The sequence and structural integrity of the SARS-CoV-2 Spike protein transmembrane domain is crucial for viral entry

Ortiz-Mateu J[1], Belda D[1], Avilés-Alía AI[2], Alonso-Romero J[1], García-Murria MJ[1], Mingarro I[1], Geller R[2], Martinez-Gil L[1].

[1]Departament de Bioquímica i Biologia Molecular, Institut Universitari de Biotecnologia i Biomedicina (BIOTECMED), Universitat de València, Burjassot, Spain.

[2] Institute for Integrative Systems Biology (I2SysBio), UV-CSIC, Paterna, Spain.

Abstract

The Spike protein of SARS-CoV-2 is a type I membrane protein that mediates target cell recognition and membrane fusion. Its transmembrane domain (TMD) is usually considered to be a passive physical anchor that attaches the protein to the viral envelope. However, current knowledge of the dynamics and biogenesis of integral membrane proteins indicates that TMDs generally play key functions beyond membrane anchoring. Here, we explored the role of the SARS-CoV-2 Spike protein TMD during viral entry. To do so, we generated a series of amino acid substitutions and insertions within the hydrophobic core of the TMD and analyzed their impact on protein function. Our results indicate that the SARS-CoV-2 Spike protein is highly susceptible to changes in its TMD. Furthermore, the TMD sequence and structural parameters relevant for viral entry are distributed across the entire segment, albeit with a stronger implication of the N-terminal portion. The relative orientation of the regions before and after the TMD was also found to be relevant for particle entry. Finally, our results suggest that the SARS-CoV-2 Spike protein TMD can form homo-oligomers through a motif in which small residues are necessary.

Introduction

It is becoming increasingly clear that a deep understanding of the molecular mechanisms occurring during viral infection is required for an effective rational design of antiviral agents and vaccine immunogens. In the case of SARS-CoV-2, special attention has been given to the entry process and the role that the Spike (S) protein plays in it[1]. The SARS-CoV-2 S protein is a trimeric class I fusion glycoprotein that mediates target cell recognition and membrane fusion[2,3,4]. During post-translational maturation, the full-length protein is cleaved into two subunits, S1 and S2. S1 contains the receptor-binding domain (RBD) and facilitates attachment to the angiotensin-converting enzyme 2 (ACE2) on the host cell surface. S2 drives the fusion of the viral and host membranes and encompasses the fusion peptide, two heptad repeat regions, and a C-terminal transmembrane domain (TMD).

Often, the TMD is seen as a passive physical anchor whose sole function is to attach the protein to the viral envelope. However, current knowledge of the structure, dynamics, and biogenesis of integral membrane proteins demonstrates that TMDs, alongside membrane-proximal regions, play key functions beyond membrane anchoring[5–7]. Furthermore, in the recent years, significant advances in understanding the role of the TMDs of viral fusion proteins have been made highlighting their critical role in the fusion process, and also demonstrating their involvement in several aspects of the viral lifecycle[8,9].

Broer et al. demonstrated that substituting the N-terminal membrane proximal region, the TMD, and the endodomain of the SARS-CoV-1 S protein with the corresponding domains of the vesicular stomatitis virus (VSV) Glycoprotein greatly affect the protein's function[10]. A few years later, a mutagenesis-based study, once again using VSV pseudo-particles, revealed that substituting Gly_{1201}, Ala_{1204}, Val_{1210}, and Leu_{1216} by Lys residues in the core hydrophobic region of the SARS-

CoV-1 S protein resulted in reduced infectivity[11] likely by altering the hydrophobic nature of the TMD.

The SARS-CoV-2 S protein TMD has been linked with trimerization. A chimera including the RBD followed by the TMD was identified as a trimer in which the RBD was properly folded, suggesting that the TMD acts as an oligomerization domain[12]. However, there is some controversy regarding the oligomerization motif within the TMD. Abery et al. identified a GxxxG motif (a sequence known to promote oligomerization of TMDs[13]) within the TMD of the SARS-CoV-1 S protein and demonstrated its importance in the trimerization of the protein[14]. Shortly after, Corver et al.'s results indicated that these residues were not important for the trimerization of the protein or for viral entry[15]. Recently, it has been suggested that a new motif based on a hydrophobic zipper mediates the trimerization of SARS-CoV-2 S protein TMD[16].

Several structures of the S protein have been solved. However, only one includes the TMD[17]. In this structure, the S protein was found in a post-fusion state where the fusion peptide (FP) forms a hairpin-like wedge spanning the entire lipid bilayer and the TMD wraps around it. Two contact points between the FP and the TMD were described involving Phe_{1220} and Leu_{1234} on the TMD.

Here, we explore the role of the SARS-CoV-2 S protein TMD during viral entry. To do so we generate a series of variations within the hydrophobic core of the TMD and analyze their impact on the protein function. Furthermore, we investigate the potential of the TMD to induce oligomerization as well as the structural and sequence parameters responsible for it. Our results demostrate that the SARS-CoV-2 S protein TMD plays a key role in the protein function beyond membrane anchoring. The sequence and structural parameters of the TMD relevant for viral entry are distributed across the entire segment, which might explain its strong sequence conservation and

the literature controversy. Additionally, our results suggest that the relative orientation of the regions before and after the TMD is important for particle entry. Finally, we identify that the SARS-CoV-2 S protein TMD is capable of forming homo-oligomers and these occur through a motif in which small residues such as glycine and alanine are relevant, results that were corroborated *in silico*.

Results and Discussion

The SARS-CoV-2 S protein contains a single TMD in the C-terminal region of the protein (Fig. 1A). It spans 23 amino acids from Trp_{1214} to Cys_{1236} with a predicted ΔG_{app} of -3,4 kcal/mol for its insertion into the membrane, according to the *ΔG Prediction Server*[18,19]. The predicted TMD includes an aromatic-rich region (Trp_{1214}-Tyr-Ile-Trp_{1217}) in its N-terminal end and a Cys pair (Cys_{1235}-Cys_{1236}) in its C-terminal end. However, the structure of the S trimer in its post-fusion state[17] shows a helical secondary structure from Leu_{1218} to Leu_{1234}. Although the aromatic region and the C-terminal Cys-rich section on the TMD most likely interact with the membrane either as an interfacial membrane proximal region[17,20] or through palmitoyl post-translational modifications[21–24], in the present study we focus exclusively on the helical hydrophobic core of the TMD, that is, from Leu_{1218} to Leu_{1234}. The sequence of all the variants and mutations included in this paper together with their insertion potential can be found in Fig.1B. Fig 1C includes a structural model and a helical wheel projection of the SARS-CoV-2 S protein TMD hydrophobic core.

Figure 1. Schematic representation of the SARS-CoV-2 spike protein and its TMD. **A**. Schematic representation of SARS-CoV-2 Spike protein domain architecture. The position of key domains is highlighted: NTD, N-terminal domain; RBD, Receptor Binding Domain; CTD1, C-terminal domain 1; CTD2, C-terminal domain 2; S1/S2, S1/S2 cleavage site; S2′, S2′ cleavage site; FP, fusion peptide; HR1, heptad repeat 1; HR2, heptad repeat 2; TM, transmembrane domain; CT, cytoplasmic tail. **B**. Sequence of the SARS-CoV-2 S protein TMD and the variants included in this study. The sequence of the SARS-CoV-2 S protein from Glu_{1206} to Thr_{1273} is shown. Membrane associated regions, including

an aromatic rich section and a cysteine rich domain, are highlighted with light letters on a gray background. The TMD hydrophobic core is shown with dark letters on a pink background. The TMD, as predicted by the ΔG Prediction Server (WT), and the variations within included in this work are highlighted. A heat-map (right of the sequence) shows the predicted ΔG_{app} for each TMD variant. **C**. Structural model of the SARS-CoV-2 S protein TMD. AlphaFold3 model of residues 1218-1234 (top) and helical wheel projection (bottom) of the same sequence. The surface of the helix containing small side chain residues is highlighted in orange, while the opposite surface containing highly hydrophobic residues is highlighted in purple.

To study the role of the SARS-CoV-2 S protein TMD during viral entry we utilized pseudotyped vesicular stomatitis virus (VSV) particles coated with different S protein variants (Fig. 2A)[10,15,25,26]. Briefly, HEK293T cells were transfected with plasmids encoding the different S variants and subsequently infected with a recombinant VSV in which the glycoprotein gene has been replaced with GFP (VSVΔG-GFP). Next, the infectivity of the resulting particles was assayed on VeroE6 cells expressing the TMPRSS2 cofactor by counting GFP-positive cells. VSVΔG-GFP particles were pseudotyped with the native SARS-CoV-2 Wuhan-Hu-1 S protein as a positive control and reference value (WT). No infection was observed when the TMD of the SARS-CoV-2 S protein was eliminated (ΔTMD) or was substituted by a scrambled version of the TMD (SCRBL) where the composition of the residues included in the TMD was preserved but their position was randomly varied, confirming the importance of the SARS-CoV-2 S protein TMD in the entry process and indicating a role beyond membrane anchoring (Fig 2B). Expression of the chimeras included in this work can be found in Supp Fig 1.

Figure 2. VSVΔG-GFP pseudotyped virus assay. **A**. Generation of VSVΔG-GFP pseudotyped particles. HEK293T cells were transfected with the corresponding SARS-CoV-2 S protein variant and infected with a VSV lacking the G gene and expressing GFP (VSVΔG-GFP). The resulting virus, incorporates the S variants on its surface (VSVΔG-GFP -S), was recovered from the media and used to infect Vero-TMPRSS2 cells. Viral infectivity was then analyzed by counting GFP-positive cells. **B-E**. Focus Forming Units (FFU) measured after infection with VSVΔG-GFP

pseudotyped with different variants of the SARS-CoV-2 S protein TMD. All samples were normalized to the wild type SARS-CoV-2 S protein (WT). The significative p-values (<0.05) for individual one-sample t-tests (vs WT) are indicated above each bar. **B**. FFU from VSVΔG-GFP particles decorated with the wild type SARS-CoV-2 S protein (WT), the SARS-CoV-2 S protein where TMD was eliminated (ΔTMD), and a particles bearing a scrambled version of the SARS-CoV-2 S protein TMD (SCRBL). **C**. FFU from pseudotyped VSVΔG-GFP with the SARS-CoV-2 S protein with substitutions of amino acid stretches, Phe_{1220} to Gly_{1223}, Leu_{1224} to Ile_{1227}, and Val_{1228} to Thr_{1231}, with alanine (Phe_{1220}-Gly_{1223}A, Leu_{1224}-Ile_{1227}A, and Val_{1228}-Thr_{1231}A respectively). **D**. FFU from pseudotyped VSVΔG-GFP with the SARS-CoV-2 S protein including alanine insertions at positions 1221, 1226, 1228, 1230, and 1232 ($InsA_{1221}$, $InsA_{1226}$, $InsA_{1228}$, $InsA_{1230}$, and $InsA_{1232}$). **E**. FFU from pseudotyped VSVΔG-GFP with the SARS-CoV-2 S protein bearing the following point mutations Gly_{1223} to Ile, Gly_{1219} to Ile and Gly_{1223} to Ile, Ile_{1221} to Tyr, Ile_{1225} to Tyr, Met_{1229} to Tyr, Ala_{1222} to Ile and Gly_{1223} to Ile, Ala_{1222} to Tyr and Gly_{1223} to Tyr, Leu_{1224} to Tyr, Leu_{1224} to Tyr and Val_{1228} to Tyr (G_{1223}I, G_{1219}I G_{1223}I, I_{1221}Y, I_{1225}Y, M_{1229}Y, A_{1222}I G_{1223}I, A_{1222}Y G_{1223}Y, L_{1224}Y, L_{1224}Y V_{1228}Y, respectively).

Supplementary Figure 1. Expression levels of SARS-CoV-2 S protein variants. Analysis by western blotting of SARS-CoV-2 S protein variants. **A.** Expression levels of WT and ΔTMD, and SCBRL variants. **B.** Expression levels of SARS-CoV-2 S protein variants including alanine substitutions Phe_{1220}-Gly_{1223}, Leu_{1224}-Ile_{1227}, and Val_{1228}-Thr_{1231}. **C.** Expression levels of SARS-CoV-2 S protein variants including alanine insertions in positions 1221, 1226, 1228, 1230, and 1232 ($InsA_{1221}$, $InsA_{1226}$, $InsA_{1228}$, $InsA_{1230}$, and $InsA_{1232}$ respectively). **D.** Expression levels of SARS-CoV-2 S protein variants including point mutations, G_{1223}I, G_{1219}I G_{1223}I, I_{1221}Y, I_{1225}Y, M_{1229}Y, A_{1222}I G_{1223}I, A_{1222}Y G_{1223}Y, L_{1224}Y, L_{1224}Y V_{1228}Y. Endogenous GAPDH was used as a loading control for all experiments.

Next, stretches of four amino acids (one helical turn) on the SARS-CoV-2 S protein TMD, starting at Phe_{1220}, were substituted with alanine residues (Fig 2C). Substitution of the section covering Phe_{1220}-Gly_{1223} and Leu_{1224}-Ile_{1227} (F_{1220}-G_{1223}A and L_{1224}-I_{1227}A respectively) greatly diminishes the infectivity of the pseudo-typed VSVΔG-GFP particles. Substitution of Val_{1228}-Thr_{1231} (V_{1228}-T_{1231}A) to alanines did not decrease viral entry, on the contrary, a non-significant increase was observed. These results indicate that the sequence and structural parameters of the TMD relevant for viral entry are distributed across the entire segment, with a stronger implication of the N-terminal portion

of the TMD. It is important to note that none of the modifications included in the study significantly modified the insertion potential of the SARS-CoV-2 S protein TMD based on the predicted ΔG_{app} for their insertion (Fig 1B).

To further analyze the role of SARS-CoV-2 S protein TMD in the viral entry process we inserted alanine residues in key positions of the hydrophobic segment (Fig 2D). Inserting a residue not only increases the length of the TMD but also changes the relative orientation of the residues positioned before and after the insertion point due to the α-helical nature of this segment[27]. We first inserted an alanine between Phe_{1220} and Ile_{1221} (Ins-A_{1221}) which alters the relative orientation of the aromatic membrane proximal region and the TMD. Additionally, Ins-Ala_{1221} disrupts the $G_{1223}xxxG_{1229}$ motif found in the TMD[14], forcing a 100º rotation for Gly_{1219} relative to Gly_{1223} (Supp Fig 2A vs 2B). As shown in Fig 2D, Ins-Ala_{1221} greatly reduced the infectivity of pseudo-typed VSVΔG-GFP particles suggesting that either the relative orientation of the aromatic stretch vs the TMD, the GxxxG interaction motif, or both are important for particle entry.

Supplementary Figure 2. Helical wheel diagram for SARS-CoV-2 S protein TMD hydrophobic core. **A-F.** The SARS-CoV-2 S protein TMD hydrophobic core sequence (top, residues 1218-1234), AlphaFold3 models (left) and helical wheel representations (right). Residues are represented with a color code based on their predicted ΔG of insertion relative to its position within the membrane (https://dgpred.cbr.su.se/index.php?p=home).

Insertion of an alanine residue in positions 1226 or 1228 (Ins-A_{1226} and Ins-A_{1228}) negatively affected viral entry, indicating that the relative orientation of the N-terminal and C-terminal sections of the TMD might be important for infectivity (Fig 2D). These two insertion points were selected with the work of Fu and Chou[16] in mind, who proposed a trimerization hydrophobic zipper involving residues 1221, 1225, 1229, and 1233. Thus, the observed reduction in viral entry could also result from disrupting the oligomerization motif they described.

Insertions at positions 1230 and 1232 (Ins-A_{1230} and Ins-A_{1232}) also disturb the hydrophobic trimerization motif proposed by Fu and Chou (Supp Fig. 2), but reduced particle entry to a lesser extent than Ins-A_{1226} or Ins-A_{1228} (Fig 2D). Of these two insertions, Ins-A_{1230} reduced pseudo-typed VSVΔG-GFP entry more than Ins-A_{1232}, suggesting that the relative orientation of the N-terminal and C-terminal sections of the TMD is more important for the protein function than the trimerization hydrophobic zipper. Note that while the authors[16] included leucine residues on positions 1229 and 1233 we incorporated the methionine residues found in the SARS-CoV-2 S protein consensus sequence (Uniprot: P0DTC2).

To further analyze the impact of the GxxxG and the hydrophobic zipper as oligomerization motifs, we selectively mutated key residues in the SARS-CoV-2 S protein. First, we replaced Gly_{1223} with isoleucine ($G_{1223}I$) since the equivalent substitution is sufficient to break the dimerization of Glycophorin A (GpA) TMD[28,29]. The GFP count associated with pseudo-typed VSVΔG-GFP entry revealed that the $G_{1223}I$ mutation is sufficient to reduce viral entry (Fig 2E). Substituting Gly_{1219} and Gly_{1223} with isoleucine ($G_{1219}I$ $G_{1223}I$) further decreases pseudo-typed VSVΔG-GFP infectivity. Next, we replaced Ile_{1221}, Ile_{1225}, or Met_{1229} with tyrosine ($I_{1221}Y$, $I_{1225}Y$, and $M_{1229}Y$), changes that were previously suggested to modify TMD homo-oligomerization[16]. While the $I_{1221}Y$ substitution slightly reduced viral entry, the $I_{1225}Y$, and $M_{1229}Y$ ones did not (Fig 2E).

A closer look at the SARS-CoV-2 TMD reveals a highly hydrophobic surface composed of Leu_{1218}, Ile_{1221}, Leu_{1224}, Ile_{1225}, Val_{1228}, and Ile_{1232} and an opposing surface rich in small residues including Gly_{1219}, Ala_{1222}, Gly_{1223}, and Ala_{1226} which could participate in TMD-TMD interactions[15,28–31] (Fig 1C). To analyze the importance of these regions, we first replaced Ala_{1222} and Gly_{1223} with isoleucine ($A_{1222}I$ $G_{1223}I$) and measured the impact on the pseudo-typed VSVΔG-GFP infectivity.

Our results revealed that the $A_{1222}I$ $G_{1223}I$ mutant is less able to promote viral entry than the previously tested $G_{1223}I$ mutant (Fig 2E). When Ala_{1222} and Gly_{1223} were replaced with tyrosine ($A_{1222}Y$ $G_{1223}Y$), a larger and less hydrophobic amino acid, we observed a greater reduction on VSVΔG-GFP infectivity. Similarly, substituting Leu_{1224} with tyrosine ($L_{1224}Y$) on the hydrophobic surface also reduced VSVΔG-GFP entry. Double substitution of Leu_{1224} and Val_{1228} to tyrosine ($L_{1224}Y$ $V_{1228}Y$) diminished VSVΔG-GFP infectivity further. According to our results, both the small residues and the hydrophobic bulky residues aligned on opposing faces of the helix (Fig 2E) are relevant for the entry process, probably because they participate in protein-protein interactions or in TMD-lipid interactions. Protein levels for all the chimeras were probed by western blotting to avoid misinterpretation of the data (Supp. Fig 1).

In order to incorporate the SARS-CoV-2 S protein into VSVΔG-GFP pseudo-particles it must be located at the plasma membrane. Thus, we investigated whether those events in which no VSVΔG-GFP pseudo-particle infectivity was observed were the consequence of absence of SARS-CoV-2 S protein at the cell surface. We determined whether the SARS-CoV-2 S protein was located at the plasma membrane by surface-staining with an anti-RBD antibody followed by flow cytometry analysis (Fig 3). All SARS-CoV-2 S protein variants that showed a decreased VSVΔG-GFP infectivity were included in this assay. Additionally, the WT S protein was used as a positive control while the ΔTMD variant was included as negative control (Fig 3A). Our results indicate that all SARS-CoV-2 S protein variants were present at the plasma membrane at similar levels except for Ins-A1228 (Fig 3B). Thus validating the results of the VSVΔG-GFP pseudo-particle assay. The absence of Ins-A_{1228} at the cell surface suggests that the SARS-CoV-2 S protein TMD might play a role in protein trafficking.

Figure 3. Surface expression of Spike variants. **A.** Spike surface expression of WT and ΔTMD controls analyzed by flow cytometry. HEK293T cells were transfected with plasmids encoding the corresponding S protein alongside a constitutively GFP-expressing plasmid. After 48 hours, cells were stained with an anti-RBD antibody. Levels of the surface SARS-CoV-2 S protein (red) and GFP (green) were analyzed by flow cytometry. **B.** Relative surface expression of all Spike variants tested in this assay. GFP was used as a transfection control. SARS-CoV-2 S protein positive cells were gated within GFP positive cells. The figure shows the percentage of double-positive events. Samples were normalized to the SARS-CoV-2 S protein WT. The significative p-values (<0.05) for individual one-sample t-tests vs. WT are indicated above each bar.

The role of the SARS-CoV-2 S protein TMD during the viral and cellular membrane fusion process was corroborated with a syncytia-based assay known as BiMuC[30,31]. Briefly, the Venus Fluorescent Protein (VFP) can be split into two fragments, VN and VC, neither of which is fluorescent. However, when these two fragments are fused to a pair of interacting proteins that bring them together, such as cJun and bFos, the structure of the VFP is reconstituted and the fluorescence is recovered. The VN-cJun and VC-bFos chimeras were expressed in separated cell pools together with the viral machinery required for membrane fusion. Therefore, the two chimeras would only meet in the event of membrane fusion (Fig 4A). The complete SARS-CoV-2 S WT protein was used as a positive control and reference value. The ΔTMD and SCRBL variants were included as negative controls (Fig 4B-C). Measurements of the area of the syncytia can be found in Supp. Fig 3.

Figure 4. Analysis of SARS-CoV-2 S protein membrane fusion properties by BiMuC. **A**. Schematic representation of the BiMuC assay. The Venus fluorescence protein (VFP) can be split into two fragments, VN and VC respectively, neither of which is fluorescent. These two fragments were fused to cJun and bFos (VN-cJun and VC-bFos) and transfected into HEK 293T cells in separated cell pools. Cells were co-transfected with the viral machinery required for membrane fusion, the corresponding SARS-CoV-2 S protein and the ACE2 receptor. Additionally both cell pools were transfected with the Renilla Luciferase (luciferase) to normalized the fluorescent values. Only a functional SARS-CoV-2 S proteins facilitated the S - ACE2 recognition and membrane fusion allowing VFP reconstitution and fluorescence.

B. Representative images of the assay were cells have been transfected with the SARS-CoV-2 S protein (WT) or the SARS-CoV-2 S protein bearing a scrambled version of the TMD (SCRBL). **C-E**. Fluorescence/luciferase signal ratio (GFP/Luc ratio) measured for the SARS-CoV-2 S protein variants tested on this assay. Samples were normalized to the SARS-CoV-2 S protein WT. Significative p-values (<0.05) for individual one-sample t-tests are indicated above each bar. **C**. GFP/Luc ratio for the SARS-CoV-2 S protein (WT), the SARS-CoV-2 S protein where TMD was eliminated (ΔTMD), and a chimera bearing a scrambled version of the SARS-CoV-2 S protein TMD (SCRBL). As a negative control we also included cells that did not incorporate the SARS-CoV-2 S protein (ΔS). **D**. GFP/Luc ratio for the SARS-CoV-2 S protein with substitutions of 4 amino acid stretches Phe_{1220} to Gly_{1223}, and Val_{1228} to Thr_{1231}, by alanines (Phe_{1220}-Gly_{1223}A and Val_{1228}-Thr_{1231}A), and the SARS-CoV-2 S protein including alanine insertions in positions 1221, 1228, and 1232 ($InsA_{1221}$, $InsA_{1228}$, and $InsA_{1232}$). **E**. GFP/Luc ratio for the SARS-CoV-2 S protein bearing the following point mutations: Gly_{1219} to Ile and Gly_{1223} to Ile, Ile_{1221} to Tyr, Ile_{1225} to Tyr, Met_{1229} to Tyr, Ala_{1222} to Tyr and Gly_{1223} to Tyr, and Leu_{1224} to Tyr and Val_{1228} to Tyr (G_{1219}I G_{1223}I, I_{1221}Y, I_{1225}Y, M_{1229}Y, A_{1222}Y G_{1223}Y, and L_{1224}Y V_{1228}Y, respectively). All panels include the WT values used to normalized the data (dotted line) and the SCRBL samples. **F**. Correlation of FFU measured in the VSVΔG-GFP infection assay against values of reconstituted GFP/Luc ratio in the BiMuC fusion assay. Colors show samples that have lower signals than WT values in both systems (pink background), higher in both systems (green background) or different between the two assays (white background).

Supplementary Figure 3. ***Area of the S protein-induced syncytia***. The area of the syncytia was calculated from fluorescent images. Each dot represents a single measurement. Bars show the area average and st. deviation of four images. P-value was assessed by an unpaired t-test against WT condition. Values higher than 0.05 were considered non-significant (ns).

Next, we selected some of the previously described variants of the S protein and tested them using the BiMuC methodology. Substituting the Phe_{1220}-Gly_{1223} helical turn with alanines (F_{1220}-G_{1223}A) diminished syncytia formation, though not significantly (Fig 4D). On the other hand, V_{1228}-T_{1231}A did not reduce syncytia formation (Fig 4D). Alanine insertions Ins-A_{1221} altered the observed syncytia-derived fluorescence (Fig 4D). However, only a small non-significant reduction in syncytia formation was observed for Ins-A_{1232}. We also tested the influence of point mutations on the ability

of the SARS-CoV-2 S protein to induce syncytia formation (Fig 4E). In this case, the $G_{1219}I$ $G_{1223}I$, $I_{1221}Y$, or the $M_{1229}Y$ substitutions did not perturb the SARS-CoV-2 S protein function. However, $I_{1225}Y$, $A_{1222}Y$ $G_{1223}Y$, and $L_{1224}Y$ $V_{1228}Y$ reduced the formation of syncytia. Altogether, the BiMuC-derived results correlate with the pseudo-typed VSVΔG-GFP assay (Fig 4F).

Some of the modifications included in the BiMuC and the pseudo-typed VSVΔG-GFP assay could alter the previously described homo-oligomerization of the SARS-CoV-2 TMD[12,16]. To directly analyze the SARS-CoV-2 TMD potential oligomer, we used a bimolecular fluorescent complementation (BiFC) approach adapted to study intramembrane interactions[32–35]. Briefly, the tested TMDs were fused to a split VFP, either to its N-terminal (VN) or its C-terminal (VC). If interaction between the tested TMDs occurs, it will bring the VN and VC ends together, facilitating the reconstitution of the VFP structure and causing fluorescence (Fig 5A). The TMD of GpA was used as a positive control and normalization value across experimental replicates. The TMD of Tomm20, a monomeric hydrophobic segment, was used as a negative control[34,36].

Figure 5. Analysis of TMD oligomerization by BiFC assay. **A**. Reconstitution of the Venus Fluorescent Protein (VFP) mediated by the SARS-CoV-2 S protein TMD oligomerization. Different S TMD mutants were fused to both the N and C terminal sections of the VFP (VN and VC, respectively) and expressed in HEK 293T cells together with the Renilla Luciferase. Only if the SARS-CoV-2 S protein TMD is able to oligomerize the VFP will be reconstituted **B.** GFP/Luc ratios measured for different chimeras bearing the, Glycophorin A TMD (GpA), wild type SARS-CoV-2 S protein TMD (WT), and the SARS-CoV-2 S protein TMD with the following point mutations $G_{1219}I$ Gly_{1223} to Ile, Gly_{1219} to Tyr and Gly_{1223} to Tyr, Ala_{1222} to Tyr and Gly_{1223} to Tyr, Ala_{1222} to Tyr, Gly_{1223} to Ile, le_{1221} to Tyr, Ile_{1225} to Tyr, Met_{1229} to Tyr, and Leu_{1224} to Tyr and Val_{1228} to Tyr ($G_{1219}I$ $G_{1223}I$, $G_{1219}Y$ $G_{1223}Y$, $A_{1222}Y$ $G_{1223}Y$, $A_{1222}Y$, $G_{1223}I$, $I_{1221}Y$, $I_{1225}Y$, $M_{1229}Y$ and $L_{1224}Y$ $V_{1228}Y$ respectively). All values are normalized against GpA homo-oligomer. The significative p-values (<0.05) for individual one-sample t-tests are indicated above each bar.

Using this BiFC assay, we first tested the SARS-CoV-2 TMD. Our results indicate that this TMD is sufficient for inducing oligomerization (Fig 5B). Next, we focused on the role of the surface with small residues on oligomerization of the TMD. The $G_{1219}I$ $G_{1223}I$ substitution did not decrease the SARS-CoV-2 S protein's TMD oligomerization potential, however, the $G_{1219}Y$ $G_{1223}Y$ did, a result that correlates with the outcome of the BiMuC assay. The $A_{1222}Y$ $G_{1223}Y$ double substitution further reduced oligomerization. Interestingly, the substitution of only Ala_{1222} or Gly_{1223} with Y ($A_{1222}Y$ and $G_{1223}Y$), did not alter the oligomerization capabilities of the SARS-CoV-2 TMD, suggesting a synergistic effect of these two residues.

We also tested the role of the bulky hydrophobic helix interface on the SARS-CoV-2 S protein TMD on oligomerization. The single substitutions $I_{1221}Y$, $I_{1225}Y$, or $M_{1229}Y$ did not alter oligomerization. Furthermore, the double substitution $L_{1224}Y$ $V_{1228}Y$, which had proven fundamental for VSVΔG-GFP entry and membrane fusion in the BiMuC assay, did not decrease the observed fluorescence, suggesting that the hydrophobic surface does not participate in the oligomerization of the SARS-CoV-2 S protein TMD.

We modeled the potential SARS-CoV-2 S protein TMD homo-trimer using TMHOP[37]. TMHOP uses Rosetta symmetric all-atom ab initio fold-and-dock simulations in an implicit membrane environment to predict low-energy conformations based on the empirical measurement of amino acid insertion propensities into *E. coli* inner membrane[38]. These conformations are clustered based on energy and structural properties (Supp Fig. 4A) and a representative model of each cluster is shown (Fig 6A). In each model, Gly_{1216}, Ala_{1222}, Gly_{1223}, and Ala_{1226} are located on the interaction surface of the potential trimer, confirming our experimental results. We also modeled the potential TMD trimer using AlphaFold3[39], a recent evolution of the AlphaFold2 architecture and training procedure capable of predicting the joint structure of complexes. Once again, the predicted models

present, in most cases, Gly_{1216}, Ala_{1222}, Gly_{1223}, and Ala_{1226} at the interior of a TMD homo-trimer (Fig. 6B and Supp Fig. 4). The AlphaFold3 server allowed us to model larger sequences, including the full-length S protein. When the TMD region was modeled in the context of the full-length protein accuracy was low and thus the models were discarded (Supp. Fig 5). However, the server modeled with sufficient confidence the TMD flanked by the N-terminal aromatic-rich region and the Cys-rich C-terminal end. The retrieved models for this section included, once again, the Gly_{1216}, Ala_{1222}, Gly_{1223}, and Ala_{1226} on the internal surface of a TMD trimer (Fig. 6C and Supp Fig. 4).

Supplementary Figure 4. **Statistical output of *SARS-CoV-2 S protein* TMD trimer models. A.** Scatterplot distribution for the S TMD homodimer modeling by TMHOPP. Dots represent each of the models for the S TMD's homotrimers. rmsd is calculated from the lowest-energy model. The selected models are highlighted in five different colors. **B-C.** Predicted alignment error (top) for different AlphaFold3 models. Each residue is colored by pIDDT values on the predicted structural models (bottom). Two inputs were applied: one covering just the TMD (B, residues 1214-1236) or including the N-terminal TMD flanking region up to the C-terminal end of the protein (C, residues 1207-1273).

Figure 6. SARS-CoV-2 S protein TMD **trimer models. A and B.** SARS-CoV-2 S protein TMD trimer modeled by the TMHOP server (**A**) and by the AlphaFold3 server (**B**). The top five models are shown. The input sequence was $_{1212}$WYIWLGFIAGLIAIVMVTIMLCC$_{1235}$. Models show the TMD hydrophobic core, residues 1218-1234 shown in pink. The secondary structure as well as the surface of the peptide is shown. Gly_{1216}, Ala_{1222}, Gly_{1223}, and Ala_{1226} are highlighted in red on the secondary structure representation. **C**. AlphaFold3 server model of the C-terminal region of the SARS-CoV-2 S protein, including the aromatic-rich section, the TMD, and the cysteine-rich domain. The five top models are shown. Models include just the TMD hydrophobic core, residues 1218-1234 shown in pink. The secondary structure as well as the surface of the TMD hydrophobic core is shown. Gly_{1216}, Ala_{1222}, Gly_{1223}, and Ala_{1226} are highlighted in red.

Supplementary Figure 5. **Statistical output of full-length SARS-CoV-2 S protein trimer models.** Predicted alignment error (top) for different Alphafold3 models. Each residue is colored by pIDDT values on the predicted structural models.

In sum, our results indicate that the SARS-CoV-2 S protein TMD is a key region for viral entry. A variety of small changes across its sequence can render a crippled S protein. One could imagine that a passive membrane anchoring domain would accept mutations as long as the overall hydrophobicity and length are not compromised. However, a closer look at the genome of all the variants of SARS-CoV-2 reveals that no mutation whatsoever has been found in the S protein TMD, and few variations are found between SARS-CoV-2, the SARS-CoV-1, or MERS-CoV (Supp Fig 6A), while TMDs of other SARS-CoV-2 proteins present multiple mutations (Supp Fig 6B).

Supplementary Figure 6. Mutational landscape of SARS-CoV-2 transmembrane proteome. **A**. Sequence alignment of the SARS-CoV-2, the SARS-CoV-1, and MERS-CoV S proteins TMD. **B**. Top. Schematic representation of SARS-CoV-2 genome. ORFs that include membrane proteins are colored. The regions that encode TMDs are highlighted in gray. Bottom. Schematic representation of SARS-CoV-2 non-structural, structural and accessory membrane proteins. Variations over the TMDs consensus sequence that are or have been circulating are indicated. Mutations in single isolates were not consider.

Within the S protein TMD, we identified at least two subdomains important for the protein's function. One is a highly hydrophobic surface that most likely works in coordination with the lipid milieu, and the other is a surface rich in small residues (alanine and glycine) which might participate in the homo-oligomerization of the TMD. In the non-polar membrane environment, the hydrophobic effect is responsible for insertion into the membrane and does not usually contribute to the tertiary and quaternary structure[40]. Furthermore, the nature of the residues that facilitate membrane insertion impede salt bridges and hydrogen bonds between TMDs[41]. In this context, the

tertiary and quaternary structure of TMDs is dictated by a delicate balance of the remaining low energy forces, where van der Waals interactions play a crucial role[42]. The nature of van der Waals forces require a large contact area between the associating protein segments. Amino acids with small side chains such as glycine and alanine facilitate the helix-helix contact increasing van der Waals forces. In this scenario, the inclusion of bulky residues that diminish the helix-helix contact area could break TMD-TMD interactions. Our results, together with the structural and sequence analysis, suggest that the SARS-CoV-2 S protein TMD is capable of forming homo-trimers held together by van der Waals forces. However, the structure of the full-length S protein in a post-fusion conformation did not reveal the TMD in a trimeric disposition[17]. In this structure, the TMDs surround the FPs and make contact with them through the Leu_{1234} and Phe_{1120} residues, which were not involved in either of the subdomains described in our work. Nonetheless, given that different oligomeric states coexist for the SARS-CoV-2 S protein[43], we believe that the TMDs could be found in different oligomerization states as well. In the early stages of the SARS-CoV-2 S protein biogenesis, the TMD-TMD interactions responsible for the TMD trimer could facilitate formation of a full-length protein trimer[8], participate in the post-translational processing of the S protein[44], or modulate its trafficking[45]. Next, the low energy nature of the forces holding the TMD trimer together would allow the transition between the different oligomeric states of the protein required during the complex membrane fusion process.

Materials and Methods

Plasmid constructs

Sequence encoding SARS-CoV-2 S protein (Addgene #164433) was a gift from Alejandro Balazs. All subsequent mutants were obtained with the Pfu Plus! mutagenesis kit (EurX, Gdańsk, Poland) according to the manufacturer's instructions. For the generation of BiFC chimeric plasmids, the sequence encoding for the TMD was PCR amplified and subcloned to generate a fusion with the Nt

or Ct of the Venus Fluorescent Protein (VN, VC, respectively: Addgene #27097 and #22011, a gift from Chang-Deng H). The TMD was cloned at the Ct of the VFP. These constructs were generated using In-Fusion HD cloning kit (Takara, Japan) also according to the manufacturer's instructions. Sequences were verified by sequencing the plasmid DNA at Macrogen (Seoul, South Korea).

Human cells and culture conditions

Human embryonic kidney (HEK293T) (from American Type Cell Collection, Manassas, VA, USA) were maintained in Dulbecco's modified Eagle's medium (DMEM, Gibco) supplemented with 10 % fetal bovine serum, 1 % penicillin-streptomycin (Sigma Aldrich, St. Louis, MO, USA), and 1 % MEM non-essential amino acids. Cells were maintained at 37°C and 5 % CO2 conditions. VeroE6-TMPRSS2 were obtained from the JCRB Cell Bank (catalog JCRB1819). Cells were cultured in DMEM with either 10% or 2% FBS for maintenance or infection, respectively. Selection of VeroE6-TMPRSS2 was performed using media containing 1 mg/mL G418 (Cat. no. A1720, Sigma).

Transient DNA transfections

HEK 293T were seeded in 6, 12 or 24-well plates and grown in DMEM (Gibco) supplemented with 10 % fetal bovine serum (FBS, Gibco) the day before transfection. For the pseudo-typed VSVΔG-GFP assay, transfections were carried out with calcium phosphate. Briefly, DNA was mixed with $CaCl_2$ 125 mM in HBS buffer (NaCl 140 mM, KCl 5 mM, Na_2HPO_4 750 μM, dextrose 6 mM, HEPES 25 mM pH 7.05), incubated for 15 minutes and added to the HEK293 cells. For the rest of transfections, DMEM medium was mixed with plasmid DNA (up to a total of 4 μg DNA for 6-well, 2 μg for 12-well plates) and transfected into HEK293T cells adding 3 μl of PEI 1 mg/ml and 100 μl of DMEM per μg of DNA. For the surface expression analysis, 2-3 x 10^6 HEK293T cells were transfected in 6 well-plates with the corresponding 1 μg Spike plasmid and 1 μg of a GFP encoding plasmid. In the BiMuC assay, 2×10^6 HEK 293T cells in 6-well plates were transfected either with

1 μg of S and 1 μg of VN-Jun or with 1 μg ACE2 and 1 μg VC-Fos, plus 100 ng pRL-CMV. Transfection mixtures were added dropwise onto cells and the media was changed after 24 hours. For BiFC chimeras, 1 μg of cMyc-VN-TMD and 1 μg HA-VC-TMD were transfected into 2×10^6 HEK 293T cells in 12-well plates. A plasmid expressing Renilla luciferase under the CMV promoter (pRL-CMV) (50 ng) was also added to the mix to normalize the signal. For western blot (WB) analysis, 2 μg of Spike plasmid was transfected into 6-well plates containing 6×10^6 HEK 293T cells.

Pseudo-typed VSVΔG-GFP assay

Pseudo-typed VSVΔG-GFP were produced by transfection of HEK 293T cells with lipofectamine 2000 (ThermoFisher Scientific) with a plasmid encoding the indicated S genes as indicated above in a 24-well plate using 750 ng DNA and 1.875 μL of lipofectamine per well. Following 24 hours, cells were infected at a multiplicity of infection of three with VSV lacking the VSV-G protein and encoding both GFP and firefly luciferase[25] that was previously pseudotyped with the VSV-G protein to yield infectious particles. Following infection, a monoclonal antibody targeting VSV-G (a kind gift of Rafael Sanjuan, University of Valencia, Spain) was added to neutralize any remaining VSV-G bearing viruses, and infected cells were incubated for 18 hours. The supernatant was collected, clarified and frozen at -80ºC. The viral titer (focus forming units [FFU] per mL) was obtained by serial dilution on VeroE6-TMPRSS2 cells in 96 well plates and counting of GFP-expressing cells using a live cell microscope following 24 hours (Incucyte SX5, Sartorius).

Flow cytometry

Two days post-transfection, cells were washed once with PBS and detached by pipetting in 500 μl of FACS buffer (0.1% sodium azide, 0.5% BSA in PBS). Then, 200 μl of cells were collected by centrifugation at 1500 rpm for 5 min at 4ºC and stained with anti-SARS-CoV-2 RBD antibody (Invitrogen, PA5-114529). After centrifugation, cells were incubated with Goat anti-Rabbit IgG

(H+L) AlexaFluor 568 (ThermoFisher, A11011). Antibodies were diluted to 1:1000 in FACS buffer and 100 μl per sample were used for resuspending cells, following a 30 min incubation on ice in both cases. After three final washes with FACS buffer, surface SARS-CoV-2 spike levels were analysed using a LSRFortessa flow cytometer (BD Biosciences). Typically, 10,000 live cells were selected based on morphology and acquired. SARS-CoV-2 spike levels were assessed by gating AlexaFluor 568 (excitation at 561 nm, collection at 586/15 nm) positive cells within the gate of GFP positive cells (excitation at 488 nm, collection at 530/30 nm). Analysis was carried out using BDFACS Diva 8.0.2 and FlowJo X softwares.

BiMuC assay

One day post-transfection, media was discarded and cells rinsed with PBS. Then, 1 mL DMEM was added to each well and one well expressing S and VN-Jun was pooled with other well expressing hACE2 and VC-Fos. Cells were counted and 500 μl containing 3 x 10^5 cells were seeded into 4 wells of a 24-well plate. The next day, media was discarded in three of the wells, cells rinsed with PBS, collected in 100 μl PBS, and seeded into 96-well black plates for fluorescence read. Fluorescence was measured in a VictorX multi-plate reader (Perking Elmer, Waltham, MA, USA). The luminescence was measured using the Renilla Luciferase Assay kit from Sigma following the manufacturer protocol on 96-well white cell-culture plates. At least three independent experiments were done. The ratio between VFP and luciferase was registered as the relative fluorescence units (RFU) for each well. Each of those three wells was accounted as a technical replicates, and the mean of the three values was taken as a biological replicate. RFU values were then normalized against the wild-type condition. The extra well was used for imaging syncytia in a fluorescence microscope.

Fluorescence imaging

For each condition, one well of a 24-well plate was used for imaging 24 h after co-culture of S and hACE2 expressing cells. Images were acquired with a Zeiss Axiovert 5 fluorescence microscope both with transmitted light and GFP channels with a 10X objective. Both images were overlayed and the GFP area was quantified using Fiji[46].

Bimolecular fluorescent complementation (BiFC) assay

Two days after transfection, PBS washed and collected for fluorescence and luciferase measurements (Victor X3 plate reader). For the Renilla luciferase readings used as signal normalization, we used the Renilla Luciferase Glow Assay Kit (Pierce, Thermofisher) according to the manufacturer's protocol. In each experiment, the fluorescence/luminescence ratio obtained with the GpA-GpA homodimer was used for normalization. All experiments were done at least in triplicates.

Protein expression and western blotting

Two days after transfection, media was aspirated, cells rinsed with PBS and, lysed in 300 µl of radioimmunoprecipitation assay (RIPA) buffer [150 mM NaCl, 50 mM tris-HCl (pH 8), 1% NP-40, 0.5 % sodium deoxycholate, 0.4% SDS, and 1 mM EDTA] supplemented with cOmplete EDTAfree protease inhibitors (Roche). Lysates were sonicated with three pulses of 1 s in a VCX-500 Vibra Cell sonicator (Sonics) following addition of 5X sample buffer (final concentration 62.5 mM Tris-HCl [pH 6.8], 2 % sodium dodecyl sulfate [SDS], 0.01 % bromophenol blue, 10 % glycerol and 5 % β-mercaptoethanol). Protein samples were boiled for 4 min at 95 ºC, subjected to 10 % SDS-polyacrylamide gel electrophoresis (PAGE), and transferred to nitrocellulose membranes (Cytiva). Membranes were blocked for 30 min at room temperature in Tris-buffered saline supplemented with 0.05 % Tween 20 (TBS-T) containing 5% non-fat dry milk and later incubated with primary antibodies diluted in the same buffer at 4 °C overnight. Antibodies used in this study were SARS-

CoV-2 Spike S1 (Invitrogen PA5-81795), rabbit anti c-Myc (Sigma PLA0001) and mouse anti-GAPDH (Santa Cruz sc-47724). Then, membranes were washed with TBS-T and incubated with goat anti-rabbit or sheep anti-mouse IgG horseradish peroxidase conjugate (Sigma DC02L or GE Healthcare NXA931, respectively) for 1 h at room temperature and washed again. All antibodies were used at a 1:10,000 dilution in TBS-T with 5 % non-fat dry milk. Detection of immunoreactive proteins was carried out using the enhanced chemiluminescence reaction (Amersham ECL Prime, Cytiva) and detected by the Image Quant LAS 4000 Mini (GE Healthcare).

Statistical analysis

Unpaired t-tests assuming Gaussian distribution and equal standard deviation (SD) for all conditions were applied. The p-value was two-tailed. One sample t-test was performed to compare different wild-type-normalized conditions, comparing means with a 'hypothetical value' of 100. Significance assessment between test samples and controls was performed using GraphPad Prism for P values < 0.05.

Acknowledgments

We thank P. Selvi and S. Pinto for their excellent technical assistance and V. Febrer for reviewing the manuscript. This work was supported by the Generalitat Valenciana (CIPROM/2022/62) and grant PID2020-119111GB-I00, PID2023-152568NB-I00 and CNS2022-135100 by MCIN/AEI/10.13039/501100011033 and European Union NextGenerationEU/PRTR. J. O-M. has a predoctoral contract granted by the Universisty of Valencia's "Atracció de Talent" programme (UV-INV_PREDOC 1911519). D. B. had a research contract funded by CIPROM/2022/62 project. Flow Cytometry experiments were carried out in the Cell Culture and Flow Cytometry facility at the SCSIE (University of Valencia).

References

1. Yadav, R. *et al.* Role of Structural and Non-Structural Proteins and Therapeutic Targets of SARS-CoV-2 for COVID-19. *Cells* **10**, 821 (2021).
2. Gallagher, T. M. & Buchmeier, M. J. Coronavirus spike proteins in viral entry and pathogenesis. *Virology* **279**, 371–374 (2001).
3. Simmons, G., Zmora, P., Gierer, S., Heurich, A. & Pöhlmann, S. Proteolytic activation of the SARS-coronavirus spike protein: cutting enzymes at the cutting edge of antiviral research. *Antiviral Res*. **100**, 605–614 (2013).
4. Hoffmann, M. *et al*. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **181**, 271-280.e8 (2020).
5. Ellaithy, A., Gonzalez-Maeso, J., Logothetis, D. A. & Levitz, J. Structural and Biophysical Mechanisms of Class C G Protein-Coupled Receptor Function. *Trends Biochem. Sci*. **45**, 1049–1064 (2020).
6. Bocharov, E. V. *et al.* Helix-helix interactions in membrane domains of bitopic proteins: Specificity and role of lipid environment. *Biochim. Biophys. Acta Biomembr.* **1859**, 561–576 (2017).
7. Duart, G., Mingarro, I. & Martinez-Gil, L. The importance of transmembrane domain interactions in the viral control of apoptosis. *Mol. Cell. Oncol*. **8**, 1911290 (2021).
8. Barrett, C. T. & Dutch, R. E. Viral Membrane Fusion and the Transmembrane Domain. *Viruses* **12**, 693 (2020).
9. Mravic, M. *et al*. De novo-designed transmembrane proteins bind and regulate a cytokine receptor. *Nat. Chem. Biol.* **20**, 751–760 (2024).
10. 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*. **80**, 1302–1310 (2006).
11. 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*. **6**, 230 (2009).

12. Azad, T. *et al.* Implications for SARS-CoV-2 Vaccine Design: Fusion of Spike Glycoprotein Transmembrane Domain to Receptor-Binding Domain Induces Trimerization. *Membranes* **10**, 215 (2020).
13. Mackenzie, K. R. Folding and stability of alpha-helical integral membrane proteins. *Chem. Rev.* **106**, 1931–1977 (2006).
14. Arbely, E., Granot, Z., Kass, I., Orly, J. & Arkin, I. T. A trimerizing GxxxG motif is uniquely inserted in the severe acute respiratory syndrome (SARS) coronavirus spike protein transmembrane domain. *Biochemistry* **45**, 11349–11356 (2006).
15. Corver, J., Broer, R., van Kasteren, P. & Spaan, W. GxxxG motif of severe acute respiratory syndrome coronavirus spike glycoprotein transmembrane domain is not involved in trimerization and is not important for entry. *J. Virol*. **81**, 8352–8355 (2007).
16. Fu, Q. & Chou, J. J. A Trimeric Hydrophobic Zipper Mediates the Intramembrane Assembly of SARS-CoV-2 Spike. *J. Am. Chem. Soc*. **143**, 8543–8546 (2021).
17. Shi, W. *et al*. Cryo-EM structure of SARS-CoV-2 postfusion spike in membrane. *Nature* **619**, 403–409 (2023).
18. Hessa, T. *et al*. Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature* **450**, 1026–1030 (2007).
19. Hessa, T. *et al.* Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* **433**, 377–381 (2005).
20. Sainz, B., Rausch, J. M., Gallaher, W. R., Garry, R. F. & Wimley, W. C. The aromatic domain of the coronavirus class I viral fusion protein induces membrane permeabilization: putative role during viral entry. *Biochemistry* **44**, 947–958 (2005).
21. Mesquita, F. S. *et al*. S-acylation controls SARS-CoV-2 membrane lipid organization and enhances infectivity. *Dev. Cell* **56**, 2790-2807.e8 (2021).
22. Wu, Z. *et al.* Palmitoylation of SARS-CoV-2 S protein is essential for viral infectivity. *Signal Transduct. Target. Ther.* **6**, 231 (2021).

23. McBride, C. E. & Machamer, C. E. Palmitoylation of SARS-CoV S protein is necessary for partitioning into detergent-resistant membranes and cell-cell fusion but not interaction with M protein. *Virology* **405**, 139–148 (2010).
24. Ujike, M. *et al.* Two palmitylated cysteine residues of the severe acute respiratory syndrome coronavirus spike (S) protein are critical for S incorporation into virus-like particles, but not for M-S co-localization. *J. Gen. Virol.* **93**, 823–828 (2012).
25. Bäuerl, C. *et al.* Assessment of SARS-CoV-2 neutralizing antibody titers in breastmilk from convalescent and vaccinated mothers. *iScience* **26**, (2023).
26. Berger Rentsch, M. & Zimmer, G. A vesicular stomatitis virus replicon-based bioassay for the rapid and sensitive determination of multi-species type I interferon. *PloS One* **6**, e25858 (2011).
27. Mingarro, I., Whitley, P., Lemmon, M. A. & von Heijne, G. Ala-insertion scanning mutagenesis of the glycophorin A transmembrane helix: a rapid way to map helix-helix interactions in integral membrane proteins. *Protein Sci. Publ. Protein Soc.* **5**, 1339–1341 (1996).
28. Lemmon, M. A., Flanagan, J. M., Treutlein, H. R., Zhang, J. & Engelman, D. M. Sequence specificity in the dimerization of transmembrane alpha-helices. *Biochemistry* **31**, 12719–12725 (1992).
29. Orzáez, M., Lukovic, D., Abad, C., Pérez-Payá, E. & Mingarro, I. Influence of hydrophobic matching on association of model transmembrane fragments containing a minimised glycophorin A dimerisation motif. *FEBS Lett.* **579**, 1633–1638 (2005).
30. García-Murria, M. J. *et al.* Identification of small molecules capable of enhancing viral membrane fusion. *Virol. J.* **20**, 99 (2023).
31. García-Murria, M. J., Expósito-Domínguez, N., Duart, G., Mingarro, I. & Martinez-Gil, L. A Bimolecular Multicellular Complementation System for the Detection of Syncytium Formation: A New Methodology for the Identification of Nipah Virus Entry Inhibitors. *Viruses* **11**, E229 (2019).
32. Kerppola, T. K. Design and implementation of bimolecular fluorescence complementation (BiFC) assays for the visualization of protein interactions in living cells. *Nat. Protoc.* **1**, 1278–1286 (2006).

33. Andreu-Fernández, V. *et al.* Bax transmembrane domain interacts with prosurvival Bcl-2 proteins in biological membranes. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 310–315 (2017).
34. García-Murria, M. J. *et al.* Viral Bcl2s' transmembrane domain interact with host Bcl2 proteins to control cellular apoptosis. *Nat. Commun.* **11**, 6056 (2020).
35. Duart, G., Grau, B., Mingarro, I. & Martinez-Gil, L. Methodological approaches for the analysis of transmembrane domain interactions: A systematic review. *Biochim. Biophys. Acta Biomembr.* **1863**, 183712 (2021).
36. Duart, G. *et al.* Computational design of BclxL inhibitors that target transmembrane domain interactions. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2219648120 (2023).
37. Weinstein, J. Y., Elazar, A. & Fleishman, S. J. A lipophilicity-based energy function for membrane-protein modelling and design. *PLoS Comput. Biol.* **15**, e1007318 (2019).
38. Elazar, A. *et al.* Mutational scanning reveals the determinants of protein insertion and association energetics in the plasma membrane. *eLife* **5**, e12125 (2016).
39. Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
40. Martinez-Gil, L. & Mingarro, I. Viroporins, Examples of the Two-Stage Membrane Protein Folding Model. *Viruses* **7**, 3462–3482 (2015).
41. Baeza-Delgado, C., Marti-Renom, M. A. & Mingarro, I. Structure-based statistical analysis of transmembrane helices. *Eur. Biophys. J. EBJ* **42**, 199–207 (2013).
42. Toward understanding driving forces in membrane protein folding - ScienceDirect. https://www.sciencedirect.com/science/article/pii/S0003986114002823.
43. Webb, S. R., Smith, S. E., Fried, M. G. & Dutch, R. E. Transmembrane Domains of Highly Pathogenic Viral Fusion Proteins Exhibit Trimeric Association In Vitro. *mSphere* **3**, e00047-18 (2018).
44. Miyauchi, K. *et al.* The membrane-spanning domain of gp41 plays a critical role in intracellular trafficking of the HIV envelope protein. *Retrovirology* **7**, 95 (2010).
45. Popa, A. *et al.* Residues in the hendra virus fusion protein transmembrane domain are critical for endocytic recycling. *J. Virol.* **86**, 3014–3026 (2012).

46. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).

Figure 1

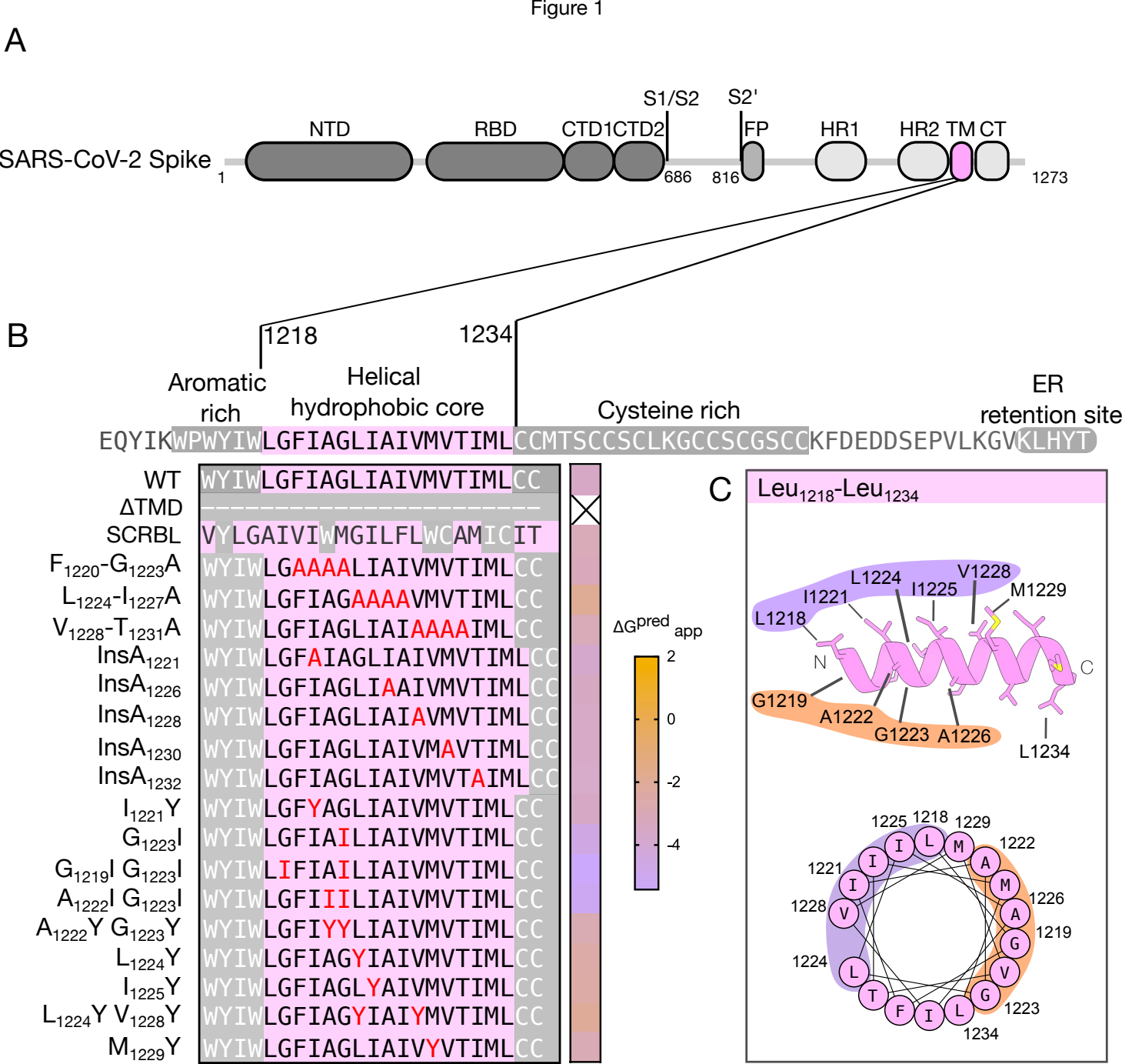

A
SARS-CoV-2 Spike
NTD
RBD
CTD1
CTD2
S1/S2
S2'
FP
HR1
HR2
TM
CT
1
686
816
1273
B
1218
1234
Aromatic rich
Helical hydrophobic core
Cysteine rich
ER retention site
EQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDDSEPVLKGVKLHYT
WT WYIWLGFIAGLIAIVMVTIMLCC
ΔTMD
SCRBL VYLGAIVIWMGILFLWCAMICIT
F1220-G1223A WYIWLGAAAALIAIVMVTIMLCC
L1224-I1227A WYIWLGFIAGAAAAVMVTIMLCC
V1228-T1231A WYIWLGFIAGLIAIAAAAIMLCC
InsA1221 WYIWLGFAIAGLIAIVMVTIMLCC
InsA1226 WYIWLGFIAGLIAAIVMVTIMLCC
InsA1228 WYIWLGFIAGLIAIAVMVTIMLCC
InsA1230 WYIWLGFIAGLIAIVMAVTIMLCC
InsA1232 WYIWLGFIAGLIAIVMVTAIMLCC
I1221Y WYIWLGFYAGLIAIVMVTIMLCC
G1223I WYIWLGFIAILIAIVMVTIMLCC
G1219I G1223I WYIWLIFIAILIAIVMVTIMLCC
A1222I G1223I WYIWLGFIIILIAIVMVTIMLCC
A1222Y G1223Y WYIWLGFIYYLIAIVMVTIMLCC
L1224Y WYIWLGFIAGYIAIVMVTIMLCC
I1225Y WYIWLGFIAGLYAIVMVTIMLCC
L1224Y V1228Y WYIWLGFIAGYIAIYMVTIMLCC
M1229Y WYIWLGFIAGLIAIVYVTIMLCC
ΔGpred app
2
0
-2
-4
C
Leu1218-Leu1234
L1218
I1221
L1224
I1225
V1228
M1229
N
C
G1219
A1222
G1223
A1226
L1234
1225
1218
1229
1222
1221
1226
1228
1219
1224
1223
1234

Figure 2

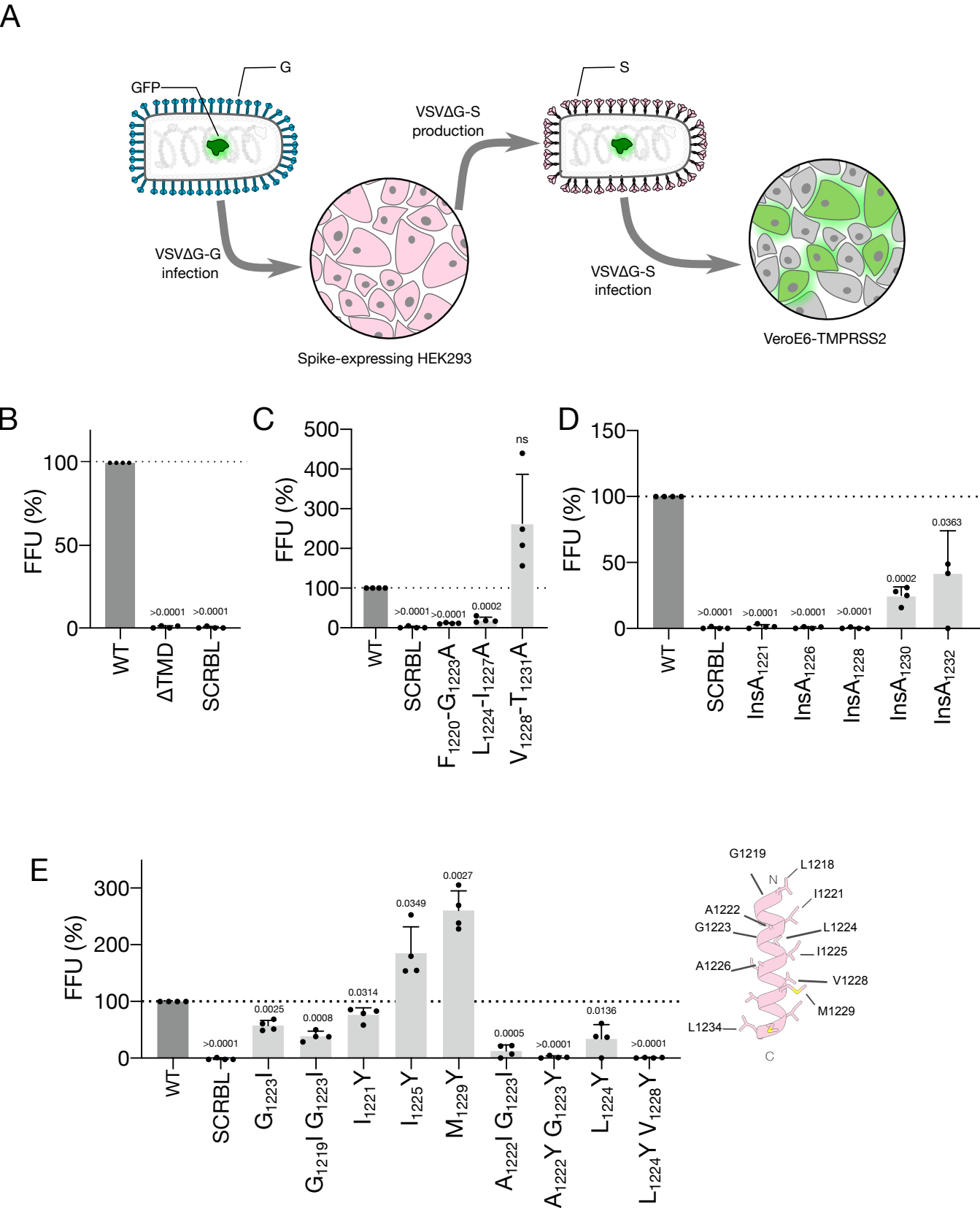

A
GFP
G
VSVΔG-G infection
Spike-expressing HEK293
VSVΔG-S production
S
VSVΔG-S infection
VeroE6-TMPRSS2
B
FFU (%)
100
50
0
>0.0001
>0.0001
WT
ΔTMD
SCRBL
C
FFU (%)
500
400
300
200
100
0
>0.0001
>0.0001
0.0002
ns
WT
SCRBL
F1220-G1223A
L1224-I1227A
V1228-T1231A
D
FFU (%)
150
100
50
0
>0.0001
>0.0001
>0.0001
>0.0001
0.0002
0.0363
WT
SCRBL
InsA1221
InsA1226
InsA1228
InsA1230
InsA1232
E
FFU (%)
300
200
100
0
>0.0001
0.0025
0.0008
0.0314
0.0349
0.0027
0.0005
>0.0001
0.0136
>0.0001
WT
SCRBL
G1223I
G1219I G1223I
I1221Y
I1225Y
M1229Y
A1222I G1223I
A1222Y G1223Y
L1224Y
L1224Y V1228Y
G1219
L1218
N
I1221
A1222
G1223
L1224
I1225
A1226
V1228
M1229
L1234
C

Figure 3

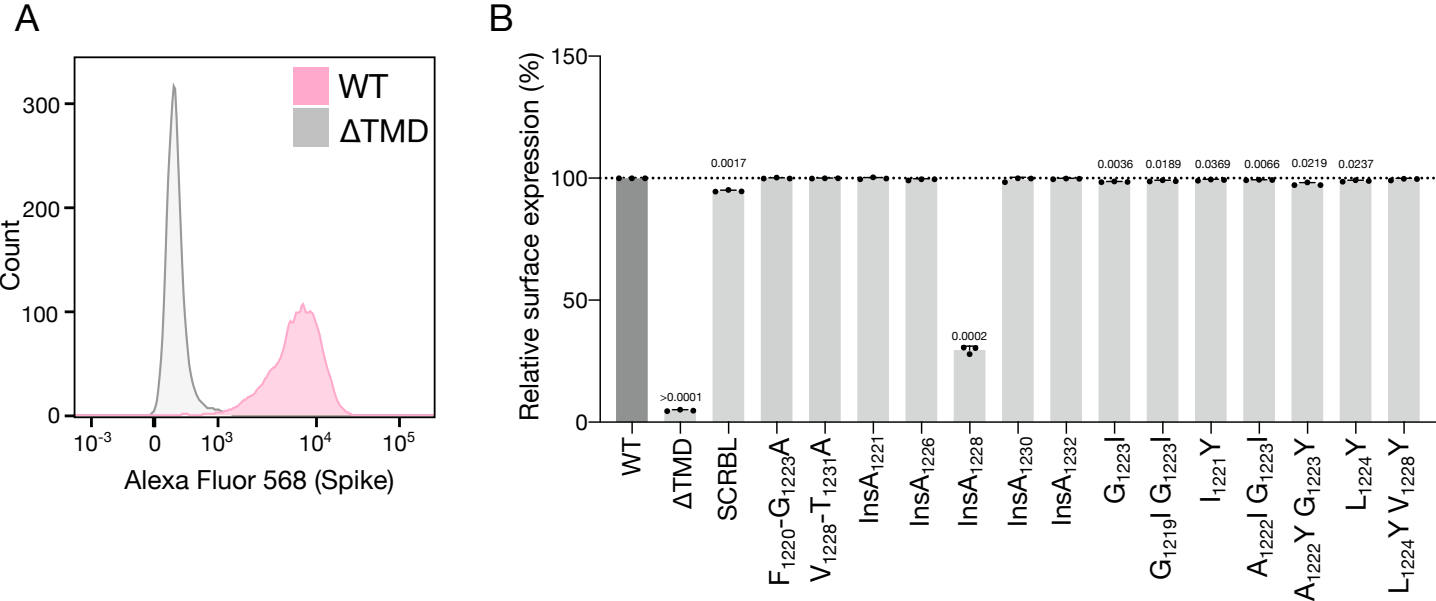

A
WT
ΔTMD
Count
300
200
100
0
10^{-3}
0
10^{3}
10^{4}
10^{5}
Alexa Fluor 568 (Spike)
B
Relative surface expression (%)
150
100
50
0
0.0017
>0.0001
0.0002
0.0036
0.0189
0.0369
0.0066
0.0219
0.0237
WT
ΔTMD
SCRBL
F_{1220}-G_{1223}A
V_{1228}-T_{1231}A
InsA_{1221}
InsA_{1226}
InsA_{1228}
InsA_{1230}
InsA_{1232}
G_{1223}I
G_{1219}I G_{1223}I
I_{1221}Y
A_{1222}I G_{1223}I
A_{1222}Y G_{1223}Y
L_{1224}Y
L_{1224}Y V_{1228}Y

Figure 4

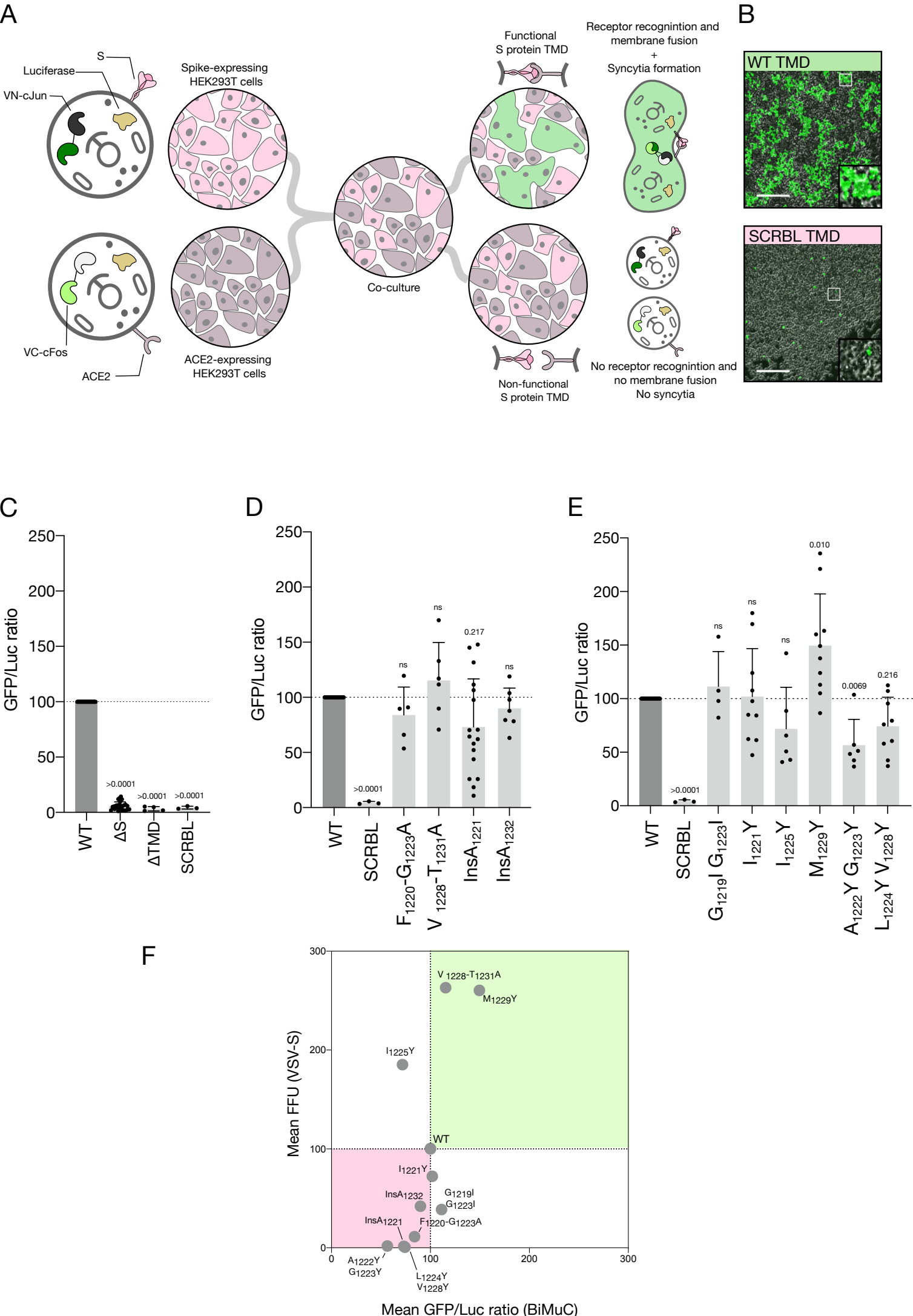
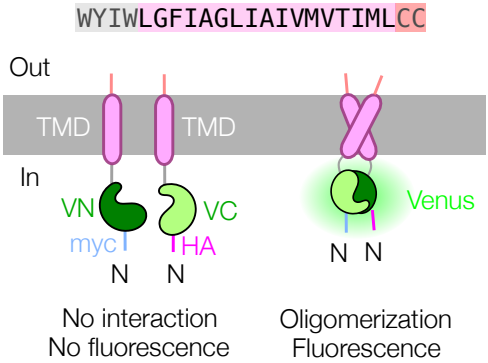

A
S
Luciferase
VN-cJun
Spike-expressing HEK293T cells
VC-cFos
ACE2
ACE2-expressing HEK293T cells
Co-culture
Functional S protein TMD
Non-functional S protein TMD
Receptor recognintion and membrane fusion + Syncytia formation
No receptor recognintion and no membrane fusion No syncytia
B
WT TMD
SCRBL TMD
C
GFP/Luc ratio
250
200
150
100
50
0
>0.0001
>0.0001
>0.0001
WT
ΔS
ΔTMD
SCRBL
D
GFP/Luc ratio
>0.0001
ns
ns
0.217
ns
WT
SCRBL
F1220-G1223A
V 1228-T1231A
InsA1221
InsA1232
E
GFP/Luc ratio
>0.0001
ns
ns
ns
0.010
0.0069
0.216
WT
SCRBL
G1219I G1223I
I1221Y
I1225Y
M1229Y
A1222Y G1223Y
L1224Y V1228Y
F
Mean FFU (VSV-S)
Mean GFP/Luc ratio (BiMuC)
300
200
100
0
V 1228-T1231A
M1229Y
I1225Y
WT
I1221Y
InsA1232
G1219I
G1223I
InsA1221
F1220-G1223A
A1222Y
G1223Y
L1224Y
V1228Y

Figure 5

A


WYIWLGFIAGLIAIVMVTIMLCC
Out
TMD
TMD
In
VN
VC
myc
HA
N
N
Venus
N
N
No interaction
No fluorescence
Oligomerization
Fluorescence

B

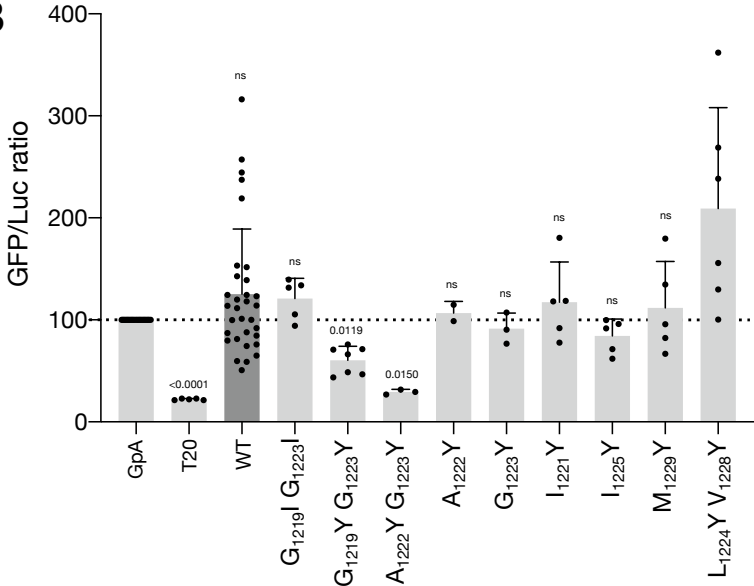

GFP/Luc ratio
400
300
200
100
0
ns
<0.0001
ns
0.0119
0.0150
ns
ns
ns
ns
ns
GpA
T20
WT
G1219I G1223I
G1219Y G1223Y
A1222Y G1223Y
A1222Y
G1223Y
I1221Y
I1225Y
M1229Y
L1224Y V1228Y

Figure 6

A

1214 TMD hydrophobic core 1236

Input sequence: WYIWLGFIAGLIAIVMVTIMLCC

Model 1 | Model 2 | Model 3 | Model 4 | Model 5

TMHOP

B

Alphafold3

C

Input sequence: 1207 TMD hydrophobic core 1273

EQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDDSEPVLKGVKLHYT

Alphafold3