suggests that the D614G may influence mobility in this important region. There is some evidence that local flexibility and local sequence context can affect the fraction of occupancy of glycosylation sites in different proteins [74]. Spike glycosylation occurs co-translationally and it is also thought that LA binds co-translationally. LA and the D614G mutation may affect glycosylation [75].

The trans-interface interactions of the carboxylate of D614 in the wild-type involve four potential candidate residues, K854, K835, Q836 and T859. In the equilibrium MD simulations, T859 came within 5 Å of D614 across an interface and made transient, weak interactions throughout the simulations of the apo and LA-bound spike systems (Figure S22). The carboxylate of D614 and NZ of K854 remains within salt-bridging distance throughout the wild-type simulations (Figure S23). The D614–K854 trans-subunit interface interaction dominates throughout the simulation time, regardless of the presence of bound LA (Figure S22). In the 200 ns open wild-type apo spike simulations, the trans-subunit interaction between D614 and K854 persisted for 99% (standard deviation 0.57%) of the frames across eight of the nine subunit interfaces. This contact was lost at only one interface and that was between an open and closed chain in one of the repeats (Figure S22).

An analogous analysis was performed on the D614G mutant to establish whether K854 makes alternative hydrogen-bond or salt-bridge contacts across the 3 subunit interfaces (averaged over 3 × 200 ns replicates) in the absence of a partnering D614

carboxylate. In the D614G spike, K854 fails to find any alternative salt-bridge and only occasionally comes within hydrogen-bonding distance of residues Q613 and N317 (Figure S24). This supports the inference drawn from cryo-EM structures of the head region of the D614G spike that this mutation disrupts the inter-monomer salt-bridge and hydrogen bond networks in this region, which may cause reduced stability of the trimer. This corresponds to the observation that the D614G mutant was mostly in an open conformation on the EM grids and suggests that loss of the D614-K854 interaction somehow destabilises the closed conformation (e.g. [76,77]).

Dynamical-nonequilibrium simulations of the D614G variant were also performed to test whether the D614G mutation affects the response of the spike to LA. A distribution of conformations taken from the equilibrium simulations of the locked form of the unglycosylated and uncleaved D614G spike with LA bound was used as the starting point for the dynamical-nonequilibrium simulations. The same perturbation as for the wild-type spike was applied to the system, namely LA removal. The Kubo–Onsager approach [34–37] was again used to extract the response of the system (Figure S25) and determine the statistical significance of the observed responses (Figure S26). In the D614G variant, there is notably less symmetry across the monomers in the response of the spike to LA removal, compared to the wild-type (Fig. 6 and S27–S34). For instance, the amplitude of the structural response of the V266–L629 and furin cleavage/recognition site regions in monomer C (Figure S33) of the D614G is substantially smaller than in monomers B and A (Fig. 6 and S34).

The conformational responses of the wild-type and D614G spikes can be directly compared because the same perturbation was used for both in the dynamical-nonequilibrium simulations. The conformational response of the NTDs and RBDs to LA removal is generally similar in the wild-type and D614G spike (Figs. S27–S33) with small variations in the amplitude of the structural rearrangements of some functional motifs, e.g. RBMs. However, the D614G mutation significantly affects inter-monomer communication, with reduction of signal transmission from the furin cleavage/recognition site and V622–L629 of one monomer to the FPPR of another (Fig. 6 and S33–S34) compared to the wild-type (Fig. 5 and S18–S19). In the D614G spike, only minor deviations of the FPPR are observed. Furthermore, the region located upstream of the fusion peptide, namely D808–S813, also shows different rates of signal propagation between the wild-type and D614G proteins (Fig. 6 and S33–S34). The differences identified here may relate to functionally important differences between the wild-type and D614G spikes. The results here show that the D614G mutation alters the allosteric networks connecting the FA site to the regions surrounding the FP, particularly the FPPR. There is reduced communication between the monomers in the D614G spike. As noted above, the response of the D614G spike to LA is also less symmetrical than the wild-type.

3. Conclusions

Our findings show that changes in ligand occupancy at the FA site influence the dynamic behaviour of functionally important motifs distant from the FA site. The simulations identify a complex network of structural pathways connecting the FA sites to key structural motifs within the SARS-CoV-2 spike. These networks extend far beyond the regions surrounding the FA sites, with structural responses being observed in the RBM, NTD, furin cleavage/recognition site and FP-adjacent regions (Movie 1). The results also show strong crosstalk between the furin cleavage/recognition site, V622–L629 and the regions adjacent to the FP. Disrupting or alter-

ing these communication networks may be a novel strategy for drug development against COVID-19.

Simulations of the D614G spike show that this mutation affects communication between the FA site and the FPPR and the S2' cleavage site. The D614G mutant shows reduced response of the FPPR and a slower rate of signal propagation to the S2' cleavage site compared to the wild-type protein (Movie 2). These results indicate that the D614G mutation affects the allosteric behaviour and the response to LA of the spike, which may be related to the changes in viral fitness associated with this mutation [78].

The results here further highlight the potential of dynamical-nonequilibrium simulations for identifying pathways of allosteric communication [38–40] and suggest that this approach may be useful in analysing mutations and differences in functionally important dynamical behaviour, and possibly different effects of LA, between SARS-CoV-2 spike variants of clinical relevance, such as the Alpha, Beta, Gamma, and Delta, and (now) Omicron.

4. Data access statement

The data is openly available from MolSSI/BioExcel COVID-19 public data repository for biomolecular simulations (<https://covid-molssi.org/simulations/>).

Declaration of Competing Interest

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

Acknowledgements

AJM and ASFO thank EPSRC (grant number EP/M022609/1) and BBSRC (grant number BB/R016445/1) for support. MD simulations were carried out using the computational facilities of the Advanced Computing Research Centre, University of Bristol (<http://www.bris.ac.uk/acrc>) under an award for COVID-19 research. We thank Bris-SynBio, a BBSRC/EPSC Synthetic Biology Research Centre (Grant Number:BB/L01386X/1) for funding DKS and ASFO and the Blue-Gem HPC system and EPSRC via HECBIOSIM (hecbiosim.ac.uk) for providing ARCHER/ARCHER2 time through a COVID-19 rapid response call. We also thank Oracle Research for Oracle Public Cloud Infrastructure (https://cloud.oracle.com/en_US/iaas) time under an award for COVID-19 research, the [Bristol UNCOVER Group](#) and the University of Bristol, for their support. I.B. acknowledges support from the EPSRC Future Vaccine Manufacturing and Research Hub (EP/R013764/1). C.S. and I.B. are Investigators of the Wellcome Trust (210701/Z/18/Z; 106115/Z/14/Z).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.csbj.2021.12.011>.

References

- [1] Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis* 2020;20(5):533–4.
- [2] T. economist. (The economist, <https://www.economist.com/briefing/2021/05/15/there-have-been-7m-13m-excess-deaths-worldwide-during-the-pandemic>, 2021).
- [3] Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 2020;579(7798):270–3.
- [4] Lu R, Zhao X, Li J, Niu P, Yang Bo, Wu H, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* 2020;395(10224):565–74.

- [5] Yan R, Zhang Y, Li Y, Xia Lu, Guo Y, Zhou Q. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* 2020;367(6485):1444–8.
- [6] Wang Q, Zhang Y, Wu L, Niu S, Song C, Zhang Z, et al. Structural and functional basis of SARS-CoV-2 entry by using human ACE2. *Cell* 2020;181(4):894–904.e9.
- [7] Rothan HA, Byrareddy SN. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *J Autoimmun* 2020;109:102433. <https://doi.org/10.1016/j.jaut.2020.102433>.
- [8] Lake MA. What we know so far: COVID-19 current clinical knowledge and research. *Clin Med (Lond)* 2020;20(2):124–7.
- [9] Puntmann VO, Carerj ML, Wieters I, Fahim M, Arendt C, Hoffmann J, et al. Outcomes of cardiovascular magnetic resonance imaging in patients recently recovered from coronavirus disease 2019 (COVID-19). *JAMA Cardiol* 2020;5(11):1265. <https://doi.org/10.1001/jamacardio.2020.3557>.
- [10] Varga Z, Flammer AJ, Steiger P, Haberecker M, Andermatt R, Zinkernagel AS, et al. Endothelial cell infection and endotheliitis in COVID-19. *Lancet* 2020;395(10234):1417–8.
- [11] Puelles VG, Lütgehetmann M, Lindenmeyer MT, Spherhake JP, Wong MN, Allweiss L, et al. Multiorgan and renal tropism of SARS-CoV-2. *N Engl J Med* 2020;383(6):590–2.
- [12] Lee M-H, Perl DP, Nair G, Li W, Maric D, Murray H, et al. Microvascular Injury in the Brains of Patients with Covid-19. *N Engl J Med* 2021;384(5):481–3.
- [13] Daly JL, Simonetti B, Klein K, Chen K-E, Williamson MK, Antón-Plágaro C, et al. Neupilin-1 is a host factor for SARS-CoV-2 infection. *Science* 2020;370(6518):861–5.
- [14] Cantuti-Castelvetri L, Ojha R, Pedro LD, Djannatian M, Franz J, Kivivanen S, et al. Neupilin-1 facilitates SARS-CoV-2 cell entry and infectivity. *Science* 2020;370(6518):856–60.
- [15] Oliveira ASF, Ibarra AA, Bermudez I, Casalino L, Gaieb Z, Shoemark DK, et al. A potential interaction between the SARS-CoV-2 spike protein and nicotinic acetylcholine receptors. *Biophys J* 2021;120(6):983–93.
- [16] W. Soh et al., The N-terminal domain of spike glycoprotein mediates SARS-CoV-2 infection by associating with L-SIGN and DC-SIGN. *bioRxiv* 10.1101/2020.11.05.369264, (2020).
- [17] Bosch BJ, van der Zee R, de Haan CAM, Rottier PJM. The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. *J Virol* 2003;77(16):8801–11.
- [18] Walls AC, Park Y-J, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* 2020;181(2):281–292.e6.
- [19] Toelzer C, Gupta K, Yadav SKN, Borucu U, Davidson AD, Kavanagh Williamson M, et al. Free fatty acid binding pocket in the locked structure of SARS-CoV-2 spike protein. *Science* 2020;370(6517):725–30.
- [20] Cai Y, Zhang J, Xiao T, Peng H, Sterling SM, Walsh RM, et al. Distinct conformational states of SARS-CoV-2 spike protein. *Science* 2020;369(6511):1586–92.
- [21] Davidson AD, Williamson MK, Lewis S, Shoemark D, Carroll MW, Heesom KJ, 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* 2020;12(1). <https://doi.org/10.1186/s13073-020-00763-0>.
- [22] Shoemark DK, Colenso CK, Toelzer C, Gupta K, Sessions RB, Davidson AD, 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* 2021;60(13):7098–110.
- [23] Zhang S, Qiao S, Yu J, Zeng J, Shan S, Tian L, et al. Bat and pangolin coronavirus spike glycoprotein structures provide insights into SARS-CoV-2 evolution. *Nat Commun* 2021;12(1). <https://doi.org/10.1038/s41467-021-1767-3>.
- [24] Bangaru S, Ozorowski G, Turner HL, Antanasijevic A, Huang D, Wang X, et al. Structural analysis of full-length SARS-CoV-2 spike protein from an advanced vaccine candidate. *Science* 2020;370(6520):1089–94.
- [25] Casalino L, Gaieb Z, Goldsmith JA, Hjorth CK, Dommer AC, Harbison AM, et al. Beyond shielding: the roles of glycans in the SARS-CoV-2 spike protein. *ACS Cent Sci* 2020;6(10):1722–34.
- [26] Casalino L, Dommer AC, Gaieb Z, Barros EP, Sztain T, Ahn S-H, et al. AI-driven multiscale simulations illuminate mechanisms of SARS-CoV-2 spike dynamics. *Int J High Perform Comput Appl* 2021;35(5):432–51.
- [27] Yu A, Pak AJ, He P, Monje-Galvan V, Casalino L, Gaieb Z, et al. A multiscale coarse-grained model of the SARS-CoV-2 virion. *Biophys J* 2021;120(6):1097–104.
- [28] Barros EP, Casalino L, Gaieb Z, Dommer AC, Wang Y, Fallon L, et al. The flexibility of ACE2 in the context of SARS-CoV-2 infection. *Biophys J* 2021;120(6):1072–84.
- [29] Amaro RE, Mulholland AJ. Biomolecular Simulations in the Time of COVID19, and After. *Comput Sci Eng* 2020;22(6):30–6.
- [30] Ray D, Le L, Andricioaei I. Distant residues modulate conformational opening in SARS-CoV-2 spike protein. *Proc Natl Acad Sci USA* 2021;118.
- [31] Zimmerman MI, Porter JR, Ward MD, Singh S, Vithani N, Meller A, et al. SARS-CoV-2 simulations go exascale to predict dramatic spike opening and cryptic pockets across the proteome. *Nat Chem* 2021;13(7):651–9.
- [32] Triveri A, Serapian SA, Marchetti F, Doria F, Pavoni S, Cinquini F, et al. SARS-CoV-2 Spike Protein Mutations and Escape from Antibodies: A Computational Model of Epitope Loss in Variants of Concern. *J Chem Inf Model* 2021;61(9):4687–700.
- [33] A. Dommer et al., #COVIDisAirborne: AI-Enabled Multiscale Computational Microscopy of Delta SARS-CoV-2 in a Respiratory Aerosol. *bioRxiv* 10.1101/2021.11.12.468428, (2021).
- [34] Ciccotti G, Jacucci G, McDonald IR. Thought-experiments by molecular dynamics. *J Stat Phys* 1979;21(1):1–22.
- [35] G. Ciccotti, in *Computer simulation in material science*, P. V. Meyer M, Ed. (Kluwer Academic Publishers, 1991), pp. 119–137.
- [36] Ciccotti G, Ferrario M. Non-equilibrium by molecular dynamics: a dynamical approach. *Mol Simul* 2016;42(16):1385–400.
- [37] Oliveira ASF, Ciccotti G, Haider S, Mulholland AJ. Dynamical nonequilibrium molecular dynamics reveals the structural basis for allostery and signal propagation in biomolecular systems. *Eur Phys J B* 2021;94:144.
- [38] Oliveira ASF, Shoemark DK, Campello HR, Wonnacott S, Gallagher T, Sessions RB, et al. Identification of the initial steps in signal transduction in the $\alpha 4\beta 2$ nicotinic receptor: insights from equilibrium and nonequilibrium simulations. *Structure* 2019;27(7):1171–1183.e3.
- [39] Oliveira ASF, Edsall CJ, Woods CJ, Bates P, Nunez GV, Wonnacott S, et al. A general mechanism for signal propagation in the nicotinic acetylcholine receptor family. *J Am Chem Soc* 2019;141(51):19953–8.
- [40] Galdadas I, Qu S, Oliveira ASF, Olehnovics E, Mack AR, Mojica MF, et al. Allosteric communication in class A β -lactamases occurs via cooperative coupling of loop dynamics. *eLife* 2021;10. <https://doi.org/10.7554/eLife.66567>.
- [41] Damas JM, Oliveira ASF, Baptista AM, Soares CM. Structural consequences of ATP hydrolysis on the ABC transporter NBD dimer: molecular dynamics studies of HlyB. *Protein Sci* 2011;20(7):1220–30.
- [42] Abreu B, Lopes EF, Oliveira ASF, Soares CM. F508del disturbs the dynamics of the nucleotide binding domains of CFTR before and after ATP hydrolysis. *Proteins* 2020;88(1):113–26.
- [43] Sztain T, Ahn S-H, Bogetti AT, Casalino L, Goldsmith JA, Seitz E, et al. A glycan gate controls opening of the SARS-CoV-2 spike protein. *Nat Chem* 2021;13(10):963–8.
- [44] Papaleo E, Saladino G, Lambrughi M, Lindorff-Larsen K, Gervasio FL, Nussinov R. The Role of Protein Loops and Linkers in Conformational Dynamics and Allostery. *Chem Rev* 2016;116(11):6391–423.
- [45] Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh C-L, Abiona O, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* 2020;367(6483):1260–3.
- [46] Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* 2020;581(7807):215–20.
- [47] Thomson E et al. The circulating SARS-CoV-2 spike variant N439K maintains fitness while evading antibody-mediated immunity. *Cell* 2020;184:1171–87.
- [48] Robbiani DF, Gaebler C, Muecksch F, Lorenzi JCC, Wang Z, Cho A, et al. Convergent antibody responses to SARS-CoV-2 in convalescent individuals. *Nature* 2020;584(7821):437–42.
- [49] Rogers TF, Zhao F, Huang D, Beutler N, Burns A, He W-T, et al. Isolation of potent SARS-CoV-2 neutralizing antibodies and protection from disease in a small animal model. *Science* 2020;369(6506):956–63.
- [50] Wec AZ, Wrapp D, Herbert AS, Maurer DP, Haslwanter D, Sakharkar M, et al. Broad neutralization of SARS-related viruses by human monoclonal antibodies. *Science* 2020;369(6504):731–6.
- [51] Taguchi F, Hirai-Yuki A. Mouse hepatitis virus receptor as a determinant of the mouse susceptibility to MHV infection. *Front Microbiol* 2012;3:68.
- [52] Lu G, Wang Q, Gao GF. Bat-to-human: spike features determining 'host jump' of coronaviruses SARS-CoV, MERS-CoV, and beyond. *Trends Microbiol* 2015;23(8):468–78.
- [53] Zhou H, Chen Y, Zhang S, Niu P, Qin K, Jia W, et al. Structural definition of a neutralization epitope on the N-terminal domain of MERS-CoV spike glycoprotein. *Nat Commun* 2019;10(1). <https://doi.org/10.1038/s41467-019-10897-4>.
- [54] Chi X, Yan R, Zhang J, Zhang G, Zhang Y, Hao M, et al. A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. *Science* 2020;369(6504):650–5.
- [55] Wang H, Wu X, Zhang X, Hou X, Liang Te, Wang D, et al. SARS-CoV-2 proteome microarray for mapping COVID-19 antibody interactions at amino acid resolution. *ACS Cent Sci* 2020;6(12):2238–49.
- [56] Cerutti G, Guo Y, Zhou T, Gorman J, Lee M, Rapp M, et al. Potent SARS-CoV-2 neutralizing antibodies directed against spike N-terminal domain target a single supersite. *Cell Host Microbe* 2021;29(5):819–833.e7.
- [57] Behloul N, Baha S, Shi R, Meng J. Role of the GTNGTKR motif in the N-terminal receptor-binding domain of the SARS-CoV-2 spike protein. *Virus Res* 2020;286:198058. <https://doi.org/10.1016/j.virusres.2020.198058>.
- [58] Hoffmann M, Kleine-Weber H, Pöhlmann S. A multibasic cleavage site in the spike protein of SARS-CoV-2 is essential for infection of human lung cells. *Mol Cell* 2020;78(4):779–784.e5.
- [59] Matsuyama S, Nao N, Shirato K, Kawase M, Saito S, Takayama I, et al. Enhanced isolation of SARS-CoV-2 by TMPRSS2-expressing cells. *Proc Natl Acad Sci U S A* 2020;117(13):7001–3.
- [60] Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* 2020;181(2):271–280.e8.
- [61] K. Gupta et al., Structure and mechanism of SARS-CoV-2 Spike N679-V687 deletion variant elucidate cell-type specific evolution of viral fitness. *bioRxiv* 10.1101/2021.05.11.443384, (2021).

- [62] Peacock TP, Goldhill DH, Zhou J, Baillon L, Frise R, Swann OC, et al. The furin cleavage site in the SARS-CoV-2 spike protein is required for transmission in ferrets. *Nat Microbiol* 2021;6(7):899–909.
- [63] Farrera-Soler L, Dagher J-P, Barluenga S, Vadas O, Cohen P, Pagano S, et al. Identification of immunodominant linear epitopes from SARS-CoV-2 patient plasma. *PLoS ONE* 2020;15(9):e0238089.
- [64] Poh CM, Carissimo G, Wang B, Amrun SN, Lee C-P, Chee R-L, et al. Two linear epitopes on the SARS-CoV-2 spike protein that elicit neutralising antibodies in COVID-19 patients. *Nat Commun* 2020;11(1). <https://doi.org/10.1038/s41467-020-16638-2>.
- [65] Callaway E. The coronavirus is mutating - does it matter? *Nature* 2020;585(7824):174–7.
- [66] Zhang J, Zhang Y, Kang J-Y, Chen S, He Y, Han B, et al. Potential transmission chains of variant B.1.1.7 and co-mutations of SARS-CoV-2. *Cell Discov* 2021;7(1). <https://doi.org/10.1038/s41421-021-00282-1>.
- [67] Tang JW, Tambyah PA, Hui DSC. Emergence of a new SARS-CoV-2 variant in the UK. *J Infect* 2021;82(4):e27–8.
- [68] Daniloski Z, Jordan TX, Ilmain JK, Guo X, Bhabha G, tenOever BR, et al. The spike D614G mutation increases SARS-CoV-2 infection of multiple human cell types. *Elife* 2021;10. <https://doi.org/10.7554/elife.65365>.
- [69] Zhang L, Jackson CB, Mou H, Ojha A, Peng H, Quinlan BD, et al. SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity. *Nat Commun* 2020;11(1). <https://doi.org/10.1038/s41467-020-19808-4>.
- [70] Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell* 2020;182(4):812–827.e19.
- [71] Volz E et al. Evaluating the effects of SARS-CoV-2 spike mutation D614G on transmissibility and pathogenicity. *Cell* 2021;184:64–75.e11.
- [72] Jiang M et al. Epitope Profiling Reveals the Critical Antigenic Determinants in SARS-CoV-2 RBD-Based Antigen. *Front Immunol* 2021;12:707977.
- [73] Yi C et al. Key residues of the receptor binding motif in the spike protein of SARS-CoV-2 that interact with ACE2 and neutralizing antibodies. *Cell Mol Immunol* 2020;17:621–30.
- [74] Reusch D, Tejada ML. Fc glycans of therapeutic antibodies as critical quality attributes. *Glycobiology* 2015;25:1325–34.
- [75] Allen JD et al. Site-specific steric control of SARS-CoV-2 spike glycosylation. *Biochemistry* 2021;60:2153–69.
- [76] Yurkovetskiy L et al. Structural and functional analysis of the D614G SARS-CoV-2 spike protein variant. *Cell* 2020;183:739–751.e738.
- [77] Gobeil SM et al. D614G mutation alters SARS-CoV-2 spike conformation and enhances protease cleavage at the S1/S2 junction. *Cell Rep* 2021;34:108630.
- [78] Plante JA et al. Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* 2021;592:116–21.