

Predicted 3D models of the SARS-CoV-2 spike protein membrane proximal external region and transmembrane domain

Alexander Izvorski

Postdoctoral Institute for Computational Studies, San Francisco, California
Correspondence: alexander.izvorski@picomps.org

Abstract

The transmembrane helix domain (TMD), membrane proximal external region (MPER) and part of heptad repeat 2 (HR2) domain in SARS-CoV-2 spike protein were modelled using a constrained fold-and-dock strategy. The resulting structures were clustered and their large scale pose variability and energy landscape is described; several representative models are discussed. The results suggest considerable flexibility in the conformation of those regions, which may have an important role in the ability of spike protein to fuse the cell and viral membranes.

Introduction

SARS-CoV-2 spike (S) glycoproteins are class I viral fusion proteins which promote viral entry into cells and are the main target of antibodies (White 2008). The C terminal end of spike protein contains a heptad repeat (HR2), a short linker region (the membrane proximal external region or MPER), a transmembrane helix domain (TMD) and a C-terminal or cytoplasmic domain (CTD).

After binding of the ACE2 receptor on the target cell to the receptor binding domain (RBD) on S protein, the heptad repeat 1 (HR1) and heptad repeat 2 (HR2) domains form a six-helix bundle fusion core (6HB), bringing the viral and cellular membranes together for fusion and cell entry (Walls 2017; Xia 2020). Thus HR2 is essential for viral entry into cells.

The short membrane proximal external region (MPER) connects the HR2 and transmembrane domain, and contains an aromatic-rich fusogenic peptide sequence which destabilizes the membrane during fusion (Mahajan 2015, Guillén 2008). This region may sometimes be

referred to as "membrane proximal ectodomain region" or "pre-transmembrane region" (PTM).

The MPER peptide 1185-LGKYEQYIKWPWYVWLG-1202 from SARS-CoV-1 has been shown to intercalate into lipid membranes and to be highly surface active; the corresponding sequence in SARS-CoV-2 is identical except for a I to V substitution at 1216.

The transmembrane domain (TMD) serves to anchor the spike protein to the viral membrane. The exact structure of this is not yet known, but it is predicted to be a transmembrane alpha helix between residues 1214-1234 (UniProt), 1214-1238 (TMSEG), 1216-1235 (TMPRED) or 1213-1237 (Xia 2020). Since the spike protein exists as a trimer, the transmembrane helices may also form a trimeric bundle, by analogy with other viral fusion proteins; additionally they may serve as the trimerization nucleus for the whole protein.

The TMD sequence in SARS-CoV-2 is 1214-WYIWLGFIAGLIAIVMTIML-1234; this differs in only two residues from SARS-CoV-1 (V to I and L to M).

Immediately adjacent to the TMD is a conserved cysteine-rich region (residues 1234-1254) which contains multiple palmitoylation sites. In SARS-CoV-1, palmitoylation at those sites is essential for cell fusion (Petit 2007).

The short cytoplasmic domain or C-terminal domain (CTD) located inside the virion also contains a conserved charged region (1255-KFDEDDSE) and a KxxKxH motif.

Despite the very important functions of these domains, there is little experimental information on the detailed 3D structure of HR2, MPER and TMD. Published cryo-EM structures for SARS-CoV-2 spike protein in PDB: [6VXX](#), [6VYB](#), [6VSB](#) end at residue 1147 or 1148 and do not resolve the C terminal regions beyond that.

Another structure, PDB: [6LXT](#) shows HR2 in association with HR1 rather than in the pre-fusion state, but only extends to residue 1202, and may be unfolded after residue 1197.

There are also two available structures for parts of the SARS-CoV-1 C-terminal region including heptad repeat (HR2) in PDB: [2FXP](#) (Hakansson-McReynolds 2006) and membrane proximal external region (MPER) in PDB: [2RUN](#) (Mahajan 2015), which would be used as the basis of the present modeling work. 2FXP is a trimeric helix bundle from solution NMR using 30% trifluoroethanol as co-solvent. 2RUN is a monomer associated with a lipid bicelle also from solution NMR; it assumes a bent-helix shape. The sequences are near-identical to SARS-CoV-2 (with one substitution, V1216 to I).

Recent models for spike protein (Zhang lab; Baker lab; SWISS-MODEL) do not have a plausible structure for the C terminal region of spike protein, either because they do not include that region, or because the position of the transmembrane helices is such that the plane of the

membrane would intersect other domains. The Zhang lab most recently published a refined model with improved helix positions, but HR2 and the MPER structure is still not resolved. The Feig lab group produced a first ever computational model and molecular dynamics simulation of the transmembrane helix bundle of spike protein and portion of CTD in association with a number of units of M protein. The fragment used unfortunately did not include the MPER or HR2 regions.

There has not been to date a thorough investigation of the possible folds of the C-terminal including the HR2, MPER and TMD with a realistic structure, thus motivating the present work.

Methods

The selected sequence for modelling was UniProt: [PODTC2](#), residues 1172-1240:

```
INASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGLFIAGLIAIVMVTIMLCCMTSC
```

This sequence was selected since the beginning of 2FXP contains a disordered region and a short helix and turn; the long continuous helix starts at the homolog to 1172-INAS. Including only the long helix was expected to simplify folding and produce stable helix bundle structures. Only a part of the HR2 domain would be included, but it would be more likely to assemble into a trimer in the right orientation. The exact end of the TMD on the C-terminal side is unknown and would depend on helix tilt; it was desirable to include at least the first palmitoylation site at 1235, and several hydrophilic residues to anchor the end of the peptide.

The sequence was first uploaded to Robetta server (Kim 2004) to generate fragments.

The fragment file was then modified to restrict residues 1172-1196 to the HR2 structure in 2FXP, and residues 1205-1216 to the MPER structure in 2RUN. The goal was to explore starting poses close to those native conformations, but possibly including bends or folds in between.

The Symmetric Fold-and-Dock (Das 2009) protocol in Rosetta (version 2020.11) was used to generate 69800 folds with C3 symmetry, which were refined with the Fast-Relax protocol.

Structures were filtered to have an overall length of at least 75 angstrom since shorter structures would not be long enough to include the two helical bundles (HR2 and TMD) along the axis of symmetry.

Because the initial folding was done without any additional position constraints and without any membrane-specific changes to the scoring function, a large number of structures folded

into compact rather than extended shapes, typically by direct contact between the HR2 and transmembrane helices. These were filtered out, leaving 1940 structures.

The remaining extended structures were inserted in a membrane and re-scored using the RosettaMP protocol (Koehler Leman 2017) with the mpframework_smooth_fa_2014 all-atom energy function for membrane proteins (Barth 2007). This allows for a membrane optimally placed at a depth that produces the best score, and potentially at a tilt to the axis of symmetry.

Pose parameters were measured using a custom python (Jupyter) notebook to process the PDB files. The tilt angle of helices was measured from the center of mass of the C-alpha of the first and last turns. Positive values correspond to a right handed twist bundle, negative to left handed. The helix turn angle around its own axis was measured based on the position of the C-alpha of the first residue relative to the trimer's axis of symmetry (0 degrees corresponding to first residue towards center, 180 degrees to first residue away from center). The overall length was measured from C-alpha of residue 1172 to C-alpha of residue 1234.

The structures were clustered using Rosetta based on a 3 angstrom C-alpha RMSD cutoff for the MPER structure 1197-1217 only. It was necessary to do this as when clustering the whole model, small changes in the relative position of the major helices predominated in the RMSD calculation over large changes in the structure of the MPER backbone.

The structures were aligned to the C-alpha of the 2FXP (residue 1172-1196) and 2RUN structures using PyMol, and the RMSD was calculated.

Finally, the best structures from the top four largest and the one best scoring clusters were selected as representative models. These were submitted to the PPM server (Lomize 2012) to determine the transfer free-energy and helix-helix interaction energy, and to the SWISS-MODEL QMEAN4 structure assessment server (Benkert 2011) to determine quality.

Results

Selected models

The five selected models are shown in Figure 1. All the structures are mostly helical, but with substantial bends or kinks. The HR2 region always assembles in a twisted bundle which is an excellent fit to the twisted bundle in 2FXP, up to the first major bend around residue 1200-1204. The transmembrane helices also form a twisted bundle, which may be either left handed (model 1, 3) or right handed (model 3, 5) and typically contains significant bends where the helices cross over. The depth of the protein in the membrane varies, from very shallow (model 2, starting at residue 1216) to very deep (model 1, starting at residue 1212).

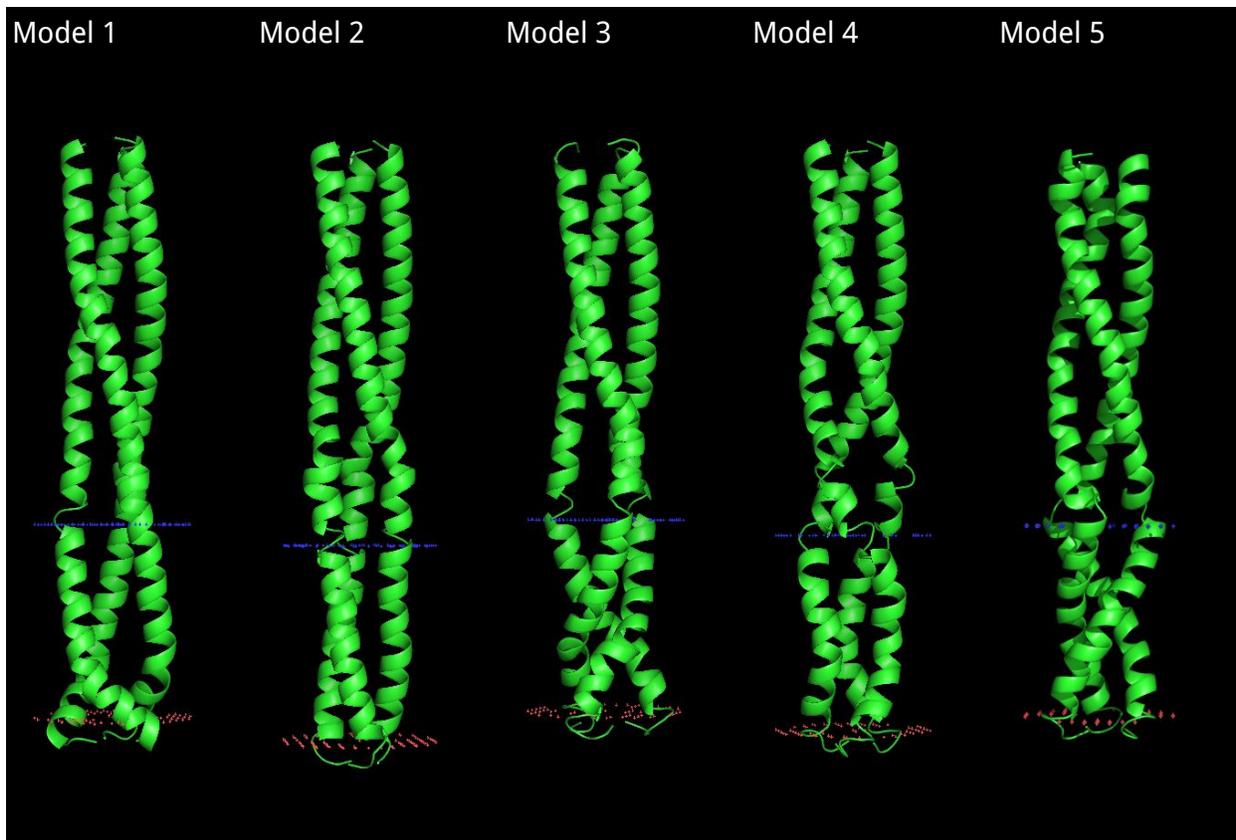


Figure 1. Structure of the selected models. Membrane outer face is indicated with blue dots, inner face with red dots. Models have been aligned to the same pose so the changes in membrane position/depth are evident.

Some of the properties of the selected models are summarized in Table 1. All models have a reasonably good Q-MEAN score, an excellent fit to the structure of 2FXP, and a reasonable fit to the structure of 2RUN. Model 1 has a surprisingly smaller transfer free-energy, indicating less good binding to the membrane (reported energies for other viral fusion proteins are in the range of -20 to -60 kcal/mol), as well as a smaller energy of association between the helices. Model 5 has a sharp left hand twist of the MPER around the axis of symmetry, and steeply crossed right hand transmembrane helix bundle; it has the best Rosetta score, but represents a small cluster.

Property	Model 1	Model 2	Model 3	Model 4	Model 5
Score	-821.7 REU	-832.7 REU	-813.4 REU	-801.5 REU	-839.7 REU
Cluster size	339	386	100	66	4
Quality, QMEAN4 Z-score	-0.68	-1.27	-1.74	-1.24	-2.04
Fit to 2FXP	0.89 A	0.67 A	0.61A	0.80 A	1.68 A
Fit to 2RUN	1.81 A	2.59 A	1.47 A	2.21 A	3.51 A
Membrane $\Delta G_{\text{transfer}}$	-24.2 kcal/mol	-40.6 kcal/mol	-38.2 kcal/mol	-33.3 kcal/mol	-37.8 kcal/mol
Transmembrane helices $\Delta G_{\text{association}}$	-1.6 kcal/mol	-3.2 kcal/mol	-4.3 kcal/mol	-3.7 kcal/mol	-6.4 kcal/mol
Embedded residues	1212-1234, 1236	1216-1237	1212-1237	1212-1237	1215-1238, 1240
MPER twist angle	-59 deg	-68 deg	-21 deg	-128 deg	-84 deg
Transmembrane helix tilt angle	-18.4 deg	-10.6 deg	+30.8 deg	-4 deg (bent)	+31.2 deg

Table 1. Properties of the selected models. Model score from RosettaMP. Free energies calculated by the PPM server, QMEAN score by SWISS-MODEL.

The relative orientation and packing of some key aromatic residues in the MPER is shown in Figure 2, including tryptophans W1212, W1214, W1217 and tyrosines Y1209 and Y1215.

In Model 1, W1212, Y1209 and Y1215 are facing out, turned up towards the surface of the bilayer; W1214 and W1217 from all chains are packed together in a pocket between the chains in approximate T configuration. Because this model is embedded deeper in the membrane, these tryptophans may be deep enough that they wouldn't interact with the headgroups/bilayer surface if they were turned outward, and it is more favorable for them to be in the interior.

In Model 2, W1214, W1217 and W1212, Y1215 on the adjacent chain are all packed together (more loosely) and all facing sideways and out. This model is shallower than model 1, and all aromatics appear to be at the right depth to interact with the lipid headgroups.

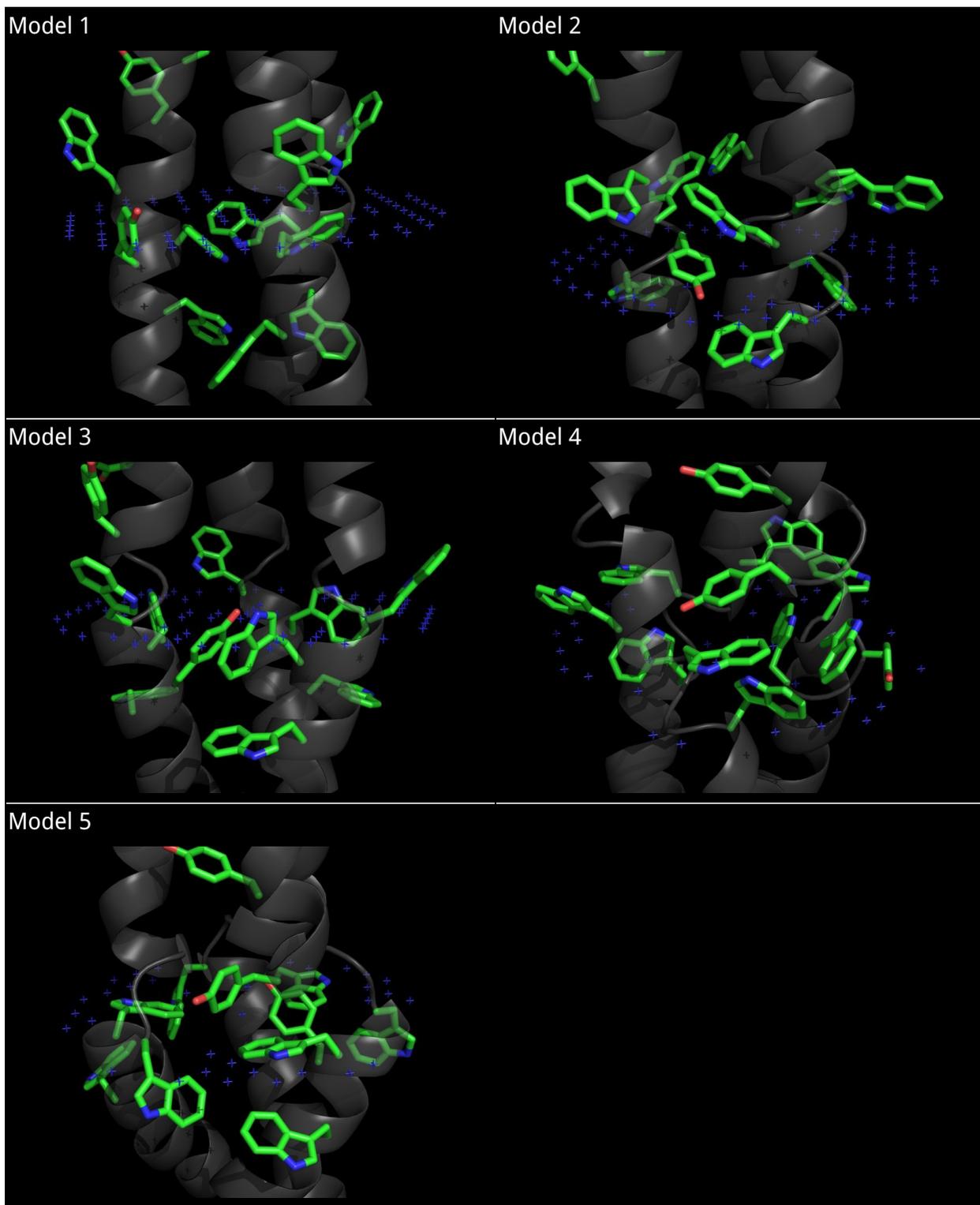


Figure 2. Aromatic residue orientation and packing in the MPER of the selected models.

In Model 3, W1212 and Y1209 are packed together, turned up and facing out. W1214, W1217 and Y1215 on the adjacent chain are packed together, but Y1215 is in the interior while the tryptophans are partially exposed.

In Model 4, Y1209, W1212 and W1214, W1217 on the adjacent chain are all packed together and facing sideways and out. This is another configuration seated shallowly in the membrane, relatively similar to Model 2, although it packs Y1209 instead of Y1215 which is fully on the exterior.

In Model 5, Y1209, W1214, W1217 and W1212 on the adjacent chain are all packed together and facing sideways and out (and slightly down). Uniquely, the three tyrosines Y1215 are packed in the interior and hydrogen bonded in a triangle configuration.

It appears that there are multiple ways for these key aromatic residues to be arranged depending on the depth of the protein in the membrane and the bends of the backbone, but in general they all tend to pack together, preferably in an approximately right-angle (T) orientation, and possibly together with aromatic residues from neighboring chains. All the tryptophans are usually exposed or partly exposed (unless the protein is seated very deeply in the membrane); the tyrosines may be hidden or exposed. The typical configuration has three or four aromatic residues packed together in each groove between chains, and at least one tryptophan (typically W1212) fully exposed to the exterior.

Distribution of poses

The distribution of the score vs large-scale pose parameters for all the extended conformations (N=1940) is described in this section, in order to identify preferred poses and to study how large a score (or energy) penalty non-optimal conformations have. The score is separated into score contributions from each region (HR2, MPER and TMD), for clarity; the distribution is similar but less tightly clustered when considering total scores.

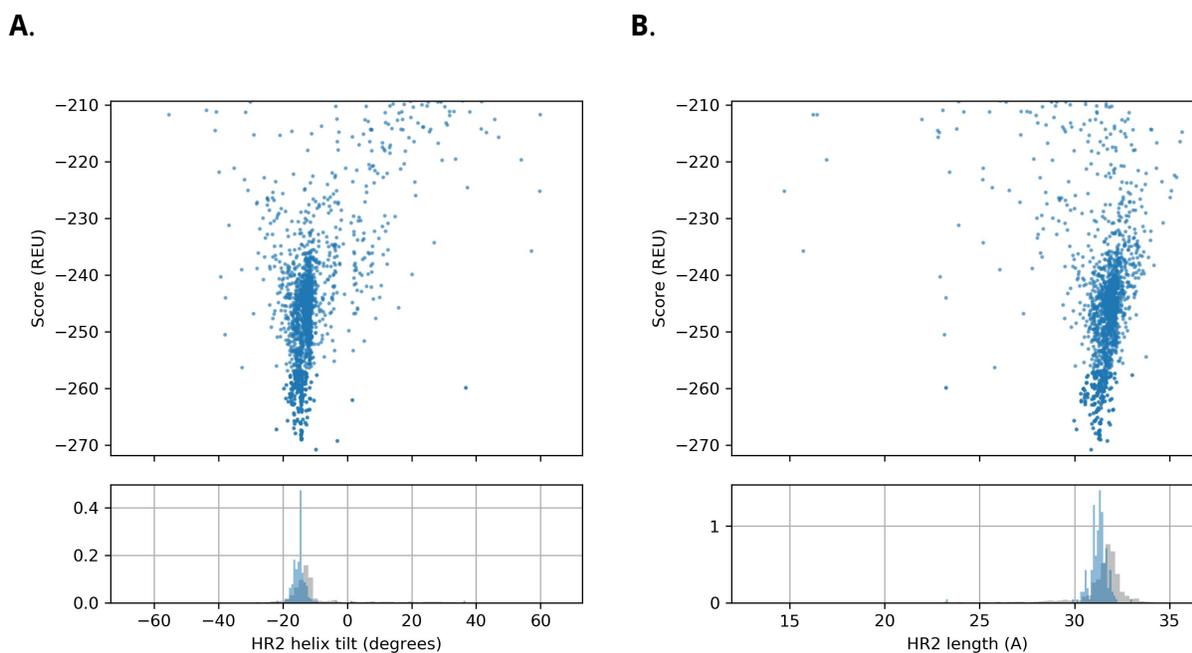


Figure 3. Distribution of HR2 conformations vs scores. Each dot is one model. Score in Rosetta Energy Units (REU) is the contribution from HR2 residues 1172-1196 only; tilt and length are measured from 1172-1175 to 1192-1196, length is along axis of symmetry. Histogram below scatter graph shows distribution of all poses (gray) and top 10 percent of poses (blue).

The HR2 segment 1172-1196 is most commonly in a twisted helix bundle configuration with a -14.7 (-16.0 to -14.2) degree left hand tilt (all measurements reported as median and interquartile range, negative numbers correspond to left handed), and length of 31.3 (31.0 to 31.5) angstroms measured along the axis of rotation, with very little variation in relative position (Figure 3A and 3B). Poses with slightly lesser twist and slightly greater length have significantly worse scores, indicating a steep energy well.

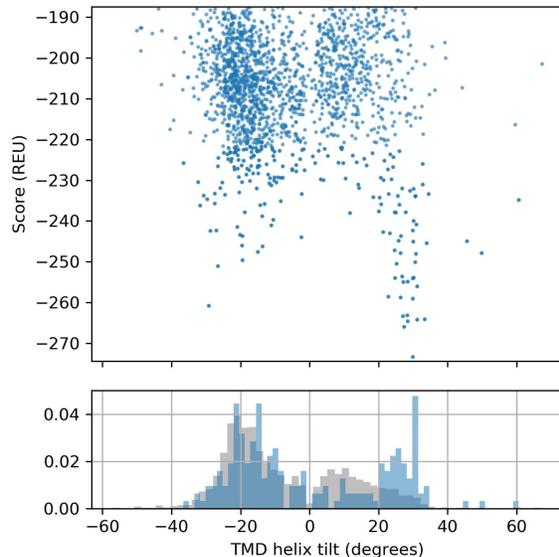
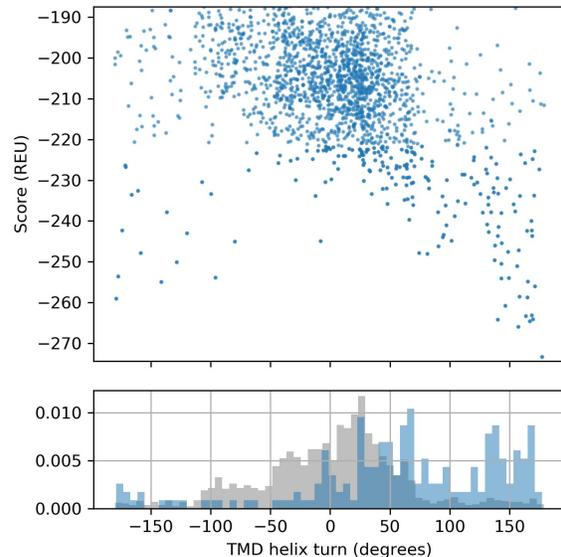
A.**B.**

Figure 4. Distribution of transmembrane helix (TMD) conformations vs scores. Score is the contribution from TMD residues 1214-1234, tilt is measured from 1214-1217 to 1231-1234 and turn is orientation of 1214 relative to the axis of symmetry.

The transmembrane helix has a relatively wide range of available tilts (Figure 4A), but most commonly has a -16.9 (-21.6 to -11.4) degree left hand or +25.0 (18.3 to 29.8) degree right hand tilts. Near-zero tilt angles are avoided. The lowest energy configurations are, perhaps surprisingly, right hand +30 degree tilts with a bend in each helix where it crosses over its neighbor.

The transmembrane helix is also apparently free to rotate around its axis (Figure 4B), with almost any orientation possible for a small energy cost (around 10-30 REU). This is likely due to the central region of the helix having only hydrophobic side chains which are easy to pack (1221-IAGLIAIVMV), and two glycines (G1219 and G1223) which enable bends of the helix. The bulkier F1220 and T1231 are near the ends where the distance between helices may be wider, and also may favorably interact with lipid headgroups. The most common orientation is 25 degrees (W1214 C-alpha towards center of symmetry), but the most favorable is around 150-180 degrees (W1214 directly outwards).

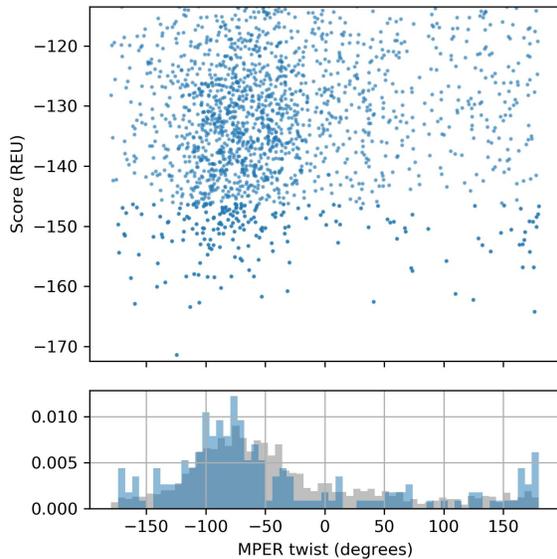
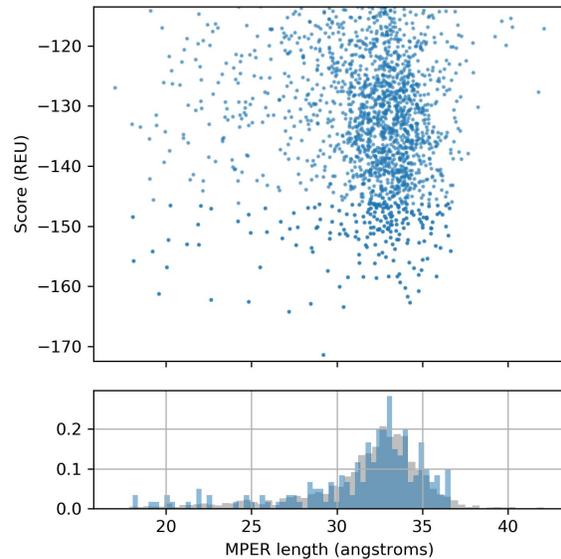
A.**B.**

Figure 5. Distribution of membrane-proximal external region (MPER) conformations vs scores. Score is the contribution from MPER residues 1197-1213 only; twist and length are measured from 1192-1196 to 1214-1217, where twist is rotation around the axis of symmetry, and length is distance along the axis of symmetry.

The MPER or hinge region appears to have a large number of roughly equally favorable large-scale conformations. The region typically spans around 33 (31 to 34) angstroms in length along the axis of symmetry (Figure 5B), but there are outlier structures between 20 and 40 angstroms; some of the shortest structures score as well as the typical length. Left hand twists around the axis of symmetry around -75 degrees are more common, but substantially any twist angle between 0 and +180 is possible (Figure 5A). As the hinge twists, the end of the HR2 helix bundle opens slightly and the transmembrane helix tilts, and at higher twist angles the HR2 helical structure may unravel partially producing a wider hinge.

Discussion

Model quality

The overall quality of these models is reasonable, as seen from good fit to known structure fragments, favorable membrane free energies and fairly good QMEAN scores. However, there are some features which suggest that these may not be the same as the most common native conformation. All of these models expose the hydrophobic I1198 entirely to solvent, which does not lead to a very large Rosetta score penalty (and matches 2FXP) but appears unlikely for a native structure. Secondary structure prediction software (RaptorX-structure, PSIPRED, PROFSEC) predicts turns or loops at 1196-1199 and 1204-1211, whereas the models here only have moderate bends at those residues in an overall helical structure. It is possible that there is a more favorable conformation which involves sharper turns (perhaps allowing a helical segment that lies flat to the surface of the membrane) that was simply not sampled in the current folding strategy. It is also possible that the best conformation(s) are not axially symmetric - the models here may be near-native, but only in the case when external mechanical forces act on the ectodomain to pull or bend it so that the HR2 helix is perpendicular to the membrane, whereas the best conformation may be asymmetric and tilted at a preferred angle.

It is also unclear to what extent the 2FXP structure itself is native-like. The fragment used in that work, as a pure peptide, tends to form insoluble aggregates in water. However, there are glycosylation sites at residue 1158, 1173 and 1194 (Watanabe 2020) which may be able to prevent aggregation of these domains of the native glycoprotein.

Flexibility and function of MPER

The MPER may be an inherently flexible or hinge region of the protein, since an excessively rigid connection could cause the spike protein to deform the membrane and/or detach from the membrane when external forces (for example hydrodynamic) act on the bulky ectodomain. The MPER may thus be expected to allow both a significant degree of bending and (as seen here) rotation. These large-scale motions should be accommodated without denaturation; the MPER hinge should be able to recover its original structure after repeated bending and straightening cycles.

There are a number of structural features which appear to allow this hinge action, including bend-promoting residues including proline P1213 and glycines G1204, G1219 and G1223; and the multiple possible arrangements of tryptophans W1212, W1214, W1217 and tyrosines Y1209 and Y1215, which may be rotated to the outside of the MPER hinge or packed inside;

may pack in an approximately T-shaped (edge to face) or offset planar configuration; and may also rotate or "snorkel" depending on depth in the membrane, while in general favoring the interfacial region.

In addition to acting as a hinge, the MPER may induce negative curvature in membranes, as well as have a detergent-like effect in lowering the rupture tension to allow pore formation. However, this may be an effect that is only triggered during fusion, as destabilizing the viral envelope under other conditions may be disadvantageous. The pathway of steps during fusion is shown in Figure 6 (reproduced from White 2008).

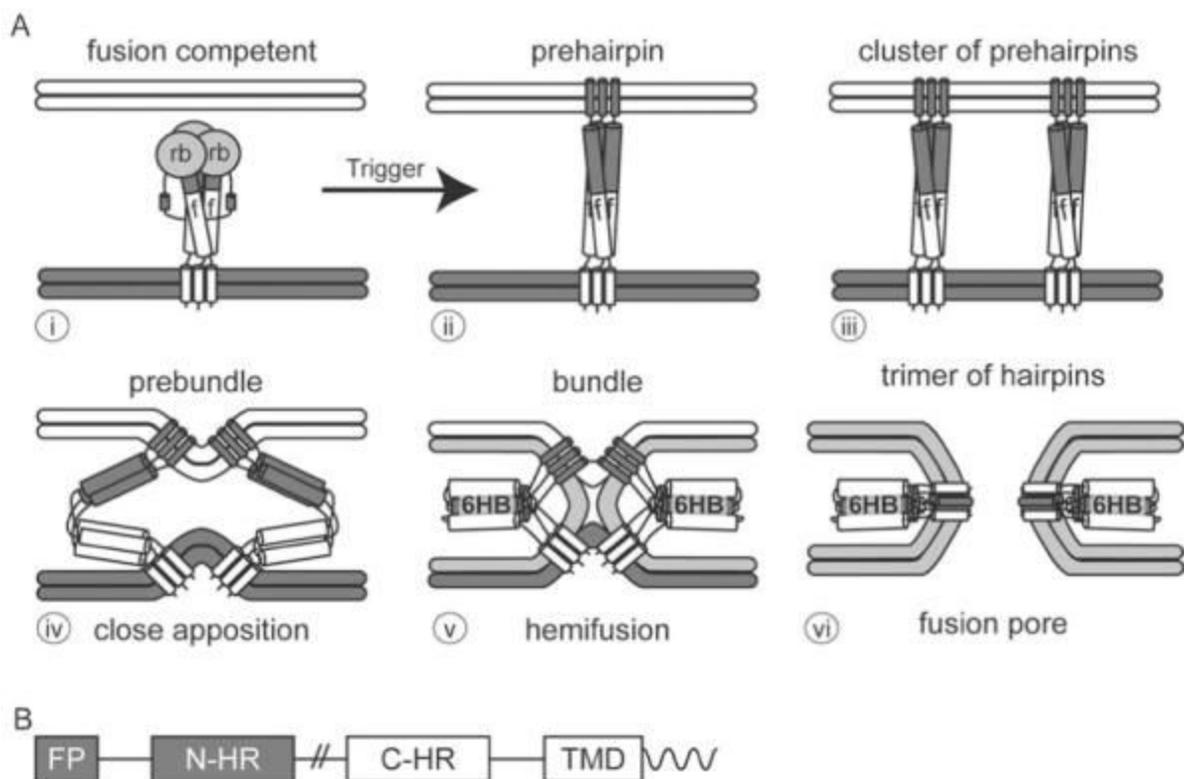


Figure 6. The common trimer-of-hairpins pathway of membrane fusion (reproduced from White 2008).

According to that model, during the pre-bundle (close apposition) step, the HR2 still exists as a trimer, but the MPER is bent sharply. During the next step, bundle (hemifusion), the HR2 and HR1 form a six-helix bundle structure (6HB). In the 6HB, the trimer of helices which form HR2 separate from each other and reassemble along the outside of the helices of HR1. It is unclear if the transmembrane domains also separate at that stage, but the interior HR1 helices would be expected to force at least part of the MPER to a wider conformation. During the last step, the trimer-of-hairpins (fusion pore formation), the transmembrane domain and N-terminal fusion peptide are shown assembled together in the newly formed pore edge. This sequence

suggests that the negative curvature inducing effect of the MPER may be triggered by mechanical bending. The rupture-tension lowering effect of the MPER may be produced by a distinct MPER configuration formed only when the HR2 helices dissociate and are arranged around the outside of HR1; or it may also be triggered by more severe bending forces.

Considering this mechanism of membrane fusion together with the observation of possible dynamic repacking of aromatic residues suggests the possibility that bending or flexing of the MPER may force aromatic side chain rearrangements which may then act to destabilize the membrane; thus the fusogenic activity of the MPER may be driven by mechanical bending forces.

The transmembrane domain and MPER may also play a role during trimerization of the spike protein. There may exist a native monomer conformation which is stable but readily forms dimers and or trimers, possibly by starting from association of the TMD helices and then "zipping up" the other domains. The flexibility of the MPER may also be important in allowing assembly of the external domains in the right orientation once the TMD helices have bound together.

Most recently, additional cryo-EM evidence has emerged which confirms the flexibility of the MPER junction (after the modelling work described here was completed). In (Ke 2020), a large number of cryo-EM images of spike on intact virions show a range of angles relative to the membrane, with roughly equal number of spike proteins imaged at any angle between 0 (perpendicular) and ± 50 degrees. A smaller fraction of spike proteins is seen with very sharp bends, up to ± 90 degrees. Unfortunately the exact structure of the hinge cannot be resolved in those images, but the data suggests the stem (HR2) is mostly straight, and the bending occurs at one or two specific regions (probably the MPER identified here, and another region near the other end of the stem). Utilizing that data to build improved models will be the subject of follow-on work.

Conclusion

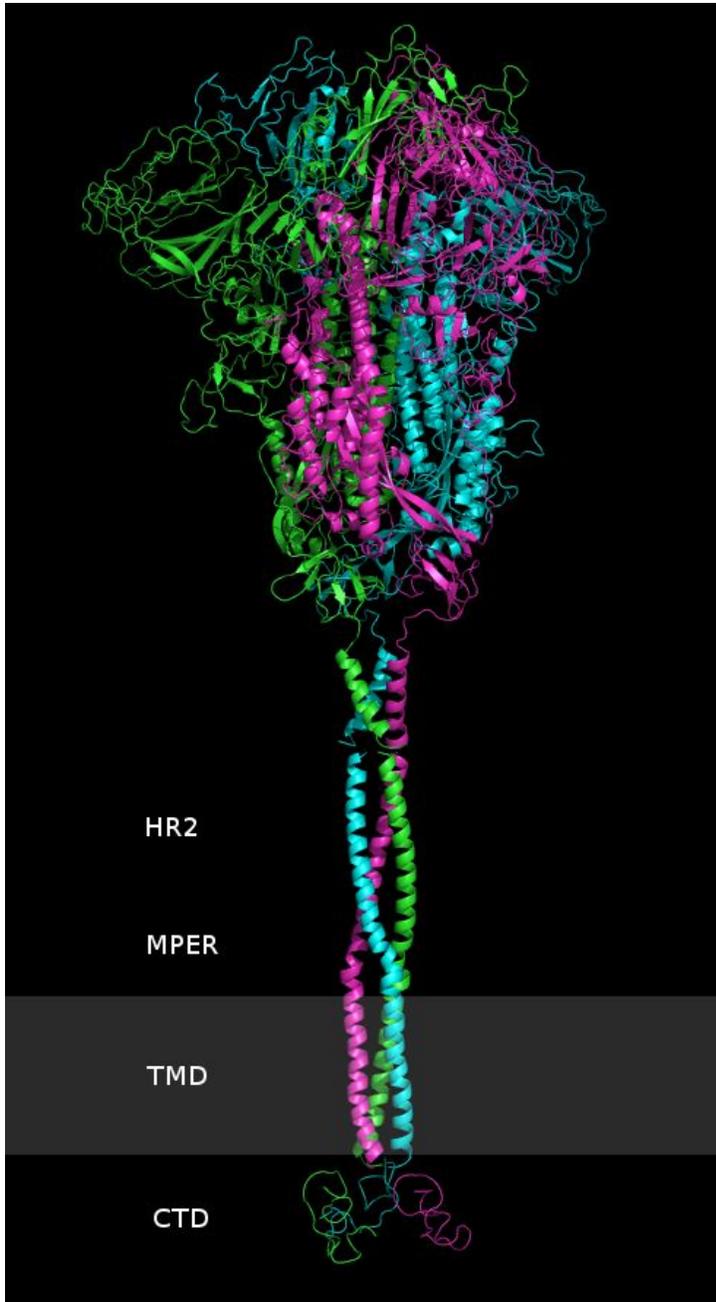


Figure 7. Whole spike protein (predicted). HR2, MPER and TMD residues 1171-1240 from Model 2 in this paper; ectodomain including residues 1-1160 and CTD residues 1241-1272 from Zhang lab's 2020-05-13 model. There are two small gaps: residues 1161-1171 are

missing, and 1240-1241 are not connected. The gray region is the approximate position of the membrane.

If there is a single, strongly preferred conformation of the MPER and TMD, the present work could not identify it.

It is perhaps better to understand the MPER not as a single stable structure, but as an ensemble of structures which transition into one another depending on the forces acting on that region and on the relative orientation of the ectodomain and the membrane. On one end, the HR2 appears to assemble into a fixed conformation, with the three helices bound tightly with hydrophobic interactions and constrained not to rotate; on the other end, the transmembrane helices are anchored in the membrane by their hydrophobic surfaces and likely by multiple palmitoyl groups embedded in the interior leaflet, but appear free to rotate, bend and tilt while staying in close contact. In between, the three chains bend to fold compactly into an aromatic exterior / hydrophobic core region, where - depending on the position/depth in the membrane and the bend angle - the aromatic residues may pack into several aromatic clusters.

In follow on work, tilts or bends away from the axis of symmetry should be explored, as well as the structural rearrangements during gradual bending. The stability and dynamic behavior of these structures should be investigated with molecular dynamics, including the effect of the MPER on the membrane curvature and stability, the way that different configurations transition into one another as the MPER bends, and possible hemi- or post-fusion configurations. Hopefully, additional experimental data from NMR or cryo-EM would be available to refine these models.

The HR2-MPER region of coronavirus spike protein is worth studying in greater detail as it would be important for developing successful therapies, including vaccines, broadly neutralizing antibodies (bNAbs) and fusion inhibitor peptides. This region is highly conserved, is critical for the action of the fusion machinery, and is already being explored as a therapeutic target (Xia 2020, Elshabrawy 2012). A complete understanding of the structure and dynamic changes of the MPER is essential for rational design of therapeutics.

Data availability

All data and code is available from: <https://osf.io/76fa2/>

References

Ke Z, Oton J, Qu K. *et al.* Structures and distributions of SARS-CoV-2 spike proteins on intact virions. *Nature* (2020). <https://doi.org/10.1038/s41586-020-2665-2>

Zhang Lab (University of Michigan). Modeling of the SARS-CoV-2 Genome using I-TASSER. Retrieved: May 13, 2020. Website: <https://zhanglab.ccmb.med.umich.edu/COVID-19/>

Baker Lab (Institute for Protein Design, University of Washington). Modeling the viral proteome. Retrieved: May 13, 2020. Website: <http://new.rosetta.org/results.php?id=15652>

Feig Lab (Michigan State University). Spike/M protein complex model. 2020. Website: <https://github.com/feiglab/sars-cov-2-proteins>

SWISS-MODEL. Spike glycoprotein (S glycoprotein). Retrieved: May 13, 2020. Website: <https://swissmodel.expasy.org/interactive/7dVLxC/models/01>

Jumper J *et al.* (DeepMind). Computational predictions of protein structures associated with COVID-19 . Website: <https://deepmind.com/research/open-source/computational-predictions-of-protein-structures-associated-with-COVID-19>

White JM, Delos SE, Brecher M, Schornberg K. Structures and mechanisms of viral membrane fusion proteins: multiple variations on a common theme. *Critical reviews in biochemistry and molecular biology*. 2008 Jan 1;43(3):189-219.

Xia S, Zhu Y, Liu M, Lan Q, Xu W, Wu Y, Ying T, Liu S, Shi Z, Jiang S, Lu L. Fusion mechanism of 2019-nCoV and fusion inhibitors targeting HR1 domain in spike protein. *Cellular & molecular immunology*. 2020 Feb 11:1-3.

Walls AC, Tortorici MA, Snijder J, Xiong X, Bosch BJ, Rey FA, Veesler D. Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion. *Proceedings of the National Academy of Sciences*. 2017 Oct 17;114(42):11157-62.

Guillén J, Kinnunen PK, Villalaín J. Membrane insertion of the three main membranotropic sequences from SARS-CoV S2 glycoprotein. *Biochimica et Biophysica Acta (BBA)-Biomembranes*. 2008 Dec 1;1778(12):2765-74.

Petit CM, Chouljenko VN, Iyer A, Colgrove R, Farzan M, Knipe DM, Kousoulas KG. Palmitoylation of the cysteine-rich endodomain of the SARS-coronavirus spike glycoprotein is important for spike-mediated cell fusion. *Virology*. 2007 Apr 10;360(2):264-74.

Hakansson-McReynolds S, Jiang S, Rong L, Caffrey M. Solution structure of the severe acute respiratory syndrome-coronavirus heptad repeat 2 domain in the prefusion state. *Journal of Biological Chemistry*. 2006 Apr 28;281(17):11965-71.

Mahajan M, Bhattacharjya S. NMR structures and localization of the potential fusion peptides and the pre-transmembrane region of SARS-CoV: Implications in membrane fusion. *Biochimica et Biophysica Acta (BBA)-Biomembranes*. 2015 Feb 1;1848(2):721-30.

Liao Y, Zhang SM, Neo TL, Tam JP. Tryptophan-dependent membrane interaction and heteromerization with the internal fusion peptide by the membrane proximal external region of SARS-CoV spike protein. *Biochemistry*. 2015 Mar 10;54(9):1819-30.

Lu Y, Neo TL, Liu DX, Tam JP. Importance of SARS-CoV spike protein Trp-rich region in viral infectivity. *Biochemical and biophysical research communications*. 2008 Jul 4;371(3):356-60.

Basso LG, Vicente EF, Crusca Jr E, Cilli EM, Costa-Filho AJ. SARS-CoV fusion peptides induce membrane surface ordering and curvature. *Scientific reports*. 2016 Nov 28;6:37131.

Sainz B, Rausch JM, Gallaher WR, Garry RF, Wimley WC. The aromatic domain of the coronavirus class I viral fusion protein induces membrane permeabilization: putative role during viral entry. *Biochemistry*. 2005 Jan 25;44(3):947-58.

Benton DJ, Nans A, Calder LJ, Turner J, Neu U, Lin YP, Ketelaars E, Kallewaard NL, Corti D, Lanzavecchia A, Gamblin SJ. Influenza hemagglutinin membrane anchor. *Proceedings of the National Academy of Sciences*. 2018 Oct 2;115(40):10112-7.

Li W, Minova-Foster TJ, Norton DD, Muggeridge MI. Identification of functional domains in herpes simplex virus 2 glycoprotein B. *Journal of virology*. 2006 Apr 15;80(8):3792-800.

Albertini A, Bressanelli S, Lepault J, Gaudin Y. Structure and working of viral fusion machinery. In *Current topics in membranes* 2011 Jan 1 (Vol. 68, pp. 49-80). Academic Press.
<https://www.sciencedirect.com/science/article/pii/B9780123858917000039>

Liu S, Jing W, Cheung B, Lu H, Sun J, Yan X, Niu J, Farmer J, Wu S, Jiang S. HIV gp41 C-terminal heptad repeat contains multifunctional domains relation to mechanisms of action of anti-HIV peptides. *Journal of Biological Chemistry*. 2007 Mar 30;282(13):9612-

Hollingsworth IV LR, Lemkul JA, Bevan DR, Brown AM. HIV-1 Env gp41 Transmembrane Domain Dynamics Are Modulated by Lipid, Water, and Ion Interactions. *Biophysical journal*. 2018 Jul 3;115(1):84-94.

Watanabe Y, Bowden TA, Wilson IA, Crispin M. Exploitation of glycosylation in enveloped virus pathobiology. *Biochimica et Biophysica Acta (BBA)-General Subjects*. 2019 May 20.

Kim DE, Chivian D, Baker D. Protein structure prediction and analysis using the Robetta server. *Nucleic acids research*. 2004 Jul 1;32(suppl_2):W526-31.

Das R, André I, Shen Y, Wu Y, Lemak A, Bansal S, Arrowsmith CH, Szyperski T, Baker D. Simultaneous prediction of protein folding and docking at high resolution. *Proc Natl Acad Sci U S A*. 2009 Nov 10;106(45):18978-83. Epub 2009 Oct 28.

Koehler Leman J, Mueller BK, Gray JJ. Expanding the toolkit for membrane protein modeling in Rosetta. *Bioinformatics*. 2017 Mar 1;33(5):754-6.

Lomize M.A., Pogozheva I,D, Joo H., Mosberg H.I., Lomize A.L. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res.*, 2012, 40(Database issue):D370-6

Yifan Song, Frank DiMaio, Ray Yu-Ruei Wang, David Kim, Chris Miles, TJ Brunette, James Thompson and David Baker. High resolution comparative modeling with RosettaCM. *Structure*. 2013 Oct 8;21(10):1735-42.

Benkert P, Biasini M, Schwede T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*. 2011 Feb 1;27(3):343-50.

Epand RM. Fusion peptides and the mechanism of viral fusion. *Biochimica et Biophysica Acta (BBA)-Biomembranes*. 2003 Jul 11;1614(1):116-21.

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. *Virology journal*. 2009 Dec 1;6(1):230.

Watanabe Y, Allen JD, Wrapp D, McLellan JS, Crispin M. Site-specific analysis of the SARS-CoV-2 glycan shield. *BioRxiv*. 2020 Jan 1.

Elshabrawy HA, Coughlin MM, Baker SC, Prabhakar BS. Human monoclonal antibodies against highly conserved HR1 and HR2 domains of the SARS-CoV spike protein are more broadly neutralizing. *Plos one*. 2012;7(11).