JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JACS

Communication

A Trimeric Hydrophobic Zipper Mediates the Intramembrane Assembly of SARS-CoV-2 Spike

Qingshan Fu and James J. Chou*

Cite This: J. Am. Chem. Soc. 2021, 143, 8543−8546

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The S protein of SARS-CoV-2 is a type I membrane protein that mediates membrane fusion and viral entry. A vast amount of structural information is available for the ectodomain of S, a primary target by the host immune system, but much less is known regarding its transmembrane domain (TMD) and its membrane-proximal regions. Here, we determined the NMR structure of the S protein TMD in bicelles that closely mimic a lipid bilayer. The TMD structure is a transmembrane α-helix (TMH) trimer that assembles spontaneously in a membrane. The trimer structure shows an extensive hydrophobic core along the 3-fold axis that resembles that of a trimeric leucine/isoleucine zipper, but with tetrad, not heptad, repeats. The trimeric core is strong in bicelles, resisting hydrogen−deuterium exchange for weeks. Although highly stable, structural guided mutagenesis identified single mutations that can completely dissociate the TMD trimer. Multiple studies have shown that the membrane anchors of viral fusion proteins can form highly specific oligomers, but the exact function of these oligomers remains unclear. Our findings should guide future experiments to address the above question for SARS coronaviruses.

The SARS-CoV-2 virion is decorated with a large number of membrane-anchored spike proteins (S) responsible for target recognition, membrane fusion, and virus entry;[1,2] it is also the dominant antigen on the virion surface used for vaccine development.[3] The full-length S is a type I membrane protein that is first expressed as a precursor that trimerizes (S_3) and then cleaved into two fragments ($(S1/S2)_3$): the receptor-binding fragment S1 and the fusion fragment S2.[4]

The processed $(S1/S2)_3$ comprises the crown-shaped ectodomain that contains the receptor-binding domain (RBD), a transmembrane domain (TMD), and a cytoplasmic tail (CT) (Figure 1a). Since the availability of the SARS-CoV-2 genetic code in January of 2020, structural biology of the SARS-CoV-2 spike has progressed at lightning speed owing to cryo-electron microscopy (cryo-EM); that is, over 26 structures of the S1/S2 ectodomain have been published, most of them covering residues 14−1162 (Table S1). However, as has been the case for the spike proteins of many enveloped viruses, the membrane region of the coronavirus spike remains unknown. In a cryo-EM study that thus far provided the most complete view of the S protein, structural details could be seen up to the HR2 region of the S2 fragment (Figure 1a), but the membrane-proximal and transmembrane regions were not resolved.[5]

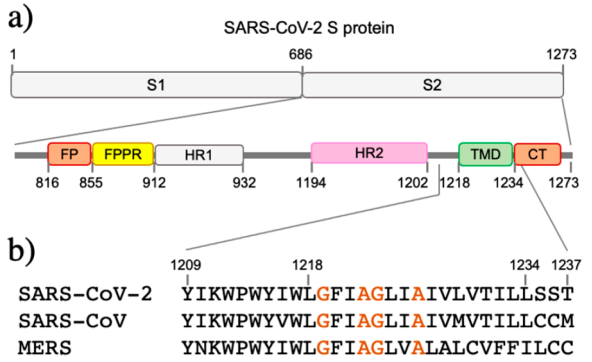


Figure 1. Sequence arrangement of the membrane-interacting regions of SARS-CoV-2 S. (a) Overall domain organization of S2. (b) Sequence alignment of the TMDs of S2 of SARS-CoV-2 (QII57161.1), SARS-CoV (AAS75868.1), and MERS (QDI73610.1).

Previous studies on SARS-CoV, however, suggest that the S TMD has important functional roles other than membrane anchoring. One study showed that swapping the TMD of SARS-CoV S with that of vesicular stomatitis virus (VSV) G protein resulted in 3−25% activity compared to the wild type.[6] Another study reported that insertion of a residue in the TMD resulted in a complete block of viral entry.[7] Further, a recent study on SARS-CoV-2 showed that directly fusing the RBD to the TMD could induce trimerization, suggesting the ability of the TMD to trimerize.[8]

In this study, we used NMR to investigate the structural properties of the TMD of SARS-CoV-2 S in bicelles to fill the knowledge gap. We find that the TMD of the S protein forms a strong trimer in bicelles by a previously unknown mode of transmembrane helix (TMH) assembly.

To determine the TMD structure using NMR, we used an S2 fragment (residues 1209−1237; Figure 1b), derived from a

Received: March 3, 2021
Published: June 4, 2021

https://doi.org/10.1021/jacs.1c02394
J. Am. Chem. Soc. 2021, 143, 8543−8546

SARS-CoV-2 isolate QII57161.1. This construct, designated $S2^{1209-1237}$, contains a short stretch of the membrane-proximal region (residues 1209−1217) and the TM segment (residues 1218−1234). $S2^{1209-1237}$ was reconstituted in DMPC-DH_6PC bicelles with q = 0.55 (Figure S1a,b), where q = [DMPC]/[DH_6PC]. At q = 0.55, the diameter of the planar bilayer region of the bicelles is ∼46 Å.[9] The bicelle-reconstituted $S2^{1209-1237}$ ran on SDS-PAGE as trimers, whereas unreconstituted peptide migrated as monomers (Figure S1c). Further, OG-label and SEC-MALS analyses independently showed that $S2^{1209-1237}$ forms trimers in bicelles (Figure S1d) and detergent micelles (Figure S2), respectively.

The trimeric $S2^{1209-1237}$ in bicelles generated good NMR spectra (Figure S3), and its NMR structure was determined using a published protocol,[10] involving (1) construction of a preliminary monomer structure with local nuclear Overhauser effect (NOE) restraints and backbone dihedral angles derived from chemical shift values (using TALOS+[11]), (2) obtaining a unique structural solution (using ExSSO[12]) that satisfies interchain NOE restraints derived from mixed isotopically labeled sample (Figure S4), and (3) refinement of the trimer structure by further assignment of self-consistent NOE restraints (overall procedure in Figure S5; refinement summary in Table S2; PDB ID: 7LC8).

In bicelles, the TMD of the SARS-CoV-2 S protein folds into a regular α-helix (residues 1218−1234) that assembles into a parallel homotrimer (Figure 2a). Residues 1209−1217 are unstructured in our sample, likely due to N-terminal truncation. The trimeric complex is held together by an extensive hydrophobic core along the 3-fold axis, and the core comprises four layers of hydrophobic interactions involving I1221, I1225, L1229, and L1233, respectively (Figure 2a). Despite the presence of signature sequences for driving TMH oligomerization such as Gly-xxx-Gly and Ala-xxx-Ala,[13,14] our TMD structure does not show direct involvement of the glycine or alanine in forming close van der Waals (VDW) contacts. In this regard, the new TMH trimerization mode is different from the known TMH structures that require one or two small amino acids in establishing intimate helix−helix contact, e.g., the G690 for HIV-1 gp41,[15] the G221 for TNFR1,[16] a central proline for Fas,[17] a central alanine for DR5,[18] and the A794 for HSV gB.[19]

The hydrophobic core of the TMD trimer shows an unusual pattern of tetrad repeat, i.e., I1221, I1225, L1229, and L1233, each occupying position "*a*" of the *abcd* repeat (Figure 2b), and this is very different from the coiled coil mode of assembly of TMH with a heptad repeat.[20] Since each turn of an α-helix consists of 3.6 residues, a four-residue repeat overshoots the i + 4 hydrophobic residues past a helical turn by 40°, diverting the hydrophobic ridge from the 3-fold axis by ∼17° (Figure 2c). Thus, tilting the TMH by 17° would align the hydrophobic ridges of the three TMHs with the 3-fold axis to allow intimate hydrophobic packing (Figure 2c). Indeed, the tilt angle in our experimentally determined structure is ∼18.6 ± 2°, in close agreement with the theoretical analysis.

To examine the S TMD independently by mutagenesis, we generated seven single mutations—G1219Y, G1223Y, I1221Y, I1225Y, A1226Y, L1229Y, and L1233Y—and evaluated their effect on TMH trimerization (Figure 3a). Mutating the characteristic glycine/alanine in the Gly^{1219}-xxx-Gly^{1223} or Ala^{1222}-xxx-Ala^{1226} signature sequence to tyrosine has no effect on TMH trimerization, further supporting the structural conclusion that the relatively conserved glycine and alanine are not directly involved in helix−helix packing. As shown in

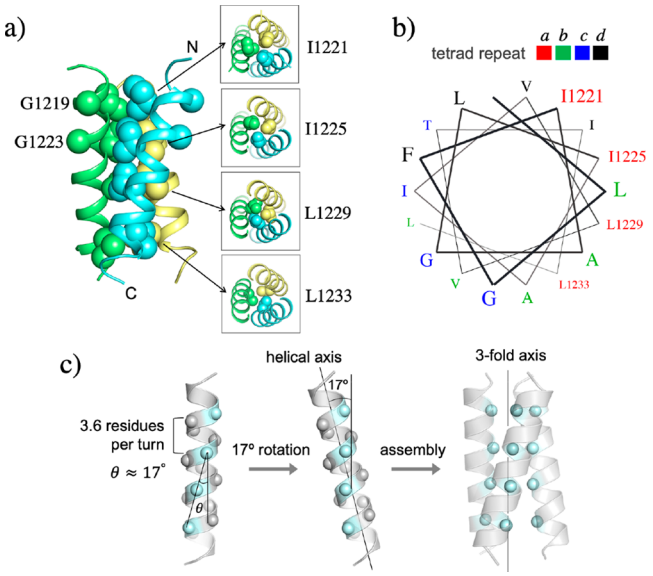


Figure 2. NMR structure of the TMH trimer of SARS-CoV-2 S in the DMPC-DH_6PC bicelle with q = 0.55. (a) Ribbon representation (left) of the TMH trimer structure with the side chain heavy atoms of the core residues shown as spheres; the Cα atoms of G1219 and G1223 are also shown as spheres. The side chain packing at four different levels along the 3-fold axis is illustrated with sectional top views of the trimer (right). (b) Helical wheel representation of an α-helix (3.6 residues per turn) showing that the core hydrophobic residues occupy the position "*a*" of the "*abcd*" tetrad repeat. (c) Theoretical analysis of the trimeric hydrophobic zipper with a tetrad repeat. The line formed by the Cα atoms of the residues at position "*a*" is tilted by ∼17° relative to the helical axis. Rotating the helix by 17° places the position-*a* residues in line with the 3-fold axis for optimal hydrophobic core formation.

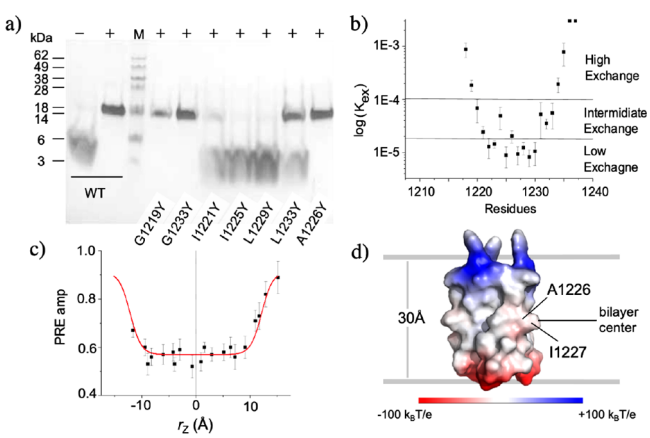


Figure 3. Stability and localization of the TMH trimer of SARS-CoV-2 S in bicelles. (a) SDS-PAGE of bicelle-reconstituted $S2^{1209-1237}$ and its mutants for showing the effect of single mutations on trimerization. Samples were run under nondenaturing conditions. "−" and '+' indicate unreconstituted and bicelle reconstituted, respectively. (b) Residue-specific amide k_{ex} at pH 6.8 determined by H−D exchange measurements. (c) PRE_{amp} vs r_Z best fitted to the symmetric sigmoidal equation (eq S2), where r_Z = 0 corresponds to the bilayer center. (d) Position of the TMH trimer (electrostatic surface representation) relative to the center and boundaries of the planar region of the bicelle.

https://doi.org/10.1021/jacs.1c02394
J. Am. Chem. Soc. 2021, 143, 8543−8546

Figure 2a, G1219 and G1223 are entirely lipid facing and are not expected to participate in interhelical VDW contacts. A1226 is closer to the packing interface but is still not interior enough to be in VDW contact with I1225 or L1229 of the neighboring chain. In contrast, mutating each of the four hydrophobic residues (I1221, I1225, L1229, L1233) that constitute the hydrophobic core to tyrosine all led to severe disruption of the trimer, further consolidating the conclusion that the tetrad repeat of bulky hydrophobic residues is important for the TMH trimerization. Further, the I1225Y or L1229Y mutation almost completely abolished trimerization, while the I1221Y and L1233Y near the N- and C-terminal ends of the TMH, respectively, are less disruptive, probably due to increased dynamics of the helix-ends (Figure S6). This is also consistent with the hydrogen−deuterium (H−D) exchange in which the core residues 1225−1229 exhibited the lowest k_{ex} of all residues (Figure 3b; Figure S7). Overall, the oligomeric properties of the seven mutants agree well with the NMR structure.

To determine the membrane partition of the TMD structure, we performed the paramagnetic probe titration (PPT) analysis[10,21] using $S2^{1209-1237}$ reconstituted in bicelles with q = 0.6. Soluble (Gd-DOTA) and lipophilic (16-DSA) probes were used to provide reciprocal paramagnetic relaxation enhancement (PRE) information (Figure S8). The analysis of residue-specific PRE amplitude (PRE_{amp}) in the context of the TMD structure shows that the lipid bilayer center is between A1226 and I1227 (Figure 3c,d; Figure S9; Table S3). Further, PRE_{amp} reaches the maximal values at about 15 Å away from the center on either side, indicating that the bilayer thickness around the TMD trimer is ∼30 Å. Thus, the S protein TMD caused substantial thinning of the membrane around it.

We have shown that the TMD of the SARS-CoV-2 fusion protein spontaneously trimerizes in the lipid bilayer, and the trimeric assembly is achieved with a previously unknown hydrophobic zipper motif with a tetrad repeat, not with the usual suspects of TMH oligomerization motifs containing glycine or alanine. The role of small amino acids in mediating TMH oligomerization has been observed in several type I/II membrane proteins including glycophorin A,[13] growth factor receptors,[22,23] and receptors in the tumor necrosis factor receptor (TNFR) superfamily.[16−18] The Gly-xxx-Gly is a well-known motif that drives TMH dimerization.[13,14,22,24] There have been no reports, however, of the Gly-xxx-Gly involvement in TMH trimerization. In the trimer structure of the HIV-1 Env TMD, which contains a highly conserved Gly-xxx-Gly, only the first glycine is involved in helix−helix packing; the second glycine is lipid facing.[15,25] The SARS-CoV-2 S TMD also contains highly conserved small amino acids (G1219, A1222, G1223, A1226), which we thought initially to be important for TMD oligomerization. But, neither the glycines nor alanines in the trimer structure appear to be important for the hydrophobic core formation. The purpose of the Gly-xxx-Gly motif remains unknown. If the trimer structure presented here represents the prefusion state, a possible role of the glycine motif is in later steps of the fusion mechanism.

Although the TMH is relatively short (∼16 residues), it can have an extensive hydrophobic core with four layers of hydrophobic interaction. This is attributed to the tetrad repeat of hydrophobic residues, as opposed to the heptad repeat in a classic coiled coil structure. On the basis of the extensive hydrophobic packing, we believe the TMD trimer is stable in the membrane. A potential implication is that the TMD trimer is unlikely to dissociate in the membrane unless significant force is applied during the unfolding and refolding steps of the fusion component.

Although functional mutagenesis of the SARS-CoV-2 S TMD has not been reported, a previous study on the SARS-CoV reported that inserting an amino acid between G1201 and F1202 of the S TMD completely blocked viral entry.[7] G1201 and F1202 in SARS-CoV correspond to G1219 and F1220 in SARS-CoV-2, respectively (Figure 1b). In the context of our TMH trimer structure, such insertion might not disrupt trimerization but could place the tetrad repeat out of register relative to the still unknown membrane-proximal structure and thus prevent proper TMH trimerization.

In conclusion, the TM anchor of the SARS-CoV-2 fusion protein spontaneously trimerizes in the membrane. The trimeric complex is stabilized by an extensive hydrophobic core along the 3-fold axis, formed by the bulky hydrophobic amino acids repeated every four residues. This mode of TMH trimerization is significantly different from the known TMH trimer structures of fusion proteins from other viruses. Strong intramembrane oligomerization appears to be a recurring theme for viral fusion proteins, but its functional roles remain unclear. The reported structure of the TMD of SARS-CoV-2 fusion protein allowed us to identify single mutations that can completely dissociate the trimeric assembly. We believe these mutations are valuable information for guiding future functional experiments for addressing the above question.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.1c02394.

Tables S1−S4 and Figures S1−S10, as well as the description of sample preparation and NMR and biochemical analyses (PDF)

■ AUTHOR INFORMATION

Corresponding Author

James J. Chou − *Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, United States;* orcid.org/0000-0002-4442-0344; Email: james_chou@hms.harvard.edu

Author

Qingshan Fu − *Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.1c02394

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Bo OuYang and Bing Chen for insightful discussions. This study was supported by NIH grants AI127193 and GM140887 to J.J.C. NMR data were collected at the MIT-Harvard CMR (supported by NIH grants P41 GM132079 and S10 OD0253523-01A1).

REFERENCES

(1) Klein, S.; Cortese, M.; Winter, S. L.; Wachsmuth-Melm, M.; Neufeldt, C. J.; Cerikan, B.; Stanifer, M. L.; Boulant, S.; Bartenschlager, R.; Chlanda, P. SARS-CoV-2 structure and replication characterized by in situ cryo-electron tomography. *Nat. Commun.* **2020**, *11* (1), 5885.

(2) Benton, D. J.; Wrobel, A. G.; Xu, P.; Roustan, C.; Martin, S. R.; Rosenthal, P. B.; Skehel, J. J.; Gamblin, S. J. Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion. *Nature* **2020**, *588* (7837), 327−330.

(3) Lv, Z.; Deng, Y. Q.; Ye, Q.; Cao, L.; Sun, C. Y.; Fan, C.; Huang, W.; Sun, S.; Sun, Y.; Zhu, L.; Chen, Q.; Wang, N.; Nie, J.; Cui, Z.; Zhu, D.; Shaw, N.; Li, X. F.; Li, Q.; Xie, L.; Wang, Y.; Rao, Z.; Qin, C. F.; Wang, X. Structural basis for neutralization of SARS-CoV-2 and SARS-CoV by a potent therapeutic antibody. *Science* **2020**, *369* (6510), 1505−1509.

(4) Bosch, B. J.; van der Zee, R.; de Haan, C. A.; Rottier, P. J. The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. *J. Virol.* **2003**, *77* (16), 8801−11.

(5) Cai, Y.; Zhang, J.; Xiao, T.; Peng, H.; Sterling, S. M.; Walsh, R. M., Jr.; Rawson, S.; Rits-Volloch, S.; Chen, B. Distinct conformational states of SARS-CoV-2 spike protein. *Science* **2020**, *369* (6511), 1586−1592.

(6) 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. *J. Virol.* **2006**, *80* (3), 1302−10.

(7) Corver, J.; Broer, R.; van Kasteren, P.; Spaan, W. Mutagenesis of the transmembrane domain of the SARS coronavirus spike glycoprotein: refinement of the requirements for SARS coronavirus cell entry. *Virol. J.* **2009**, *6*, 230.

(8) Azad, T.; Singaravelu, R.; Crupi, M. J. F.; Jamieson, T.; Dave, J.; Brown, E. E. F.; Rezaei, R.; Taha, Z.; Boulton, S.; Martin, N. T.; Surendran, A.; Poutou, J.; Ghahremani, M.; Nouri, K.; Whelan, J. T.; Duong, J.; Tucker, S.; Diallo, J. S.; Bell, J. C.; Ilkow, C. S., Implications for SARS-CoV-2 Vaccine Design: Fusion of Spike Glycoprotein Transmembrane Domain to Receptor-Binding Domain Induces Trimerization. *Membranes (Basel, Switz.)* **2020**, *10* (9), 215

(9) Sanders, C. R., 2nd; Schwonek, J. P. Characterization of magnetically orientable bilayers in mixtures of dihexanoylphosphatidylcholine and dimyristoylphosphatidylcholine by solid-state NMR. *Biochemistry* **1992**, *31* (37), 8898−905.

(10) Fu, Q.; Piai, A.; Chen, W.; Xia, K.; Chou, J. J. Structure determination protocol for transmembrane domain oligomers. *Nat. Protoc.* **2019**, *14* (8), 2483−2520.

(11) Shen, Y.; Delaglio, F.; Cornilescu, G.; Bax, A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **2009**, *44* (4), 213−23.

(12) Yang, J.; Piai, A.; Shen, H. B.; Chou, J. J. An Exhaustive Search Algorithm to Aid NMR-Based Structure Determination of Rotationally Symmetric Transmembrane Oligomers. *Sci. Rep.* **2017**, *7* (1), 17373.

(13) MacKenzie, K. R.; Prestegard, J. H.; Engelman, D. M. A transmembrane helix dimer: structure and implications. *Science* **1997**, *276* (5309), 131−3.

(14) Trenker, R.; Call, M. E.; Call, M. J. Crystal Structure of the Glycophorin A Transmembrane Dimer in Lipidic Cubic Phase. *J. Am. Chem. Soc.* **2015**, *137* (50), 15676−9.

(15) 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* (6295), 172−175.

(16) Zhao, L.; Fu, Q.; Pan, L.; Piai, A.; Chou, J. J. The Diversity and Similarity of Transmembrane Trimerization of TNF Receptors. *Front. Cell Dev. Biol.* **2020**, *8*, 569684.

(17) Fu, Q.; Fu, T. M.; Cruz, A. C.; Sengupta, P.; Thomas, S. K.; Wang, S.; Siegel, R. M.; Wu, H.; Chou, J. J. Structural Basis and Functional Role of Intramembrane Trimerization of the Fas/CD95 Death Receptor. *Mol. Cell* **2016**, *61* (4), 602−613.

(18) Pan, L.; Fu, T. M.; Zhao, W.; Zhao, L.; Chen, W.; Qiu, C.; Liu, W.; Liu, Z.; Piai, A.; Fu, Q.; Chen, S.; Wu, H.; Chou, J. J. Higher-Order Clustering of the Transmembrane Anchor of DR5 Drives Signaling. *Cell* **2019**, *176* (6), 1477−1489.

(19) Cooper, R. S.; Georgieva, E. R.; Borbat, P. P.; Freed, J. H.; Heldwein, E. E. Structural basis for membrane anchoring and fusion regulation of the herpes simplex virus fusogen gB. *Nat. Struct. Mol. Biol.* **2018**, *25* (5), 416−424.

(20) Rout, A. K.; Strub, M. P.; Piszczek, G.; Tjandra, N. Structure of transmembrane domain of lysosome-associated membrane protein type 2a (LAMP-2A) reveals key features for substrate specificity in chaperone-mediated autophagy. *J. Biol. Chem.* **2014**, *289* (51), 35111−23.

(21) Piai, A.; Fu, Q.; Dev, J.; Chou, J. J. Optimal Bicelle Size q for Solution NMR Studies of the Protein Transmembrane Partition. *Chem. - Eur. J.* **2017**, *23* (6), 1361−1367.

(22) Bocharov, E. V.; Mineev, K. S.; Volynsky, P. E.; Ermolyuk, Y. S.; Tkach, E. N.; Sobol, A. G.; Chupin, V. V.; Kirpichnikov, M. P.; Efremov, R. G.; Arseniev, A. S. Spatial structure of the dimeric transmembrane domain of the growth factor receptor ErbB2 presumably corresponding to the receptor active state. *J. Biol. Chem.* **2008**, *283* (11), 6950−6.

(23) Endres, N. F.; Das, R.; Smith, A. W.; Arkhipov, A.; Kovacs, E.; Huang, Y.; Pelton, J. G.; Shan, Y.; Shaw, D. E.; Wemmer, D. E.; Groves, J. T.; Kuriyan, J. Conformational coupling across the plasma membrane in activation of the EGF receptor. *Cell* **2013**, *152* (3), 543−56.

(24) Chen, W.; Gamache, E.; Rosenman, D. J.; Xie, J.; Lopez, M. M.; Li, Y. M.; Wang, C. Familial Alzheimer's mutations within APPTM increase Abeta42 production by enhancing accessibility of epsilon-cleavage site. *Nat. Commun.* **2014**, *5*, 3037.

(25) Chen, B.; Chou, J. J. Structure of the transmembrane domain of HIV-1 envelope glycoprotein. *FEBS J.* **2017**, *284* (8), 1171−1177.

https://doi.org/10.1021/jacs.1c02394
J. Am. Chem. Soc. 2021, 143, 8543−8546

Supporting Information

A trimeric hydrophobic zipper mediates the intramembrane assembly of SARS-CoV-2 spike

Qingshan Fu and James J. Chou*

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA. *Corresponding author: james_chou@hms.harvard.edu

Materials and Methods

Protein expression and purification

DNA of a fragment of SARS-CoV-2 S2 (isolate QII57161.1) containing residues 1209-1236 was synthesized by GenScript (Piscataway, NJ). The expression construct, designated $S2^{1209-1237}$, was created as a N terminal trpLE fusion in the plasmid pMM-LR6 following a procedure described previously[1]. Mutant constructs were generated by standard PCR mutation protocols and confirmed by DNA sequencing. Transformed *E. coli* strain BL21 (DE3) cells were grown in M9 minimal media supplemented with stable isotopes. Cultures were grown at 37°C to reach an absorbance of 0.6~0.8 at 600 nm and cooled to 22°C before induction with 150 μM isopropyl β-D-thiogalatopyranoside at 22° C for overnight (12 hours). For completely deuterated proteins, bacterial culture in 1 ml LB medium was spun down and the cells were adapted in 99.8% D_2O medium (100 ml, with deuterated glucose) over night. Then collected cells were grown in 99.8% D_2O (4 L) (Sigma Aldrich, St. Louis, MO) with deuterated glucose (99.8% deuterium, Cambridge Isotope Laboratories, Tewksbury, MA). The expressed fusion protein was extracted in a denaturing buffer (1% Triton X-100, 6 M guanidine hydrochloride, 50 mM Tris, pH 8.0, and 200 mM NaCl), purified with a Ni-NTA affinity column, cleaved by cyanogen bromide (CNBr) to separate trpLE and $S2^{1209-1237}$. The final product was purified RP-HPLC in a Zorbax SB-C3 column (Agilent Technologies, Santa Clara, CA) with a gradient from 95% dH_2O, 5% isopropanol (IPA), 0.1% trifluoroacetic acid (TFA) (buffer A) to 75% IPA, 25% acetonitrile, 0.1% TFA (buffer B). Fractions containing pure $S2^{1209-1237}$ was confirmed by MALDI-TOF mass spectrometry and SDS-PAGE.

Reconstitution

Approximately 2 mg of lyophilized $S2^{1209-1237}$ was mixed with 10 mg 1,2-dimyristoyl-sn-Glycero-3-Phosphocholine (DMPC; protonated or deuterated from Avanti Ploar Lipids, Alabaster, AL) and dissolved in 1,1,1,3,3,3-hexafluoro-2-propanol (HFIP). The mixture was slowly dried to a thin film under nitrogen stream, followed by overnight lyophilization. The dried thin film was redissolved in 3 ml of 8 M urea containing ~20 mg 1,2-dihexanoyl-sn-Glycero-3-Phosphocholine (DH_6PC; protonated or deuterated from Avanti Polar Lipids), then dialyzed (MWCO 3.5 kDa) against 20 mM Tris buffer, pH 6.8 to remove the denaturant. During and after the dialysis, additional DH_6PC was added to make up for the DH_6PC loss due to dialysis, adjusting the DMPC:DH_6PC ratio (q) to approximately 0.5. The bicelle q was quantified by signal integration of the DMPC and DH_6PC methyl peaks in the 1D ^1H NMR spectrum. The sample was concentrated in a Centricon (EMD Millipore, Billerica, MA) to ~350 μl. The final NMR sample contained ~0.8 mM MPER-TMD (monomer), ~55 mM DMPC, ~100 mM DH_6PC, 20 mM Tris (pH 6.8), 0.02% NaN3 and 5% D_2O.

Analysis of oligomeric state of the TMD of SARS-CoV-2 S2 in bicelles by SDS-PAGE

Wild type (WT) and mutants of $S2^{1209-1237}$ were first reconstituted in bicelles (q = 0.5) and then mixed with SDS-PAGE sample loading buffer (Invitrogen) without boiling, followed by SDS-PAGE at 200 volts for 30 minutes and Commassie blue staining. Purified WT $S2^{1209-1237}$ (without reconstitution) migrated as monomer at ~5 kDa (theoretical MW = 3.4 kDa; see Fig. S1c). The WT $S2^{1209-1237}$ (after reconstitution) migrated at ~15 kDa (theoretical MW = 10.2 kDa; see Fig. S1c), consistent with the size of a trimer.

Analysis of oligomeric state of the TMD of SARS-CoV-2 S2 in bicelles by OG-label

In the OG-label method[2], each protomer of the oligomer to be studied is non-covalently labeled with a soluble cross-linkable protein (SCP), so that the latter can be cross-linked with Lomant's reagents to read out the sample oligomeric state. The small Ig-fold protein named GB1 (MW = 8.4 kDa) has been proven to serve as the SCP very effectively[1-3]. A TriNTA molecule is linked via PEG-2-SMCC (succinimidyl 4-(N-maleimidomethyl)cyclohexane-1-carboxylate) to the N-terminus of GB1, while a His_6-tag is added to the C-terminus of the oligomeric protein, so that the TriNTA-GB1 conjugate can strongly attach to the protomer (the binding affinity of TriNTA to His_6-tag is 20 ± 10 nM). The GB1s are then cross-linked to report the oligomeric state of the protein, as the local concentration of stoichiometric amount of GB1 to the oligomer allows for more efficient cross-linking than for the free GB1 in solution. Finally, the cross-linked GB1s are released from the oligomer by addition of EDTA and analyzed by SDS-PAGE.

To implement the OG-label method for the $S2^{1209-1237}$, a His_6-tag was added at the C-terminus of the protein. The His_6-tagged protein was expressed, purified and reconstituted in bicelles (q = 0.5) as described above, except the sample buffer was a phosphate buffer (pH 7.4) for better cross-linking efficiency. To prevent unwanted cross-linking between the membrane protein and GB1, the free primary amines of the $S2^{1209-1237}$ were blocked by reacting with 100-fold molar excess of Sulfo-NHS acetate (Thermo Fisher Scientific) at room temperature for 1.5 hours. Excess Sulfo-NHS acetate was removed by dialysis while tightly controlling the bicelle q. After dialysis, the sample was concentrated to 30 μM and mixed with 60 μM TriNTA-GB1 to ensure that all the His_6-tags were saturated with TriNTA-GB1. The mixtures were then incubated at room temperature with various concentration of BS3(PEG9) (0.5, 1.0, and 2.0 mM; see lanes 4, 5, and 6 in the gel in Fig. S1d) for 30 minutes, followed by a second incubation at room temperature with 0.6 mM glutaraldehyde for 3 minutes. The cross-linking reactions were quenched with 20 mM Tris buffer (pH 7.5) upon incubation at room temperature for 15 minutes. As negative control, 2.0 mM BS3(PEG9) and 0.6 mM glutaraldehyde were sequentially added to 60 μM TriNTA-GB1 in the absence of the His_6-tagged $S2^{1209-1237}$. Lane 3 in the gel in Fig. S1d shows that the TriNTA-GB1 alone remained mostly monomeric in the cross-linking condition used. The cross-linked GB1s were then released from the samples by adding 50 mM EDTA and examined by SDS-PAGE using a 12% Bis-Tris protein gel (Thermo Fisher Scientific) (Fig. S1d).

Analysis of oligomeric state of the TMD of SARS-CoV-2 S2 by SEC-MALS

To independently address $S2^{1209-1237}$ oligomerization using a non-crosslinking method, we performed SEC-MALS of the His_6-tagged $S2^{1209-1237}$ (used above for OG-label) reconstituted in n-Dodecyl-β-D-Maltopyranoside (DDM) micelles before and after mixing with GB1-TriNTA to examine whether the difference in mass account for three GB1-TriNTAs. The analytical size exclusion chromatography (SEC) was performed at room temperature using a Superdex 200 10/300GL column (Sigma) equilibrated with a mobile phase buffer containing 20mM Tris, pH 7.5, 50mM NaCl, 0.5mM DDM, 1mM $NiSO_4$. 100 μL of 0.4 mM His_6-tagged $S2^{1209-1237}$ in 60 mM DDM with 20mM Tris, pH 7.5, 50mM NaCl, and 1mM $NiSO_4$ was injected into the column and eluted at a flow rate of 0.4 ml/min. Then in a separate run, 100 μL of the same His_6-tagged $S2^{1209-1237}$ sample containing 0.6 mM GB1-TriNTA was applied. The instrument setup used for the SEC-MALS experiment consists of an Agilent 1260 Infinity Isocratic Liquid Chromatography System connected in series with a Wyatt Dawn Heleos II Multi-Angle Light Scattering (MALS) detector (Wyatt Technology) and a Wyatt Optilab T-rEX Refractive Index Detector (Wyatt Technology).

The column effluent was monitored in line with three detectors that simultaneously monitored UV absorption, light scattering, and refractive index. The data from the three detectors were imported by the ASTRA software package, and the three-detector method was used to determine the molecular mass (Fig. S2).

NMR resonances and NOE assignment

NMR data were collected at 303K on Bruker spectrometers operating at ^{1}H frequency of 900 MHz, 700 MHz, or 600 MHz equipped with cryogenic probes. NMR data sets were processed using nmrPipe[4]. NMR spectra were analyzed using Sparky (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco) and XEASY[5]. Peak intensities were measured at peak local maxima using quadratic interpolation to identify peak centers. Origin (OriginLab, Northampton, MA) was used to fit the experimental data.

Sequence-specific assignment of backbone $^{1}H^{N}$, ^{15}N, $^{13}C^{\alpha}$ and ^{13}C' resonances were accomplished using 3D TROSY-based HNCA, HN(CO)CA, HN(CA)CO and HNCO spectra[6-7], recorded using a (^{15}N, ^{13}C, 90% ^{2}H)-labeled sample. The aliphatic and aromatic resonances of the protein sidechains were assigned using a 3D ^{15}N-edited NOESY-TROSY-HSQC (τ_{NOE} = 60 ms) and a 3D ^{13}C-edited NOESY-HSQC (τ_{NOE} = 100 ms) spectra, recorded at ^{1}H frequency of 900 MHz. These NOESYs were performed using a (^{15}N, ^{13}C)-labeled protein sample in bicelles (q = 0.55) made of DMPC and DH_6PC with deuterated acyl chains (Avanti Lipids). Specific stereo assignments of the methyl groups of valines and leucines were obtained from a ^{1}H-^{13}C HSQC spectrum recorded with 28 ms ^{13}C constant-time evolution, on a 700 MHz spectrometer, using a 15% ^{13}C-labeled protein that was expressed with 15% uniformly ^{13}C-labeled glucose and 85% regular glucose[8]. The same 3D ^{15}N-edited NOESY-TROSY-HSQC and ^{13}C-edited NOESY-HSQC spectra were used to assign local NOE-derived distance restraints.

For assigning inter-chain distance restraints, a J_{CH}-coupled NOE experiment was performed to exclusively detect inter-chain NOEs between the ^{15}N-attached protons of one chain and the ^{13}C-attached protons of the neighboring chains, using a mixed sample containing 50% (^{15}N, ^{2}H)-labeled and 50% ^{13}C-labeled protein. In this experiment, two interleaved 3D ^{15}N-edited NOESY-TROSY spectra (τ_{NOE} = 150 ms) were recorded, at ^{1}H frequency of 900 MHz, in which one has ^{13}C decoupling during ^{1}H evolution before NOE mixing and the other does not (Fig. S4b).

Independent of the J_{CH}-coupled NOE experiment, we also performed the J_{CH}-modulated NOE experiment[1], in which the ^{1}H evolution period before the NOE mixing is changed to a "mixed constant-time" evolution to introduce ^{1}H-^{13}C J evolution. Two interleaved spectra were recorded at ^{1}H frequency of 900 MHz with different J_{CH} evolution (J_{CH} = 0 ms, and J_{CH} = 8 ms) before the NOE mixing (τ_{NOE} = 150 ms). For J_{CH} = 0 ms, the inter- and intra- molecular NOE peaks are both positive, whereas for J_{CH} = 8 ms, the inter- and intra- molecular NOE peaks are negative and positive, respectively (Fig. S4c). The ^{1}H transverse relaxation during 8 ms mixed constant-time evolution resulted in significant loss of sensitivity, and hence, a shorter S2 fragment with better relaxation property ($S2^{1215-1237}$) was used for this experiment.

NMR structure calculation

The structures were generated using the program XPLOR-NIH[9]. First, the monomer structure was generated using local NOE restraints and the backbone dihedral restraints derived from the

backbone ^{15}N, $^{1}H^{N}$, $^{13}C\alpha$, and $^{13}C'$ chemical shifts (using the TALOS+ program[10]). Second, the monomer structure and inter-chain NOE restraints were used with the ExSSO program[11] to generate a unique solution of trimeric assembly (Fig. S5a, S5b). Finally, the initial trimer solution was fed to the XPLOR-NIH for iterative refinement against all NMR restraints, while assigning more self-consistent inter-chain NOEs in both ^{13}C-edited NOESY-HSQC and isotopically mixed NOE spectra after each iteration (Fig. S5c, S5d, S10).

For each inter-chain restraint between two adjacent chains, three identical distance restraints were assigned respectively to all pairs of neighboring chains to satisfy the condition of C3 rotational symmetry. The XPLOR refinement used a simulated annealing (SA) protocol in which the temperature in the bath was cooled from 1000 to 200 K with steps of 20 K. The NOE restraints were enforced by flat-well harmonic potentials, with the force constant ramped from 2 to 30 kcal/mol $Å^{-2}$ during annealing. Backbone dihedral angle restraints were taken from the 'GOOD' dihedral angles from TALOS+, all with a flat-well (±the corresponding uncertainties from TALOS+) harmonic potential with force constant ramped from 5 to 1000 kcal/mol rad^{-2}. A total of 75 structures were calculated and 15 lowest energy structures were selected as the final structural ensemble (Fig. S5e; Table S2).

Analysis of transmembrane partition

The paramagnetic probe titration (PPT) method[1, 12] was used to determine the membrane partition of the S2 TMD. This method is based on the notion that if the bicelle is sufficiently wide (q > 0.5), the lateral solvent PRE becomes negligible, thus allowing the use of measurable solvent paramagnetic relaxation enhancement (PRE) to probe residue-specific depth immersion of the protein in the bilayer region of the bicelle. To ensure that our bicelles are wide enough, we reconstituted (^{15}N, ^{2}H)-labeled S2$^{1209\text{-}1237}$ in bicelles with q = 0.6 (Fig. S1b). The water-soluble and membrane-inaccessible paramagnetic agent, Gd-DOTA (Sigma), was used to generate solvent PRE. Gd-DOTA stock solution (200 mM) was titrated into the bicelle sample to reach final concentrations of 0.0, 1.0, 2.0, 4.0, 6.0, 8.0, 10.0 and 15.0 mM. At each concentration, a 2D ^{1}H-^{15}N TROSY-HSQC spectrum was recorded at 600 MHz to measure residue-specific PRE, defined here as the ratio of peak intensity in the presence (I) and absence (I_0) of the paramagnetic agent. Peak intensities were measured at peak local maxima using quadratic interpolation to identify peak centers. For each of the residues, we used *Origin* (OriginLab, Northampton, MA) to fit the PRE titration curve to exponential decay

$$\frac{I}{I_0} = 1 - PRE_{amp}\left(1 - e^{-\frac{[Gd-DOTA]}{\tau}}\right) \qquad \text{(Eq. S1)}$$

to derive the residue-specific PRE amplitude (PRE_{amp}) (Fig. S8a). To determine the position of the trimeric TMD relative to the bilayer center, we calculated, for each residue i, the distance (r_Z) along the protein symmetry axis from the amide proton to an arbitrary reference point based on the structure of the trimeric TMD. This calculation converted PRE_{amp} vs. (residue number) to PRE_{amp} vs. r_Z, which was then analyzed using the sigmoidal fitting method. Briefly, the trimer structure was moved along the 3-fold axis in increment of 0.5 Å (Fig. S9a, left) to achieve the best fit to the symmetric sigmoid equation:

$$PRE_{amp} = PRE_{amp}^{min} + \frac{\left(PRE_{amp}^{max} - PRE_{amp}^{min}\right)}{1 + e^{\left(r_Z^I - |r_Z|\right)/SLOPE}} \qquad \text{(Eq. S2)}$$

where PRE_{amp}^{min} and PRE_{amp}^{max} are the limits within which PRE_{amp} can vary for a particular protein system, r_Z^I is the inflection point (the distance from the bilayer center at which PRE_{amp} is halfway between PRE_{amp}^{min} and PRE_{amp}^{max}), and *SLOPE* is a parameter which reports the steepness of the curve at the inflection point. The best fit (Fig. S9a, right) gave an adjusted coefficient of determination (R^2_{adj}) of 0.92 and was used to determine the position of the trimer structure with respect to the bilayer center (r_Z = 0).

To complete the Gd-DOTA titration, a set of lipophilic PRE data was also acquired using a lipophilic paramagnetic agent 16-Doxyl-stearic acid (16-DSA). Using the identical approach to that employed for the solvent PRE, the bicelle-reconstituted S2$^{1209\text{-}1237}$ was titrated with the membrane-embedded paramagnetic agent 16-DSA at various known concentrations. The titrant stock solution (25 mM 16-DSA) was prepared in the same buffer as that of the protein sample, i.e., using q = 0.6 bicelle solution to solubilize the paramagnetic agent to prevent changes in the sample bicelle q during titration. The 16-DSA was added in small aliquots (8~10 µL per step) to minimize sample dilution. The progress of the titration was monitored by measuring a 2D ^{1}H-^{15}N TROSY-HSQC spectrum at each of the following 16-DSA concentrations: 0, 0.5, 1.0, 2.0, 3.0, 4.0, 5.0 and 7.5 mM. The residue-specific PRE_{amp} was determined by fitting the peak intensity decay as a function of [16-DSA] to the exponential decay equation Eq. S1. The PRE_{amp} vs. (residue number) plot for 16-DSA shows a reciprocal profile of that of Gd-DOTA (Fig. S8b).

Hydrogen-deuterium (H-D) exchange

For measuring H-D exchange of the S2 TMD in bicelles, S2$^{1209\text{-}1237}$ was reconstituted in protonated solvent in the beginning and was flash-frozen in liquid nitrogen and then thoroughly lyophilized to get rid of protonated water. The dried sample was dissolved in 360 µL of 99.9% D_2O. The progress of the H-D exchange was monitored by measuring a 2D ^{1}H-^{15}N TROSY-HSQC spectrum at uniform time intervals of ~3 hours up to ~5 days. The residue-specific exchange constant, k_{ex} (=1/τ_{ex}), was determined by fitting the fractional peak intensity vs. time to the following exponential decay equation:

$$\frac{I}{I_0} \propto \left(e^{-\frac{t}{\tau ex}}\right) \qquad \text{(Eq. S3)}$$

where I_0 and I are the peak intensities before and after the H-D exchange, t is the time passed from the beginning of the exchange, and τ_{ex} is the time constant of the decay.

NMR Dynamics Measurements

Residue-specific rotational correlation time (τ_C) was measured using a 2D version of the 1D experiment known as TRACT (TROSY for rotational correlation times)[13]. For this experiment, a sample of ~0.9 mM (^{15}N, ^{13}C, 90% ^{2}H)-labeled S2$^{1209\text{-}1237}$ reconstituted in bicelles with q = 0.55 was used. In the TRACT experiment, ^{15}N relaxation delays were set to 0, 8, 16, 32, 40, 56, 80 ms for the TROSY component, and set to 0, 8, 16, 20, 24, 32, 50 ms for anti-TROSY component. Due to the fast relaxation of the anti-TROSY component, 4-fold more scans were used for recording the anti-TROSY spectra. For each of the assigned peaks, the R_2 relaxation rate of the

TROSY (R_2^{α}) and anti-TROSY (R_2^{β}) components were determined by fitting the measurements to the exponential decay function:

$$I/I_0 = y + Ae^{-R*Delay} \qquad \text{(Eq. S4)}$$

where *I* and I_0 represent the peak intensity with and without relaxation delay, respectively. Data fitting was done using the Origin program (OriginLab, Northampton, MA). τ_C was derived from the following equations:

$$R_\beta - R_\alpha = 2p\delta_N\big(4\,J(0) + 3\,J(\omega_N)\big)(3cos^2\theta - 1) \qquad \text{(Eq. S5)}$$

$$J(\omega) = 0.4\tau_C/[1 + (\tau_C\omega)^2] \qquad \text{(Eq. S6)}$$

where *p* is the dipole-dipole coupling between ^{1}H and ^{15}N and δ_N is the chemical shift anisotropy of the ^{15}N nucleus with the parameter θ = 17°, and ω is the spectrometer ^{15}N frequency (60 MHz).

Table S1. Reported structures of SARS-CoV-2 spike protein in Year 2020

PDB ID	Method	Construct	Structured region	state	Reference
6LXT	X-ray	910-1206	913-988, 1163-1202	Post-fusion state	[14]
6VSB	EM	1-1208	27-1146	Ecto-domain trimer in pre-fusion state	[15]
6VXX	EM	14-1211	27-1147	Ecto-domain trimer in close state	[16]
6VYB	EM	14-1211	27-1147	Ecto-domain trimer in open state	[16]
6X29	EM	16-1208	27-1147	Ecto-domain trimer in rS2d Down State	[17]
6X2A	EM	16-1208	27-1147	Ecto-domain trimer in u1S2q 1 Up State	[17]
6X2B	EM	16-1208	27-1147	U1S2q 2 RBD Up Spike Protein Trimer	[17]
6X2C	EM	16-1208	27-1147	U1S2q All Down RBD State	[17]
6X79	EM	14-1211	27-1140	Ecto-domain trimer in close state	[18]
6XR8	EM	1-1273	14-1162	S protein trimer in prefusion state	[19]
6XRA	EM	1-1273	703-1197	S protein trimer in postfusion state	[19]
6XS6	EM	1-1213	27-1147	Spike D614G variant	[20]
6ZGE	EM	1-1208	27-1146	S protein in closed state	[21]
6ZGF	EM	1-1208	27-1146	S protein in closed state	[21]
6ZGG	EM	1-1208	27-1146	Furin Cleaved Spike Protein	[21]
6ZGH	EM	1-1208	27-1146	Intermediate state	[21]
6ZGI	EM	1-1208	27-1146	Closed state	[21]
6ZOX	EM	14-1211	27-1146	Disulphide-stabilized Spike Protein Trimer	[22]
6ZOY	EM	14-1211	27-1146	Disulphide-stabilized Spike Protein Trimer	[22]
6ZOZ	EM	14-1211	27-1146	Disulphide-stabilized Spike Protein Trimer	[22]
6ZP0	EM	14-1211	27-1146	Disulphide-stabilized Spike Protein Trimer	[22]
6ZP1	EM	14-1211	27-1146	Disulphide-stabilized Spike Protein Trimer	[22]
6ZP2	EM	14-1211	27-1146	Disulphide-stabilized Spike Protein Trimer	[22]
6ZWV	EM	1-1273	27-1151	Spike Proteins on intact virions	[23]
7JJI	EM	1-1273	14-1146	Prefusion spike trimer	[24]
7JJJ	EM	1-1273	14-1146	Dimers of spike trimers	[24]

Table S2. NMR and refinement statistics for the TMD of SARS-CoV-2 S protein

NMR distance and dihedral constraints [a]	TMD (1217-1237)
Distance constraints from NOE	387
Short-range intramolecular ($\|i-j\| \leq 4$)	109 x 3
Long-range intramolecular ($\|i-j\| \geq 5$)	0
Intermolecular	20 x 3
Total dihedral angle restraints [b]	120
ϕ (TALOS)	20 x 3
ψ (TALOS)	20 x 3
Structure statistics [c]	
Violations (mean ± s.d.)	
Distance constraints (Å)	0.091 ± 0.005
Dihedral angle constraints (°)	0.210 ± 0.042
Deviations from idealized geometry	
Bond lengths (Å)	0.006 ± 0.000
Bond angles (°)	0.683 ± 0.020
Impropers (°)	0.367 ± 0.020
Average pairwise r.m.s. deviation (Å) [d]	
Heavy	0.958
Backbone	0.579

[a] The numbers of constraints are summed over all three subunits.

[b] Backbone ϕ and ψ restraints and their respective uncertainties were obtained from the "GOOD" dihedrals generated by the TALOS+ program[10] based on the backbone chemical shift values.

[c] Statistics are calculated and averaged over an ensemble of the 15 lowest energy structures out of 75 calculated structures.

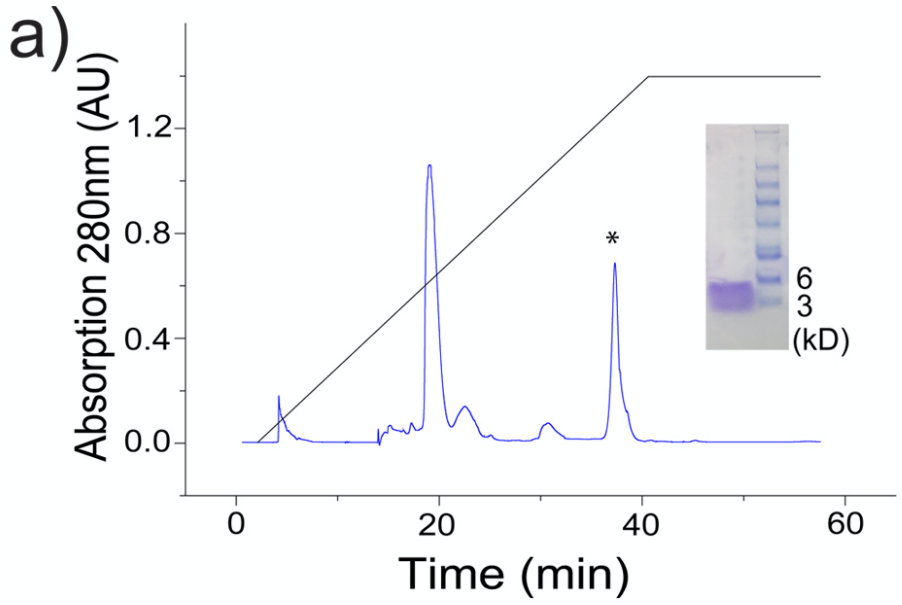
[d] The precision of the atomic coordinates is defined as the average r.m.s. difference between the 15 final structures and their mean coordinates. The atomic structure coordinate and structural constraints have been deposited in the Protein Data Bank (PDB), accession number 7LC8. The chemical shift values have been deposited in the Biological Magnetic Resonance Data Bank (BMRB), accession number 30842.

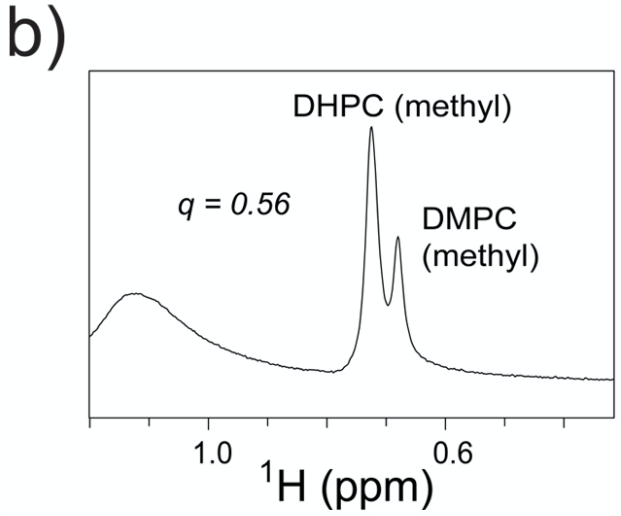
Table S3. Residue specific PRE_{amp} of S2 TMD in q = 0.6 bicelles

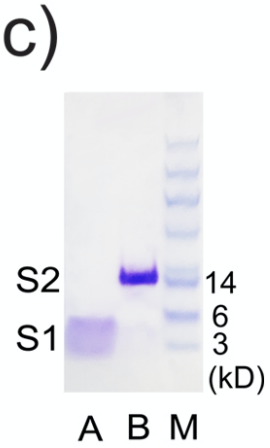
Residue	PRE_{amp} (Gd-DOTA)	PRE_{amp} (16-DSA)
1218	0.67±0.028	0.79±0.047
1219	0.61±0.034	0.84±0.044
1220	0.53±0.045	0.83±0.052
1221	0.56±0.036	0.84±0.051
1222	0.57±0.042	0.87±0.048
1223	0.58±0.037	0.84±0.036
1224	0.53±0.029	0.85±0.026
1225	0.59±0.044	0.88±0.042
1226	0.52±0.047	0.86±0.044
1227	0.54±0.038	0.89±0.043
1228	0.60±0.021	0.85±0.046
1229	0.57±0.027	0.82±0.049
1230	0.58±0.035	0.87±0.045
1231	0.60±0.042	0.82±0.044
1232	0.56±0.051	0.85±0.032
1233	0.60±0.046	0.84±0.055
1234	0.71±0.043	0.81±0.053
1235	0.73±0.056	0.78±0.052
1236	0.82±0.052	0.68±0.058
1237	0.89±0.067	0.58±0.062

Table S4. Membrane localization of S2 TMD trimer by PPT analysis

Residue (H^N)	r_z (Å)
L1218	-11.63689
G1219	-9.42969
F1220	-8.89779
I1221	-8.22165
A1222	-6.05935
G1223	-4.217
L1224	-3.97906
I1225	-3
A1226	-0.73905
(Membrane center)	0.0
I1227	0.90939
V1228	1.53238
L1229	2.89463
V1230	5.13462
T1231	6.3745
I1232	7.21047
L1233	9.06871
L1234	10.98703
S1235	11.6717
S1236	12.88747
T1237	15.15529


a)
Absorption 280nm (AU)
1.2
0.8
0.4
0.0
*
6
3
(kD)
0
20
40
60
Time (min)


b)
DHPC (methyl)
q = 0.56
DMPC
(methyl)
1.0
0.6
1H (ppm)


c)
S2
S1
14
6
3
(kD)
A B M

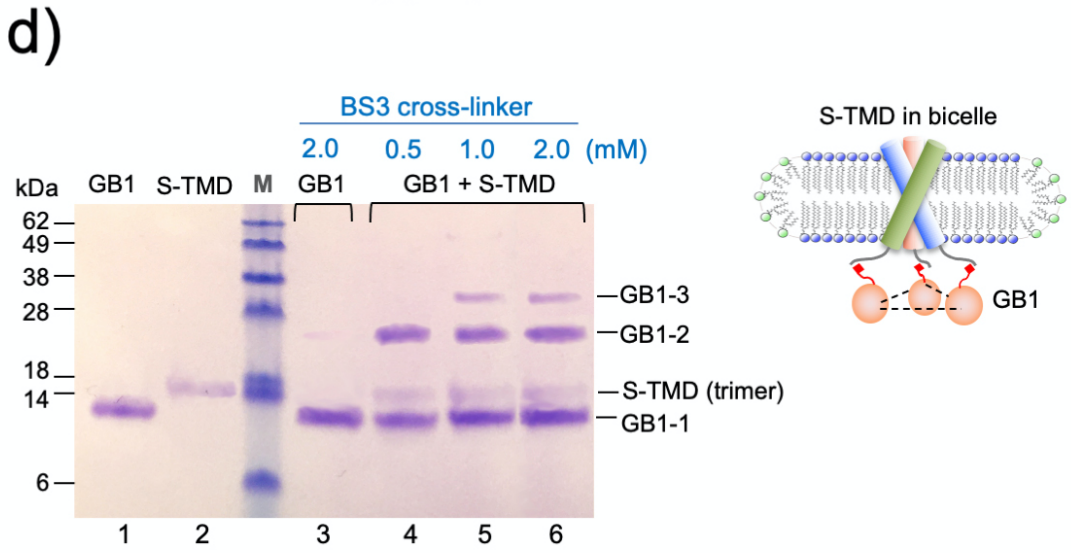

d)
BS3 cross-linker
2.0 0.5 1.0 2.0 (mM)
kDa
GB1
S-TMD
M
GB1
GB1 + S-TMD
62
49
38
28
18
14
6
GB1-3
GB1-2
S-TMD (trimer)
GB1-1
1 2 3 4 5 6
S-TMD in bicelle
GB1

Figure S1. Purification and bicelle reconstitution of SARS-CoV-2 $S2^{1209\text{-}1237}$

(**a**) Reverse phase HPLC purification of $S2^{1209\text{-}1237}$ from CNBr-cleaved trpLE-$S2^{1209\text{-}1237}$ fusion protein on Zorbax SB-C3 column with a gradient from 95% dH_2O, 5% isopropanol (IPA), 0.1% trifluoroacetic acid (TFA) (buffer A) to 75% IPA, 25% acetonitrile, 0.1% TFA (buffer B). The product was verified by SDS-PAGE and Mass Spectrometry.

(**b**) ^1H NMR spectrum of the reconstituted bicelle sample recorded at 600 MHz, showing the molar ratio of DMPC to DH_6PC to be ~0.55.

(**c**) Oligomerization of $S2^{1209\text{-}1237}$ in bicelles analyzed by SDS-PAGE. The gel lanes from left to right are: A – purified $S2^{1209\text{-}1237}$ powder without reconstitution; B – $S2^{1209\text{-}1237}$ reconstituted in DMPC-DH_6PC bicelles (q = 0.55); M – MW markers. Both samples were dissolved in gel loading buffer prior to SDS-PAGE. S1: Monomer band at ~5 kDa (theoretical MW = 3.4 kDa). S2: Trimer band at ~15 kDa (theoretical MW = 10.2 kDa).

(**d**) Oligomerization of His_6-tagged $S2^{1209\text{-}1237}$ in bicelles analyzed by the OG-label method[2] that uses the GB1 protein as reporter of $S2^{1209\text{-}1237}$ oligomerization. The SDS-PAGE gel lanes are: 1 – GB1-TriNTA; 2 – His_6-tagged $S2^{1209\text{-}1237}$ in DMPC-DH_6PC bicelles with q = 0.55 (trimeric in SDS-PAGE); 3 – GB1-TriNTA alone cross-linked sequentially with 2.0 mM BS3(PEG9) and 0.6 mM glutaraldehyde; 4-6 – GB1-TriNTA cross-linked as in lane 3 in the presence of His_6-tagged $S2^{1209\text{-}1237}$ at increasing BS3(PEG9) concentration.

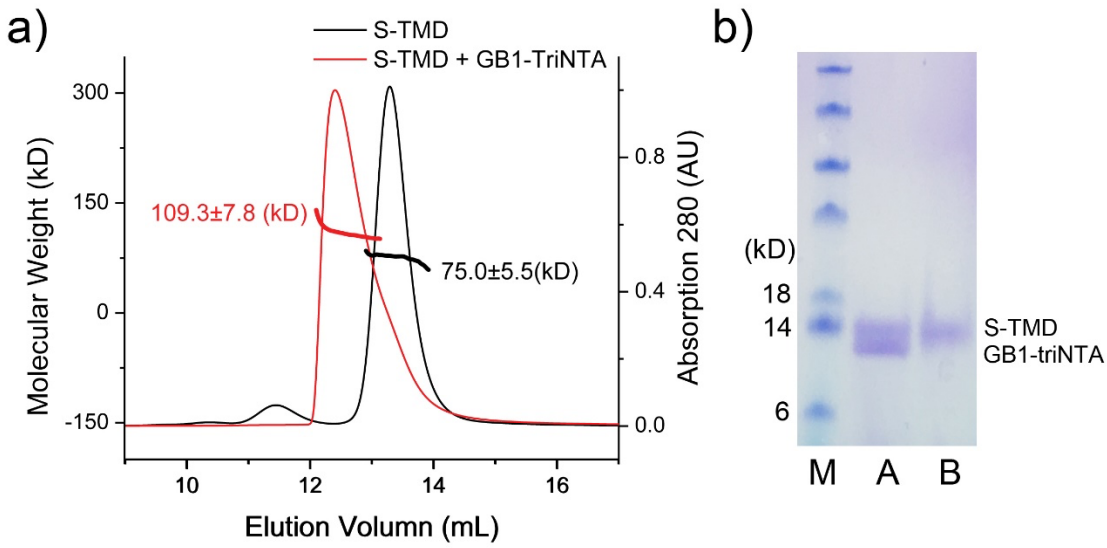


Figure S2. SEC-MALS analysis of the His_6-tagged $S2^{1209-1237}$ with and without GB1-TriNTA

(**a**) Elution profiles (UV 280nm absorption) for the His_6-tagged $S2^{1209-1237}$ alone (black) and 1:1.5 mixture of the His_6-tagged $S2^{1209-1237}$ and GB1-TriNTA (red). The elution peak for the His_6-tagged $S2^{1209-1237}$ alone in DDM has an average mass of 75.0±5.5 kD (TMD oligomer + DDM micelle). After mixing with GB1-TriNTA, the elution peak shifted significantly to the left, with an average mass of 109.3±7.8 kD. The difference of 34 kD agrees closely with three GB1-TriNTAs (~11kD each) recruited by the TMD oligomer, indicating that the $S2^{1209-1237}$ forms homogeneous trimer.

(**b**) SDS-PAGE analysis of protein components in SEC elution peaks in (a). The three lanes are: M – protein marker; A – His_6-tagged $S2^{1209-1237}$ + GB1-TriNTA; B – His_6-tagged $S2^{1209-1237}$ alone.

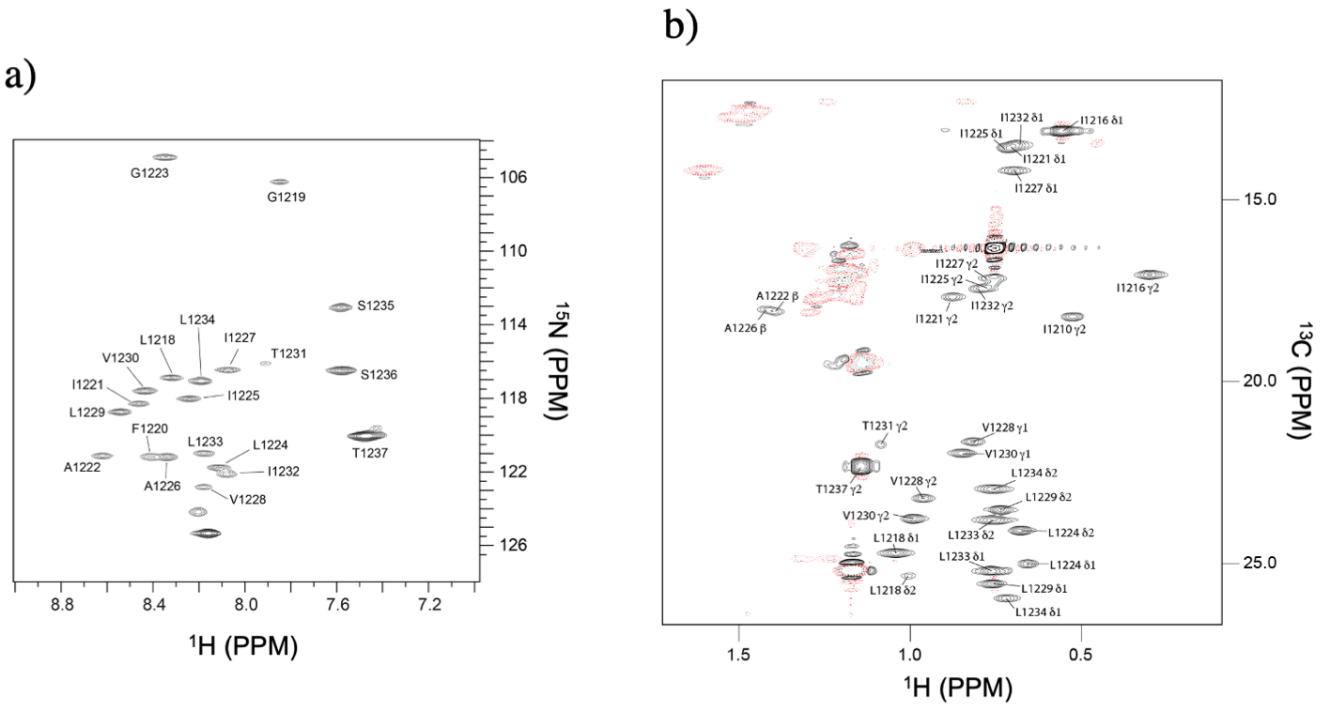
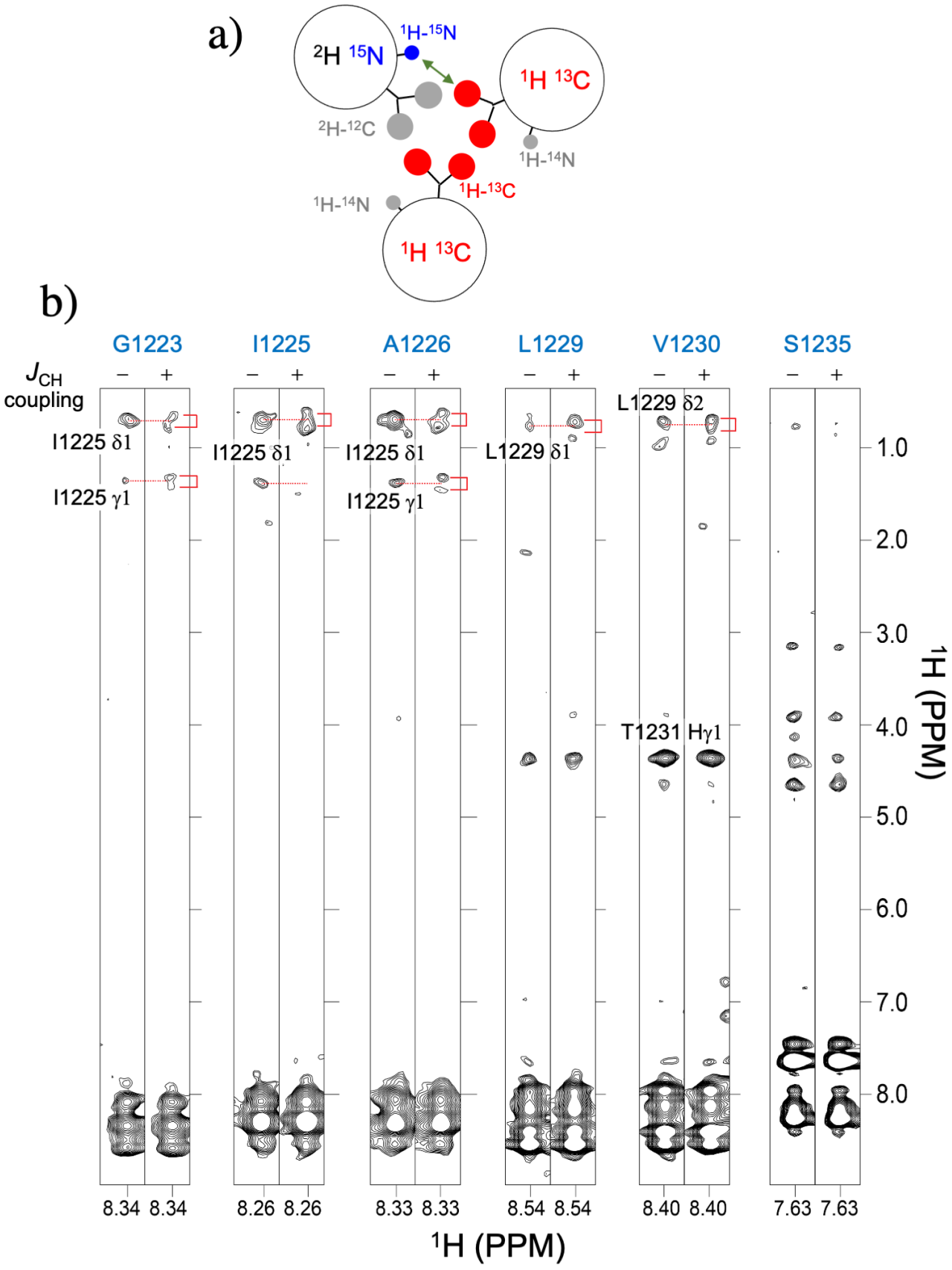


Figure S3. **NMR spectra of S2$^{1209\text{-}1237}$ reconstituted in DMPC-DH$_6$PC bicelles with *q* = 0.55**

(**a**) The ^{1}H-^{15}N TROSY-HSQC spectrum of (^{15}N,^{13}C, ^{2}H)-labeled S2$^{1209\text{-}1237}$, recorded at ^{1}H frequency of 600 MHz at 303 K.

(**b**) The methyl group region of the 2D ^{1}H-^{13}C HSQC spectrum (28 ms constant-time ^{13}C evolution), recorded at ^{1}H frequency of 700 MHz using (^{15}N, ^{13}C)-labeled S2$^{1209\text{-}1237}$ reconstituted in DMPC and DH$_6$PC with deuterated acyl chains.


a)
^{2}H ^{15}N
^{1}H-^{15}N
^{1}H ^{13}C
^{2}H-^{12}C
^{1}H-^{14}N
^{1}H-^{13}C
^{1}H-^{14}N
^{1}H ^{13}C
b)
G1223
I1225
A1226
L1229
V1230
S1235
J_{CH} coupling
− +
I1225 δ1
I1225 γ1
I1225 δ1
I1225 δ1
I1225 γ1
L1229 δ1
L1229 δ2
T1231 Hγ1
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
^{1}H (PPM)
8.34 8.34
8.26 8.26
8.33 8.33
8.54 8.54
8.40 8.40
7.63 7.63
^{1}H (PPM)

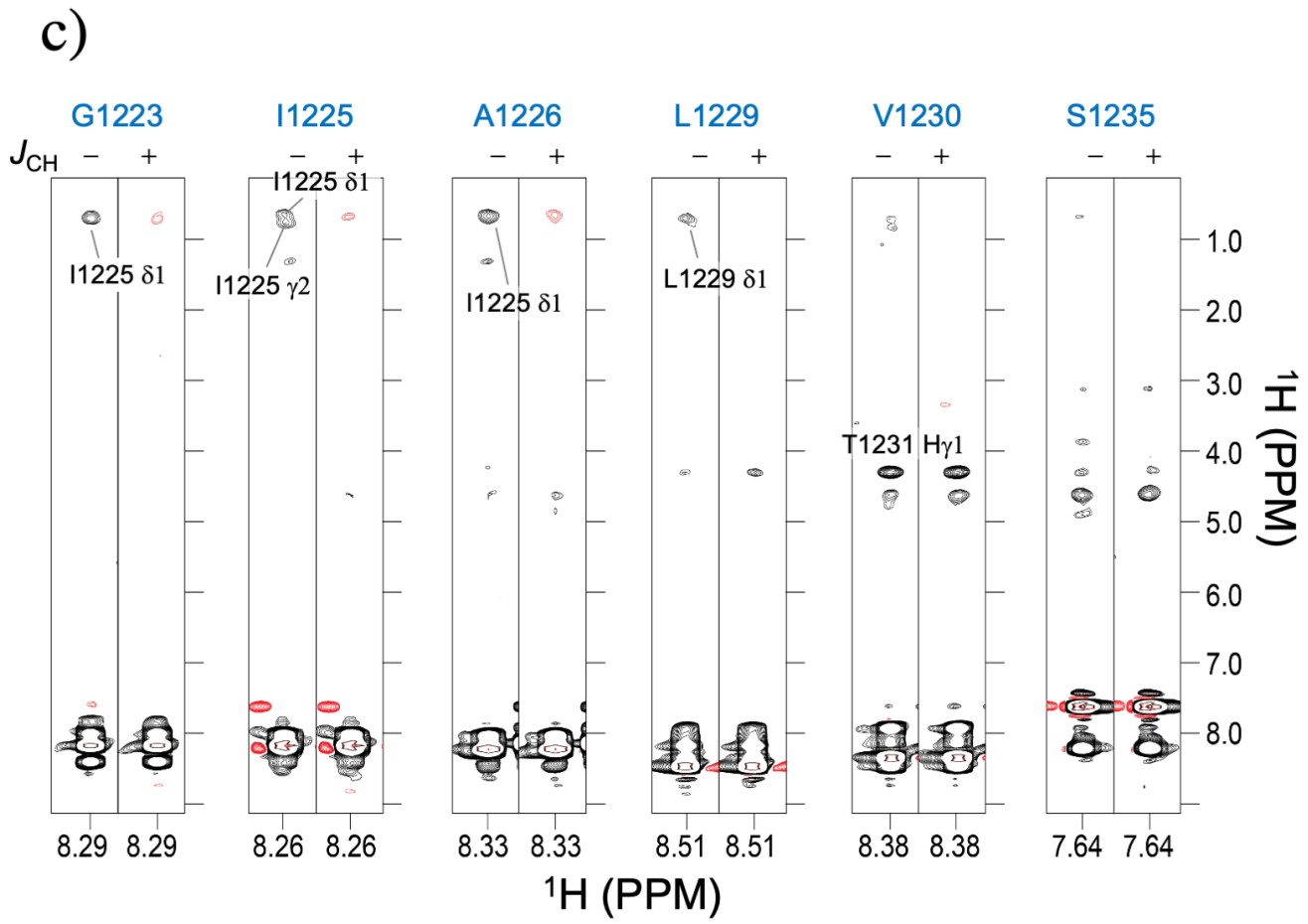


Figure S4. **Detection of inter-chain NOEs**

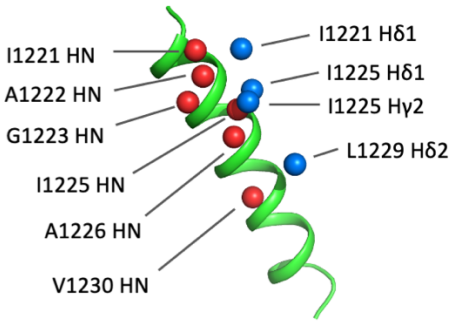
(**a**) Schematic illustration of the mixed sample comprising 1:1 ratio of (^{15}N, ^{2}H)-labeled S2$^{1209\text{-}1237}$ to ^{13}C-labeled S2$^{1209\text{-}1237}$. ^{1}H-^{15}N and ^{1}H-^{13}C groups are shown in blue and red, respectively, whereas the undetectable groups (e.g., ^{1}H-^{14}N and ^{2}H-^{12}C) are shown in gray.

(**b**) Residue-specific stripes from a 3D ^{15}N-edited NOESY-TROSY (τ_{NOE} = 150 ms), recorded in an interleaved fashion with and without ^{13}C decoupling during ^{1}H evolution before NOE mixing. The spectrum was recorded at 900 MHz and 303 K using the mixed sample in (a). For each of the selected residues, two stripes are shown: left – ^{13}C decoupled; right – ^{13}C coupled. The red brackets indicate peak splitting due to J_{CH}.

(**c**) Independent validation of inter-chain NOEs in (b) by the J_{CH}-modulated NOESY[1]. Residue-specific stripes from the 3D J_{CH}-modulated NOESY (NOE mixing time = 150 ms) recorded at 900 MHz and 303 K. For each of the selected residues, two stripes are shown: left – J_{CH} = 0, inter- and intra- chain NOE peaks are both positive (black); right – J_{CH} = 8 ms, inter-chain NOE peaks are negative (red). Note that the negative NOE peaks are significantly weaker due to cross correlation effects resulting from C-H dipolar interaction.

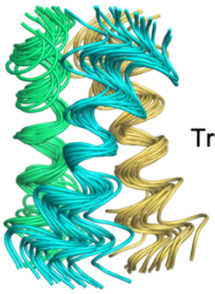
a

Amide	Aliphatic	dist (Å)
I1221 HN	I1221 Hδ1	4.0 ± 1
A1222 HN	I1221 Hδ1	4.0 ± 1
G1223 HN	I1225 Hδ1	4.0 ± 1
G1223 HN	I1225 Hγ2	4.0 ± 1
I1225 HN	I1225 Hδ1	4.0 ± 1
I1225 HN	I1225 Hγ2	4.0 ± 1
A1226 HN	I1225 Hδ1	4.0 ± 1
V1230 HN	L1229 Hδ2	4.0 ± 1



ExSSO

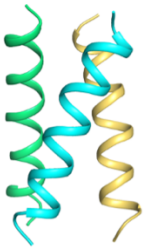
b



Convergent Trimer Solutions

Representative Solution

c



Self Consistent NOEs

XPLOR Refinement

More Inter-Chain NOEs

d

Proton 1	Proton 2	Protomer	dist (Å)
L1218 Hδ1	I1221 Hδ1	i, i - 1	2.5 ± 1
L1218 Hδ1	F1220 Hε	i, i - 1	3.0 ± 1
I1221 Hδ1	L1218 Hδ1	i, i + 1	2.5 ± 1
I1225 Hδ1	A1222 Hβ	i, i + 1	3.0 ± 1
I1225 Hγ2	A1226 Hα	i, i + 1	2.5 ± 1
I1227 HN	L1229 Hδ2	i, i - 1	4.0 ± 1
V1228 Hγ1	A1226 Hβ	i, i + 1	3.0 ± 1
L1229 Hδ1	A1226 Hβ	i, i + 1	2.5 ± 1
L1229 Hδ2	A1226 Hβ	i, i + 1	2.5 ± 1
L1229 Hδ1	I1227 HN	i, i + 1	4.0 ± 1
L1229 Hδ2	I1227 Hα	i, i + 1	4.0 ± 1
V1230 HN	I1232 Hγ1	i, i - 1	4.0 ± 1

e

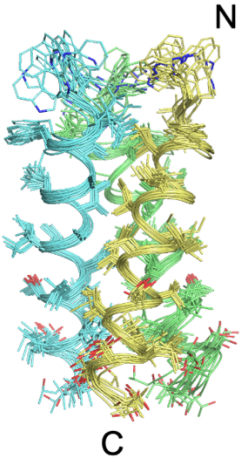


Figure S5. Overall procedure of structure determination by NMR.

(**a**) *Left*: Initial set of inter-chain NOE restraints identified in the NOE spectra of mixed isotope labeled sample (Fig. S4). *Right*: Protons associated with these NOEs are indicated by spheres in the context of the monomer structure generated with local NOE restraints and backbone dihedral restraints from TALOS+[10].

(**b**) The monomer structure and inter-chain NOE restraints (with ambiguous directionality) were subject to an exhaustive conformational search algorithm (ExSSO[11]), which found an ensemble of convergent solutions of trimeric assembly.

(**c**) A representative trimer structure from (b) was fed to the XPLOR-NIH[9] for iterative refinement against all NMR restraints, while assigning more self-consistent NOEs after each iteration.

(**d**) Additional list of inter-chain NOEs assigned during the iterative refinement step in (c) (also see Fig. S10 below).

(**e**) Ensemble of 15 lowest energy structures from 75 structures calculated in the final refinement step in (c). Protons are not displayed. The region displayed include residues 1217-1237. Residues 1209-1216 is unstructured in our sample and essentially no NMR data was collected for this region.

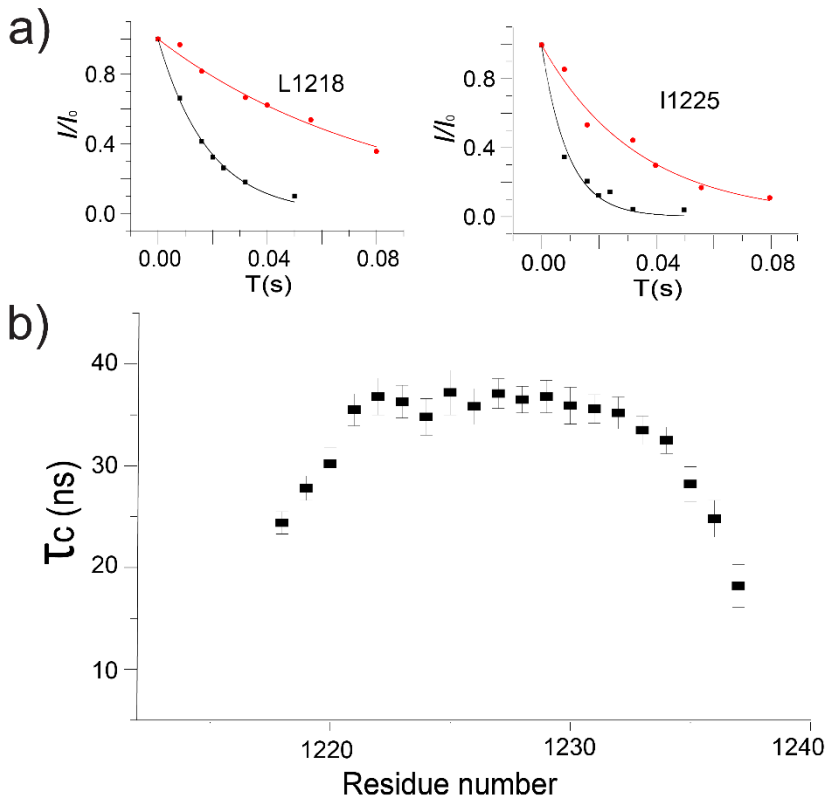


Figure S6. Measurement of residue-specific τ_C for the S2$^{1209-1237}$ in bicelles (q = 0.55) using the TRACT experiment. The TRACT experiment[13] was performed in 2D ^{1}H-^{15}N correlation mode to resolve resonance overlap. The data was collected at 600 MHz and 303K.

(**a**) Example plots showing decay of relative peak intensity, I/I_0, due to ^{15}N transverse relaxation for the TROSY (red) and anti-TROSY (black) components. The TROSY relaxation rate (R_α) and the anti-TROSY relaxation rate (R_β) were obtained by exponential fitting. τ_C was calculated using the difference R_β - R_α as described in Ref [13].

(**b**) Residues-specific τ_C calculated for the S2$^{1209-1237}$ in bicelles.

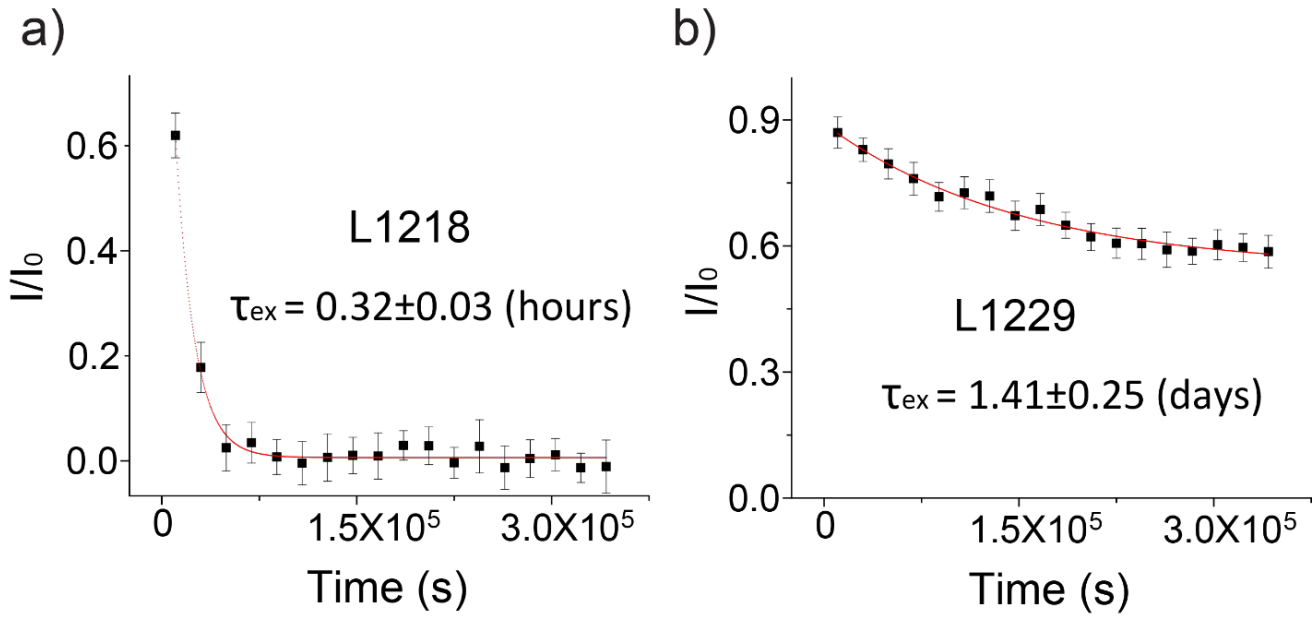


Figure S7. H-D exchange of the S2$^{1209\text{-}1237}$ in bicelles with *q* = 0.55.

(**a**) Signal decay over time for L1218, representing the high exchange regime (the N-terminal end of the TM helix with higher solvent accessibility).

(**b**) Signal decay over time for L1229, representing the low exchange regime (the hydrophobic core).

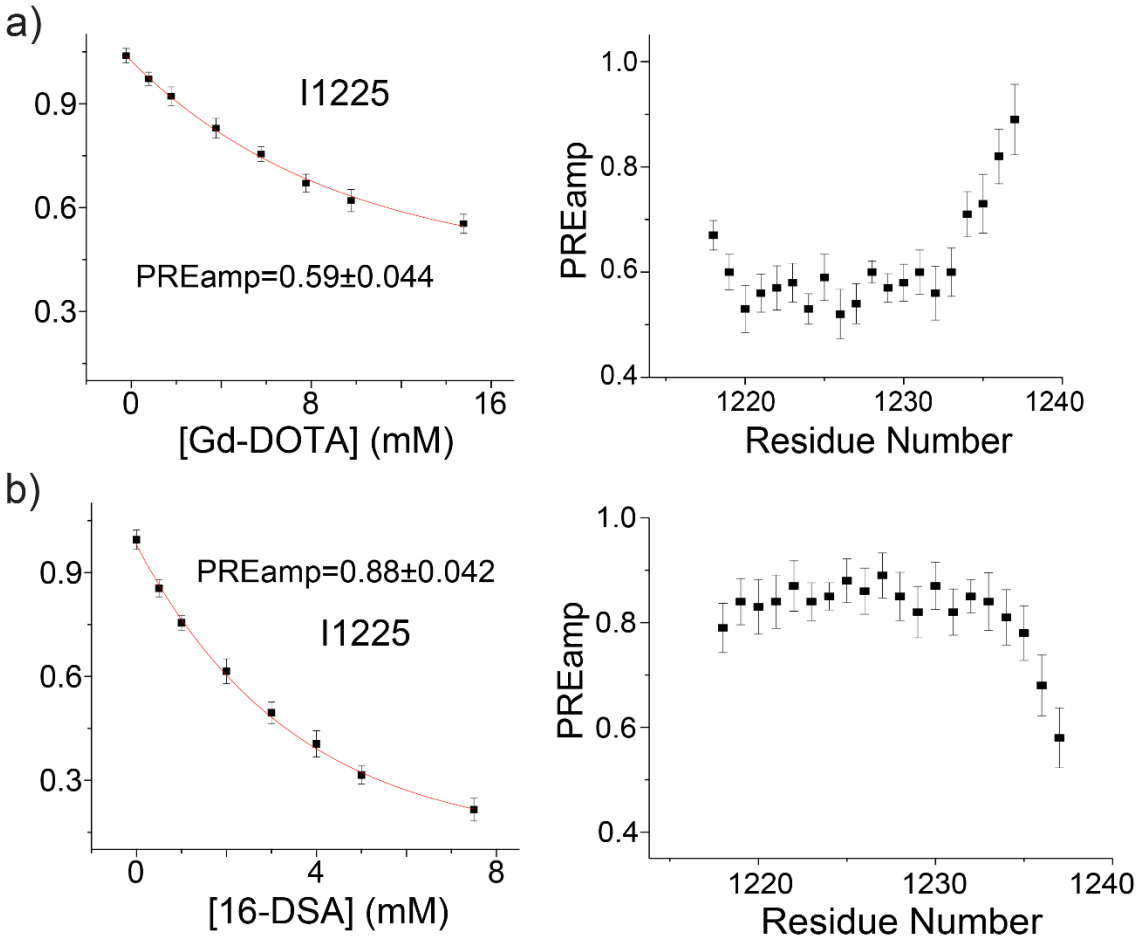


Figure S8. Membrane partition analysis of the S2$^{1209\text{-}1237}$ trimer

(**a**) Solvent PRE titration with Gd-DOTA. ^{1}H-^{15}N TROSY-HSQC signal decay vs. Gd-DOTA concentration were fitted to Eq. S1 to determine the residue-specific PRE_{amp} as shown for residue I1225 (left) and all residues (right).

(**b**) Lipophilic PRE titration with 16-DSA was performed to provide a reciprocal profile of PRE_{amp} vs. (residue number), as shown for I1225 (left) and for all residues (right).

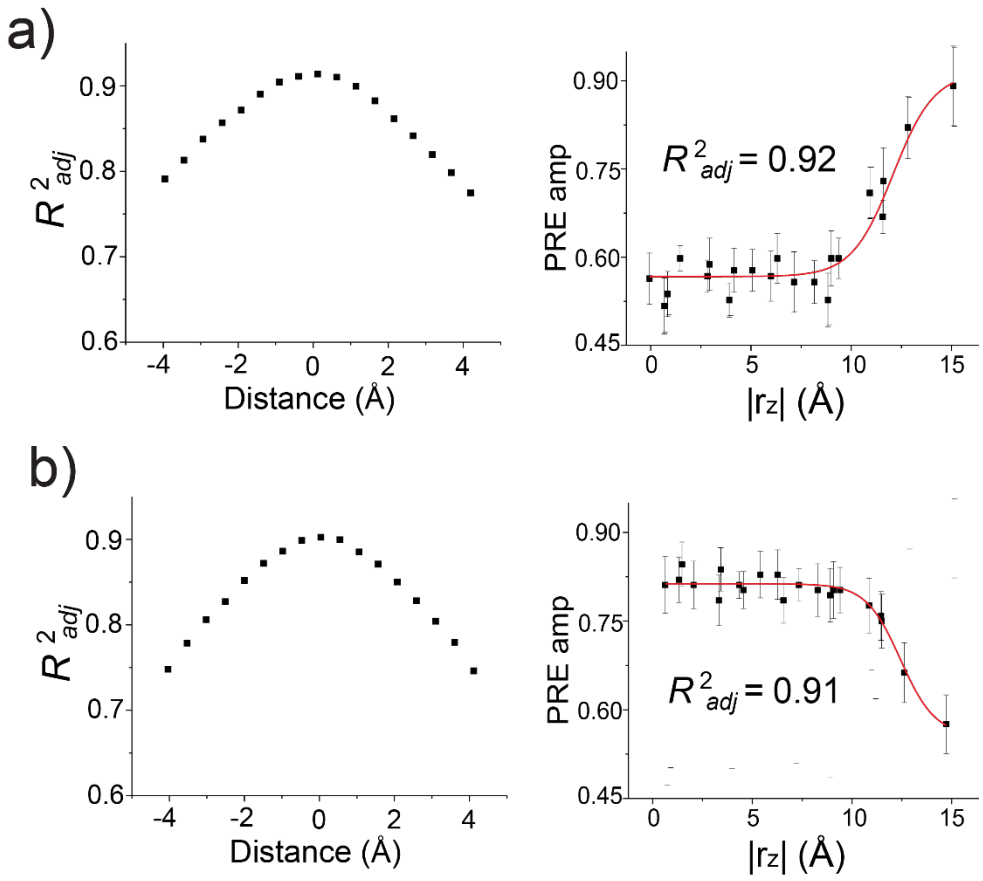


Figure S9. Assignment of the bilayer center to the S2 TMD trimer using the paramagnetic probe titration (PPT) method

(**a**) *Left*: Sliding of the S2 TMD trimer structure along the 3-fold axis (or bilayer normal) to yield best fit of Gd-DOTA generated PRE_{amp} to the symmetric sigmoidal function (Eq. S2). *Right*: The PRE_{amp} vs. r_Z profile from the best fit, showing the adjusted coefficient of determination (R^2_{adj}).

(**b**) Same as in (a), but for the 16-DSA data set.

The plots show that R^2_{adj} is a reliable indicator of the protein position with an error of about ± 0.5 Å. Both fittings show very close bilayer center position, which corresponds approximately to A1226/I1227.

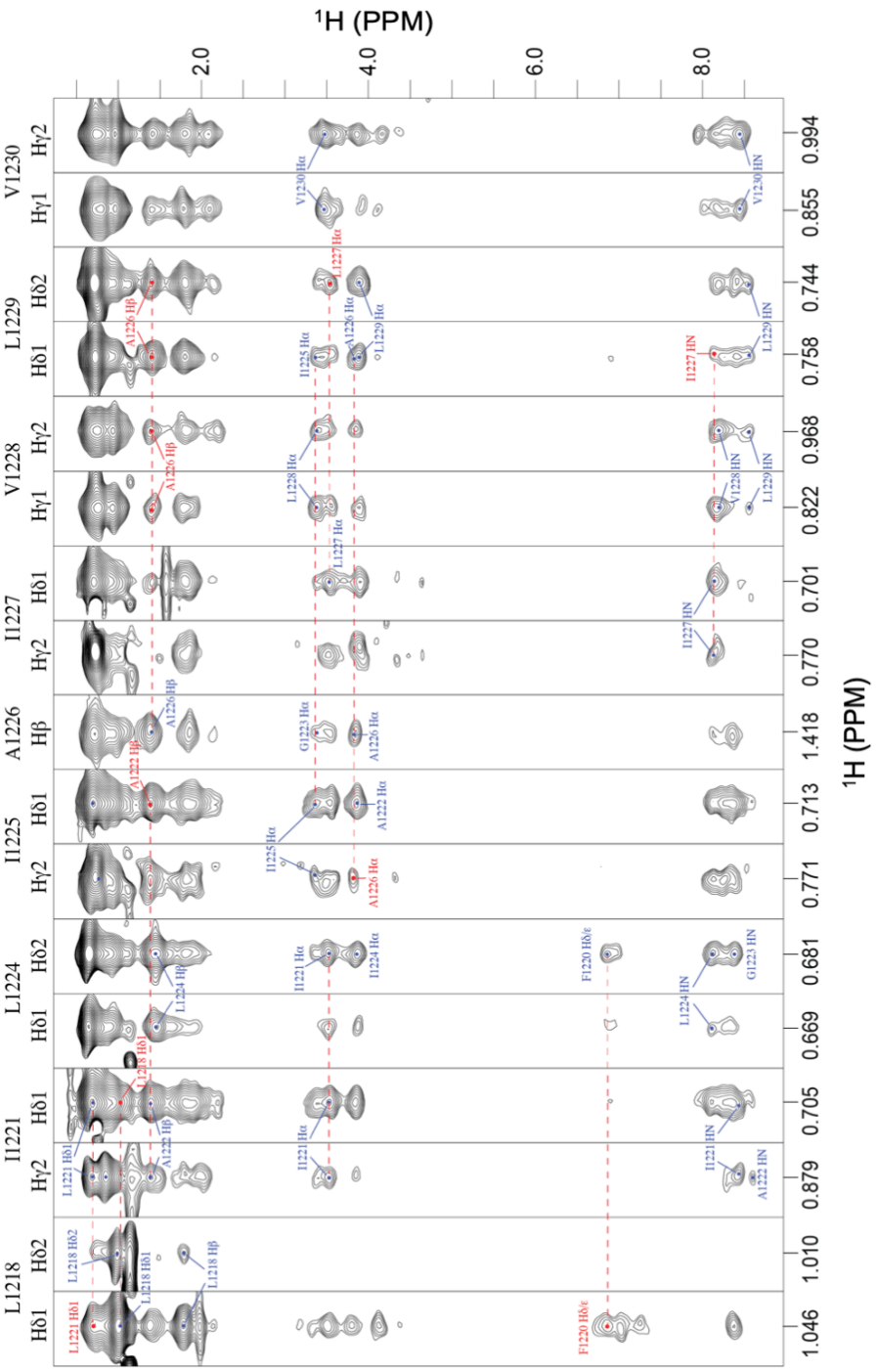


Figure S10. Additional inter-chain NOEs consistent with the trimer solution in Fig. S5b. Example stripe from 3D ^{13}C-edited NOESY-HSQC spectrum recorded at 700 MHz (^{1}H frequency) and 303 K with a NOE mixing time of 100 ms. The red dashed lines indicate matching chemical shift between intra-monomer (blue) and inter-monomer (red) NOEs. The spectrum was acquired using (^{15}N,^{13}C)-labeled S2$^{1209\text{-}1237}$ reconstituted in bicelles with deuterated DMPC and DH_6PC acyl chains.

1. Fu, Q.; Piai, A.; Chen, W.; Xia, K.; Chou, J. J., Structure determination protocol for transmembrane domain oligomers. *Nat Protoc* **2019,** *14* (8), 2483-2520.
2. Chen, W.; Dev, J.; Mezhyrova, J.; Pan, L.; Piai, A.; Chou, J. J., The Unusual Transmembrane Partition of the Hexameric Channel of the Hepatitis C Virus. *Structure* **2018,** *26* (4), 627-634 e4.
3. Piai, A.; Fu, Q.; Cai, Y.; Ghantous, F.; Xiao, T.; Shaik, M. M.; Peng, H.; Rits-Volloch, S.; Chen, W.; Seaman, M. S.; Chen, B.; Chou, J. J., Structural basis of transmembrane coupling of the HIV-1 envelope glycoprotein. *Nat Commun* **2020,** *11* (1), 2317.
4. Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A., NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR* **1995,** *6* (3), 277-293.
5. Bartels, C.; Xia, T. H.; Billeter, M.; Guntert, P.; Wuthrich, K., The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. *J Biomol NMR* **1995,** *6* (1), 1-10.
6. Kay, L. E.; Ikura, M.; Tschudin, R.; Bax, A., Three-dimensional triple-resonance NMR Spectroscopy of isotopically enriched proteins. 1990. *J Magn Reson* **1990,** *213* (2), 423-41.
7. Salzmann, M.; Wider, G.; Pervushin, K.; Senn, H.; Wuthrich, K., TROSY-type triple-resonance experiments for sequential NMR assignments of large proteins. *Journal of the American Chemical Society* **1999,** *121* (4), 844-848.
8. Szyperski, T.; Neri, D.; Leiting, B.; Otting, G.; Wuthrich, K., Support of 1H NMR assignments in proteins by biosynthetically directed fractional 13C-labeling. *J. Biomol. NMR* **1992,** *2* (4), 323-334.
9. Schwieters, C. D.; Kuszewski, J.; Tjandra, N.; Clore, G. M., The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **2002,** *160*, 66-74.
10. Shen, Y.; Delaglio, F.; Cornilescu, G.; Bax, A., TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *Journal of Biomolecular NMR* **2009,** *44* (4), 213-23.
11. Yang, J.; Piai, A.; Shen, H. B.; Chou, J. J., An Exhaustive Search Algorithm to Aid NMR-Based Structure Determination of Rotationally Symmetric Transmembrane Oligomers. *Sci Rep* **2017,** *7* (1), 17373.
12. Piai, A.; Fu, Q.; Dev, J.; Chou, J. J., Optimal Bicelle Size q for Solution NMR Studies of the Protein Transmembrane Partition. *Chemistry* **2017,** *23* (6), 1361-1367.
13. Lee, D.; Hilty, C.; Wider, G.; Wuthrich, K., Effective rotational correlation times of proteins from NMR relaxation interference. *J Magn Reson* **2006,** *178* (1), 72-6.
14. Xia, S.; Liu, M.; Wang, C.; Xu, W.; Lan, Q.; Feng, S.; Qi, F.; Bao, L.; Du, L.; Liu, S.; Qin, C.; Sun, F.; Shi, Z.; Zhu, Y.; Jiang, S.; Lu, L., Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion. *Cell Res* **2020,** *30* (4), 343-355.
15. 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* (6483), 1260-1263.
16. Walls, A. C.; Park, Y. J.; Tortorici, M. A.; Wall, A.; McGuire, A. T.; Veesler, D., Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **2020,** *181* (2), 281-292 e6.
17. Henderson, R.; Edwards, R. J.; Mansouri, K.; Janowska, K.; Stalls, V.; Gobeil, S. M. C.; Kopp, M.; Li, D.; Parks, R.; Hsu, A. L.; Borgnia, M. J.; Haynes, B. F.; Acharya, P., Controlling the SARS-CoV-2 spike glycoprotein conformation. *Nat Struct Mol Biol* **2020,** *27* (10), 925-933.
18. McCallum, M.; Walls, A. C.; Bowen, J. E.; Corti, D.; Veesler, D., Structure-guided covalent stabilization of coronavirus spike glycoprotein trimers in the closed conformation. *Nat Struct Mol Biol* **2020,** *27* (10), 942-949.
19. Cai, Y.; Zhang, J.; Xiao, T.; Peng, H.; Sterling, S. M.; Walsh, R. M., Jr.; Rawson, S.; Rits-Volloch, S.; Chen, B., Distinct conformational states of SARS-CoV-2 spike protein. *Science* **2020,** *369* (6511), 1586-1592.
20. Yurkovetskiy, L.; Wang, X.; Pascal, K. E.; Tomkins-Tinch, C.; Nyalile, T. P.; Wang, Y.; Baum, A.; Diehl, W. E.; Dauphin, A.; Carbone, C.; Veinotte, K.; Egri, S. B.; Schaffner, S. F.; Lemieux, J. E.; Munro, J. B.; Rafique, A.;

Barve, A.; Sabeti, P. C.; Kyratsous, C. A.; Dudkina, N. V.; Shen, K.; Luban, J., Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant. *Cell* **2020,** *183* (3), 739-751 e8.
21. Wrobel, A. G.; Benton, D. J.; Xu, P.; Roustan, C.; Martin, S. R.; Rosenthal, P. B.; Skehel, J. J.; Gamblin, S. J., SARS-CoV-2 and bat RaTG13 spike glycoprotein structures inform on virus evolution and furin-cleavage effects. *Nat Struct Mol Biol* **2020,** *27* (8), 763-767.
22. Xiong, X.; Qu, K.; Ciazynska, K. A.; Hosmillo, M.; Carter, A. P.; Ebrahimi, S.; Ke, Z.; Scheres, S. H. W.; Bergamaschi, L.; Grice, G. L.; Zhang, Y.; Collaboration, C.-N. C.-B.; Nathan, J. A.; Baker, S.; James, L. C.; Baxendale, H. E.; Goodfellow, I.; Doffinger, R.; Briggs, J. A. G., A thermostable, closed SARS-CoV-2 spike protein trimer. *Nat Struct Mol Biol* **2020,** *27* (10), 934-941.
23. Ke, Z.; Oton, J.; Qu, K.; Cortese, M.; Zila, V.; McKeane, L.; Nakane, T.; Zivanov, J.; Neufeldt, C. J.; Cerikan, B.; Lu, J. M.; Peukes, J.; Xiong, X.; Krausslich, H. G.; Scheres, S. H. W.; Bartenschlager, R.; Briggs, J. A. G., Structures and distributions of SARS-CoV-2 spike proteins on intact virions. *Nature* **2020,** *588* (7838), 498-502.
24. Bangaru, S.; Ozorowski, G.; Turner, H. L.; Antanasijevic, A.; Huang, D.; Wang, X.; Torres, J. L.; Diedrich, J. K.; Tian, J. H.; Portnoff, A. D.; Patel, N.; Massare, M. J.; Yates, J. R., 3rd; Nemazee, D.; Paulson, J. C.; Glenn, G.; Smith, G.; Ward, A. B., Structural analysis of full-length SARS-CoV-2 spike protein from an advanced vaccine candidate. *Science* **2020,** *370* (6520), 1089-1094.