

Post-Translational Modifications Optimize the Ability of SARS-CoV-2 Spike for Effective Interaction with Host Cell Receptors

Karan Kapoor[†], Tianle Chen[†], and Emad Tajkhorshid*

Theoretical and Computational Biophysics Group, NIH Center for Macromolecular Modeling and Bioinformatics, Beckman Institute for Advanced Science and Technology, Department of Biochemistry, and Center for Biophysics and Quantitative Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

[†]These authors contributed equally to this work.

E-mail: emad@illinois.edu

Significance

SARS-CoV2 Spike protein, which forms the basis for high pathogenicity and transmissibility of the virus, is also prime target for the development of both diagnostics and vaccines for the debilitating disease caused by the virus. We present a full model of spike methodically crafted and used to study its atomic-level dynamics by multiple- μ s simulations. The results shed new light on the impact of posttranslational modifications in the pathogenicity of the virus. We show how glycan-glycan and glycan-lipid interactions broaden the protein's dynamical range, and thereby, its effective interaction with the surface receptors on the host cell. Palmitoylation of spike membrane domain, on the other hand, results in a unique deformation pattern that might prime the membrane for fusion.

Key Words

Spike protein | coronaviruses | structural dynamics | glycosylation | molecular dynamics | lipid bilayer

Abstract

SARS-CoV2 spike glycoprotein is prime target for vaccines and for diagnostics and therapeutic antibodies against the virus. While anchored in the viral envelope, for effective virulence, the spike needs to maintain structural flexibility to recognize the host cell surface receptors and bind to them, a property that can heavily hinge upon the dynamics of the unresolved domains, most prominently the stalk. Construction of the complete, membrane-bound spike model and the description of its dynamics remain critical steps in understanding the inner working of this key element in viral infection. Using a hybrid approach, combining homology modeling, protein-protein docking and MD simulations, guided by biochemical and glycomics data, we have developed a full-length, membrane-bound, palmitoylated and fully-glycosylated spike structure in a native membrane. Multi-microsecond MD simulations of this model, the longest known trajectory of the full-spike, reveals conformational dynamics employed by the protein to explore the crowded surface of the host cell. In agreement with cryoEM, three flexible hinges in stalk allow for global conformational heterogeneity of spike in the fully-glycosylated system mediated by glycan-glycan and glycan-lipid interactions. Dynamical range of spike is considerably reduced in its non-glycosylated form, confining the area explored by the spike on the host cell surface. Furthermore, palmitoylation of the membrane domain amplify the local curvature that may prime the fusion. We show that the identified hinge regions are highly conserved in SARS coronaviruses, highlighting their functional importance in enhancing viral infection, and thereby provide novel points for discovery of alternative therapeutics against the virus.

Introduction

The pandemic caused by SARS-CoV-2 continues to impact nearly all aspects of human lives across the globe.^{1,2} SARS-CoV-2 belongs to the beta coronavirus genus, same as SARS-CoV and MERS-CoV responsible for the severe acute respiratory syndrome outbreak in 2003 and Middle East respiratory syndrome in 2012, respectively.^{3,4} The first step in viral entry and infection in SARS-CoV-2, and in other coronaviruses in general, is the binding of its extended spike glycoproteins located on the viral envelope to specific human cell surface receptors.^{5,6} These envelope spike proteins contain all key components necessary for the infection of human cells; after binding to the surface receptors they initiate the fusion of the viral and human membranes and the subsequent release of the viral genome into the host cell.⁷ Characterizing the mechanism of action of the spike protein thus remains key to our understanding of the critical steps involved in viral infection, paving way for development of novel therapeutics against the virus.

Major efforts have been spent recently in characterizing the structural features of the spike, and as such, quite a few SARS-CoV-2 spike cryo-EM structures have been resolved.⁸⁻¹⁵ Though providing essential information on the structural details of the spike globular domain (spike head), which contains the receptor-binding domain (RBD) recognizing and binding to the ACE2 host cell receptors,⁶ the resolved structures are still missing a number of functionally important regions in the full spike (Fig. 1A). These include the highly conserved fusion-peptide segment involved in initiating the fusion of the viral and host cell membranes,^{16,17} the stalk heptad repeat 2 (HR2) domain undergoing refolding during the membrane fusion,¹⁸ and the transmembrane (TM) domain containing multiple palmitoylation sites shown to stabilize the spike in the viral envelope and thus essential for assembly and activity in other viruses.^{19,20} The structures also lack critical information about the glycosylation of different residues, known to be essential in mediating protein folding, shaping viral tropism, as well as shielding the virus from immune recognition.²¹ Additionally, the assembly and budding of the virus are known to take place in the endoplasmic reticulum–Golgi intermediate compartment

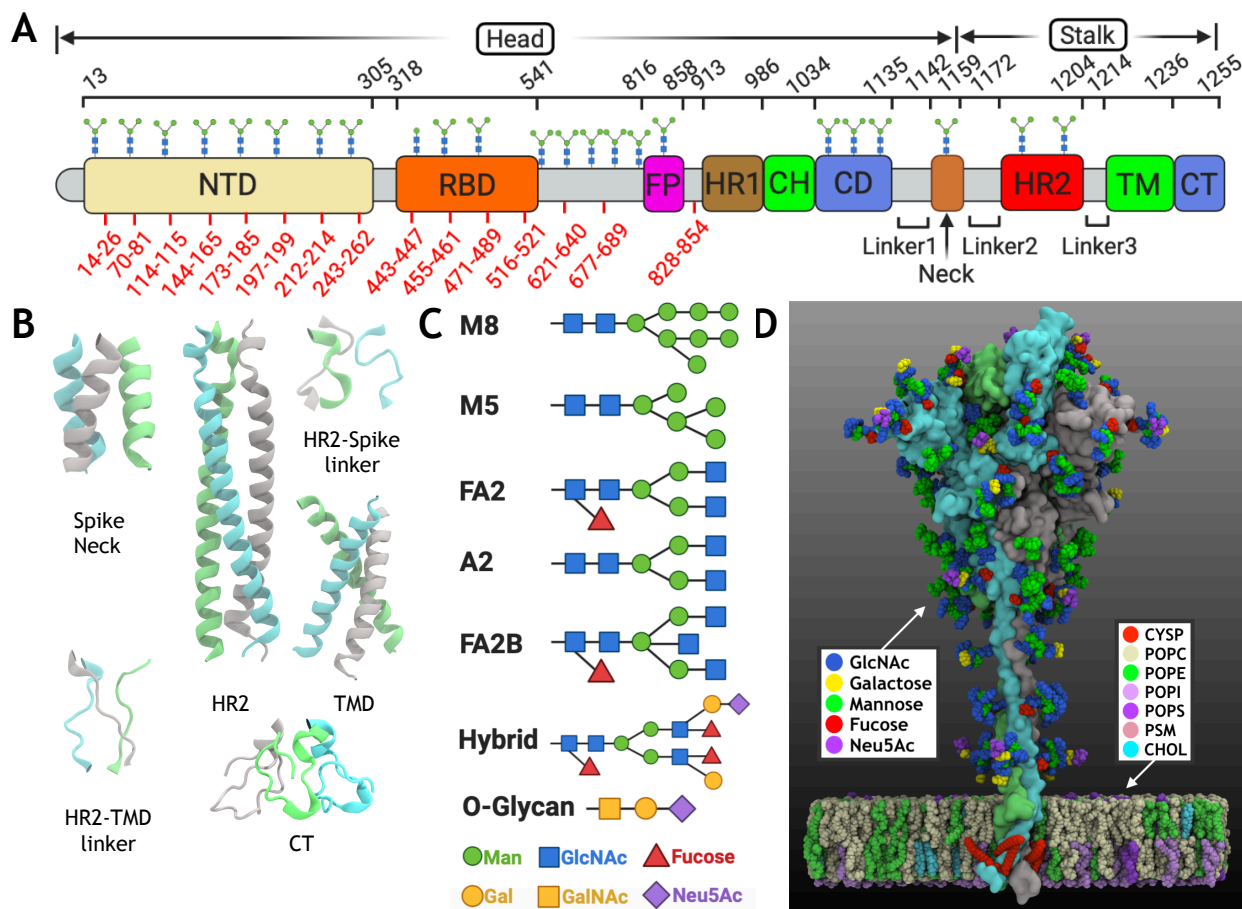


Figure 1: Overview of full spike system: (A) Schematic of the different functional domains of the spike, which can be largely divided into the head and stock domains: N-terminal domain (NTD), receptor binding domain (RBD), fusion peptide (FP), heptad repeat 1 (HR1), central helix (CH), connecting domain (CD), spike neck, heptad repeat 2 (HR2), transmembrane domain (TM) and cytoplasmic tail (CT). The residue ranges of these domains are provided along with the experimentally identified glycosylation sites marked on top of the domains. Out of these, cryo-EM structures are only available for the spike head, with the missing regions in the structure shown in red. (B) The modeled structures in the spike neck, HR2, TM domain, CT and the 2 connecting linkers are shown. These missing regions in the spike stalk were modeled using a hybrid approach (described in methods). (C) The types of N-linked and O-linked glycans modeled in the full spike structure are shown. (D) The modeled structure of the full-length, palmitoylated and glycosylated spike is shown. Each spike monomer is depicted in surface representation in a different color. The protein is embedded in a membrane bilayer with a lipid composition from endoplasmic reticulum-Golgi intermediate compartment (ERGIC). The constituting glycan components and lipid/palmitoylation types are provided in the legends.

(ERGIC),^{22,23} an environment that is not taken into account in any of the above cryo-EM studies which have resolved only the ectodomain structure stabilized using C-terminal foldon trimerization motif.⁸ Further resolution of these structural and environmental details thus

becomes a necessary step to study the behavior of native spike proteins in a physiological context.

Computational simulations offer an effective technique for modeling complex systems like the spike at a detailed level. As such, recent efforts in applying classical molecular dynamics (MD) simulations to the resolved structures of the spike head^{24–27} as well as constructed full-length spike models^{28–30} have provided powerful insights into various functionally relevant features, e.g., how glycans can shield the spike from the immune system and neutralizing antibodies,^{24,25,29,30} or the intra-domain dynamics of the spike head that may facilitate the RBD to switch between the 'up' and 'down' (active or inactive) conformations.^{26,27,29} However, we are still far from fully understanding how the global dynamics of the spike may contribute to its function, specifically in relation to effectively locating its target receptor in crowded cellular environments. So far, characterization of the effect of the spike glycosylation has been focused on their global shielding effects, leaving the possible roles that the glycans may play in the protein's conformational dynamics unexplored. Furthermore, the role of palmitoylations present in the TM/endodomain, known to participate in the modulation of the membrane curvature and in mediating cell fusion in other viruses,^{31–34} also remains lacking in SARS-CoV-2.

To obtain further insights into the workings of this important component of the virus machinery, we aimed to construct a full-length, membrane-embedded spike for investigation of its conformational dynamics at the atomic-scale using MD simulations. In the first step, we use a hybrid approach, combining homology modeling, protein-protein docking and MD simulations of the modeled regions, with the available experimental data,^{35–38} to develop a full-length, palmitoylated and fully-glycosylated spike. After validating the modeled parts, multi-microsecond MD simulations of the full-length spike, embedded in a lipid bilayer of composition derived from ERGIC, were used to characterize global dynamics and inter-domain motions, focusing specifically on the potential role of glycans in these processes. Control, multi-microsecond simulations of the non-glycosylated spike highlighted the impor-

tance of the glycans in modulating the global motion of the spike. We also characterized the specific role played by palmitoylations in modulating the membrane curvature, potentially
60 regulating the membrane fusion with the host cell. This study provides key insight into the dynamics used by the spike to sample the crowded surface of its host cell and the direct role of the glycans in regulating this process while successfully allowing the virus to evade the host immune response.

Methods

65 **Modeling of full-length spike**

Topologically, the spike can be divided into two main regions - the spike head, which includes the RBD that binds to the host cell receptor, and an extended stalk consisting of a short spike neck, the HR2 domain, and the TM domain tethering the spike to the viral envelope. In addition, the spike also contains a cysteine-rich, disordered C-terminal region inside the
70 envelope. Since the experimental structures have only been resolved for the spike head, as the first step, we model the missing regions to construct a full-length spike structure, which is then palmitoylated at the TM/endodomain and fully-glycosylated, before performing MD simulations. The different steps involved in this process are described below.

Constructing missing regions of the spike head

75 The recent cryo-EM structures of the spike head don't provide coordinates for a number of missing regions including a short peptide segment called the fusion peptide, known to be important for membrane binding and insertion,^{16,17} as well as a number of loops. These missing regions were generated using either template- or fragment-based modeling as described below. Firstly, the cryo-EM structures of the spike head from SARS-CoV-2 (PDB: 6VYB and 6VXX⁹), SARS-CoV (PDB: 5X58,³⁹ 5XLR,⁴⁰ 6CRW⁴¹) and MERS-CoV (PDB: 6Q04⁴²), along with the SARS-CoV-2 RBD structure in complex with human receptor ACE2
80 (PDB: 6M17⁴³) were aligned and using sequence- and structure-based alignment in MOE.⁴⁴ The missing regions were then constructed in MOE by grafting a suitable template region from the aligned structures and using local superimposition based on the flanking residues
85 of the template region. Residues in the grafted templates from other CoV structures were mutated back to those in the SARS-CoV-2 sequence. Missing loop regions in the spike head without a suitable template were constructed using Rosetta ab-initio fragment assembly in the Robetta protein structure prediction server.^{45,46}

Modeling the neck and the HR2 domain

90 Due to lack of a suitable template, the spike neck was modeled using ab-initio modeling in Robetta. The HR2 domain was homology-modeled using the NMR-resolved structure of homologous SARS-CoV (PDB: 2FXP⁴⁷) as the template in MOE. These modeled regions were validated against secondary structure predictions by JPred4.⁴⁸

Developing a trimeric TM model

95 The structural information about the TM region of SARS-CoV-2 spike, which is highly conserved among coronaviruses and known to play a role in the spike assembly and stabilization,^{19,49} is limited, as no structure for this domain has been resolved. The available structure from the HIV virus displays a low sequence homology, making it unsuitable for template-based modeling.⁵⁰ The TM domain of SARS-CoV-2 was thus built in an ab initio
100 manner. Firstly, the location of the TM region in the spike sequence was predicted using TMHMM,⁵¹ a bioinformatics-based approach for predicting membrane spanning regions in protein sequences. Using the predicted TM sequence, a TM helix monomer was constructed in MOE. The monomer was then placed in a membrane patch using CHARMM-GUI^{52,53} with a lipid composition based on the lipoprofiles of the ERGIC,^{22,54} where coronaviruses
105 assemble.²³ The specific lipid compositions of the two leaflets are provided in Table S1. After solvating the system, 1,000 minimization steps and 100 ns equilibration were performed under NPT conditions to relax the monomeric structure in the membrane. Next, in order to construct the trimeric assembly of the TM domain, multiple trimeric configurations of the TM helix were generated using *Multimer Docking* in ClusPro,⁵⁵ a fast Fourier transform-
110 based, rigid-body docking method. Briefly, a total of 10^9 docking models were generated out of which a subset of 10^3 models were selected based on the predicted binding-energy score of Cluspro. We then clustered these models using RMSD and ranked them according to their populations. The models in the top 10 clusters were visually inspected for the expected symmetry between the individual TM monomers and their membrane orientations, and the two

115 best trimeric TM configurations were selected for additional refinement in MD simulations.

The two selected trimeric TM configurations were similarly embedded in the membrane, solvated, energy-minimized, and equilibrated for 200 ns each. The stability of these models was evaluated by calculating the following metrics: i) RMSD of the TM domain with respect to the starting structure using C_α atoms, ii) TM tilt/inclination, which was measured as the
120 angle between the third principal axis of the moment of inertia of the heavy atoms in each TM monomer and that of the whole TM trimer, and iii) coordination number (coordNum implemented in the collective variables (COLVARS) module^{56,57} of VMD⁵⁸) quantifying the number of heavy atom pairs within a cutoff distance, between each TM monomer and the other two TM monomers, using a half-contact value of $d_0=4.5 \text{ \AA}$. Here, for $d \ll d_0$ the contact
125 value is close to 1, at $d = d_0$ the value is 0.5, and at $d \gg d_0$ the value ~ 0 . The total contact is then calculated by summing over all heavy-atom pairs in the two groups. The most stable TM trimeric conformation thus identified using the above three structural metrics was then used to construct the full-length spike structure.

Constructing the C-terminal region

130 The C-terminal region, located in the lumen of the viral envelope, is predicted to be intrinsically disordered, through secondary structure predictions by JPred4, and does not have a suitable template structure. We used Robetta to generate multiple ab initio models of this region. From these, a model was selected based on its spatial arrangement with respect to the TM trimeric domain, namely, a model without any steric clashes when connected to the
135 membrane-inserted TM trimer.

Assembling the full-length spike

The structure of the full-length spike was constructed by assembling together the spike head and the individually constructed domains of the stalk region. The structures of the individual domains were aligned using VMD, with the sequential gaps between them ranging between

140 3-5 residues, which were filled using loop modeling in MODELLER.⁵⁹

Spike glycosylation and palmitoylations

The surface of the SARS-CoV-2 spike is known to be extensively glycosylated,³⁷ but structural information on this feature is by and large missing from the available structures. Recent mass spectrometry studies have successfully identified both the spike's glycosylation sites as
145 well as their specific chemical compositions.^{36,37} Utilizing this information, we added a total of 22 N-glycans and 1 O-glycan to each of the spike monomers (Table S2).

SARS-CoV-2 spike is also known to be palmitoylated at the endodomain,³⁵ but its specific palmitoylation residues have not been determined. Since the cysteine-rich endodomain of the the SARS-CoV spike shares ~95% sequence identity with the SARS-CoV-2 spike, we
150 used the palmitoylation data from the mutagenesis analysis of the cysteine clusters in the former⁶⁰ as a reference for palmitoylation of the SARS-CoV-2 spike. Thus, we adopted 3 palmitoylation sites (residues 1,236, 1,240 and 1,241) in the first two cysteine-rich clusters of the C-terminal region in each spike monomer and covalently linked palmitoyl groups to them.

155 MD simulations of membrane-bound spike

In order to probe the conformational dynamics of the entire spike as well as the motions displayed by the individual domains, we carried out MD simulations of the full-length, membrane-embedded spike structure, both in glycosylated and non-glycosylated forms. The following sections provide more details on the simulation system construction and the MD
160 protocol used.

System preparation

The final membrane-embedded, full-length spike system was prepared with the CHARMM-GUI webserver.^{52,53} 22 N-glycosylations and 1 O-glycosylation were added to each spike

monomer using `Glycan Reader & Modeler` in CHARMM-GUI.^{28,61} Three palmitic acid tails
165 were also added to cysteine residues 1,236, 1,240 and 1,241 in the endodomain of each
monomer. A total of 15 disulfide bonds were introduced between specific cysteine pairs
(15-136, 131-166, 291-301, 336-361, 379-432, 391-525, 480-488, 538-590, 617-649, 662-671,
738-760, 743-749, 840-851, 1,032-1,043 and 1,082-1,126). The protonation states of titratable
residues were set using PROPKA.^{62,63} Following this, the spike protein was inserted into a
170 lipid bilayer with ERGIC composition described above.^{22,54} The system was then solvated
with water and ionized with 0.15 M NaCl, resulting in a simulation system with approximate
dimensions of $250 \times 250 \times 390 \text{ \AA}^3$ with ~ 2.3 million atoms.

MD simulations

MD simulations were performed using NAMD2, a highly scalable MD engine.^{64,65} The
175 CHARMM36m force field was used to represent the proteins, lipids, glycosylations, palmitoylations,
and ions,^{66,67} and the TIP3P model was used for water.⁶⁸ The simulations were
performed as an NPT ensemble with the temperature and pressure maintained at 310 K and
1 bar using Langevin thermostat and barostat,^{69,70} respectively. The SHAKE algorithm was
used to constrain all bonds with hydrogen atoms.⁷¹ For the calculation of van der Waals
180 interactions, a pairlist distance of 13.5 \AA , a switching distance of 12 \AA , and a cutoff of 10 \AA
were used. The Particle mesh Ewald (PME) method under periodic boundary conditions
was utilized for the calculation of electrostatic interactions and forces.⁷²

The system was equilibrated in 6 steps. In the first step, protein backbone and side
chain heavy atoms were restrained using harmonic potentials with force constants of 10
185 and 5 kcal/mol/\AA^2 , respectively, while the system was minimized for 10,000 timesteps using
the steepest descent algorithm, followed by 100 ps of MD simulation. The modeled loop
regions in the spike head and the linker regions connecting the different domain in the stalk
were excluded from the restraints in this step to allow their adjustment. The restraints
on the protein were sequentially reduced to half in each subsequent equilibration step, and

190 the system was further equilibrated for 2 ns in each step. Additionally, coordinates of the lipid head group heavy atoms were restrained along the z axis (membrane normal) in order to maintain the membrane thickness during the whole equilibration phase with decreasing force constants from an initial value of 5 to 0.1 kcal/mol/Å² over the 6 equilibration steps. The first 3 equilibration steps were carried out using a timestep of 1 fs and the subsequent 195 steps and the production runs with a timestep of 2 fs. After the equilibration phase, all the restraints were removed, and a production run was carried out for 5 μ s.

As a control, full-length spike without glycosylations and palmitoylations was similarly prepared, embedded in an ERGIC-like lipid bilayer, equilibrated, and then simulated for 5 μ s.

200 **Analysis**

We first evaluated the structural stability of the spike over the course of the MD simulations, specially in the modeled TM domain. The global conformational dynamics of the spike protein obtained from the simulations were characterized in terms of the bending and twisting motions of the spike head with respect to the different domains in the stalk region. 205 Additionally, the potential role of glycosylation in modulating these motions was analyzed in terms of direct interactions with the lipid bilayer as well as contacts formed between the glycans at the interfaces of different domains and the lipids. Finally, we also investigated the potential effect of palmitoylations on SARS-CoV-2 spike in modulating the membrane curvature.

210 **Stability of the spike domains**

The internal RMSD of different spike domains was calculated by superimposing the trajectory with respect to the starting structure of the respective domain using C_{α} atoms. Furthermore, we characterized the stability of the modeled TM domain by calculating the following: i) TM tilt (described earlier for the evaluation of the TM domain models). ii) TM self-rotation

215 which is calculated as the angle between the first principal axis of the moment of inertia of the heavy atoms in each TM monomer with respect to its equilibrated configuration, and
iii) residue contact map between each TM monomer and the other two, calculated for C_α atoms and using a half-contact distance of 4 Å.

Global structural dynamics

220 The global conformational dynamics of the spike protein were characterized in terms of the orientational changes in the spike head with respect to the different stalk domains. For this, the trajectory was separately aligned using either the spike neck, the HR2 domain, or the TM domain, and the motion of the spike head was quantified with respect to the superimposed region in terms of the i) head bend, i.e., bending motion of the spike head, ii) head twist, i.e.,
225 twisting motion of the spike head in the xy plane, and iii) head distance to the superimposed region. For the calculation of the angles, specific vector representations were introduced for different domains as described below (see Fig. S1). For the calculation of the head bend, the spike head was represented by a vector connecting the C_α atoms of the bottom and top residues of the central helices (i.e., residues 986 to 1,034), and the angle was calculated with
230 respect to its initial position. For the calculation of the head twist, the top of the spike head was first approximated by a triangle with the C_α atoms of residues 146 in the 3 spike monomers as vertices. The rotation of this triangle (quantified as the angle of one of its sides with regard to the initial position) was used as the twist angle. The head distance was calculated between the centroid of the triangle described above and the C_α atom of the
235 bottom residues of the different stalk domains used for superimposition (residue 1159 for the spike neck, residue 1204 for the HR2 domain, and residue 1236 for the TM domain).

Multiple sequence alignment of human coronaviruses

In order to analyze the sequence conservation of the individual spike domains, multiple sequence alignment was performed for the protein from different human coronaviruses (HCoV).

240 Currently a total of seven known HCoV exist, namely, 229E and NL63 from the alpha sub-
family of coronaviruses, and OC43, HKU1, MERS-CoV, SARS-CoV, and SARS-CoV-2, from
the beta subfamily. Additionally, the bat coronavirus RaTG13, a closely related homolog
of SARS-CoV-2, was also included in the sequence alignment. Multiple sequence alignment
for the above eight sequences was carried out using the MAFFT program with the L-INS-i
245 method⁷³ and visualized using Jalview.⁷⁴

Glycan-lipid and glycan-glycan interactions

As the glycan molecules attached to the HR2 domain are situated close to the membrane and
to the glycans on the neck region, their possible interactions were further analyzed. For this,
we identified all heavy atom contacts between glycans and other glycans or lipids (cutoff of
250 4.5 Å). The number of glycan heavy-atom contacts in each monomer for each frame in the
trajectory was denoted as the glycan contact number, calculated separately for glycan-glycan
and glycan-lipid interactions.

Additionally, the bending of the HR2 domain with respect to the membrane as well as
the angle between the neck region and the HR2 domain was also quantified. For the former,
255 the trajectory was first aligned by rotating the first principle axes of the membrane patch
to that in the first frame so that the membrane normal vector lied in the direction of the z .
The HR2 bending was then calculated as the angle between the third principal axis of the
HR2 domain, determined using the C_α atoms, and the z axis representing the membrane
normal. The neck-HR2 bending was determined similarly by obtaining the angle between
260 the third principal axis of the HR2 domain and the neck region.

Correlated motions in the spike

In order to identify potential correlations in the motions of the different domains of the spike,
the Gram-Schmidt orthogonalization process was carried out, a widely used mathematical
method to orthonormalize a set of vectors, in this case vectors representing orientations of

265 the spike head, the HR2 domain, and the TM domain, generating 3 linearly independent, perpendicular vectors. For given vectors \vec{v}_1 (HR2), \vec{v}_2 (TM) and \vec{v}_3 (spike head), representing the C_α atoms of the HR2, TM and the central helices of the spike head, respectively, we define a plane formed by two vectors, e.g., \vec{v}_1 and \vec{v}_2 , and obtain the components of \vec{v}_3 on this plane and a perpendicular plane by performing orthogonalization. Thus, the spike
270 head's coplanar and orthogonal motions with respect to the HR2-TM domain bending can be derived to investigate the inter-correlation between different modes of bending.

Calculation of membrane curvature

Since the palmitoyl tails have been shown to contribute to the bending of the membrane in other viruses,³² the curvature of the membrane around the fully-palmitoylated spike system
275 was examined. Briefly, the coordinates of the upper and lower leaflet headgroups were fitted with a trigonometric surface function by the least square method, to maintain the periodic condition, followed by the numerical calculation of the mean curvature for the fitted grid points. For comparison, the membrane curvature of the non-palmitoylated spike system was also reported.

280 **Results and Discussion**

SARS-CoV-2 binds to its host cell by the spike protein functioning as an extended viral antennae that recognizes the host cell surface receptors. Constructing the full spike and studying its behavior in its physiological environment remains the first and critical step in understanding the working of this most important component of the virus infection machinery. In the presented study, we first develop the full spike structure in its physiological, membrane-bound state, and subsequently, using μ s-scale MD simulations we characterize the conformational dynamics of the protein. We describe these conformational changes in terms of bending and twisting motions of the spike head with respect to the stalk, and examine possible correlations among inter-domain motions as well as the the potential role of glycan-lipid and glycan-glycan interactions in these motions. Sequence comparison with other human coronaviruses allows us to delineate multiple conserved hinge regions in the stalk domain that likely promote the conformational motions of the spike head. Furthermore, we quantify the effects of palmitoylation in the TM domain offering support for the role of this post-transnational modification in modulating the local membrane curvature.

295 **Construction of the membrane-embedded full spike protein**

We combined homology modeling utilizing multiple template structures from homologous coronaviruses like SARS-CoV and MERS-CoV with fragment-based, ab initio modeling and secondary structure prediction to construct the missing regions in the spike head and the stalk including the HR2 domain (Fig. 1A and B and Fig. S2). Additionally, we carried out an expansive procedure for deriving a stable model of the TM domain, where we combine sequence-based secondary structure prediction for the TM region (Fig. S3), with protein-protein docking for generating the initial configurations of TM trimers (Table SS3) followed by MD relaxation of suitable TM trimeric configurations in membrane (Fig. S4 and S5). The above-described modeling procedure was combined with the the biochemical data on

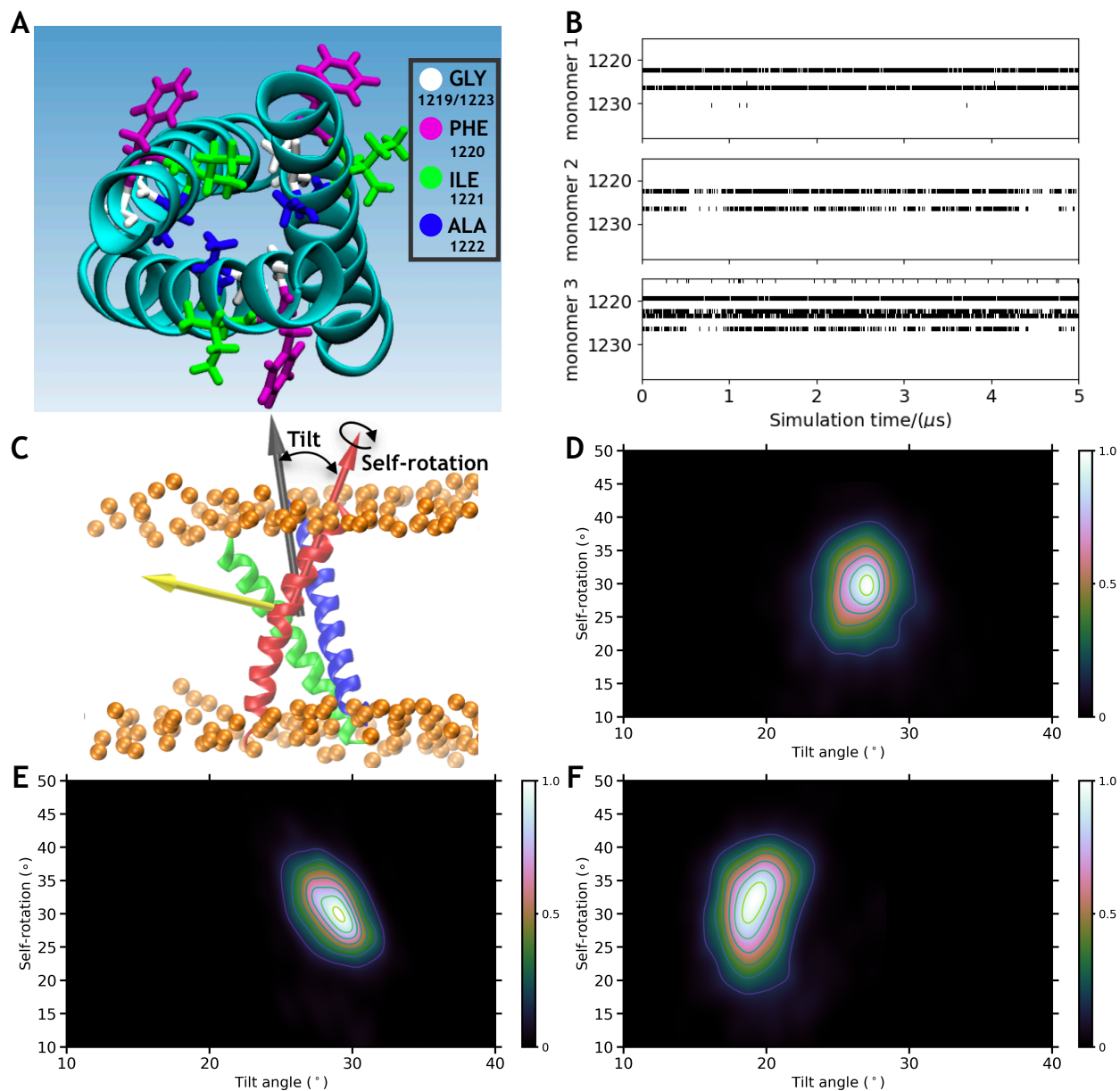


Figure 2: **Stability of the modeled TM domain:** (A) The top view of the modeled TM domain as well as the licorice representation of the "GxxxG" motif (residue 1,219-1,223) are shown. (B) Contact maps between a TM monomer and the other two monomers in the TM domain during the trajectory. A black dot is used to indicate the presence of at least one C_{α} - C_{α} contact (4.0 Å cut-off) between residues of a monomer and the other two monomers. (C) TM monomers' self-rotation and tilt motions. The self-rotation of each monomer is calculated with respect to its equilibrated configuration, and the tilt angle is measured with respect to the whole TM domain. The heat map for the distribution of the tilt angle versus the self-rotation angle is shown for the TM monomer 1 (D), monomer 2 (E), and monomer 3 (F).

305 the palmitoylation sites³⁵ and the recently reported glycomics data to determine the glycosylation compositions of specific Asn residues^{36,37} (Fig. 1C and Table S2), to develop the full-length, palmitoylated and fully-glycosylated spike structure (Fig. 1D).

We have also carried out structural comparison of our full-length spike model with previously published models.^{28,29,75} The spike head structure in our study is obtained from
310 cryo-EM data as done in the previous models. The individual monomers of the spike head are observed to display relatively stable conformations during our simulations after their initial equilibration (Fig. S6). The HR2 domain in the previous studies was modeled using either a coiled-coil template from salmonella autotransporter adhesin,²⁹ using a de novo coiled-coil modeling program,⁷⁵ or using the HR2 template from SARS-CoV.²⁸ Here, we also
315 make use of the HR2 template structure from SARS-CoV,⁴⁷ resulting in a model observed to be stable throughout the trajectory (Fig. S6). As SARS-CoV and SARS-CoV-2 display complete sequence identity in the HR2 domain, our stable model is expected to provide a closer physiological representation of the HR2 domain compared to the coiled-coil models used in the earlier studies. The TM domain in the previous spike models^{28,29} was constructed
320 using a template structure from HIV.⁵⁰ It is noteworthy that the TM helices in HIV each contain a ‘GxxxG’ motif at their N-terminal end, a motif known to be frequently involved in close inter-TM helix interactions.⁷⁶ Consequently, the TM domain assembly in HIV displays inter-helical contacts in this region with the C-terminal half only held together by polar contacts.⁵⁰ As the ‘GxxxG’ motif is present in the central region of the TM helices in the
325 SARS-CoV-2 spike (Fig. 1A), the contacts between the TM helices in the SARS-CoV-2 spike can be expected to lie in this region instead. The TM domain in our spike model is a trimeric assembly with the majority of the inter-helical contacts formed between the central regions of the helices containing the ‘GxxxG’ motifs (Fig. 2A and B). Additionally, the TM domain maintains a relatively stable and symmetric structure during the simulations (Fig. S6 and
330 Fig. 2C, D, E and F). The correct prediction of the different stalk domain structures, including the trimeric HR2 and TM domains, may be critical to correctly describing the global

motions observed in the spike head, as described below.

Bending and twisting motions of spike head around stalk

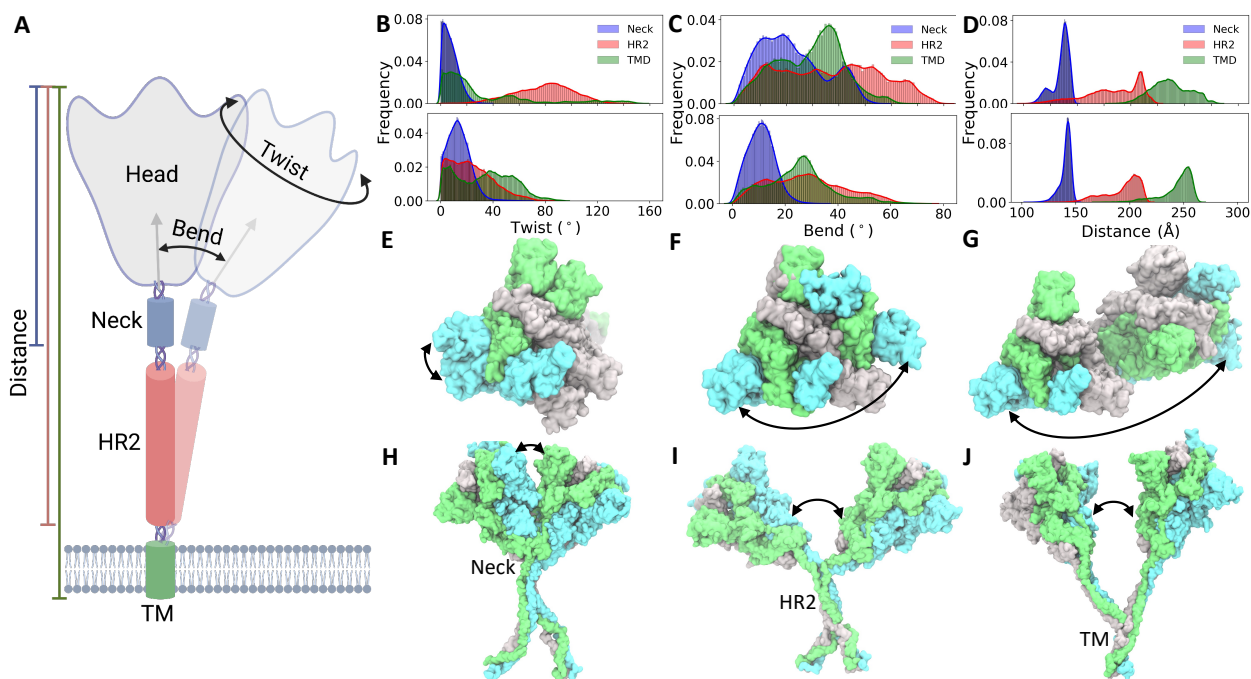


Figure 3: **Global conformational dynamics of spike:** (A) The schematic depicts the calculated bend and twist motions of the spike head (grey) around the spike neck (blue), the HR2 domain (red) and the TM domain (green), as well as the calculated head distances with respect to these domains. The probability distribution (in terms of normalized frequency) of the (B) twist, (C) bend, and (D) distance of the spike head with respect to the spike neck, HR2 domain and TM domain are provided. The top and bottom plots show the distributions observed in the fully-glycosylated spike and the non-glycosylated spike. Example snapshots of the spike (surface representation with each monomer in a different color) displaying the largest head twist with respect to the (E) spike neck, (F) the HR2 domain, and (G) the TM domain are shown for the fully-glycosylated system. Similarly, example snapshots of the spike showing the largest head bend with respect to the (H) spike neck, (I) the HR2 domain, and (J) TM domain are shown.

To further describe the the dynamics employed by the spike to sample the crowded
 335 cellular surface, we characterized its global motions in terms of the movements of the spike
 head with respect to the stalk region during the simulations (see Supplementary Video 1).
 These motions were quantified in terms of head-twist and head-bend angles as well as head-
 distances calculated separately with respect to either the spike neck, the HR2 domain, or the
 TM domain (Fig. 3A). In its fully-glycosylated, native form, the spike can sample large head-

340 twist angles (in the xy plane) with respect to both the HR2 and TM domains, highlighting the large conformational flexibility available along this degree of freedom (Fig. 3B top, F and G). A wide normal distribution of head-twist angles is observed with respect to the HR2 domain, with the highest sampled configurations turned by almost 90° relative to the starting structure. With respect to the TM domain, although lowly populated, spike-head can sample
345 conformations that are almost oppositely facing (turned by as much as 160°) compared to the starting orientation. A relatively smaller range of motion is allowed around the spike neck, with the most populous regions lying close to the starting structure (Fig. 3B top and E). These results can be directly related to the length of the linkers connecting the different domains of spike. The domains around which spike head displays larger motions (HR2 and
350 TM domains) are connected by longer linkers (linker 2: 13 residues; linker 3: 11 residues), compared to the domain around which spike head displays relatively more restricted motions (spike neck), that is connected to the spike head by a smaller linker (linker 1: 7 residues). Furthermore, direct comparison of the head-twist motions in the glycosylated form to the control non-glycosylated system shows surprisingly that these motions become restricted
355 in the latter (Fig. 3B bottom), especially with respect to the head-twist angles calculated around the HR2 and TM domains. These results point to an unexpected enhancing role of glycans in the overall dynamics of the spike.

Similarly, the spike head displays a broad range of head-bend angles with respect to the HR2 domain, sampling configurations almost orthogonal to the starting structure (Fig. 3C
360 top and I). Comparatively, smaller head-bend motions are observed around the spike neck and the TM domain, with the respective populations showing distributions over a range of $0 - 60^\circ$ angles (Fig. 3C top, H and J). The extent of these bending angles is also directly related to the changes in the head-distances with respect to these domains, with the largest distance changes observed in conformations with the largest head-bend angles (Fig. 3D top).
365 Thus, the spike head can move as much as 125 \AA with respect to the HR2 domain compared to its starting configuration. Again, comparison of the glycosylated and non-glycosylated

forms shows that the spike domains rigidify (display relatively smaller range of motion) in the absence of glycans (Fig. 3C and D bottom). The potential role of the glycans in modulating global motions of the spike is further discussed in later sections.

370 Mechanistically, these motions (bend/twist/translation) of the spike head represent major degrees of freedom available bestowed upon the receptor-binding domain (located in the spike head) to locate and bind to the target ACE2 receptors in the host cell membranes. As spike proteins have been found to be sparsely distributed on the virion envelope (recent cryo-EM and tomography studies^{12,77}), the flexibility and the allowed range of motions of
375 the spike domain may play an important role in extending the search space covered by the spikes. Cryo-EM and other computational studies have similarly reported the conformational diversity of the spike head around its stalk.^{75,78} Compared to our results, the conformational variability of the spike head was observed to a smaller extent in one of these MD studies,⁷⁵ likely due to the crowded conditions used in the simulation system, where 4 spikes were
380 packed at close proximity to each other, preventing the protein to amply sample the available space. The crowded distribution used in the above study is unlikely to represent the physiological distribution observed in the whole virion cryo-EM study where approximately one spike is observed per 1000 nm² of the viral surface area.⁷⁷ Additionally, previous studies have quantified the inter-domain motions compared to the spike head orientational changes
385 measured in our study, that are more directly related to the motions of the RBD binding to the host receptors, and hence directly related to the function of the protein.

Sequence conservation in multiple hinge regions

In order to examine potential generality of the observed global motions in the SARS-CoV-2 spike, we analyzed the sequence similarity of the spike protein, specifically focusing on
390 the flexible regions connecting the different domains. These include the small connecting linkers between the spike head and the neck (linker 1: NTVYDPL), between the spike neck and the HR2 domain (linker 2: HTSPDVLDLGDIG), and between the HR2 domain and the

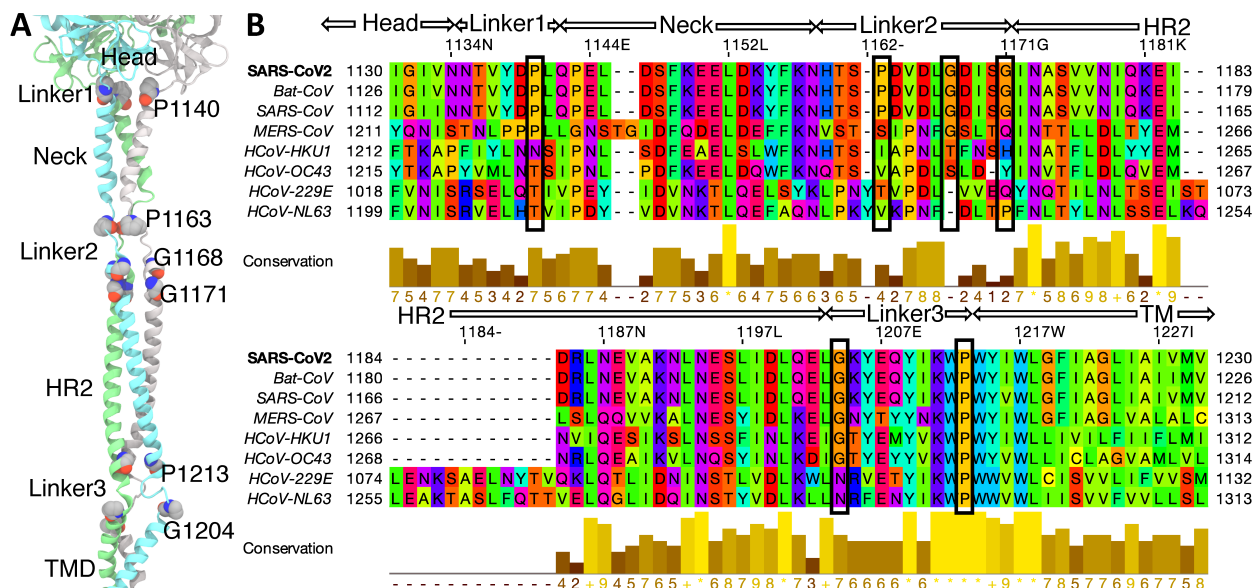


Figure 4: **Multiple sequence alignment of spike homologs.** (A) Structure of the stalk region of spike in our model is shown. Pro and Gly residues of the linkers are shown in van der Waals representation. (B) The sequence of SARS-CoV-2 spike is aligned with those from closely related viruses (bat coronavirus RaTG13 and six other human HCoV viruses). The residues are colored according to their physicochemical properties in different shades of rainbow colors (Taylor coloring scheme). The different domains and linkers are labeled at the top of the sequence. The completely and partially conserved Pro and Gly residues present in the flexible linkers are enclosed in black boxes. Conservation between the different sequences is visualized as histograms with their heights, and color variation from brown to yellow, reflecting the level of conservation of physicochemical properties. Conserved and identical amino acid columns show the highest score of 11 and are indicated by '*', whereas conserved and similar amino acid columns show a score of 10 and are indicated by '+'. Other groupings are indicated by lower scores accordingly.

TM domain (linker 3: GKYEQYIKWP). The three linkers are found to be rich in Pro and Gly, residues known to contribute to the conformational flexibility in proteins^{79,80} (Fig. 4A).

395 Additionally, the Pro residues in these regions may be important for the hinge or recoil motion suggested to take place in the refolding of the spike during the fusion process, which may result in an elongated post-fusion state of the protein.¹⁵

To further elucidate the possible functional importance of these residues in the mobility of the spike, multiple sequence alignment was carried out with other coronaviruses infecting
400 humans. Interestingly, all Pro and Gly residues present in the flexible regions are completely conserved in the homologous Bat-CoV and human SARS-CoV, and to a large extent in human MERS-CoV where only Pro-1163 is not conserved (Fig. 4B). Most of these residues are

not conserved in other human coronaviruses except Pro-1213 that is completely conserved in all coronaviruses. We propose that the presence of these conserved flexible residues in the regions connecting the different domains not only promotes the observed conformational flexibility of the spike head around the stalk domains but, compared to other coronaviruses, also provides an evolutionary advantage to SARS-CoV-2, SARS-CoV and MERS-CoV, coronaviruses known to be highly infective in humans.

Stalk bending motions modulated by glycan-lipid and glycan-glycan interactions

In addition to the important roles glycans can play in shielding the spike from neutralizing antibodies,²¹ we noted that the glycans also form contacts with each other, as well as with the lipid bilayer, during the simulations (see example of these phenomena in Supplementary Video 2). To obtain deeper insights into the impact of these glycan interactions, especially with respect to the spike global motions, we evaluated the relationship between the inclination of the different stalk domains and the glycan-lipid or glycan-glycan contacts formed, both in the native glycosylated spike simulations and the control non-glycosylated spike. In the glycosylated spike when the inclination angle of the HR2 domain exceeded a threshold of around 20°, we observed formation of more glycan-lipid or glycan-glycan contacts, indicating a correlation between the bending of the HR2 domain and the glycan interactions (Fig. 5A, B, C and D). On examining the distribution of the HR2 tilt angles towards the membrane in the glycosylated spike simulation, only ~20% of the trajectory was found to lie between 0° – 20° (vertical orientation) compared to ~80% of the non-glycosylated trajectory that lied below the 20° inclination. It can be inferred that in the fully-glycosylated spike when no glycan-lipid interactions are formed the HR2 domain can move freely within the 20° inclination (Fig. 5A and B); the glycan-lipid interactions appear when the inclination angle lies between 20° – 50°, further restraining the flexibility of the HR2 hinge motion in this range. The HR2-neck inclination shows similar properties as the HR2-membrane inclination, where

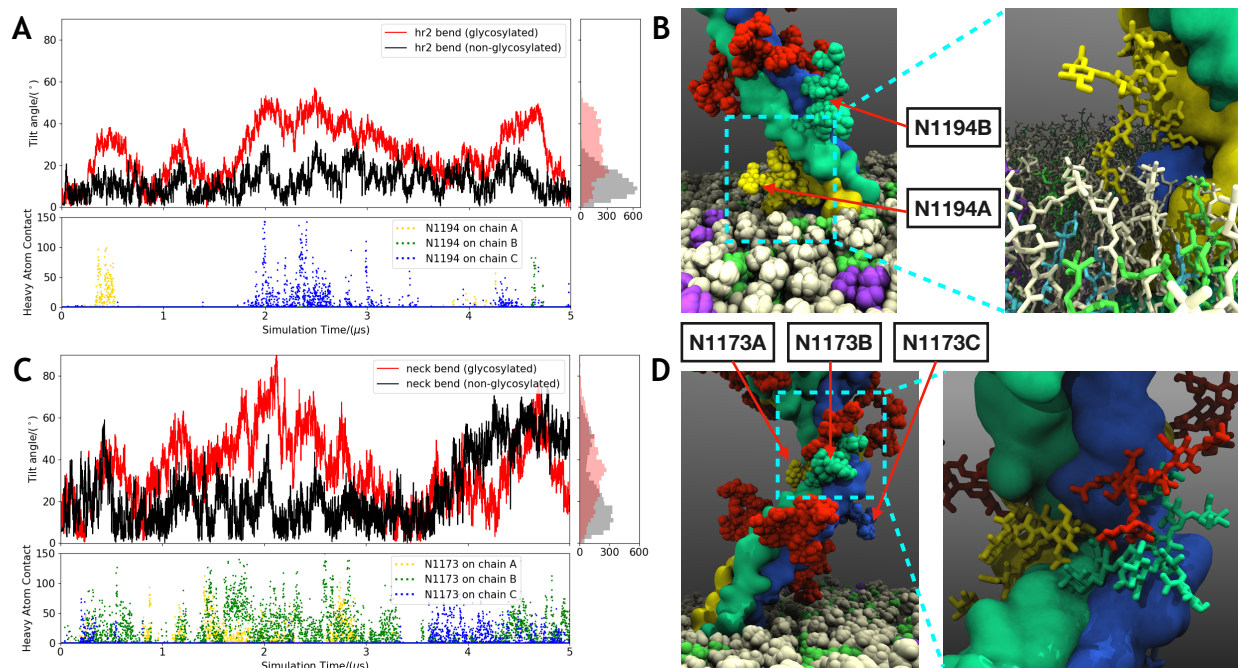


Figure 5: Glycan-glycan and glycan-lipid interactions: (A) Time series of the HR2 tilt angle towards the membrane in the glycosylated (red) and non-glycosylated (black) simulations (distribution histogram shown on the side). The number of heavy-atom contacts (within 4.5 Å) between the glycans at N1194 and lipids over the course of the trajectory is also shown (bottom). The contacts shown use the same color scheme as the corresponding monomer and glycan representations shown in (B) and (D). (B) Snapshot of the glycan at N1194 contacting with lipids in van der Waals representation and zoomed in detail shown in inset in licorice representation. Glycans in the front are labeled. (C) Similar tilt angle between the HR2 domain and the neck region of the glycosylated (red) and non-glycosylated (black) simulations along with the distribution histogram (side). The glycan-glycan heavy atom contacts between N1173 on HR2 and N1158 on the neck region is shown as points (bottom). (D) Snapshot of the glycan at N1173 (same color scheme as chains) contacting with glycans at N1168 (red) in van der Waals representation and zoomed in detail shown in inset in licorice representation.

larger inclination angles, ranging between 20° – 90°, mostly correspond with high glycan-
glycan contacts (Fig. 5C and D). Although the HR2 domain in the non-glycosylated spike
430 can show a similar range of motion (with respect to the neck) as in the native form, almost
half of the sampled configurations lie close to the starting vertical orientation (0° – 20°)
(Fig. 5A and C), suggesting coupling between glycan-glycan interactions and the dynamics
of the stalk. Overall, the data indicate that the glycans around the flexible linkers can fur-
435 ther enhance the relative motion between the stalk domains, in turn increasing the probed
space by the spike head.

N-glycans have been reported to be involved in mediating receptor recognition and facilitating cell adhesion.⁸¹ The self-recognition and conjugation has been previously demonstrated for N-linked complex glycans with multiple N-acetylglucosamine (GlcNAc) termini.^{82,83}
440 The glycans located at the bottom end of the neck region (N1158), as well as the top (N1173) and bottom (N1194) ends of the HR2 domain, are all complex glycans containing multivalent GlcNAc (Table SS2). Additionally, these glycans are completely conserved in human coronaviruses (Fig. 4B). Overall, we propose that these complex glycosylation sites not only shield the stalk from epitopes, by occupying the space around the flexible hinges connecting
445 the neck, HR2 and TM domains, but also contribute to the flexibility of the stalk, that may in turn increase the range of motion of the spike head allowing for more effective sampling of the human cell surface.

On further examination of the glycan-lipid interactions we observed that the majority of these contacts are formed with the zwitterionic lipids (POPC/POPE) followed by anionic
450 lipids (POPS) (Fig. S7). Majority of the interactions between glycans and lipids consist of hydrogen bonds formed between O/N atoms of the N-acetyl-D-neurominic acid (NE5AC) or D-galactose sugars present in the glycans and N atoms from PE head group or O atoms in the PE/PC phosphate groups. These interactions may further stabilize the inclined conformations of the HR2 domain with respect to the membrane. The contacts with cholesterol are
455 minimal, which is expected as cholesterol is known to insert more deeply in the membrane core compared to the phospholipids.^{84,85}

Correlated inter-domain motions in the spike

The global conformational dynamics of the spike were further examined for the presence of correlated motions between its different domains. A correlation between the motions of the
460 spike head and the TM domain about the HR2 domain is evident in the fully-glycosylated spike (Fig. 6A and B; see Supplementary Video 3). A high Pearson's coefficient of 0.85 between the Head1 and TM vectors (Fig. 6A and B), representing the spike head's coplanar

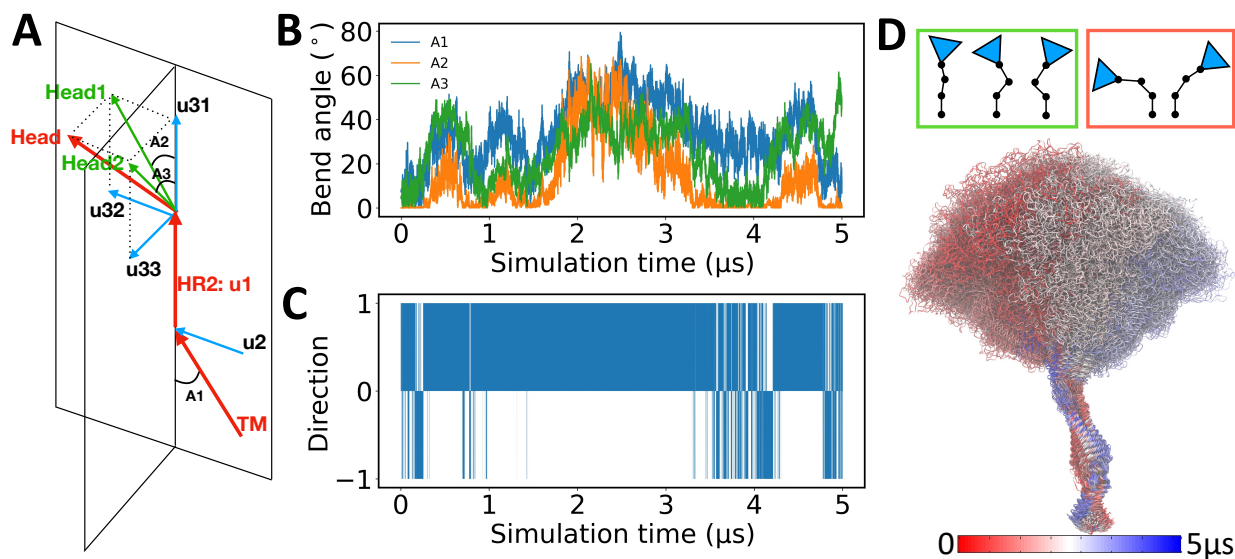


Figure 6: **Global dynamics correlation in spike:** (A) Vectors representing the different domains of the spike (arrows in red): spike head (top), HR2 (middle), TM (bottom), and their orthogonalized sub-vectors (arrows in blue) in orthogonal planes of spike head (u31, u32 and u33) and TM (u2). Head1 and Head2 (arrows in green) are respectively the portions of the head vector on the HR2-TM plane and the one perpendicular to it, representing the lateral and axial movements of the head. (B) The time-series of the A1, A2 and A3 angles introduced in (A) are shown for the fully-glycosylated spike. (C) The time-series of the Gram-Schmidt process for computing the dot product between the u2 and u32 vectors in (A) is shown. A value of +1 means the vectors are pointing in the same direction and a value of -1 means that they are pointing in the opposite direction. (D) The conformational ensemble of the spike generated by taking 100 images evenly distributed over the trajectory of the fully-glycosylated spike and superimposed on the stalk region (the HR2 and TM domains). The color changes from red to blue over the time. Cartoon representation of the spike on top (spike head depicted as triangle and different stalk domains connected by lines) in green box are the favored (highly sampled) conformations, whereas representations in red box are rarely sampled during the simulations.

motion with respect to the HR2-TM domain bending, further confirms the observed correlation in the motions of these domains. A Pearson's coefficient of 0.48 was also observed
465 between the Head2 and TM vectors (Fig. 6A and B), showing that the spike head's orthogonal motion is similarly correlated, but to a lesser degree, with respect to the HR2-TM domain bending. This is an interesting observations in that the orientational motions of the different domains of the spike appear to remain inter-related even though they are connected by flexible hinge regions. Without glycosylations, not only is the magnitude of inter-domain
470 motions dampened (Fig. S8), but, at the same time, the correlation between them is also reduced, with the non-glycosylated spike showing relatively lower Pearson's coefficient, 0.78

and 0.39, respectively, for the in-plane and orthogonal motions of the spike head with respect to the HR2-TM domain bending. In general, the correlated motions between the different spike domains can be attributed to its unique tertiary structure and can be considered an
475 intrinsic property of the protein, but, at the same time, glycosylation may further amplify these motions through the extensive glycan-lipid and glycan-glycan interactions in the system (Fig. 5). For example, interactions of the glycans located at the bottom end of the HR2 domain with the lipids seem to promote its bending towards the membrane that may in turn facilitate formation of more contacts between the glycans located at the top/opposite end of
480 the HR2 domain and glycans in the neck region, promoting bending of the spike neck and head in the opposite direction.

The Gram-Schmidt analysis (see Methods) indicates concerted motions of the spike head and the TM domain about the HR2 domain. Here, the orthonormal vectors representing the spike head and the TM domain are oriented in the same direction about the HR2 domain
485 for 88% of the simulation in the fully-glycosylated spike (Fig. 6C). Non-glycosylated spike displays a similar behavior, with the the orthonormal vectors oriented in the same direction for 89% of the time, further highlighting that these motions are largely a property of the spike structure itself. Topologically, this reflects a configuration where the spike head and the TM domain orient diametrically opposing each other around the HR2 domain, a configuration
490 which may further extend the reach of the spike head and the RBD for the target cellular receptors (Fig. 6D). Additionally, considering the fact that the inclination of TM domain in the membrane can be substantial, ranging between 0-30° (Fig. S9), the concerted, opposing bendings of the spike head and the TM domain may pull the head back to an upwards position with respect to the membrane, preventing it from severe inclination and falling
495 into the membrane and thus maintaining a proper configuration for binding with the ACE2 receptor.

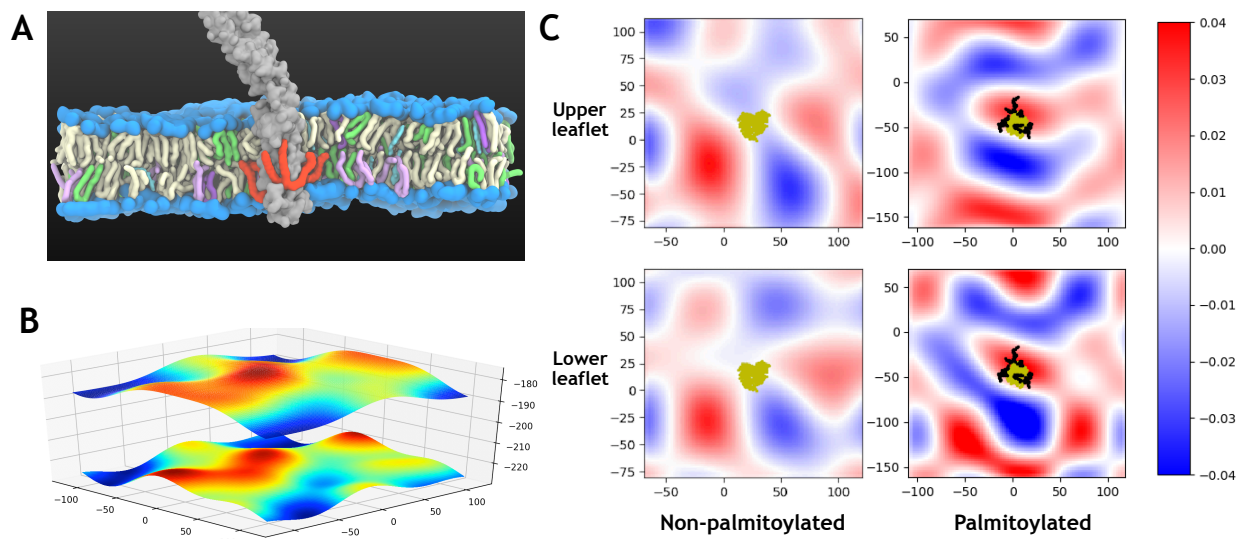


Figure 7: **Membrane curvature modulation by palmitoylation:** (A) Section view of the membrane embedded palmitoylated spike is shown. The coloring scheme for the lipid composition is identical to Fig. 1D, with the palmitoylations and lipid head groups displayed in red and blue surfaces, respectively. (B) Fitted surface (periodic) with lipid head group coordinates used for the calculation of curvature. (C) The mean curvature of the fitted membrane surface of the upper and lower leaflets as in (B). The left and right panels show the curvature profiles at the end of the simulations of non-palmitoylated and palmitoylated TM domains. The heavy atoms of the TM domain (yellow) and palmitoyls (black) are also shown.

Role of palmitoylations in modulating membrane curvature

Protein S-palmitoylation, the covalent lipid modification of the Cys residues, widely exists in viral membrane proteins and reported to contribute to membrane fusion and other processes.^{31–33} To characterize the effect of spike palmitoylations on the membrane shape, we compared the membrane curvature for the control simulation of non-palmitoylated spike with the palmitoylated (and fully glycosylated) one, taking into account the periodicity in fitting the lipid head groups (Fig. 7A, B). The two simulations share a similar membrane size, but the palmitoylated system seems to deform the membrane more than the non-palmitoylated system (Fig. 7C). Keeping in mind that the periodic boundary conditions used in the simulations can suppress long-range curvature effects, comparing the local curvature (in the range of 20-50 Å from the center of the TM domain) shows that the palmitoylated TM domain is found within a positively curved membrane, whereas the non-palmitoylated TM prefers a saddle point. Thus, the palmitic tails may induce a positive curvature locally in the vicinity

510 of the TM domain.

The lipidation of other viral proteins is relatively well studied. In influenza virus, cryo-ET measurements have shown membrane curvature induced by HA palmitoylation.³² Apart from this, non-acylated HA mutants of influenza virus show a severe negative effect on the membrane fusion step of the infection process.⁸⁶ As shown by the results above, SARS-CoV-
515 2 spike palmitoylation may similarly contribute to the bending of the membrane, especially locally, and thus play a role in the viral fusion with the host cell membrane.

Conclusions

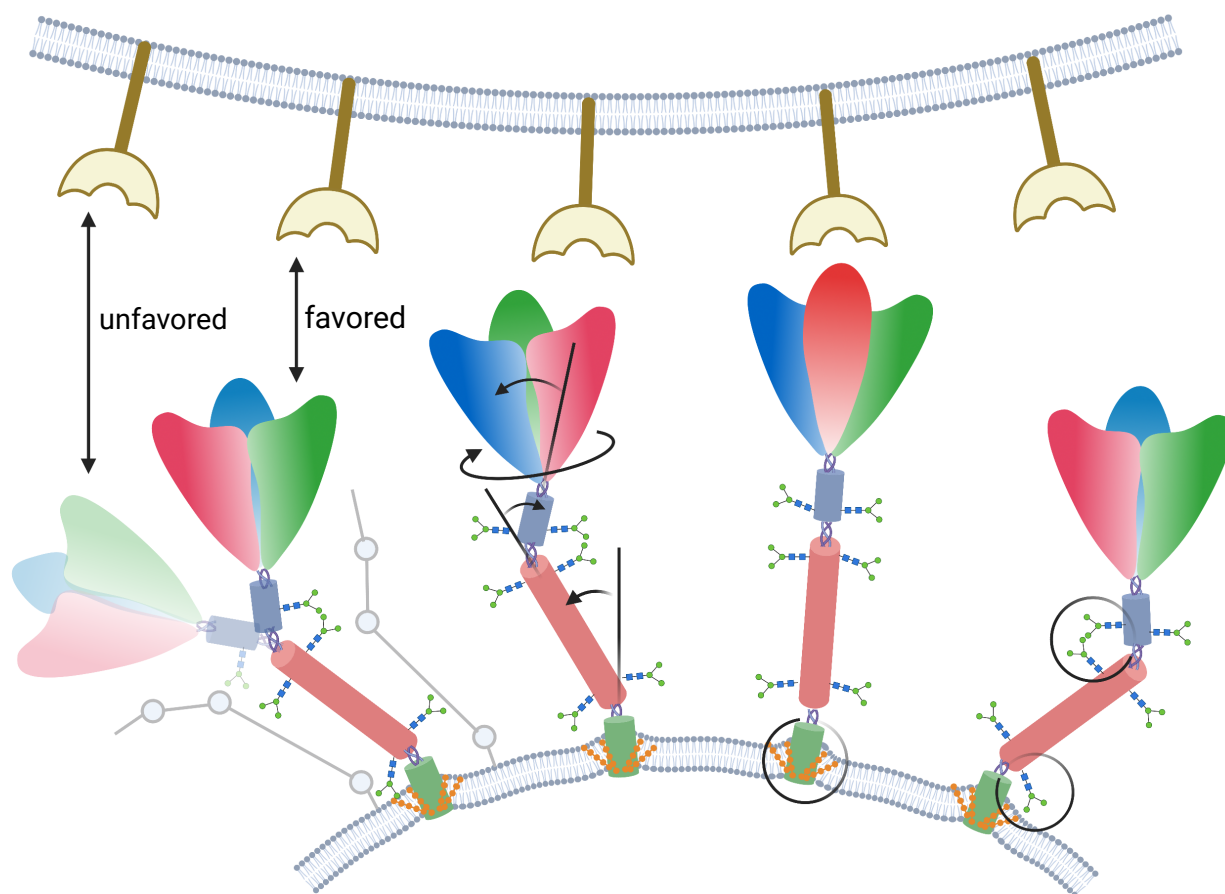


Figure 8: **Impact of post-translational modifications on spike global dynamics and membrane deformation.** The host cellular membrane with binding receptors are shown at the top, and the viral envelope containing multiple spikes at the bottom. Each spike head is shown in three colors, representing the three constituting monomers, and the different stalk domains are also shown in different colors. Spike copies shown from left to right highlight the following results: i) correlated motions of the spike head and stalk allowing reorientation of the spike head with respect to the receptor (stick representations of the spike in favorable and unfavorable orientations also shown on the side); ii) global motion of the spike head around the stalk (twisting and bending motions) allowing it to broaden its reach for the cellular receptors and optimally binding to it; iii) membrane curvature induced by palmitoylation potentially plays a role in viral fusion; iv) glycan-lipid and glycan-glycan interactions restricting/modulating the stalk motions. Created with BioRender.com.

The spike protein of SARS-CoV-2 constitutes a major component in the viral infection and, therefore, the main target for both diagnostics and vaccines developed for the disease
520 caused by the virus. To better understand the many mechanistic steps the spike is involved
in, it is imperative to characterize its structure and dynamics in the most detailed manner and

under the most realistic conditions possible. We report here a full model for the full-length spike in its native, glycosylated and palmitoylated, membrane-bound form which we use for several microseconds of atomistic simulations highlighting some of the dynamics employed by the spike to search for and locate the host cell receptors effectively (Fig. 8). The hinge regions identified in the stalk domain directly modulate the conformational landscape of the spike head, where the RBDs reside, and may thus represent important structural elements regulating the effective role of the spike in binding its receptor. These highly conserved hinges may offer novel points of target for developing alternative drugs and antibodies that bind to and modulate the structure/dynamics of these regions, and therefore the effectiveness of the spike protein. In addition we propose possible roles for posttranslational modifications of spike, specifically glycosylation and palmitoylation, in modulating its dynamics. The results provide a deeper understanding of the intricate roles and functions of the different elements of the spike, each evolved in order to maximize the chance of successful infection and transmission of the genetic material to the host cell. Future extended simulations studying the spike behavior in a full viral envelope will allow further characterization of the impact of a more crowded environment of the envelope and its proteins in modulating the structure and dynamics of the spike and will provide additional insight into the inner working of the virus.

540 **Acknowledgement**

This study was supported by the National Institutes of Health under awards P41-GM104601 and R01-GM123455 (to ET). Simulations were performed using computational resources provided by Lawrence Livermore National Laboratory, Oak Ridge Leadership Computing Facility, and Microsoft Azure, under COVID-19 HPC Consortium Grant MCB200183 (to
545 KK).

References

- (1) Huang, C.; Wang, Y.; Li, X.; Ren, L.; Zhao, J.; Hu, Y.; Zhang, L.; Fan, G.; Xu, J.; Gu, X., et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **2020**, *395*, 497–506.
- 550 (2) Chan, J. F.-W.; Yuan, S.; Kok, K.-H.; To, K. K.-W.; Chu, H.; Yang, J.; Xing, F.; Liu, J.; Yip, C. C.-Y.; Poon, R. W.-S., et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet* **2020**, *395*, 514–523.
- (3) Chan-Yeung, M.; Xu, R.-H. SARS: epidemiology. *Respirology* **2003**, *8*, S9–S14.
- 555 (4) Zumla, A.; Hui, D. S.; Perlman, S. Middle East respiratory syndrome. *Lancet* **2015**, *386*, 995–1007.
- (5) Li, W.; Moore, M.; Vasllieva, N.; Sui, J.; Wong, S.; Berne, M.; Somasundaran, M.; Sullivan, J.; Luzuriaga, K.; Greeneugh, T.; Choe, H.; Farzan, M. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **2003**, *426*, 450–454.
- 560 (6) Wan, Y.; Shang, J.; Graham, R.; Baric, R. S.; Li, F. Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus. *Journal of Virology* **2020**, *94*, e00127–20.
- (7) Tang, T.; Bidon, M.; Jaimes, J. A.; Whittaker, G. R.; Daniel, S. Coronavirus membrane fusion mechanism offers a potential target for antiviral development. *Antiviral Research*
565 **2020**, *178*, 104792.
- (8) Wrapp, D.; Wang, N.; Corbett, K. S.; Goldsmith, J. A.; Hsieh, C.-L.; Abiona, O.; Graham, B. S.; McLellan, J. S. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **2020**, *367*, 1260–1263.

- (9) Walls, A. C.; Park, Y.-J.; Tortorici, M. A.; Wall, A.; McGuire, A. T.; Veerler, D.
570 Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* **2020**,
- (10) Tortorici, M. A.; Beltramello, M.; Lempp, F. A.; Pinto, D.; Dang, H. V.; Rosen, L. E.;
McCallum, M.; Bowen, J.; Minola, A.; Jaconi, S., et al. Ultrapotent human antibodies
protect against SARS-CoV-2 challenge via multiple mechanisms. *Science* **2020**, *370*,
950–957.
- 575 (11) Henderson, R.; Edwards, R.; Mansouri, K.; Janowska, K.; Stalls, V.; Gobeil, S.;
Kopp, M.; Li, D.; Parks, R.; Hsu, A.; Borgnia, M.; Haynes, B.; Acharya, P. Control-
ling the SARS-CoV-2 spike glycoprotein conformation. *Nature Structural & Molecular*
Biology **2020**, *27*, 925–933.
- (12) Ke, Z.; Oton, J.; Qu, K.; Cortese, M.; Zila, V.; McKeane, L.; Nakane, T.; Zivanov, J.;
580 Neufeldt, C. J.; Cerikan, B., et al. Structures and distributions of SARS-CoV-2 spike
proteins on intact virions. *Nature* **2020**, *588*, 498–502.
- (13) Toelzer, C.; Gupta, K.; Yadav, S. K.; Borucu, U.; Davidson, A. D.; Williamson, M. K.;
Shoemark, D. K.; Garzoni, F.; Stauffer, O.; Milligan, R., et al. Free fatty acid binding
pocket in the locked structure of SARS-CoV-2 spike protein. *Science* **2020**, *370*, 725–
585 730.
- (14) Juraszek, J.; Rutten, L.; Blokland, S.; Bouchier, P.; Voorzaat, R.; Ritschel, T.;
Bakkers, M. J. G.; Renault, L. L. R.; Langedijk, J. P. M. Stabilizing the closed SARS-
CoV-2 spike trimer. *Nature Communications* **2021**, *12*, 244.
- (15) Cai, Y.; Zhang, J.; Xiao, T.; Peng, H.; Sterling, S. M.; Walsh, R. M.; Rawson, S.; Rits-
590 Volloch, S.; Chen, B. Distinct conformational states of SARS-CoV-2 spike protein.
Science **2020**, *369*, 1586–1592.
- (16) Tamm, L. K.; Han, X. Viral fusion peptides: a tool set to disrupt and connect biological
membranes. *Bioscience reports* **2000**, *20*, 501–518.

- (17) White, J. M.; Delos, S. E.; Brecher, M.; Schornberg, K. Structures and mechanisms
595 of viral membrane fusion proteins: multiple variations on a common theme. *Critical reviews in biochemistry and molecular biology* **2008**, *43*, 189–219.
- (18) Walls, A. C.; Tortorici, M. A.; Snijder, J.; Xiong, X.; Bosch, B.-J.; Rey, F. A.; Veesler, D. Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion. *Proceedings of the National Academy of Sciences* **2017**, *114*, 11157–11162.
- 600 (19) Broer, R.; Boson, B.; Spaan, W.; Cosset, F.-L.; Corver, J. Important role for the transmembrane domain of severe acute respiratory syndrome coronavirus spike protein during entry. *Journal of virology* **2006**, *80*, 1302–1310.
- (20) Thorp, E. B.; Boscarino, J. A.; Logan, H. L.; Goletz, J. T.; Gallagher, T. M. Palmitoylations on murine coronavirus spike proteins are essential for virion assembly and
605 infectivity. *Journal of virology* **2006**, *80*, 1280–1289.
- (21) Watanabe, Y.; Bowden, T. A.; Wilson, I. A.; Crispin, M. Exploitation of glycosylation in enveloped virus pathobiology. *Biochimica et Biophysica Acta – General Subjects* **2019**, *1863*, 1480–1497.
- (22) Schweizer, A.; Clausen, H.; Van Meer, G.; Hauri, H.-P. Localization of O-glycan initiation, sphingomyelin synthesis, and glucosylceramide synthesis in Vero cells with respect
610 to the endoplasmic reticulum-Golgi intermediate compartment. *Journal of Biological Chemistry* **1994**, *269*, 4035–4041.
- (23) Hogue, B. G.; Machamer, C. E. *Nidoviruses*; American Society of Microbiology, 2008; pp 179–200.
- 615 (24) Grant, O. C.; Montgomery, D.; Ito, K.; Woods, R. J. Analysis of the SARS-CoV-2 spike protein glycan shield reveals implications for immune recognition. *Scientific Reports* **2020**, *10*, 14991.

- (25) Serapian, S. A.; Marchetti, F.; Triveri, A.; Morra, G.; Meli, M.; Moroni, E.; Sautto, G. A.; Rasola, A.; Colombo, G. The Answer Lies in the Energy: How Simple
620 Atomistic Molecular Dynamics Simulations May Hold the Key to Epitope Prediction on the Fully Glycosylated SARS-CoV-2 Spike Protein. *The Journal of Physical Chemistry Letters* **2020**, *11*, 8084–8093.
- (26) Gur, M.; Taka, E.; Yilmaz, S. Z.; Kilinc, C.; Aktas, U.; Golcuk, M. Conformational
625 transition of SARS-CoV-2 spike glycoprotein between its closed and open states. *Journal of Chemical Physics* **2020**, *153*, 075101.
- (27) Peters, M. H.; Bastidas, O.; Kokron, D. S.; Henze, C. E. Static all-atom energetic mappings of the SARS-Cov-2 spike protein and dynamic stability analysis of “Up” versus “Down” protomer states. *PLoS One* **2020**, *15*, e0241168.
- (28) Woo, H.; Park, S.-J.; Choi, Y. K.; Park, T.; Tanveer, M.; Cao, Y.; Kern, N. R.; Lee, J.;
630 Yeom, M. S.; Croll, T. I.; Seok, C.; Im, W. Developing a Fully Glycosylated Full-Length SARS-CoV-2 Spike Protein Model in a Viral Membrane. *Journal of Physical Chemistry B* **2020**, *124*, 7128–7137.
- (29) Casalino, L.; Gaieb, Z.; Goldsmith, J. A.; Hjorth, C. K.; Dommer, A. C.; Harbison, A. M.; Fogarty, C. A.; Barros, E. P.; Taylor, B. C.; McLellan, J. S., et al. Beyond
635 shielding: the roles of glycans in the SARS-CoV-2 spike protein. *ACS Central Science* **2020**, *6*, 1722–1734.
- (30) Sikora, M.; von Bülow, S.; Blanc, F. E.; Gecht, M.; Covino, R.; Hummer, G. Computational epitope map of SARS-CoV-2 spike protein. *PLoS Computational Biology* **2021**, *17*, e1008790.
- 640 (31) Aicart-Ramos, C.; Valero, R. A.; Rodriguez-Crespo, I. Protein palmitoylation and sub-cellular trafficking. *Biochimica et Biophysica Acta – Biomembranes* **2011**, *1808*, 2981 – 2994.

- (32) Chlanda, P.; Mekhedov, E.; Waters, H.; Sodt, A.; Schwartz, C.; Nair, V.; Blank, P. S.; Zimmerberg, J. Palmitoylation contributes to membrane curvature in influenza A virus assembly and hemagglutinin-mediated membrane fusion. *Journal of Virology* **2017**, *91*, e00947–17.
- (33) Parton, D. L.; Tek, A.; Baaden, M.; Sansom, M. S. P. Formation of Raft-Like Assemblies within Clusters of Influenza Hemagglutinin Observed by MD Simulations. *PLoS Computational Biology* **2013**, *9*, e1003034.
- (34) Ernst, A. M.; Syed, S. A.; Zaki, O.; Bottanelli, F.; Zheng, H.; Hacke, M.; Xi, Z.; Rivera-Molina, F.; Graham, M.; Rebane, A. A.; Björkholm, P.; Baddeley, D.; Toomre, D.; Pincet, F.; Rothman, J. E. S-Palmitoylation Sorts Membrane Cargo for Anterograde Transport in the Golgi. *Developmental Cell* **2018**, *47*, 479 – 493.e7.
- (35) Petit, C. M.; Chouljenko, V. N.; Iyer, A.; Colgrove, R.; Farzan, M.; Knipe, D. M.; Kousoulas, K. G. Palmitoylation of the cysteine-rich endodomain of the SARS–coronavirus spike glycoprotein is important for spike-mediated cell fusion. *Virology* **2007**, *360*, 264–274.
- (36) Shajahan, A.; Supekar, N. T.; Gleinich, A. S.; Azadi, P. Deducing the N-and O-glycosylation profile of the spike protein of novel coronavirus SARS-CoV-2. *bioRxiv* **2020**,
- (37) Watanabe, Y.; Allen, J. D.; Wrapp, D.; McLellan, J. S.; Crispin, M. Site-specific glycan analysis of the SARS-CoV-2 spike. *Science* **2020**, *369*, 330–333.
- (38) Xu, W.; Wang, M.; Yu, D.; Zhang, X. Variations in SARS-CoV-2 Spike Protein Cell Epitopes and Glycosylation Profiles During Global Transmission Course of COVID-19. *Frontiers in Immunology* **2020**, *11*, 2222.
- (39) Yuan, Y.; Cao, D.; Zhang, Y.; Ma, J.; Qi, J.; Wang, Q.; Lu, G.; Wu, Y.; Yan, J.; Shi, Y.; Zhang, X.; Gao, G. F. Cryo-EM structures of MERS-CoV and SARS-CoV spike

glycoproteins reveal the dynamic receptor binding domains. *Nature Communications* **2017**, *8*, 15092.

- 670 (40) Gui, M.; Song, W.; Zhou, H.; Xu, J.; Chen, S.; Xiang, Y.; Wang, X. Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding. *Cell research* **2017**, *27*, 119–129.
- (41) Kirchdoerfer, R. N.; Wang, N.; Pallesen, J.; Wrapp, D.; Turner, H. L.; Cottrell, C. A.; Corbett, K. S.; Graham, B. S.; McLellan, J. S.; Ward, A. B. Stabilized coronavirus spikes are resistant to conformational changes induced by receptor recognition or proteolysis. *Scientific Reports* **2018**, *8*, 1–11.
- 675 (42) Park, Y.-J.; Walls, A. C.; Wang, Z.; Sauer, M. M.; Li, W.; Tortorici, M. A.; Bosch, B.-J.; DiMaio, F.; Veerler, D. Structures of MERS-CoV spike glycoprotein in complex with sialoside attachment receptors. *Nature Structural & Molecular Biology* **2019**, *26*, 1151–1157.
- 680 (43) Yan, R.; Zhang, Y.; Li, Y.; Xia, L.; Guo, Y.; Zhou, Q. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* **2020**, *367*, 1444–1448.
- (44) Vilar, S.; Cozza, G.; Moro, S. Medicinal chemistry and the molecular operating environment (MOE): application of QSAR and molecular docking to drug discovery. *Current Topics in Medicinal Chemistry* **2008**, *8*, 1555–1572.
- 685 (45) Raman, S.; Vernon, R.; Thompson, J.; Tyka, M.; Sadreyev, R.; Pei, J.; Kim, D.; Kellogg, E.; DiMaio, F.; Lange, O., et al. Structure prediction for CASP8 with all-atom refinement using Rosetta. *PROTEINS: Structure, Function, and Bioinformatics* **2009**, *77*, 89–99.
- 690 (46) Song, Y.; DiMaio, F.; Wang, R. Y.-R.; Kim, D.; Miles, C.; Brunette, T.; Thompson, J.; Baker, D. High-resolution comparative modeling with RosettaCM. *Structure* **2013**, *21*, 1735–1742.

- (47) Hakansson-McReynolds, S.; Jiang, S.; Rong, L.; Caffrey, M. Solution structure of the severe acute respiratory syndrome-coronavirus heptad repeat 2 domain in the prefusion state. *Journal of Biological Chemistry* **2006**, *281*, 11965–11971.
695
- (48) Drozdetskiy, A.; Cole, C.; Procter, J.; Barton, G. J. JPred4: a protein secondary structure prediction server. *Nucleic Acids Research* **2015**, *43*, W389–W394.
- (49) Bosch, B. J.; de Haan, C. A.; Smits, S. L.; Rottier, P. J. Spike protein assembly into the coronavirus: exploring the limits of its sequence requirements. *Virology* **2005**, *334*, 306–318.
700
- (50) Dev, J.; Park, D.; Fu, Q.; Chen, J.; Ha, H. J.; Ghantous, F.; Herrmann, T.; Chang, W.; Liu, Z.; Frey, G.; Seaman, M. S.; Chen, B.; Chou, J. J. Structural basis for membrane anchoring of HIV-1 envelope spike. *Science* **2016**, *353*, 172–175.
- (51) Krog, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E. L. L. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes.
705 *Journal of Molecular Biology* **2001**, *305*, 567–580.
- (52) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: a Web-based Graphical User Interface for CHARMM. *Journal of Computational Chemistry* **2008**, *29*, 1859–1865.
- (53) Wu, E.; Cheng, X.; Jo, S.; Rui, H.; Song, K.; Dávila-Contreras, E.; Qi, Y.; Lee, J.;
710 Monje-Galvan, V.; Venable, R.; Klauda, J.; Im, W. CHARMM-GUI Membrane Builder Toward Realistic Biological Membrane Simulations. *Journal of Computational Chemistry* **2014**, *35*, 1997–2004.
- (54) Krijnse-Locker, J.; Ericsson, M.; Rottier, P.; Griffiths, G. Characterization of the budding compartment of mouse hepatitis virus: evidence that transport from the RER to the Golgi complex requires only one vesicular transport step. *Journal of Cell Biology*
715 **1994**, *124*, 55–70.

- (55) Kozakov, D.; Hall, D. R.; Xia, B.; Porter, K. A.; Padhorny, D.; Yueh, C.; Beglov, D.; Vajda, S. The ClusPro web server for protein–protein docking. *Nature Protocols* **2017**, *12*, 255–278.
- 720 (56) Fiorin, G.; Klein, M. L.; Hémin, J. Using Collective Variables to Drive Molecular Dynamics Simulations. *Molecular Physics* **2013**, *111*, 3345–3362.
- (57) Coutsiaris, E. A.; Seok, C.; Dill, K. A. Using quaternions to calculate RMSD. *Journal of Chemical Physics* **2004**, *25*, 1849–1857.
- (58) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *Journal of*
725 *Molecular Graphics* **1996**, *14*, 33–38.
- (59) Fiser, A.; Do, R. K. G.; Šali, A. Modeling of loops in protein structures. *Protein science* **2000**, *9*, 1753–1773.
- (60) Petit, C. M.; Chouljenko, V. N.; Iyer, A.; Colgrove, R.; Farzan, M.; Knipe, D. M.; Kousoulas, K. Palmitoylation of the cysteine-rich endodomain of the SARS–coronavirus
730 spike glycoprotein is important for spike-mediated cell fusion. *Virology* **2007**, *360*, 264
– 274.
- (61) Park, S.-J.; Lee, J.; Qi, Y.; Kern, N. R.; Lee, H. S.; Jo, S.; Joung, I.; Joo, K.; Lee, J.; Im, W. CHARMM-GUI Glycan Modeler for modeling and simulation of carbohydrates and glycoconjugates. *Glycobiology* **2019**, *29*, 320–331.
- 735 (62) Søndergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. Improved treatment for ligands and coupling effects in empirical calculation and rationalization of pK_a values. *Journal of Chemical Theory and Computation* **2011**, *7*, 2284–2295.
- (63) Olsson, M. H.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: consistent treatment of internal and surface residues in empirical pK_a predictions. *Journal*
740 *of Chemical Theory and Computation* **2011**, *7*, 525–537.

- (64) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **2005**, *26*, 1781–1802.
- (65) Phillips, J. et al. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *Journal of Chemical Physics* **2020**, *153*.
745
- (66) Hart, K.; Foloppe, N.; Baker, C. M.; Denning, E. J.; Nilsson, L.; Mackerell, Jr., A. D. Optimization of the CHARMM Additive Force Field for DNA: Improved Treatment of the BI/BII Conformational Equilibrium. *Journal of Chemical Theory and Computation* **2012**, *8*, 348–362.
- (67) Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell Jr., A. D.; Pastor, R. W. Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *Journal of Physical Chemistry B* **2010**, *114*, 7830–7843.
750
- (68) Jorgensen, W.; Chandrasekhar, J.; Maudura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *Journal of Chemical Physics* **1983**, *79*, 926–935.
755
- (69) Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant Pressure Molecular Dynamics Algorithms. *Journal of Chemical Physics* **1994**, *101*, 4177–4189.
- (70) Feller, S. E.; Zhang, Y.; Pastor, R. W. Constant Pressure Molecular Dynamics Simulation: the Langevin Piston Method. *Journal of Chemical Physics* **1995**, *103*, 4613–4621.
760
- (71) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System With Constraints: Molecular Dynamics of *n*-Alkanes. *Journal of Computational Physics* **1977**, *23*, 327–341.

- (72) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: an $N \cdot \log(N)$ Method for
765 Ewald Sums in Large Systems. *Journal of Chemical Physics* **1993**, *98*, 10089–10092.
- (73) Katoh, K.; Misawa, K.; Kuma, K.; Miyata, T. MAFFT: a novel method for rapid
multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*
2002, *30*, 3059–3066.
- (74) Clamp, M.; Cuff, J.; Searle, S. M.; Barton, G. J. The Jalview Java alignment editor.
770 *Bioinformatics* **2004**, *20*, 426–427.
- (75) Turoňová, B.; Sikora, M.; Schürmann, C.; Hagen, W. J.; Welsch, S.; Blanc, F. E.;
von Bülow, S.; Gecht, M.; Bagola, K.; Hörner, C., et al. In situ structural analysis
of SARS-CoV-2 spike reveals flexibility mediated by three hinges. *Science* **2020**, *370*,
203–208.
- 775 (76) Teese, M. G.; Langosch, D. Role of GxxxG Motifs in Transmembrane Domain Interac-
tions. *Biochemistry* **2015**, *54*, 5125–5135.
- (77) Yao, H.; Song, Y.; Chen, Y.; Wu, N.; Xu, J.; Sun, C.; Zhang, J.; Weng, T.; Zhang, Z.;
Wu, Z., et al. Molecular architecture of the SARS-CoV-2 virus. *Cell* **2020**, *183*, 730–
738.
- 780 (78) Choi, Y. K.; Cao, Y.; Frank, M.; Woo, H.; Park, S.-J.; Yeom, M. S.; Croll, T. I.;
Seok, C.; Im, W. Structure, Dynamics, Receptor Binding, and Antibody Binding of
the Fully Glycosylated Full-Length SARS-CoV-2 Spike Protein in a Viral Membrane.
Journal of Chemical Theory and Computation **2021**, *17*, 2479–2487.
- (79) Betts, M. J.; Russell, R. B. In *Bioinformatics for geneticists*; Barnes, M. R., Gray, I.,
785 Eds.; Wiley New York, New York, USA, 2003; Vol. 317; pp 289–316.
- (80) Ho, B. K.; Coutsias, E. A.; Seok, C.; Dill, K. A. The flexibility in the proline ring
couples to the protein backbone. *Protein Science* **2005**, *14*, 1011–1018.

- (81) te Riet, J.; Joosten, B.; Reinieren-Beeren, I.; Figdor, C. G.; Cambi, A. N-glycan mediated adhesion strengthening during pathogen-receptor binding revealed by cell-cell
790 force spectroscopy. *Scientific Reports* **2017**, *7*, 6713.
- (82) Yoon, S.-J.; Nakayama, K.-i.; Takahashi, N.; Yagi, H.; Utkina, N.; Wang, H. Y.; Kato, K.; Sadilek, M.; Hakomori, S.-i. Interaction of N-linked glycans, having multivalent GlcNAc termini, with GM3 ganglioside. *Glycoconjugate Journal* **2006**, *23*, 639–649.
- 795 (83) Yoon, S.-J.; Ikeda, S.; Sadilek, M.; Hakomori, S.-i.; Ishida, H.; Kiso, M. Self-recognition of N-linked glycans with multivalent GlcNAc, determined as ceramide mimetic conjugate. *Glycobiology* **2007**, *17*, 1007–1014.
- (84) Marquardt, D.; Heberle, F. A.; Greathouse, D. V.; Koeppe, R. E.; Standaert, R. F.; Van Oosten, B. J.; Harroun, T. A.; Kinnun, J. J.; Williams, J. A.; Wassall, S. R.; Katsaras, J. Lipid bilayer thickness determines cholesterol's location in model membranes.
800 *Soft Matter* **2016**, *12*, 9417–9428.
- (85) Dai, J.; Alwarawrah, M.; Huang, J. Instability of Cholesterol Clusters in Lipid Bilayers and The Cholesterol's Umbrella Effect. *Journal of Physical Chemistry B* **2010**, *114*, 840–848.
- 805 (86) Sakai, T.; Ohuchi, R.; Ohuchi, M. Fatty Acids on the A/USSR/77 Influenza Virus Hemagglutinin Facilitate the Transition from Hemifusion to Fusion Pore Formation. *Journal of Virology* **2002**, *76*, 4603–4611.
- (87) Varki, A.; Cummings, R. D.; Aebi, M.; Packer, N. H.; Seeberger, P. H.; Esko, J. D.; Stanley, P.; Hart, G.; Darvill, A.; Kinoshita, T., et al. Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology* **2015**, *25*, 1323–1324.
810

Supporting Information

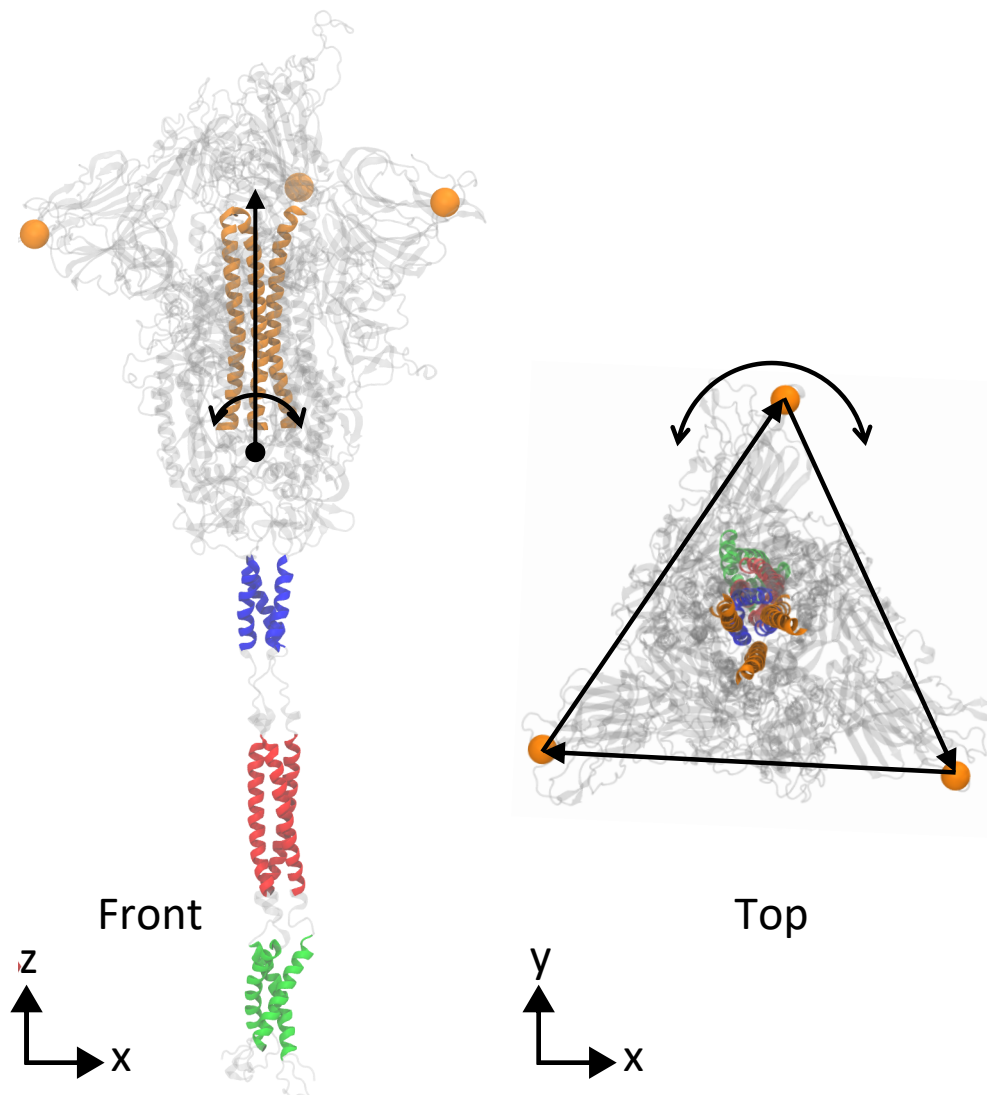


Figure S1: **Selection for global analysis.** Front and top views of describing selections used to represent the different domains in the calculation of respective vectors. TM: green, HR2: red, spike neck: blue, and spike head (represented by central helices): orange. For the calculation of the head twist, the top of the spike was approximated by a triangle formed by the C_{α} atom of residue 146 from each monomer (orange van der Waals sphere). The sides of this triangle are the vectors used to calculate the bending and twist angles with respect to the starting geometry.

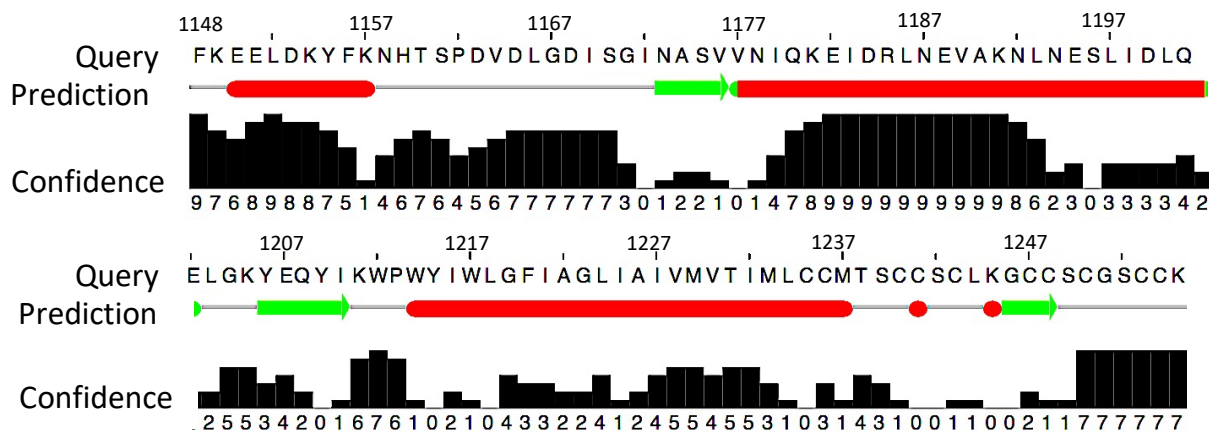


Figure S2: **Secondary structure prediction for the unresolved stalk in spike** generated using JPred4.⁴⁸ Red bars represent predicted helical structures, and green bars represent β -sheets. The confidence level for the predictions ranges between 0 (low confidence) and 9 (high confidence).

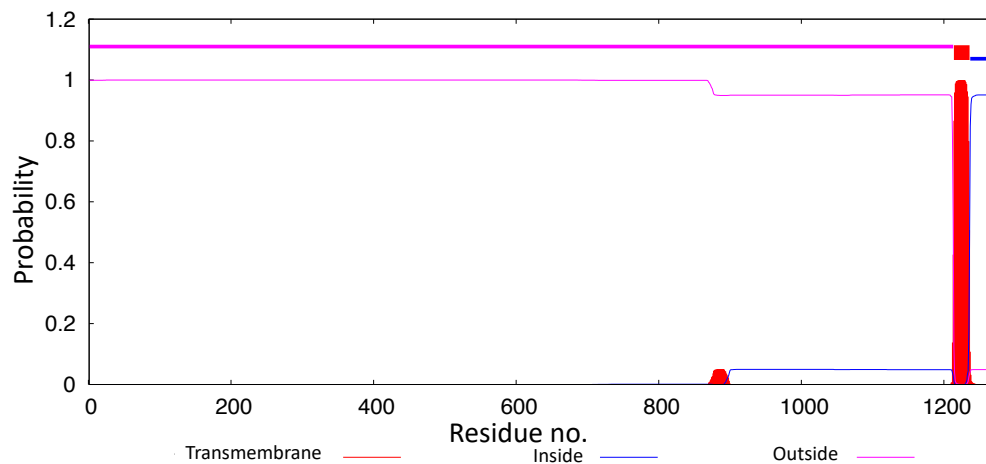


Figure S3: **Spike TM domain prediction** using TMHMM.⁵¹ The plot shows the predicted probabilities of transmembrane as well as intracellular and extracellular regions of the protein sequence. At the top of the plot (between 1.0 and 1.2) the most probable location of the protein sequence (described in the bottom legend) is shown.

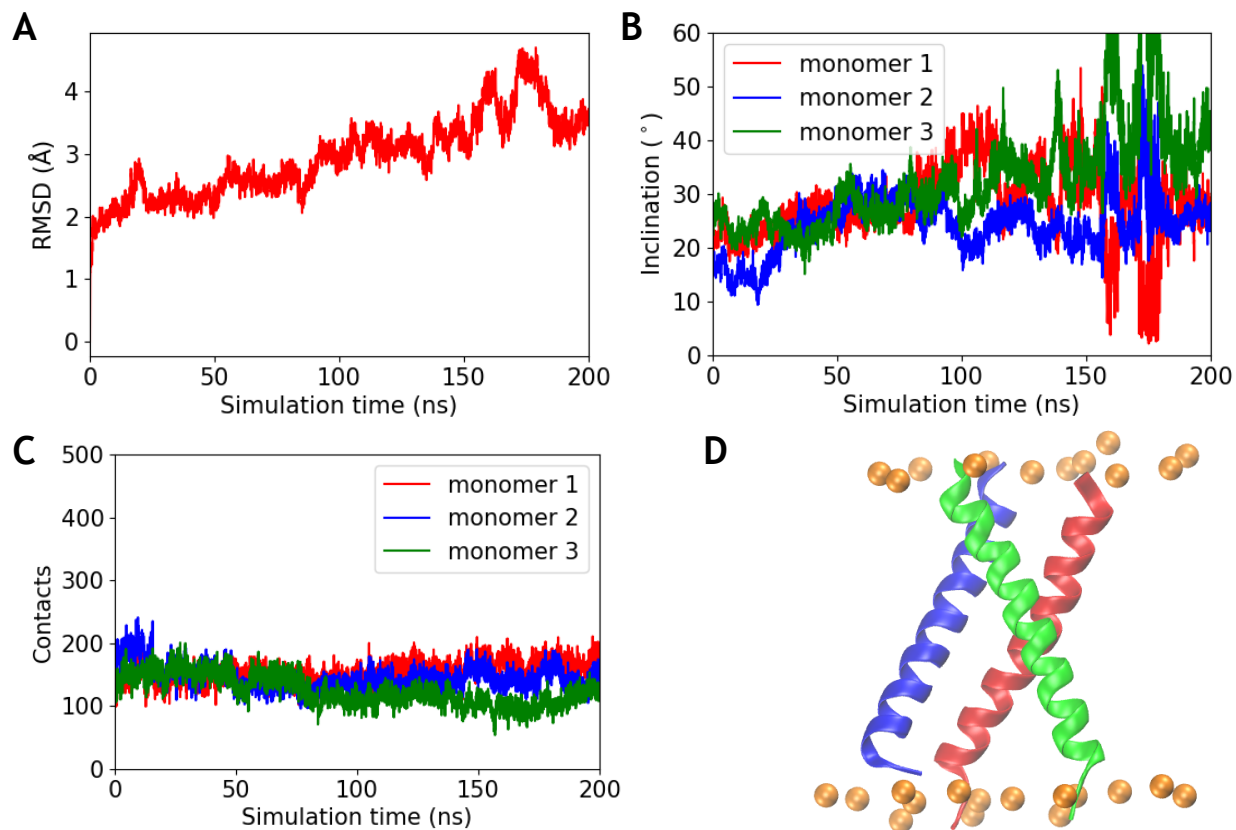


Figure S4: **Stability of predicted TM domain model 1.** (A) RMSD of the simulated TM domain model 1 (refer to Table S3), after aligning The trajectory with the starting structure using C_{α} atoms. (B) The inclination of each TM monomer calculated with respect to the whole TM domain during the simulation. (C) The number of contacts between one TM monomer and the other two monomers in the TM trimer, quantified in terms of the coordination number (described in methods). (D) Snapshot of the TM domain structure at the end of the simulation is shown. The TM monomers are shown in same colors used in (B) and (C), and the phosphorus atoms of the lipid bilayer are shown as orange spheres.

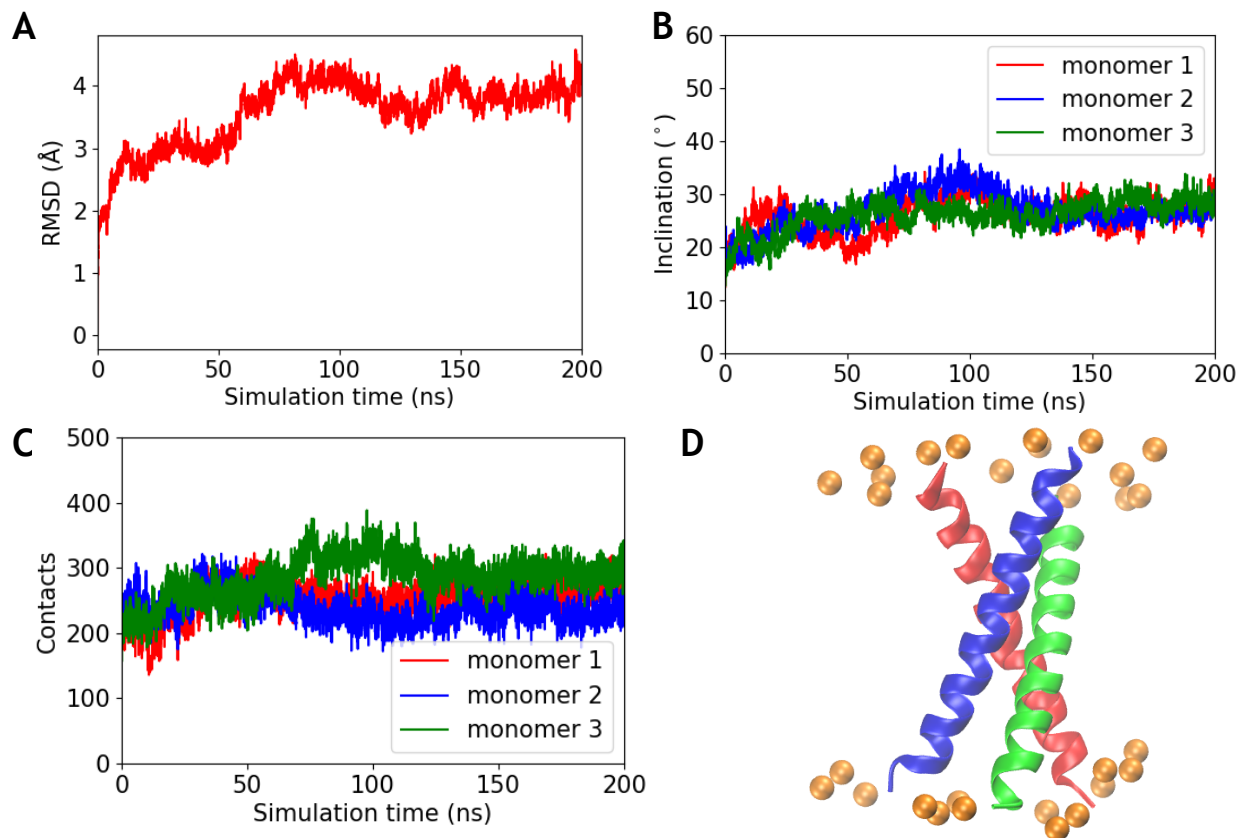


Figure S5: **Stability of predicted TM domain model 3.** (A) RMSD of the simulated TM domain model 3 (refer to Table S3), after aligning The trajectory with the starting structure using C_{α} atoms. (B) The inclination of each TM monomer calculated with respect to the whole TM domain during the simulation. (C) The number of contacts between one TM monomer and the other two monomers in the TM trimer, quantified in terms of the coordination number (described in methods). (D) Snapshot of the TM domain structure at the end of the simulation is shown. The TM monomers are shown in same colors used in (B) and (C), and the phosphorus atoms of the lipid bilayer are shown as orange spheres.

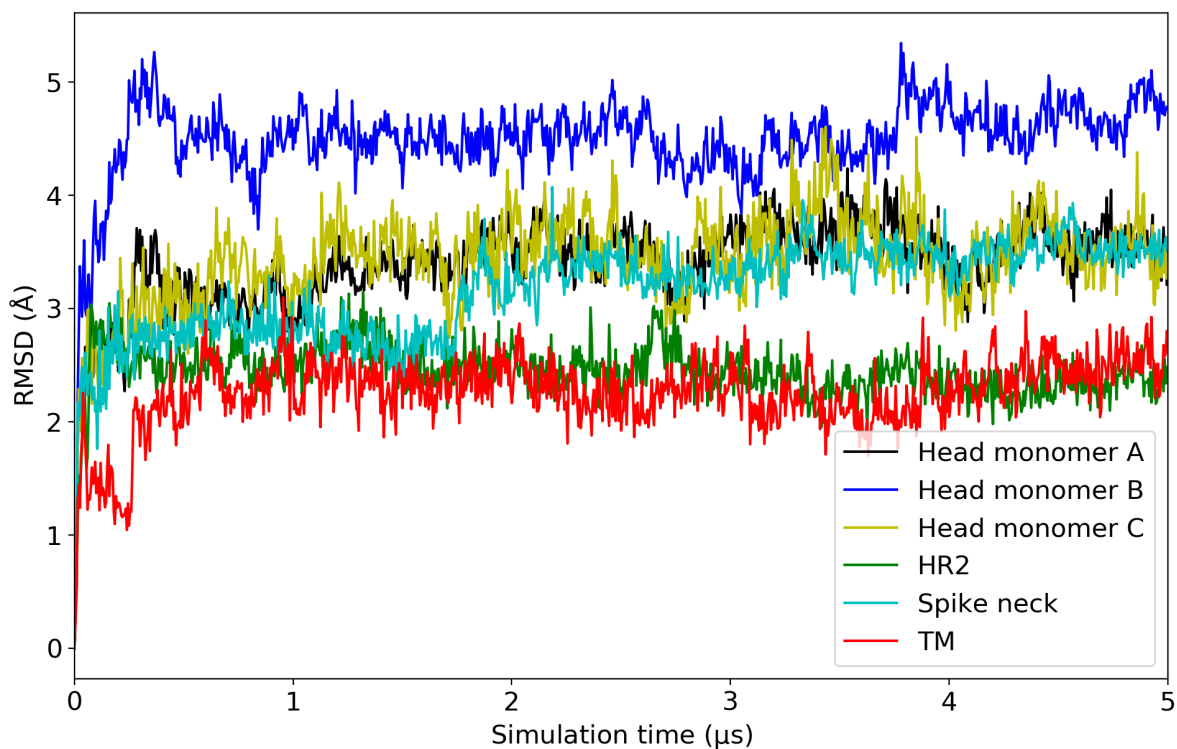


Figure S6: C_{α} RMSD of different spike domains: The RMSD is calculated individually for different stalk domains: trimeric spike neck (residues 1139-1158), HR2 (residues 1173-1200), and the TM (residues 1213-1238), as well as the individual spike head monomers (monomer A, B and C (residues 1-1134). Monomer B of the spike head contains the RBD in the "up" configuration and is relatively more mobile compared to the other monomers containing RBD in the "down" configuration.

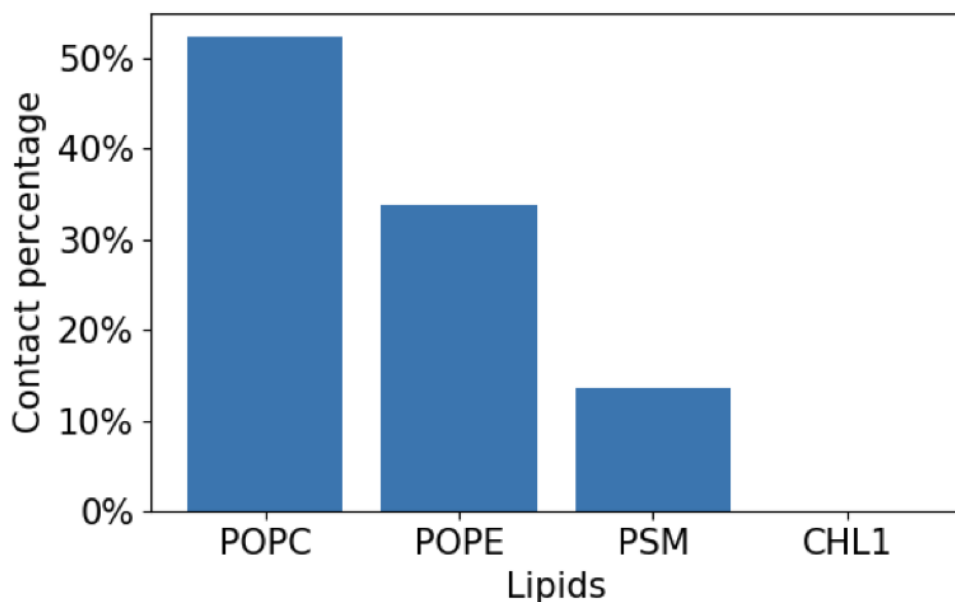


Figure S7: **Lipid-glycan contacts with different lipid types.** The contact frequency of different lipid types with the glycans at N1198 is shown. The reported percentages are normalized by the amount of POPC, POPE, PSM, and cholesterol in the leaflet proximal to the glycans.

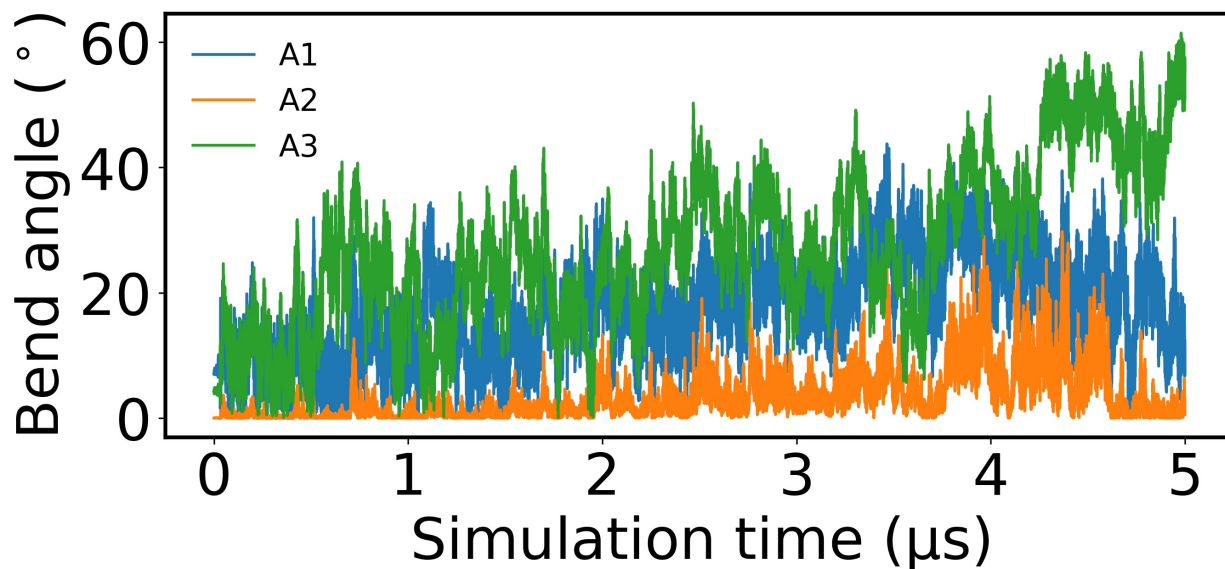


Figure S8: **Correlation of the global motions in non-glycosylated spike.** The time series of the A1, A2, and A3 angles depicted in Fig. 6A are shown for the simulation of the non-glycosylated spike.

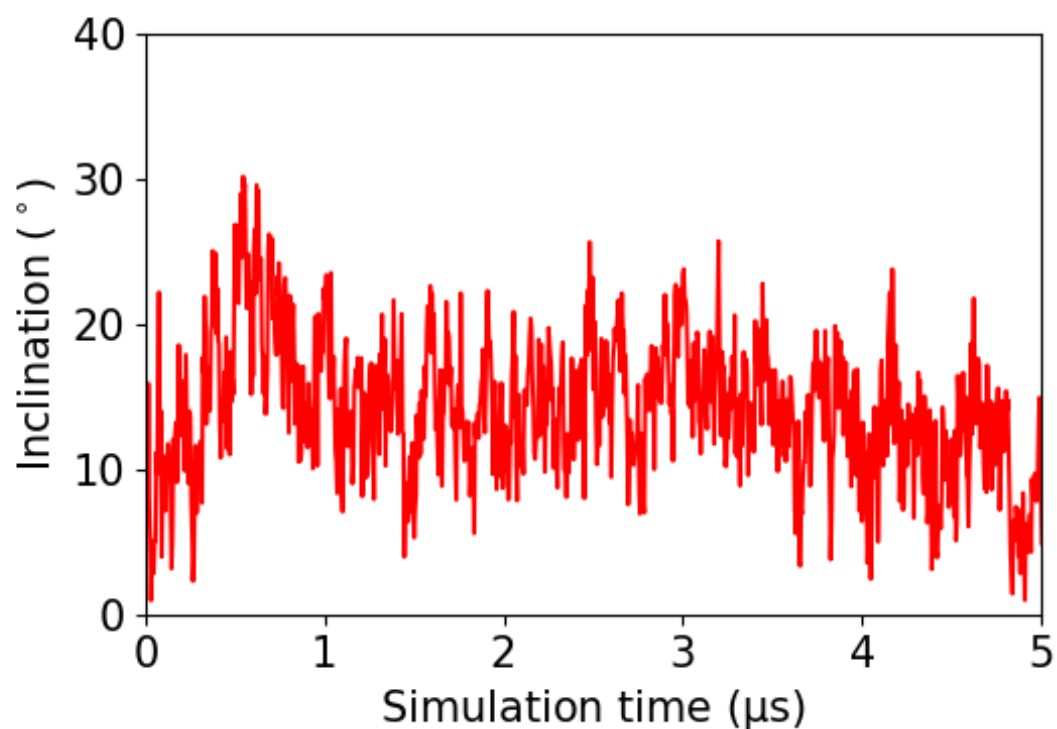


Figure S9: **Tilt angle of the TM domain in the membrane.** The tilt angle of the TM domain is calculated with respect to its initial orientation. The angle was measured between the third principal component of inertia calculated using all heavy atoms of the TM domain with respect to that for the starting structure.

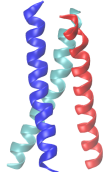
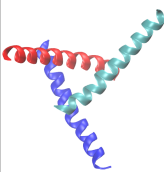
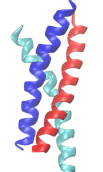
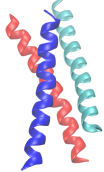
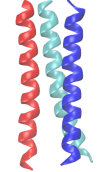
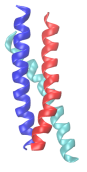
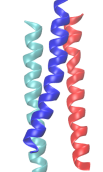
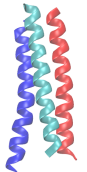
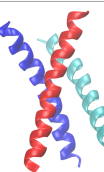
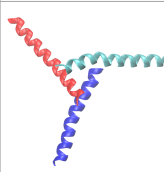
Table S1: **Lipid composition of the simulated membranes.** The lipid composition (%) of the membrane patch representing the ERGIC lipid composition.

Leflet	POPC	POPE	Cholesterol	POPS	POPI	PSM
Outer	59.0	25.5	15.0	0.0	0.0	0.5
Inner	45.0	18.5	15.0	6.0	15.0	0.5

Table S2: **Glycosylation sites and types.** The glycosylation sites in spike, as well as the respective composition of the sugar moiety. The sequence graph shown for the glycans follows the standard of Symbol Nomenclature For Glycans (SNFG).⁸⁷

Site	Composition and type	Sequence
N61	HexNAc(2)Hex(5) M5	
N122		
N603		
N709		
N717		
N801		
N1074		
N234	HexNAc(2)Hex(8) M8	
N657	HexNAc(3)Hex(6) Hybrid	
N149	HexNAc(4)Hex(3) Fuc(1) Complex	
N331		
N343		
N616		
N1134		
N1158	HexNAc(4)Hex(4) Complex	
N1098	HexNAc(4)Hex(4) Fuc(1)NeuAc(1) Complex	
N165	HexNAc(4)Hex(5) Fuc(3)NeuAc(1) Complex	
N282	HexNAc(5)Hex(3) Fuc(1) Complex	
N17	HexNAc(5)Hex(4) Fuc(1) Complex	
N1173	HexNAc(6)Hex(3) Fuc(1) Complex	
N74	HexNAc(6)Hex(4) Fuc(1)NeuAc(1) Complex	
N1194	HexNAc(6)Hex(5) Fuc(1)NeuAc(1) Complex	
T323	O-glycan	

Table S3: **TM domain trimer models predicted by Cluspro.** The predicted structures of the TM domain trimers as predicted by Cluspro⁵⁵ are shown. The 'cluster' column denotes the cluster size, which is the respective number of models present in each cluster after clustering is carried out on the selected TM domain trimer models. The 'score' shows the predicted binding energy calculated using the energy function of Cluspro.

	Model	Cluster	Score		Model	Cluster	Score
1		183	-635.1	6		60	-638.2
2		152	-611.6	7		41	-599.5
3		134	-604.7	8		37	-594.2
4		76	-642.2	9		20	-612.0
5		69	-595.1	10		19	-646.9