Dynamics and self-assembly of the SARS-CoV-2 spike transmembrane domain

Sahil Lall[1#], Padmanabhan Balaram[1,2#], Shachi Gosavi[3#] and M.K. Mathew[1#]

[1] National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bangalore - 560065, India.

[2] Molecular Biophysics Unit, Indian Institute of Science, Bangalore - 560012, India.

[3] Simons Centre for the Study of Living Machines, National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bangalore - 560065, India

[#] Corresponding authors

Abstract

The spike (S) protein is a trimeric, membrane-anchored fusion protein that enables coronaviruses, such as the SARS-CoV-2, to recognize and fuse with their hosts' cells. While the prefusion and postfusion structures of the ectomembrane domain of the spike protein are available, the corresponding organization of its transmembrane domain is obscure. Since the transmembrane and ectomembrane domains of fusion proteins are conformationally linked, an understanding of trimerization and transmembrane conformations in the viral envelope is a prerequisite to completely understand viral fusion by the spike protein. To address this, we computationally explored the self-assembly of the SARS-CoV-2 spike transmembrane domain, starting first by determining the membrane boundaries of the spike transmembrane helix. Using atomistic molecular dynamics simulations, we found the spike protein transmembrane domain to be plastic, and the transmembrane helix to be very dynamic. The observed movements of the helix changed the membrane embedded sequence, and thereby affected the conformational ensemble of the transmembrane assembly in Martini coarse grained simulations, even flipping the super-helical handedness. Analysis of the transmembrane organization of the spike transmembrane helix provided rich insights into the interfaces utilized to self-associate. Moreover, we identified two distinct cholesterol binding regions on the transmembrane helix with different affinities for the sterol. The cholesterol binding pockets overlapped with regions involved in the initiation of transmembrane protein-protein interaction. Together, the results from our multiscale simulations not only provide insight into understudied trimeric helical

interfaces in biomembranes, but also enhance our understanding of the elusive transmembrane conformational dynamics of SARS-CoV-2 spike and more generally of viral fusion proteins. These insights should enable the inclusion of the conformations of the spike protein transmembrane domain into the prevalent models of virus fusion.

Significance

Enveloped viruses rely on fusion proteins, called spike proteins in coronaviruses, to infect cells by fusing the virus envelope with the host cell membrane. The transmembrane domain (TMD) of the coronavirus spike protein is critically involved in successful viral fusion and other aspects of the virus lifecycle, but is poorly studied. Using multiscale molecular dynamics simulations of the SARS-CoV-2 spike TMD, we explore its conformational dynamics and self-assembly in different lipid environments. The results provided here improve our understanding of transmembrane stabilization of spike trimers, which are indispensable for viral infection. Exploiting this knowledge to destabilize spike trimers should facilitate design of transmembrane domain targeted viral fusion inhibitors.

Introduction

The fusion of two biological membranes is a thermodynamically favorable but kinetically limited process which is catalyzed by fusion proteins during viral fusion (1-3). Fusion proteins are divided into three classes, I to III. The spike (S) protein of coronaviruses is a class I fusion protein, a division it shares with other well established fusion proteins like hemagglutinin (HA) and envelope protein (Env/Gp160) from influenza virus and human immunodeficiency virus (HIV), respectively (4). Spike proteins are the first point of contact between coronaviruses and their host cells. Molecular recognition is followed by fusion of the viral envelope to the host plasma membrane and requires concerted action from all the domains of fusion proteins (5). The general features of viral fusion are well formulated based on decades of research and are understood to be variations on a common theme (6, 7). In coronaviruses, viral fusion is triggered upon binding of the fusion protein soluble domains to the host receptor(s) (8). Cleavage at the S1/S2 site primes the metastable, folded prefusion conformation of the spike protein for fusion (9). This new transient conformation, known as the prehairpin intermediate, exposes the fusion peptide and makes it accessible to engage with the host membrane. Irrespective of which virus they belong to, all the studied fusion proteins, including the spike protein, are believed to transition through a trimeric prehairpin structure that bridges the virus

and the host lipid bilayers (9, 10). Thereafter, the C-terminal HR2 coiled-coil domain, which is closer to the viral envelop, disassembles to refold in an antiparallel manner over the long extended trimer, forming a six-helix bundle often referred to as the fusion core (11). This refolding and reorientation pulls the two membranes together and facilitates fusion, ending with the fusion protein in a hairpin-like postfusion conformation with its fusion peptide and transmembrane domain in the same membrane (3).

A large body of evidence backs the important role of the fusion protein transmembrane domain (TMD) in successful completion of viral fusion (10). Classical mutagenesis experiments have established that a sequence-specific, proteinaceous, transmembrane anchor capable of spanning the entire envelope bilayer is a prerequisite for viral entry into cells (12). Despite its critical role in the process of viral fusion, the prefusion conformation of the TMD as well as the changes it undergoes during the process of fusion remain unknown. The dynamics of the TMD have been recalcitrant to experimental characterization partly because the fusion process takes only tens of seconds to complete (13). However, even the stable conformations of the trimeric TMD in the context of a bilayer have only recently been determined for well-studied viruses like HIV and influenza virus (14, 15). Fresh attempts towards visualizing the transmembrane domains of these viruses have made it possible to test and refine existing hypotheses of viral membrane fusion and improve vaccine candidates (11, 16, 17).

SARS-CoV-2 is a novel coronavirus which has caused the COVID-19 pandemic. The issues that remain unaddressed on established fusion proteins also apply to the SARS-CoV-2 fusion protein, better known as the spike protein, because structural efforts have focused on its soluble region (18-20). Our understanding of the transmembrane dynamics of the spike glycoprotein continues to be rudimentary because the structural information is partial towards the stable and symmetric soluble oligomeric assemblies (21, 22). No information about the transmembrane domain of any coronavirus is available either due to the design of the construct studied or because of experimental limitations (18-27). Even the juxtamembrane domains are absent or at best, poorly resolved in the characterized structures. Prior studies of the coronavirus prefusion spike have modelled a generic trimeric coiled TMD to simulate a membrane anchored spike, but overlook the transmembrane dynamics (28, 29). This disregard is unfortunate given that juxtamembrane and transmembrane mutants of SARS-CoV are unable to optimally infect host cells, emphasizing the importance of organization and dynamics of the TMD of the coronavirus spike protein in facilitating viral fusion and providing an opportunity for therapeutic intervention (30-33). Moreover, knowledge of the coordinated reorganization

of the TMD following the rearrangements of the soluble domains should enable the generation of a more comprehensive model of viral fusion.

Given that neither do we have structural information about the arrangement of SARS-CoV-2 spike TMD in the lipid bilayer, nor the effects of the surrounding lipid environment on its self-assembly is known; we examined the self-association of the spike protein TMD in a model lipid bilayer. To this end, we first located the membrane boundaries for the TMD using atomistic simulations. We found that the diffusing transmembrane domain exhibited bobbing motions in the bilayer and readily underwent bending and tilting, resulting in concomitant changes of the residues at the membrane-water interface. The most frequently occurring helical conformations were selected and coarse-grained simulations were performed without the restraints of structured, trimeric ectomembrane domains, to assess the inherent intramembrane self-association potential of the helical conformations of the TMD. The coarse-grained approach permits the characterization of ensembles of structures as homotypic interactions emerge between monomeric helical transmembrane peptides. Both symmetric and asymmetric trimers were obtained, with the ensemble being influenced by the exact membrane-embedded helical sequence. Even the extent of trimerization observed, was dependent on the transmembrane sequence. Consequently, we identified the residues that promoted self-assembly of the SARS-CoV-2 spike transmembrane helices. Furthermore, given the presence of a cholesterol binding sequence motif on the TMD, we investigated the binding of cholesterol with spike transmembrane helices and found it to interfere with the process of helical trimerization. Collectively, these results offer a step forward in understanding the spike transmembrane dynamics that accompany fusion of the viral envelope to the host plasma membrane.

Methods

Modeling the transmembrane domain

The segment corresponding to Ser1196-Ser1249 was chosen to be simulated because S1196 marks the last residue of the soluble SARS-CoV-2 spike which is included in most structures in the protein data bank (PDB). The sequence Pro1213-Cys1236 was modelled with ideal α-helical backbone torsion angles and the rest of the amino acids were designed as random coil, using PyMOL (v2.3.3; Schrodinger). The N- and C- termini of the helix were charge neutralized by amidation at the carboxyl end, while a neutral amino group was modelled at the N-terminus. This structure was embedded in a POPC (1-palmitoyl-2-oleoyl-

glycerophosphatidylcholine) bilayer using the membrane builder of CHARMM-GUI (34), solvated in TIP3P water and neutralized by increasing the ionic concentration. The four components: one copy of the 54 amino acid residue transmembrane domain, 203 lipid molecules, 0.15 mM KCl and 22,966 water molecules were assembled in a box of 8 nm x 8 nm x 14.5 nm designed to accommodate the long stretched ectomembrane residues, resulting in an initial configuration comprising 97,083 atoms (**Figure S1**).

Atomistic simulation of the single transmembrane domain

The initial configuration was sequentially equilibrated with gradual relaxation of position restraints on heavy atoms, first with 5000 steps of steepest descent and subsequently using the time resolved md integrator. The bonds involving hydrogen atoms were constrained with LINCS (35). Particle mesh Ewald (PME) summation (36) was used for treatment of long range electrostatics and the van der Waals interactions were cut off at 1.2 nm. Temperature was maintained at 30°C by coupling the system to Berendsen's thermostat (37) using a time constant (τ_t) of 1 ps. Semi-isotropic pressure on the system was controlled with the Berendsen barostat (37), set at 1 bar with a coupling constant (τ_p) of 5 ps. The integration step for this equilibration phase was kept at 1 fs. The final 0.5 ns equilibration was done using a 2 fs step free of any position restraints. The production run differed in the choice of algorithms that manage energy conservation, Nose-Hoover thermostat (38, 39) to control the temperature (30°C, τ_t = 1 ps) and Parinello-Rahman barostat (40) to maintain the pressure (semi isotropic, 1 bar; τ_p = 5 ps). Three replicates with randomized initial atomic velocities were performed for 500 ns each (**Table S1**).

Coarse-grained self-association simulations

Self-assembly simulations were performed for transmembrane segments of various lengths using the Martini force field (41, 42). We started from either an atomistic structure derived from the all atom simulations and converted it to its coarse-grained approximation using the *martinize* script or from an idealized α-helix. The DAFT protocol (43) was utilized to generate randomly rotated helices embedded in pure POPC bilayers or a POPC membrane doped with 30% cholesterol. All the helices were placed equidistant from each other with 6.5 nm between them and their periodic images. Standard Martini 2.2 parameters were used to describe all the particles. The starting conformations were energy minimized with 500 steps of steepest descent and 100 ps of NPT with a 20 fs time step to bring temperature and pressures to their set value. Temperature was maintained at 310 K by the v-rescale thermostat (44) coupled with a

temperature constant of 1 ps. Pressure was maintained semi-isotropically by the Berendsen barostat algorithm (37) set to 1 bar. Coulombic interactions were cut-off beyond 1.2 nm and the Lennard-Jones potential was decreased smoothly to zero between 0.9 and 1.2 nm. Production runs were performed for the time as mentioned in **Table S1** with the force field standard time step of 20 fs (45).

All the simulations were performed using GROMACS 4.6.7 (www.gromacs.org) on our high performance computing cluster. VMD (University of Illinois (46)) was used to visualize the trajectories and render the graphics. The analysis was performed using inbuilt Gromacs routines and PyLipID (https://github.com/wlsong/PyLipID, (47)) for cholesterol interactions. DSSP (48) was used to assign secondary structure in the simulations. Clustering of structures was done throughout using g_cluster using the Doura et al algorithm (49).

Results

Determination of the membrane-embedded helical segment

Membrane spanning helices are usually 21-25 predominantly hydrophobic residues long and are flanked by charged residues on both termini to signal the translocase-driven partitioning of these helices into biomembranes (50, 51). The spike transmembrane domain (TMD) is distinctly different in this regard, having a long stretch of 33 amino acids between two lysine residues (**Figure 1A**). This TMD has a characteristic N-terminal enrichment of aromatic residues, especially Trp (32), which we call the tryptophan-rich aromatic conserved stretch (TRACS). The TRACS is similar to an analogous juxtamembrane region observed in other enveloped viruses, referred to as the MPER (membrane proximal external region) (16). The N-terminal TRACS leads into small and hydrophobic amino acids, followed by polar residues and a lone lysine, which is surrounded by many cysteine residues of unknown function. Several algorithms that predict secondary structure, and those that predict transmembrane helices, could not unambiguously identify a C-terminus with a charged anchoring residue (**Figure 1A** & **S1**). In order to independently and confidently ascertain membrane boundaries for the transmembrane region of the spike, we performed unrestrained atomistic simulations. Starting from the last residue of the SARS-CoV-2 spike without a coordinate in the PDB - Ser1196, one copy of a 54 amino acid long segment (Ser1196-Ser1249) which included the putative 33 residue long TMD was simulated in a model POPC bilayer using classical atomistic molecular dynamics (MD) simulations (**Figure S1**).

Three replicate 500 ns simulations of membrane-embedded TMD yielded very different trajectories. The helix was highly dynamic as it readily tilted up to 55° during the course of the simulation and sometimes bent. Despite the bending, the angle of tilt of the TM helix from the membrane normal was largely anti-correlated with the distance of the N-terminal Tyr1215 from the center of the bilayer (**Figure 1B** & **1C**), further substantiated by the conformations shown (**Figure 1C**). The tilting of the helix was accompanied by bobbing motions of the helix as the α-carbon atom of Ala1226 at the helix center moved as much as one nm with respect to the bilayer midplane (**Figure S2**). Moreover, the simulated polypeptide gained and lost secondary structure at both the termini of the helix (**Figure 1D**, **S3-S5**). Using DSSP to quantify secondary structure, we found that the TRACS often lost helical secondary structure and became disordered, and seldom pushed the snorkeling Lys1211 (**Figure S6**) beyond the range of the lipid headgroups (**Figure S4B**). The non-consensus predicted helical regions around the N-terminal Glu1202 residue also gained helicity during the course of the simulation (**Figure 1D**). The C-terminus of this 54 residue peptide, predicted to be palmitoylated at one of the Cys residues (52), stayed close to the lipid headgroups despite the absence of this modification in our simulation (**Figure 1E**). The dynamics of the helical peptide created packing defects in the membrane, such that the phosphorus atoms from the phosphate groups on the lipid headgroups moved nearly 2 nm per leaflet along the axis of the membrane normal, despite the simulation box being centered to avoid translational artifacts (**Figure 1E**).

Since the simulations showed that the sequence composition and conformation of the TM helix can vary dynamically, we employed root mean square deviation (RMSD) analysis to extract some stable conformations of the TM helix and chose one dominant conformation from each of the three replicates (**Figure S7**). These three conformations had the following TM segments: Pro1213-Cys1236, Tyr1215-Ser1239 and Lys1211-Met1237. The Pro1213-Cys1236 segment remained helical across the three replicate simulations and was the most rigid of the three conformations considered (**Figure S8**). Therefore, we coarse grained this sequence first and examined its ability to associate into trimers, the arrangement of the soluble, S1 part of the prefusion spike protein of the SARS-CoV-2 (21, 22).

Transmembrane helices assemble into symmetric and asymmetric trimers

The Martini force field has been widely used to investigate oligomerization in membranes (53). Being coarse-grained, it accelerates the kinetics of the system by providing a smoothened energy landscape to facilitate studying the emergence of homotypic interactions among

monomeric helices (53). Moreover, the force field does not allow secondary structure to change over the course of the simulation (42), a feature which we used to our advantage to examine the differences in the association of TM helices formed by stretches of residues shifted in sequence. First, we modelled the 25 amino acid sequence, Trp1212-Cys1236, as an ideal α-helix, T1, and examined its potential to self-assemble into trimers (**Figure 2A**). 100 replicates of trimerizing simulations were set up and each replicate was simulated for 4 μs. The last frame of the 100 trajectories revealed that 53% ended as trimers while the remaining 47% showed a dimeric stoichiometry (**Figure 2B**). Interestingly, we continued to see dynamics and instability even upon trimerization, which can be seen from clustering the last 50 ns of all the successfully trimerized trajectories differing at 5 Å RMSD (**Figure S9**). Two of the most populated conformational clusters were dramatically different from each other, with one being a symmetric trimer and the other, an asymmetric trimer (**Figure 2C** & **2D**). The symmetric trimer employed homotypic interactions between the hydrophobic residues occupying one face of the helix and had a slight left-handed supercoil (**Figure 2C** & **S10**). In contrast, the asymmetric trimer used the GxxxG motif (54-56) to form a homodimer on which the third helix can dock, forming van der Waals interactions with the many hydrophobic residues (**Figure 2D** & **S11**). The other highly populated clusters displayed a range of conformations that were intermediate between the compact, symmetric trimer and the expanded, asymmetric trimer (**Figure S12**). Notably, the N-terminal conserved Trp residues (30, 31, 57) that comprise the TRACS did not participate in the homotypic organization as seen in the proximity maps where N-terminal interactions are conspicuously absent (**Figure S10** & **S11**). Furthermore, all clusters were observed to be more compact at the C-terminus than at the N-terminus, except the symmetric coil which was compact at both termini (**Figure 2D** & **S12**).

Three overlapping motifs facilitate the organization of transmembrane assembly

We repeated our trimerization simulations with two more constructs, T2 (Tyr1215-Cys1241) and T3 (Lys1211-Met1237), both derived from the all atom simulation results (**Figure S7**). The T2 and T3 helical transmembrane constructs were composed of 27 residues offset by 4 residues in sequence (**Figure 3A**). These simulations helped us delineate the role of the highly conserved TRACS in the trimerization of the helices as only T3 contains the TRACS. T2 simulations yielded 63 out of 100 simulations that were trimeric in the end while T3 trimerized in only 21 of the 100 simulations (**Figure 3B**). More importantly, T2 simulations displayed reduced dynamics as only 2 major structural ensembles were observed after clustering the last 50 ns of all the trimeric simulations following the same methodology as used in T1. (**Figure**

3C & **S13**). The most populated cluster of the T2 simulations was a symmetric trimer which was very similar to that seen with the T1 sequence, with a near identical interaction surface, but extended by one helical turn at the C-terminus (**Figure 3D** & **S14**). This extended interaction surface and the absence of the bulky Trp residues of the TRACS at the N-terminus of T2 reduced the left-handed supercoiling and improved the packing of the helices when compared to T1 (**Figure 3D**).

Unlike T2, which showed an increase in trimerization over T1, T3 showed a reduction. Also, trimers seen in T3 were again highly variable and dynamic as the last 50 ns from 21 simulations grouped into 8 major clusters (**Figure 3C** & **S15**). In order to understand if the sequence of T3 by itself could diminish trimerization, we systematically considered multiple alternate factors, including the length of T3, its starting conformation and the size of the simulation box. T2 and T3 are both 27 residues in length, suggesting length alone is insufficient to explain the reduction in trimerization. Analysis of the most populated trimer in T3 (**Figure S16**) indicated that the trimer had straight helices despite the bent starting conformation of T3 (**Figure S17**). The strength of the backbone dihedral restraints in the Martini force field (K_{BBBB} = 400 kJ mol^{-1}, ref (42)) is comparable to lipidic interactions and especially weaker at the ends of the helix, purely by definition (58), hence slightly bent helices can be straightened during the course of the simulation and facilitate the formation of trimers with straight helices, like in T3 (**Figure S17**). Another difference between the T1, T2 and T3 simulations was the simulation box size which differed to accommodate 6.5 nm between the helices (**Figure S18**). The simulation box was the largest in T3, and smallest in T1 (**Figure S18**), a trend that does not match the increase in trimerization seen for T2. Considering the value of the diffusion constant (**Figure S19**), it is unlikely that helical collisions were significantly reduced for the largest T3 simulation box. Moreover, even reducing the diffusion constant by half does not affect the propensity to trimerize (**Figure S24**). While we cannot explicitly rule out the unknown contribution of an increase in lipid to peptide ratio for reduced trimerization in T3, the above observations suggest that sequence is a more significant factor in the reduced trimerization of T3. Particularly, the aromatic residues of the N-terminal TRACS in T3 which preferred to interact with lipids rather than their peptide neighbors, failed to convert proximity into association beyond dimers as discussed next.

The most populated trimeric assembly observed in T3 was reminiscent of a consecutive right-handed and left-handed dimer converging on a central helix (**Figure S16**). Motivated by this and the large population of dimers in T3 (70%; **Figure 3A**), we decided to investigate the

dimerization of helices. We simulated two sequences, D1 (Trp1214-Cys1236) and D2 (Trp1212- Leu1234) offset by two residues (**Figure 4A**). These showed nearly 50% and 40% dimerization, respectively (**Figure 4B**). Strikingly, the two resulted in visually different dimers resembling the two interfaces of the T3 trimer. The two dimeric ensembles were mirrored in their super-helical handedness, which is remarkable as D1 and D2 are offset by only two amino acids. D1 displayed classical right-handed dimers stabilized by the GxxxG motif formed by Gly1219 and Gly1223 (**Figure 4C** & **S20**). The D2 sequence showed a left-handed dimer stabilized by extensive van der Waals contacts along the interaction surface utilizing the LxxVIxxT segment (**Figure 4D** & **S21**). Thus, these dimerizing simulations helped us delineate the two interaction surfaces in the spike TMD trimers. One interface was formed by the GxxxG motif, while the other was formed by the stretch of predominantly β-branched hydrophobic residues, LxxVIxxT that occupied one face of the helix (**Figure 4E**). These interfaces could combine in two ways into a symmetric or an asymmetric structure as schematically shown in **Figure 4E**. Another GASright motif (55) formed by Ala1222 and Ala1226 facilitated many intermediate conformations (e.g. **Figure S12)** and so increased the instability in the system. Reassuringly, the two interfaces identified in these dimerization simulations were also observed in the T3 simulation that ended with dimeric stoichiometry (**Figure S22**).

Cholesterol interaction discriminates between the length of CoV TM helices in the membrane and utilizes the same residues involved in helical self-assembly

Having simulated and analyzed the trimerization of the spike TMD in POPC lipid, we next addressed the role of lipid environment on these interactions. The lipids in the virus envelope are derived from the host membranes and so their lipid composition depends on the membrane that a virus buds from (59). Unlike other well-studied class I enveloped viruses (HIV, Influenza), coronaviruses such as SARS-CoV-2 do not assemble at the plasma membrane, but at ERGIC (Endoplasmic Reticulum-Golgi intermediate compartment) (60). So, its envelope is expected to be similar to the ERGIC membrane, although there has not been a thorough experimental characterization of the coronavirus lipidome. Nevertheless, apart from ~70% phosphatidylcholine (59), the coronavirus envelope should have between 8 and 10% cholesterol originating from a pre-Golgi compartment (61). On the other hand, the plasma membrane which the SARS-CoV-2 envelope fuses with upon infection, can have up to 40-50% cholesterol (61). In addition, the spike TM helix carries a cholesterol binding sequence, i.e. the inverted cholesterol recognition amino acid consensus (CRAC) motif (62), also called CARC,

at the N-terminus. Hence, we decided to repeat our trimerization simulation in the presence of an intermediate cholesterol content of 30 mol%.

The same T1, T2 and T3 were used and 50 replicates of each were simulated for 4 μs per trajectory (**Table S1**). While T1 seemed indifferent to the lipid environment as it associated equally well in the presence and absence of cholesterol, T2 and T3 both displayed a 30% reduction in trimerization (63% to 44% for T2 and 21% to 14% for T3) (**Figure 5A**). This result was surprising because only T3 carried the full CARC motif and so, had the sequence more likely to respond to cholesterol in the bilayer. But counter to our expectations, we found that T2 and T3 both interacted with cholesterol to a similar degree (**Figure 5B** & **S23**). Analyzing the residue specific contacts made by cholesterol with the monomeric helices resolved this confusing result. We found that the main site for cholesterol interaction was not the predicted CARC motif, but rather was located at the C-terminus half of the helix (**Figure 5C**), consistent with a recent experimental observation (63). Further, we did not find significant differences between the diffusion of the T1, T2 or T3 helices (**Figure S24**), the time to first interaction (**Figure S25**) or even the number and duration of interacting cholesterol molecules per helix (**Figure S26**). Overall it seems that the length of the helix (25 residue long T1; 27 residue long T2 and T3) and not the sequence distinguishes the effect of cholesterol on self-association of the spike TMD. However, more work is required to completely understand the effects of cholesterol with respect to the spike TMD.

Cholesterol makes extensive van der Waals interactions from both its α and β faces (64) to dock on to the surface of the TM helix in our simulations. Frequently observed modes of cholesterol binding at the C-terminus and N-terminus are compiled in **Figure S29** and **Figure S30**, respectively. Despite the fact that the cholesterol binding region on the helix (**Figure 5C**) overlaps with the location of self-association (**Figure 4E**), the final structures in the two lipidic milieus are comparable (**Figure S27** & **S28**). However, by occluding the C-terminal half of the helix (**Figure 5C** & **S26**), cholesterol modulates the initial events that occur upon the first encounter of TM helices by preventing the C-terminus from initiating self-association. Thus, unlike in pure POPC (**Figure 6A**), the N-terminal residues initiate self-association in the presence of cholesterol (**Figure 6B**). The C-terminal preference for cholesterol in isolated helices persists even after trimerization (**Figure S31**). This preference has implications for the instant when fusion occurs and cholesterol molecules from the host plasma membrane begin equilibrating with the viral envelop, as discussed in the next section.

Discussion

Dimerization of α-helices is frequently observed in natural membrane proteins and has been the system of choice to understand the self-association of both bitopic and multi-pass transmembrane domains (TMDs) (65). On the other hand, trimerization is primarily observed for β-barrel proteins on mitochondrial or prokaryotic membranes. Virus fusion proteins are one group where transmembrane helices naturally trimerize (10). Trimeric association of viral transmembrane helices has previously been shown to be important for stability and maturation of the fusion proteins, and mutations that interfere with TM oligomerization alter the structure of the fusion protein (66, 67). Using extensive simulations of the transmembrane domain of the SARS-CoV-2 spike, we found that it can form trimers, but these trimers exist in multiple conformations. We also discovered the underlying interactions that stabilize these trimeric conformations.

Transmembrane domains of fusion proteins, including spike protein, are known to exist in a monomer-trimer equilibrium when studied in isolation from the soluble domains (67, 68). Our atomistic simulations of the monomeric helix demonstrate it to be immensely flexible and dynamic. The TM helix was sometimes observed to lose structure at the ends and become perpendicular to the membrane while at other times, we noticed it to tilt up to 55° bury more residues in the membrane. These motions of the helix induced disturbances in the lipid bilayer, due to hydrophobic matching (69). The TRACS at the N-terminal membrane-water interface was more dynamic than the comparatively stable C-terminus and made more frequent excursions into the surrounding water. Together, the dynamics of the monomeric helix combined with the sequence at the N-terminus may be important for destabilizing the viral envelope, thereby reducing the kinetic barrier to fusion (3). This result is reminiscent of recent observations from the HIV where broadly neutralizing antibodies actually bind the MPER to prevent it from slipping into the membrane and inducing membrane fusion (16). This suggests that binders to the TRACS, which is analogous to HIV MPER, may be useful for pausing cellular entry of SARS-CoV-2. Also, given the high sequence conservation of TRACS (31, 32), such binders may work across the coronavirus family.

One of the main trimeric conformations that we encountered in our coarse-grained simulations is a symmetric trimer (**Figure 3D**). This threefold symmetric conformational ensemble, assembled from monomeric helices, likely represents the resting state of the prefusion trimer, since the membrane ectodomain of the spike protein also exhibits three-fold symmetry. This

trimer was stabilized by an extended interface along one face of the helix (**Figure S32**) that remained unchanged despite simulating different constructs which were shifted in sequence. Such robustness towards the symmetric trimer should aid in keeping the transmembrane domain of the spike protein quiescent, before fusion initiates. Thus, we postulate that oligomerization of the transmembrane domain in the resting prefusion state could work to dampen transmembrane dynamics before the process of fusion begins from the RBD at the N-terminus of the spike protein.

We could tease apart two overlapping dimerization interfaces that contribute to the trimerization of the spike transmembrane helix (**Figure 4E**). One of the two interfaces is the classical GxxxG stabilized right-handed dimer (54, 55, 65, 70). This classical motif is overrepresented in the membrane helix sequence space and is probably the most well studied dimerizing interface (56, 71, 72). It is characterized to be stabilized by a mix of hydrogen bonding and van der Waals interactions (72). The other interface that we observe is a stretch of hydrophobic residues (LxxIVxxT) that extends throughout one face of the transmembrane helix. Such long hydrophobic stretches have previously been shown to stabilize transmembrane dimers (65, 73, 74). Apart from together participating in stabilizing the symmetric trimer, these two interfaces also associate in a different manner in our simulations, by forming an asymmetric trimer which we observe as the second most prominent ensemble in our simulations (**Figure 2D**). Moreover, many of our simulations resulted in an intermediate structure with the C-terminus more compact and closer to the three-fold symmetric trimer than the open and splayed N-terminus. This intermediate state is stabilized by another GASright motif, AxxxA, which occurs in the middle of the spike transmembrane helix. Structural association through this second GxxxG-like motif becomes more prominent when the N-terminal aromatic Trp residues (in the TRACS), that favor interactions with the surrounding lipid rather than forming homotypic protein-protein interactions, are inserted in the membrane. Such an analysis contributes to our understanding of transmembrane helical interfaces in virus membranes, which can also be generally used to engineer transmembrane trimers using dimeric interfaces.

Although the symmetric TMD trimer seems naturally representative of the trimeric resting prefusion state, we posit that the asymmetric trimer that we observe is also biologically significant. Between the symmetric prefusion and postfusion conformations, fusion proteins cross an intermediate state where the symmetry is lost (6, 75). This intermediate conformation which is thought to arise from the prehairpin intermediate, displays asymmetry close to the C-

terminal coiled-coil repeat while it collapses to the postfusion state. Owing to the proximity of the C-terminal coiled coil repeat and the TMD, it is plausible that the asymmetric transmembrane trimer locally facilitates such a transition of the spike protein.

Our simulations also lead to an understanding of the influence of cholesterol on the interacting transmembrane helices. Even with the addition of 30% cholesterol which is much higher than expected in the coronavirus envelope, we do not observe a major change in the final structural ensembles when compared with those observed in the absence of the sterol, despite finding a cholesterol interaction surface on the SARS-CoV-2 spike TM helix. This is because cholesterol does not interact with the spike TM helix through the CRAC motif. However, the population of helical trimerization in T2 and T3 reduces in the presence of cholesterol. Additionally, since the cholesterol interaction surface engages the same residues that also partake in self association, the initiation of association is affected. In terms of the process of fusion, where the lipid mixing between the viral envelope and the host membrane is just beginning, cholesterol from the plasma membrane is expected to diffuse into the viral envelope. At this stage, competition with inflowing cholesterol could weaken the transmembrane oligomer and promote the overall conformational changes that occur in the soluble domains of the spike protein during fusion. Additionally, we observe that insertion of TRACS into the membrane (T3) also leads to reduced trimerization by favoring the dimeric stoichiometry. Furthermore, the dimers were either left-handed or right-handed which could also help in unwinding the transmembrane assembly.

Caveats

There is a dearth of experimental data on the composition of the coronavirus envelope. For our simulations, we used only phosphatidylcholine (POPC) lipid which has a single unsaturation in one of the lipid tails. POPC has been shown to be the most abundant lipid comprising up to 70% of phospholipids in a model coronavirus (59) and also has been used as a model lipid for long. Although common for enveloped viruses like influenza and HIV (76, 77), our POPC and cholesterol ratio is an averaged approximation for the various environments that the transmembrane domain of the SARS-CoV-2 would likely experience. Further, the curvature of the viral envelope could have a bearing on transmembrane conformations but is not studied here. Given the sequence of the TRACS, a tantalizing possibility is the Pro1213 residue sandwiched between the two Trp residues, exists in a *cis* conformation which was not

considered here. Finally, the other domains of the spike protein, not included in the simulations here may influence the equilibrium ensemble of conformations of the transmembrane domain.

Conclusions

We performed unrestrained multiscale MD simulations to study the dynamics and self-assembly of the single helix in the SARS-CoV-2 spike TMD. Starting from the amino acid sequence of a 45 residue peptide, corresponding to the TMD embedded in a model lipid membrane, we performed atomistic MD simulations in order to resolve predicted ambiguities in the boundaries of the transmembrane helix. We found that the peptide dynamically tilted with respect to the membrane normal with mainly the N-terminal residues bobbing in and out of the membrane. Three of the most frequently accessed conformations of the TM helix that were observed, differing in their length and shifted in sequence were then coarse grained using the Martini force field to investigate their potential to self-associate in membranes. We found that both the length and the sequence of the TM helix influenced the populations (degree of trimerization) and the structures of the trimeric ensembles. Specifically, the N-terminal TRACS with its aromatic Trp residues seemed to prefer lipid interactions over protein-protein interactions. Broadly, two kinds of trimers, symmetric and asymmetric, were obtained. These used the classical GxxxG motif and an extended hydrophobic stretch LxxVIxxT as the primary interaction surfaces, in addition to an AxxxA motif that served to increase the dynamics in the system. We also tested the effects of membrane cholesterol on this trimerization. Instead of binding to the inverted CRAC (CARC) motif in the TM helix, cholesterol preferred the C-terminal half of the helix, utilizing the same residues that were involved in protein-protein interaction. Together, these simulations should motivate hypotheses on how spike TMD trimerization dynamics enable viral fusion. They should also facilitate the design of fusion inhibitors targeted to bind the TMD and subdue its dynamics critical to viral fusion.

Author Contributions

SL designed the research, performed the research, analyzed the data and wrote the initial draft. PB, SG, MKM supervised the work and revised the article.

Author Information

Correspondence may be addressed to

SL: sahil.lall@gmail.com

PB: pb@iisc.ac.in

SG: shachi@ncbs.res.in

MKM: mathew@ncbs.res.in

Acknowledgements

This work was funded by the Department of Atomic Energy, Government of India through the Tata Institute of Fundamental Research from Project Identification No. RTI 4006. SG acknowledges the support of the Simons Foundation (Grant No. 287975).

References

1. Chernomordik, L. V., and M. M. Kozlov. 2003. Protein-lipid interplay in fusion and fission of biological membranes. Annual review of biochemistry 72:175-207.
2. Blumenthal, R., M. J. Clague, S. R. Durell, and R. M. Epand. 2003. Membrane fusion. Chemical reviews 103:53-69.
3. Harrison, S. C. 2008. Viral membrane fusion. Nature structural & molecular biology 15:690-698.
4. Belouzard, S., J. K. Millet, B. N. Licitra, and G. R. Whittaker. 2012. Mechanisms of coronavirus cell entry mediated by the viral spike protein. Viruses 4:1011-1033.
5. Peisajovich, S. G., and Y. Shai. 2003. Viral fusion proteins: multiple regions contribute to membrane fusion. Biochimica et Biophysica Acta (BBA) - Biomembranes 1614:122-129.
6. Harrison, S. C. 2015. Viral membrane fusion. Virology 479-480:498-507.
7. Kielian, M. 2014. Mechanisms of Virus Membrane Fusion Proteins. Annu Rev Virol 1:171-189.
8. Tang, T., M. Bidon, J. A. Jaimes, G. R. Whittaker, and S. Daniel. 2020. Coronavirus membrane fusion mechanism offers a potential target for antiviral development. Antiviral Res 178:104792.
9. White, J. M., S. E. Delos, M. Brecher, and K. Schornberg. 2008. Structures and mechanisms of viral membrane fusion proteins: multiple variations on a common theme. Crit Rev Biochem Mol Biol 43:189-219.
10. Barrett, C. T., and R. E. Dutch. 2020. Viral Membrane Fusion and the Transmembrane Domain. Viruses 12.
11. Du, L., Y. He, Y. Zhou, S. Liu, B. J. Zheng, and S. Jiang. 2009. The spike protein of SARS-CoV--a target for vaccine and therapeutic development. Nature reviews. Microbiology 7:226-236.
12. Langosch, D., M. Hofmann, and C. Ungermann. 2007. The role of transmembrane domains in membrane fusion. Cellular and molecular life sciences : CMLS 64:850-864.
13. Floyd, D. L., J. R. Ragains, J. J. Skehel, S. C. Harrison, and A. M. van Oijen. 2008. Single-particle kinetics of influenza virus membrane fusion. Proceedings of the National Academy of Sciences of the United States of America 105:15382-15387.
14. Benton, D. J., A. Nans, L. J. Calder, J. Turner, U. Neu, Y. P. Lin, E. Ketelaars, N. L. Kallewaard, D. Corti, A. Lanzavecchia, S. J. Gamblin, P. B. Rosenthal, and J. J. Skehel. 2018. Influenza hemagglutinin membrane anchor. Proceedings of the National Academy of Sciences of the United States of America 115:10112-10117.

15. Dev, J., D. Park, Q. Fu, J. Chen, H. J. Ha, F. Ghantous, T. Herrmann, W. Chang, Z. Liu, G. Frey, M. S. Seaman, B. Chen, and J. J. Chou. 2016. Structural basis for membrane anchoring of HIV-1 envelope spike. Science 353:172-175.
16. Wang, Y., P. Kaur, Z. J. Sun, M. A. Elbahnasawy, Z. Hayati, Z. S. Qiao, N. N. Bui, C. Chile, M. L. Nasr, G. Wagner, J. H. Wang, L. Song, E. L. Reinherz, and M. Kim. 2019. Topological analysis of the gp41 MPER on lipid bilayers relevant to the metastable HIV-1 envelope prefusion state. Proceedings of the National Academy of Sciences of the United States of America 116:22556-22566.
17. Fu, Q., M. M. Shaik, Y. Cai, F. Ghantous, A. Piai, H. Peng, S. Rits-Volloch, Z. Liu, S. C. Harrison, M. S. Seaman, B. Chen, and J. J. Chou. 2018. Structure of the membrane proximal external region of HIV-1 envelope glycoprotein. Proceedings of the National Academy of Sciences of the United States of America 115:E8892-E8899.
18. Wrapp, D., N. Wang, K. S. Corbett, J. A. Goldsmith, C. L. Hsieh, O. Abiona, B. S. Graham, and J. S. McLellan. 2020. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science 367:1260-1263.
19. Walls, A. C., Y. J. Park, M. A. Tortorici, A. Wall, A. T. McGuire, and D. Veesler. 2020. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell 181:281-292 e286.
20. Cai, Y., J. Zhang, T. Xiao, H. Peng, S. M. Sterling, R. M. Walsh, Jr., S. Rawson, S. Rits-Volloch, and B. Chen. 2020. Distinct conformational states of SARS-CoV-2 spike protein. Science 369:1586-1592.
21. Yao, H., Y. Song, Y. Chen, N. Wu, J. Xu, C. Sun, J. Zhang, T. Weng, Z. Zhang, Z. Wu, L. Cheng, D. Shi, X. Lu, J. Lei, M. Crispin, Y. Shi, L. Li, and S. Li. 2020. Molecular Architecture of the SARS-CoV-2 Virus. Cell 183:730-738 e713.
22. Ke, Z., J. Oton, K. Qu, M. Cortese, V. Zila, L. McKeane, T. Nakane, J. Zivanov, C. J. Neufeldt, B. Cerikan, J. M. Lu, J. Peukes, X. Xiong, H. G. Krausslich, S. H. W. Scheres, R. Bartenschlager, and J. A. G. Briggs. 2020. Structures and distributions of SARS-CoV-2 spike proteins on intact virions. Nature 588:498-502.
23. Xiong, X., K. Qu, K. A. Ciazynska, M. Hosmillo, A. P. Carter, S. Ebrahimi, Z. Ke, S. H. W. Scheres, L. Bergamaschi, G. L. Grice, Y. Zhang, C.-N. C.-B. Collaboration, J. A. Nathan, S. Baker, L. C. James, H. E. Baxendale, I. Goodfellow, R. Doffinger, and J. A. G. Briggs. 2020. A thermostable, closed SARS-CoV-2 spike protein trimer. Nature structural & molecular biology 27:934-941.
24. Fan, X., D. Cao, L. Kong, and X. Zhang. 2020. Cryo-EM analysis of the post-fusion structure of the SARS-CoV spike glycoprotein. Nature communications 11:3618.
25. Walls, A. C., M. A. Tortorici, J. Snijder, X. Xiong, B. J. Bosch, F. A. Rey, and D. Veesler. 2017. Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion. Proceedings of the National Academy of Sciences of the United States of America 114:11157-11162.
26. Walls, A. C., M. A. Tortorici, B. J. Bosch, B. Frenz, P. J. M. Rottier, F. DiMaio, F. A. Rey, and D. Veesler. 2016. Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer. Nature 531:114-117.
27. Kirchdoerfer, R. N., C. A. Cottrell, N. Wang, J. Pallesen, H. M. Yassine, H. L. Turner, K. S. Corbett, B. S. Graham, J. S. McLellan, and A. B. Ward. 2016. Pre-fusion structure of a human coronavirus spike protein. Nature 531:118-121.
28. Yu, A., A. J. Pak, P. He, V. Monje-Galvan, L. Casalino, Z. Gaieb, A. C. Dommer, R. E. Amaro, and G. A. Voth. 2020. A Multiscale Coarse-grained Model of the SARS-CoV-2 Virion. Biophysical journal.
29. Woo, H., S. J. Park, Y. K. Choi, T. Park, M. Tanveer, Y. Cao, N. R. Kern, J. Lee, M. S. Yeom, T. I. Croll, C. Seok, and W. Im. 2020. Developing a Fully Glycosylated Full-Length SARS-CoV-2 Spike Protein Model in a Viral Membrane. J Phys Chem B 124:7128-7137.

30. Corver, J., R. Broer, P. van Kasteren, and W. Spaan. 2009. Mutagenesis of the transmembrane domain of the SARS coronavirus spike glycoprotein: refinement of the requirements for SARS coronavirus cell entry. Virol J 6:230.
31. Lu, Y., T. L. Neo, D. X. Liu, and J. P. Tam. 2008. Importance of SARS-CoV spike protein Trp-rich region in viral infectivity. Biochem Biophys Res Commun 371:356-360.
32. Howard, M. W., E. A. Travanty, S. A. Jeffers, M. K. Smith, S. T. Wennier, L. B. Thackray, and K. V. Holmes. 2008. Aromatic amino acids in the juxtamembrane domain of severe acute respiratory syndrome coronavirus spike glycoprotein are important for receptor-dependent virus entry and cell-cell fusion. J Virol 82:2883-2894.
33. Broer, R., B. Boson, W. Spaan, F. L. Cosset, and J. Corver. 2006. Important role for the transmembrane domain of severe acute respiratory syndrome coronavirus spike protein during entry. J Virol 80:1302-1310.
34. Lee, J., X. Cheng, J. M. Swails, M. S. Yeom, P. K. Eastman, J. A. Lemkul, S. Wei, J. Buckner, J. C. Jeong, Y. Qi, S. Jo, V. S. Pande, D. A. Case, C. L. Brooks, 3rd, A. D. MacKerell, Jr., J. B. Klauda, and W. Im. 2016. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. J Chem Theory Comput 12:405-413.
35. Hess, B., H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. 1997. LINCS: A linear constraint solver for molecular simulations. Journal of Computational Chemistry 18:1463-1472.
36. Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald: AnN·log(N) method for Ewald sums in large systems. The Journal of chemical physics 98:10089-10092.
37. Berendsen, H. J. C., J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. 1984. Molecular dynamics with coupling to an external bath. The Journal of chemical physics 81:3684-3690.
38. Hoover, W. G. 1985. Canonical dynamics: Equilibrium phase-space distributions. Phys Rev A Gen Phys 31:1695-1697.
39. Nosé, S. 1984. A molecular dynamics method for simulations in the canonical ensemble. Molecular Physics 52:255-268.
40. Parrinello, M., and A. Rahman. 1981. Polymorphic transitions in single crystals: A new molecular dynamics method. Journal of Applied Physics 52:7182-7190.
41. de Jong, D. H., G. Singh, W. F. Bennett, C. Arnarez, T. A. Wassenaar, L. V. Schafer, X. Periole, D. P. Tieleman, and S. J. Marrink. 2013. Improved Parameters for the Martini Coarse-Grained Protein Force Field. J Chem Theory Comput 9:687-697.
42. Monticelli, L., S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman, and S. J. Marrink. 2008. The MARTINI Coarse-Grained Force Field: Extension to Proteins. J Chem Theory Comput 4:819-834.
43. Wassenaar, T. A., K. Pluhackova, A. Moussatova, D. Sengupta, S. J. Marrink, D. P. Tieleman, and R. A. Bockmann. 2015. High-Throughput Simulations of Dimer and Trimer Assembly of Membrane Proteins. The DAFT Approach. J Chem Theory Comput 11:2278-2291.
44. Bussi, G., D. Donadio, and M. Parrinello. 2007. Canonical sampling through velocity rescaling. The Journal of chemical physics 126:014101.
45. Marrink, S. J., H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries. 2007. The MARTINI force field: Coarse grained model for biomolecular simulations. Journal of Physical Chemistry B 111:7812-7824.
46. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. J Mol Graph 14:33-38, 27-38.
47. Ansell, T. B., W. Song, and M. S. P. Sansom. 2020. The Glycosphingolipid GM3 Modulates Conformational Dynamics of the Glucagon Receptor. Biophysical journal 119:300-313.
48. Kabsch, W., and C. Sander. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22:2577-2637.

49. Daura, X., K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren, and A. E. Mark. 1999. Peptide folding: When simulation meets experiment. Angew Chem Int Edit 38:236-240.
50. Hessa, T., H. Kim, K. Bihlmaier, C. Lundin, J. Boekel, H. Andersson, I. Nilsson, S. H. White, and G. von Heijne. 2005. Recognition of transmembrane helices by the endoplasmic reticulum translocon. Nature 433:377-381.
51. von Heijne, G. 1986. The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology. The EMBO journal 5:3021-3027.
52. van Berlo, M. F., W. J. van den Brink, M. C. Horzinek, and B. A. van der Zeijst. 1987. Fatty acid acylation of viral proteins in murine hepatitis virus-infected cells. Brief report. Arch Virol 95:123-128.
53. Marrink, S. J., and D. P. Tieleman. 2013. Perspective on the Martini model. Chem Soc Rev 42:6801-6822.
54. Teese, M. G., and D. Langosch. 2015. Role of GxxxG Motifs in Transmembrane Domain Interactions. Biochemistry 54:5125-5135.
55. Walters, R. F., and W. F. DeGrado. 2006. Helix-packing motifs in membrane proteins. Proceedings of the National Academy of Sciences of the United States of America 103:13658-13663.
56. MacKenzie, K. R., J. H. Prestegard, and D. M. Engelman. 1997. A transmembrane helix dimer: structure and implications. Science 276:131-133.
57. Liao, Y., S. M. Zhang, T. L. Neo, and J. P. Tam. 2015. Tryptophan-dependent membrane interaction and heteromerization with the internal fusion peptide by the membrane proximal external region of SARS-CoV spike protein. Biochemistry 54:1819-1830.
58. Li, S., B. Wu, and W. Han. 2019. Parametrization of MARTINI for Modeling Hinging Motions in Membrane Proteins. J Phys Chem B 123:2254-2269.
59. van Genderen, I. L., G. J. Godeke, P. J. Rottier, and G. van Meer. 1995. The phospholipid composition of enveloped viruses depends on the intracellular membrane through which they bud. Biochem Soc Trans 23:523-526.
60. Griffiths, G., and P. Rottier. 1992. Cell biology of viruses that assemble along the biosynthetic pathway. Seminars in Cell Biology 3:367-381.
61. van Meer, G., D. R. Voelker, and G. W. Feigenson. 2008. Membrane lipids: where they are and how they behave. Nature reviews. Molecular cell biology 9:112-124.
62. Epand, R. M. 2006. Cholesterol and the interaction of proteins with membrane domains. Prog Lipid Res 45:279-294.
63. Wei, C., L. Wan, Q. Yan, X. Wang, J. Zhang, X. Yang, Y. Zhang, C. Fan, D. Li, Y. Deng, J. Sun, J. Gong, X. Yang, Y. Wang, X. Wang, J. Li, H. Yang, H. Li, Z. Zhang, R. Wang, P. Du, Y. Zong, F. Yin, W. Zhang, N. Wang, Y. Peng, H. Lin, J. Feng, C. Qin, W. Chen, Q. Gao, R. Zhang, Y. Cao, and H. Zhong. 2020. HDL-scavenger receptor B type 1 facilitates SARS-CoV-2 entry. Nat Metab.
64. Fantini, J., and F. J. Barrantes. 2013. How cholesterol interacts with membrane proteins: an exploration of cholesterol-binding sites including CRAC, CARC, and tilted domains. Front Physiol 4:31.
65. Lall, S., and M. K. Mathew. 2017. Dynamics of Membrane Proteins. In Membrane Organization and Dynamics. A. Chattopadhyay, editor. Springer International Publishing, Cham. 219-241.
66. Shang, L., and E. Hunter. 2010. Residues in the membrane-spanning domain core modulate conformation and fusogenicity of the HIV-1 envelope glycoprotein. Virology 404:158-167.
67. Smith, E. C., S. E. Smith, J. R. Carter, S. R. Webb, K. M. Gibson, L. M. Hellman, M. G. Fried, and R. E. Dutch. 2013. Trimeric transmembrane domain interactions in paramyxovirus fusion proteins: roles in protein folding, stability, and function. The Journal of biological chemistry 288:35726-35735.
68. Webb, S. R., S. E. Smith, M. G. Fried, and R. E. Dutch. 2018. Transmembrane Domains of Highly Pathogenic Viral Fusion Proteins Exhibit Trimeric Association In Vitro. mSphere 3.

69. Killian, J. A. 1998. Hydrophobic mismatch between proteins and lipids in membranes. Bba-Rev Biomembranes 1376:401-416.
70. Lemmon, M. A., H. R. Treutlein, P. D. Adams, A. T. Brunger, and D. M. Engelman. 1994. A dimerization motif for transmembrane alpha-helices. Nat Struct Biol 1:157-163.
71. Senes, A., M. Gerstein, and D. M. Engelman. 2000. Statistical analysis of amino acid patterns in transmembrane helices: the GxxxG motif occurs frequently and in association with beta-branched residues at neighboring positions. Journal of molecular biology 296:921-936.
72. Anderson, S. M., B. K. Mueller, E. J. Lange, and A. Senes. 2017. Combination of Calpha-H Hydrogen Bonds and van der Waals Packing Modulates the Stability of GxxxG-Mediated Dimers in Membranes. J Am Chem Soc 139:15774-15783.
73. Bragin, P. E., K. S. Mineev, O. V. Bocharova, P. E. Volynsky, E. V. Bocharov, and A. S. Arseniev. 2016. HER2 Transmembrane Domain Dimerization Coupled with Self-Association of Membrane-Embedded Cytoplasmic Juxtamembrane Regions. Journal of molecular biology 428:52-61.
74. Li, E., W. C. Wimley, and K. Hristova. 2012. Transmembrane helix dimerization: beyond the search for sequence motifs. Biochimica et biophysica acta 1818:183-193.
75. Kawase, M., M. Kataoka, K. Shirato, and S. Matsuyama. 2019. Biochemical Analysis of Coronavirus Spike Glycoprotein Conformational Intermediates during Membrane Fusion. J Virol 93.
76. Gerl, M. J., J. L. Sampaio, S. Urban, L. Kalvodova, J. M. Verbavatz, B. Binnington, D. Lindemann, C. A. Lingwood, A. Shevchenko, C. Schroeder, and K. Simons. 2012. Quantitative analysis of the lipidomes of the influenza virus envelope and MDCK cell apical membrane. The Journal of cell biology 196:213-221.
77. Brugger, B., B. Glass, P. Haberkant, I. Leibrecht, F. T. Wieland, and H. G. Krausslich. 2006. The HIV lipidome: a raft with an unusual composition. Proceedings of the National Academy of Sciences of the United States of America 103:2641-2646.

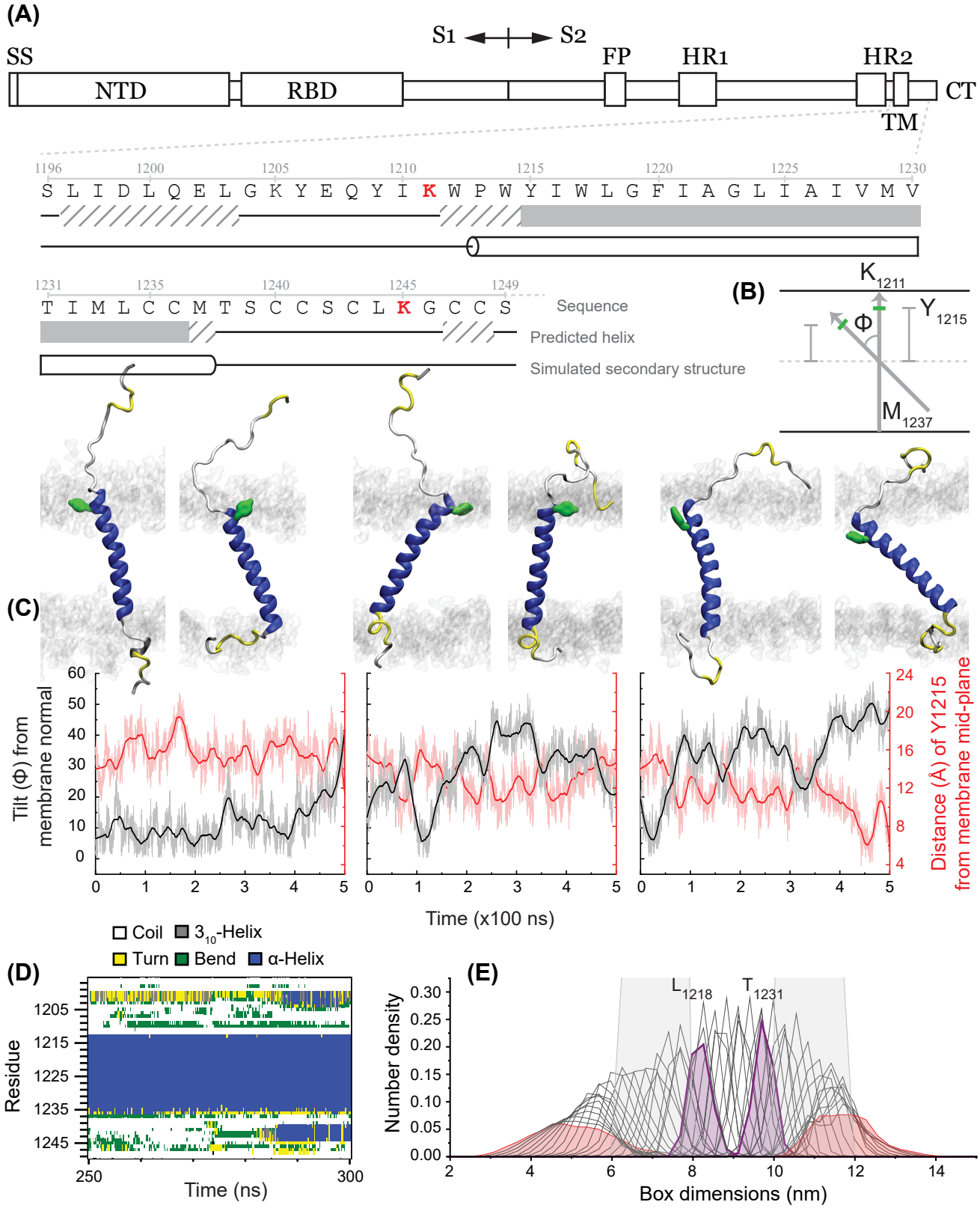

Figure 1: **Atomistic simulations of the spike transmembrane domain (TMD)**. The domain organization of the SARS-CoV-2 spike protein (**A**). S1 part of the spike, comprising of SS (signal sequence), NTD (N-terminal domain) and RBD (receptor binding domain), is soluble and binds to the cellular receptor - ACE2. The TM (Transmembrane) domain anchors the fusion machinery of the virus, encompassing the FP (fusion platform) and the HRs (heptad repeats), to the viral envelop. The sequence shown underneath was simulated atomistically to determine the membrane resident portion of the spike protein TMD. A schematic guide (**B**) shows how the helical tilt is anti-correlated with the distance of Tyr1215 from the membrane midplane across the replicate simulations (**C**). Two snapshots (300 ns & 500 ns) each from the three replicates are shown above (DSSP color scheme; Tyr1215, green) to substantiate the same. Representative 50 ns DSSP analysis is provided (**D**) to highlight the plasticity of secondary structure observed across the simulated sequence. The dynamic TMD disturbs membrane packing such that the phosphate group (gray) ranges over 2 nm on both leaflets, sandwiching the core hydrophobic segment Leu1218 to Thr1231 (violet) (**E**). Also, the C-terminal residues tend to remain closer to the membrane (terminal residues in red). (Color figure available online)

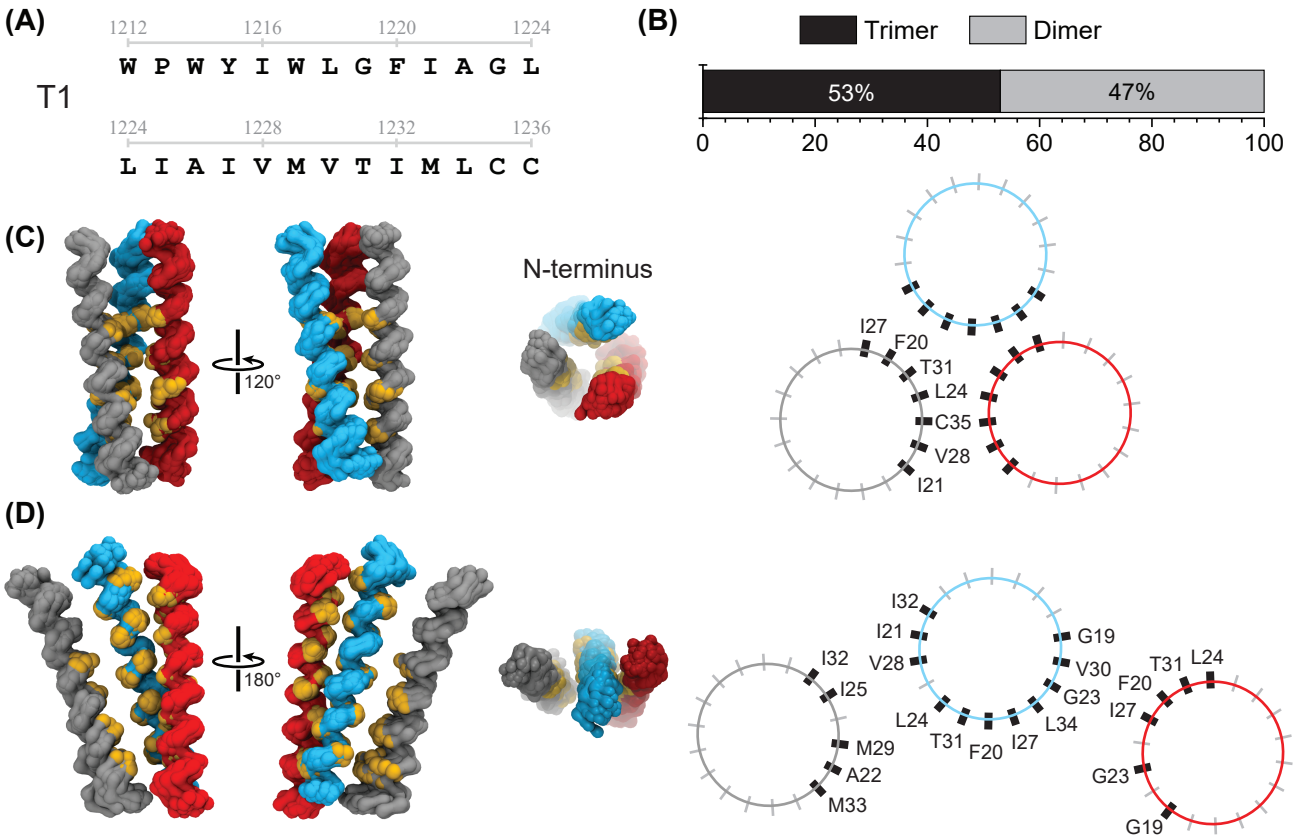

Figure 2: **Dynamics continue despite trimerization**. Coarse-grained trimerizing simulations of the spike transmembrane domain were performed with 100 replicates for 25 residue T1 (**A**). Nearly 50% of the simulations ended in trimers (**B**). The last 50 ns of all the successful trimerizing simulations could be clustered at 5 Å RMSD into either a symmetric (**C**) or asymmetric (**D**) trimer. The interaction surface for both is shown in yellow. (Color figures available online) We subtracted 1200 from the residue numbers in the helical wheel for brevity.

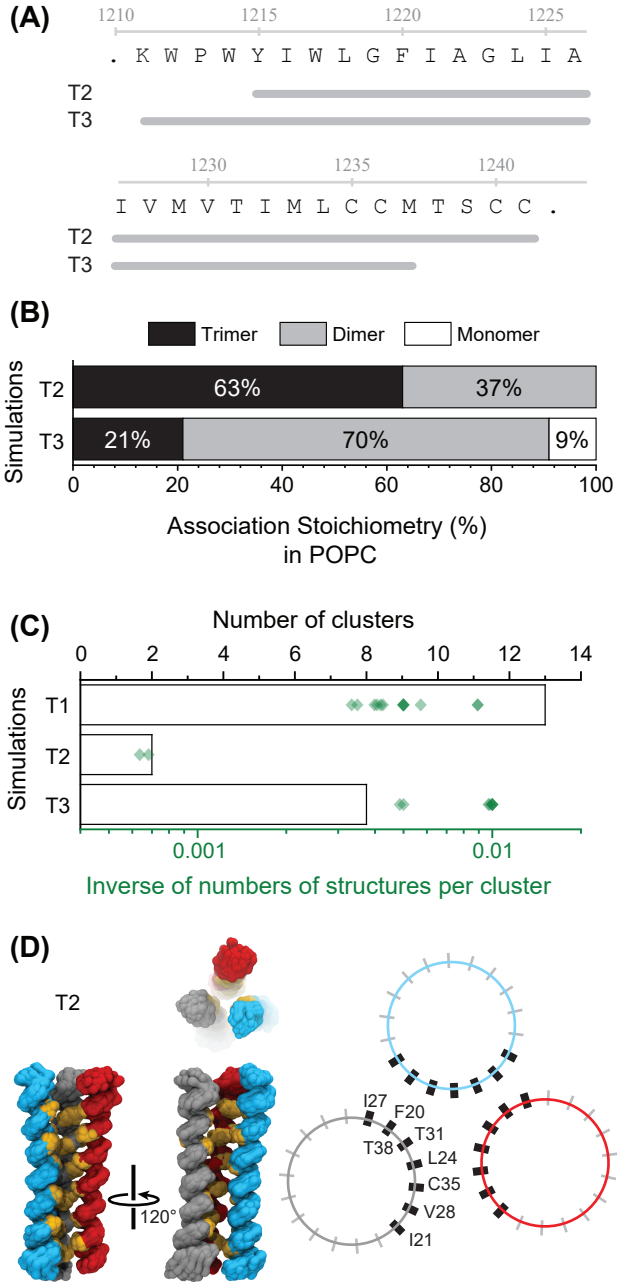

Figure 3: **Sequence displaced TM helices also utilize the same residues to interact**. Sequences of T2 and T3, both 27 residue helices (**A**) were used to ascertain the role of the conserved N-terminal aromatic residues. T2 trimerized more than T1, while T3 showed a reduction (**B**) to the extent that 9% of the simulations ended with monomeric helices. The TM helices T1, T2 and T3 differ in not just their ability to trimerize but also how dynamic the trimers are, with T1 and T3 showing greater diversity of trimers than T2 (**C**). The cluster with the largest number of structures in T2 showed a symmetric trimer with an interaction surface similar to that of the T1 symmetric trimer, but extended by one helical turn at the C-terminus. (Coloring scheme as before, color figures available online)

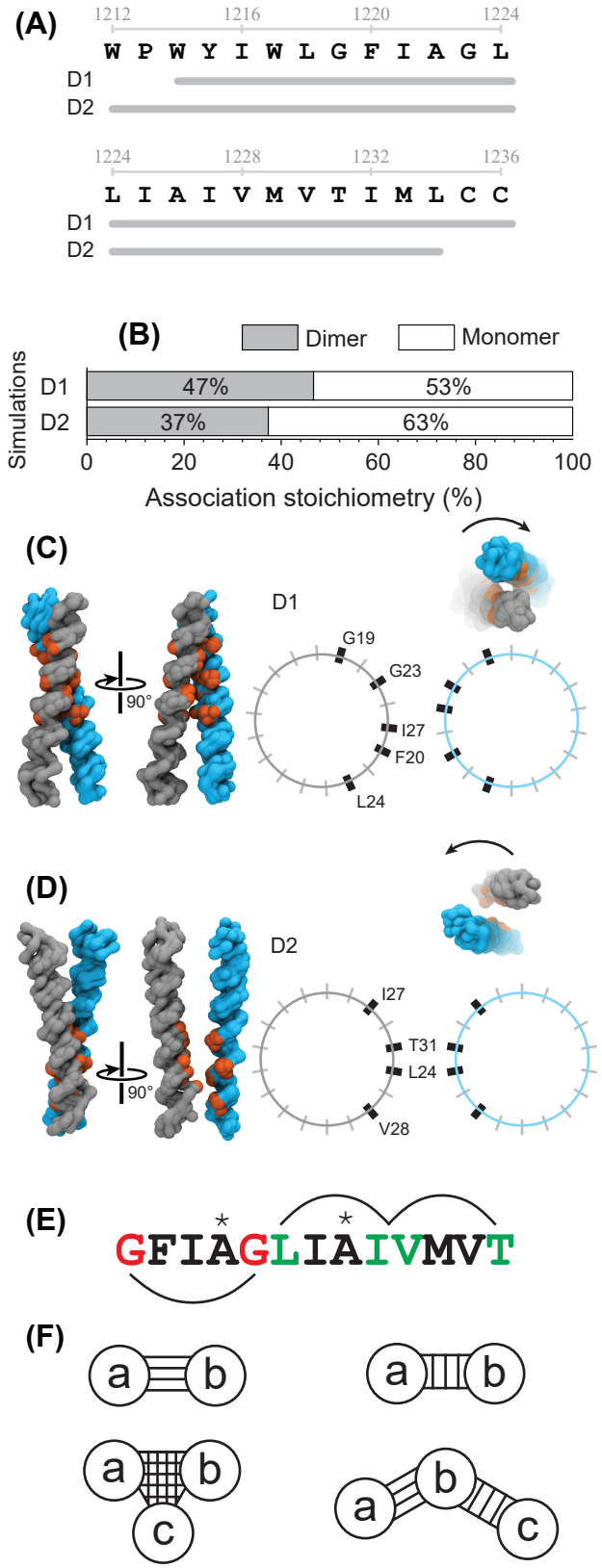

Figure 4: **Dimerization simulations reveal the basal association interfaces**. Two sequences, D1 and D2 (**A**) were used to delineate the interaction surfaces available for self-association. D1 associates 10% more than D2 (**B**). Clustering the final 50 ns of the simulations, like before, reveals that D1 uses the GxxxG motif (**C**) to dimerize and D2 uses the extended hydrophobic surface (**D**). These two dimers also exhibit mirrored handedness (interaction surface shown in orange). We summarize the results from all the self-association simulations and highlight the two interaction surfaces (red & green) (**E**) that can be schematically shown to interact in two very different ways (**F**) to give a symmetric and an asymmetric trimer. (Color figures available online)

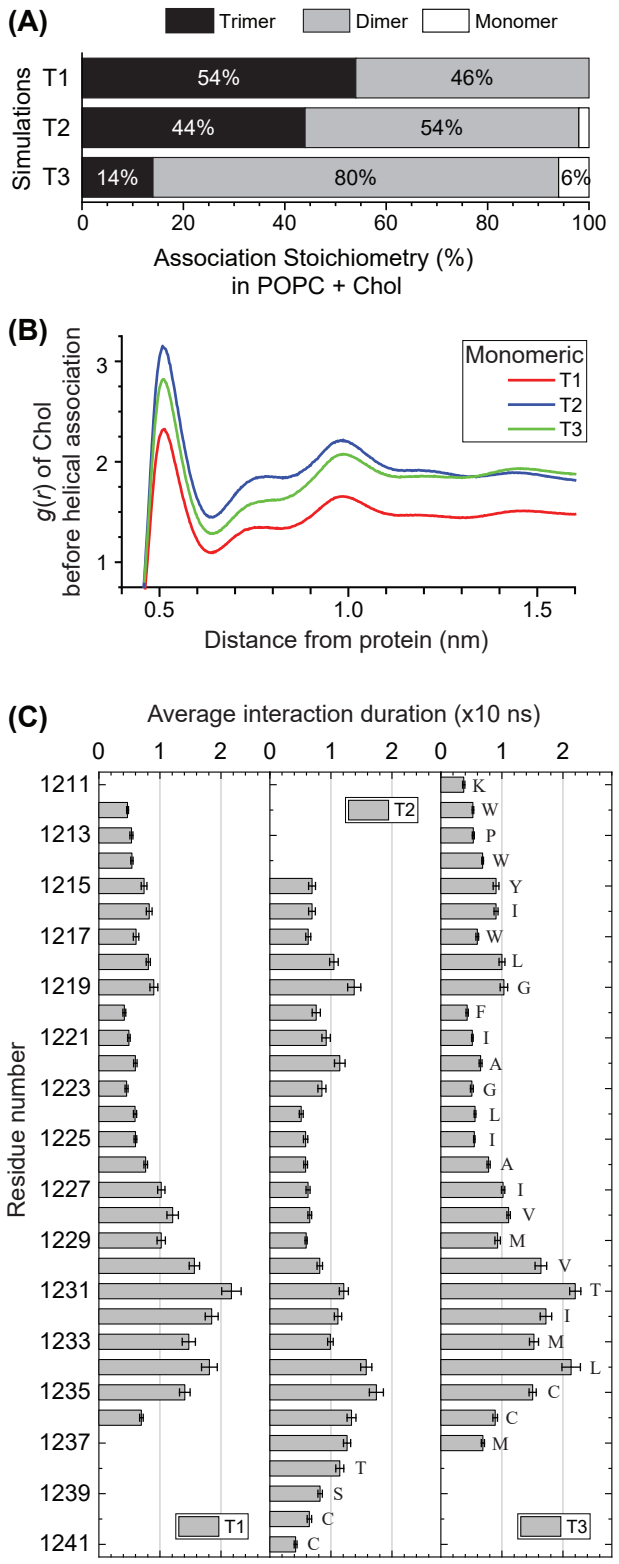

Figure 5: **Cholesterol interacts more with the C-terminus of the spike TM helices**. The three helical sequences were simulated in the presence of cholesterol to test the role of the lipid environment on self-association. While the propensity to trimerize remained unchanged for T1 (**A**), both T2 and T3 registered a 30% reduction compared to the results in pure POPC. The radial distribution function of cholesterol around monomeric helices (**B**) indicates its binding to both T2 and T3. This is only possible if the N-terminal CRAC motif (only present in T3) does not show a strong preference for binding cholesterol. This is exactly what we see upon analyzing residue wise cholesterol binding (**C**). Cholesterol spends more time closer to the C-terminal residues.

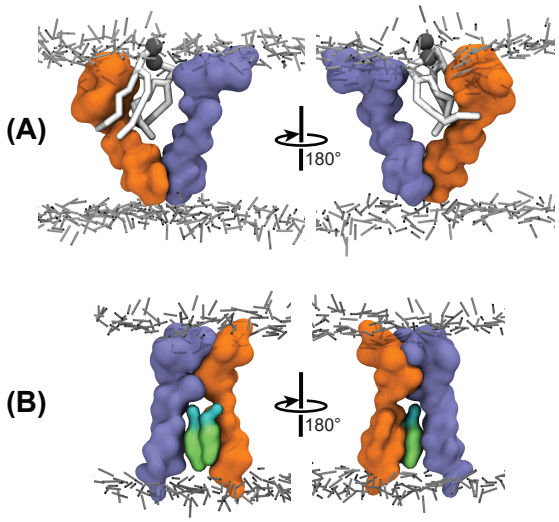

Figure 6: **Cholesterol alters the emergence of homotypic interactions**. A representative snapshot of the initiation of first self-association in POPC membranes (**A**) and in those containing cholesterol (**B**). The presence of cholesterol in the membranes changes the initiation of self-association by binding the C-terminus, effectively occluding the protein-protein interaction. Despite this initial difference, the final structures are comparable in the two lipid environments (**Figure S27** & **S28**).