Integrative structural studies of the SARS-CoV-2 spike protein during the fusion process (2022)

Jacob C. Miner [a], Paul W. Fenimore [a], William M. Fischer [a], Benjamin H. McMahon [a], Karissa Y. Sanbonmatsu [a,b], Chang-Shung Tung [a,*]

[a] *Theoretical Biology & Biophysics Group, Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM, 87545, USA*
[b] *New Mexico Consortium, Los Alamos, NM, 87544, USA*

ARTICLE INFO

ABSTRACT

SARS-CoV-2 is the virus responsible for the COVID-19 pandemic and catastrophic, worldwide health and economic impacts. The spike protein on the viral surface is responsible for viral entry into the host cell. The binding of spike protein to the host cell receptor ACE2 is the first step leading to fusion of the host and viral membranes. Despite the vast amount of structure data that has been generated for the spike protein of SARS-CoV-2, many of the detailed structures of the spike protein in different stages of the fusion pathway are unknown, leaving a wealth of potential drug-target space unexplored. The atomic-scale structure of the complete S2 segment, as well as the complete fusion intermediate are also unknown and represent major gaps in our knowledge of the infectious pathway of SAR-CoV-2. The conformational changes of the spike protein during this process are similarly not well understood. Here we present structures of the spike protein at different stages of the fusion process. With the transitions being a necessary step before the receptor binding, we propose sites along the transition pathways as potential targets for drug development.

1. Introduction

To date, more than 400 million people in more than 190 countries have been confirmed infected by the novel coronavirus disease of 2019 (COVID-19), according to the Johns Hopkins Coronavirus Resource Center. The emergence of highly contagious variants with mutations in the spike protein (e.g. delta, delta+, and omicron subvariants BA.1 and BA.2) demonstrate the importance of a detailed mechanistic understanding of the spike protein, and the viral entry process. To address this crisis, the scientific and healthcare communities are directing massive efforts to mitigate the spread of COVID-19 and develop novel treatment options for those infected. COVID-19 is caused by a β-coronavirus known as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), an enveloped virus containing a positive-sense, single-stranded RNA (Gorbalenya et al., 2020; Kim et al., 2020), with protruding spike (S) proteins on its surface that form arrangements reminiscent of 'coronas' (Li, 2015). Viral entry into host cells is facilitated by these S proteins forming trimer arrangements and binding to surface proteins of the host cell [4].

There is abundant structural information for the S protein (>250 entries deposited in PDB). Since the COVID-19 pandemic outbreak began in 2019, structures of the SARS-CoV-2 S protein in *apo* form, or complexed with antibodies or receptors, have been solved using x-ray crystallography and cryo-EM (Pinto et al., 2020; Schoof et al., 2020; Wang et al., 2020; Wrapp et al., 2020). As with other glycoproteins, the S protein has at least one receptor-binding domain (RBD) that allows the attachment of the virus to the surface of the targeted cell and a helical domain that allows a conformational change during the pre-fusion to post-fusion transition, enabling the fusion peptide region (FP) to reach the host cell surface. The polypeptide chain of S protein can be divided into separate domains (Fig. 1) based on each functional role (Wrapp et al., 2020; Xia et al., 2020a). A close inspection of the SARS-CoV-2 S protein structure (e.g., PDB ID: 6VSB) shows that this protein consists of four well-folded domains (S1a, S1ab, S1b and S2). Based on different structures of the S protein (PDB IDs: 2AJF, 3KBH, 4KR0, 4F5C and 3R4D), we observe that despite variation in the domain arrangements, each of the domains retains its fold. Here, we focus on these structures and structural changes.

Host entry of SARS-CoV-2 depends on the S protein transitioning from a pre-fusion conformation to one that draws the viral membrane closer to the host cell membrane and allows membrane fusion. These transitions

---
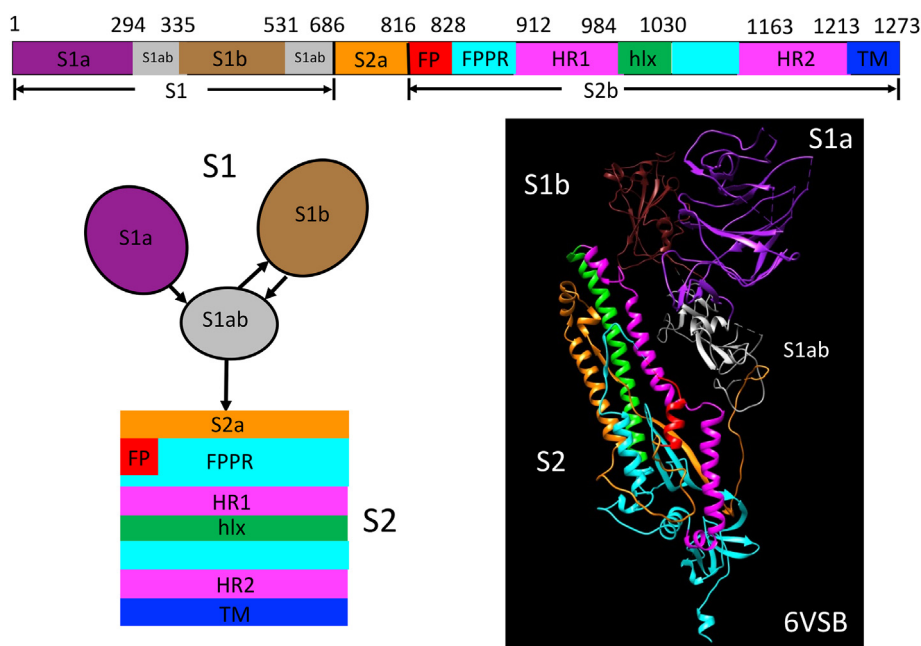
Fig. 1. Definition of folding domains of the S protein. Top, linear sequence depicting subdomains. Different segments of the protein relate to different aspects of spike protein function. Bottom left, schematic of domain organization. Bottom right, 3D structure of the S protein structure (PDB_ID: 6VSB). The domain colors are the same in the three images. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

involve (i) S protein binding of the host receptor angiotensin-converting enzyme 2 (ACE2), (ii) S protein binding to the host membrane, and (iii) S protein contraction to bring the viral envelope to the host membrane surface. Each of these steps requires that the S protein adopt a unique conformation, presenting potential drug targets: different drugs might be necessary to halt viral infection at different stages of the fusion process. Taken together, the S protein is a critical, but highly variable target for therapeutic development. Robust treatment protocols are dependent on a clear understanding of S protein structures at all stages of the viral-host fusion cycle.

Regarding the pre-fusion to post-fusion process for SARS-COV-2, the S protein must be cleaved by a host protease (e.g., furin) (Bertram et al., 2011; Hoffmann et al., 2020), resulting in segments S1 (residues 1–685) and S2 (residues 686–1273). The S1 segment is responsible for cell receptor binding, while the S2 segment undergoes most of the structural transitions required for fusion between the host cell and the virus.

The S1 segment itself can be functionally divided into the S1a, S1b, and S1ab domains. The *in vivo* function of the S1a domain of SARS-CoV-2 S protein is still unclear, though *in silico* studies have shown binding to the sialic acid receptor (Milanetti et al., 2020; Fantini et al., 2020), and cryo-EM structures of MERS-CoV S protein have also shown binding to sialic-acid ligands (Pike et al., 2019). The S1b domain of other related SARS-CoV viruses recognizes at least four different types of cell receptors: (i) angiotensin-converting enzyme 2 (ACE2) (Li et al., 2005; Wu et al., 2009; Belouzard. et al., 2012; Song et al., 2018; Lu et al., 2013; Wang et al., 2013), (ii) dipeptidyl peptidase 4 (DPP4) (Song et al., 2018; Lu et al., 2013), (iii) aminopeptidase N (APN) (Reguera et al., 2012) and (iv) carcinoembryonic antigen-related cell adhesion molecule 1a (CEA-CAM1a) host cell receptors (Peng et al., 2011). In SARS-CoV-2, S1b recognizes ACE2, but not APN or DPP4. CEACAM is upregulated upon SARS-CoV-2 infection (Sharif-Askari et al., 2021). The S1b structure is flexible and can adopt either "up" or "down" conformations in the trimer arrangement (Wrapp et al., 2020; Li et al., 2005). This structural flexibility is associated with receptor-antibody binding functional roles for the S protein. Due to potential clashes between the receptor and S protein, the S1b domain can only bind to ACE2 while in the "up" position in a trimer arrangement.

A second protease (TMPRSS2) cleavage site between residues 815 and 816 is responsible for cleaving segment S2 into S2a and S2b domains, enabling the structural transition of S2 (Belouzard et al., 2009; Jaimes et al., 2020; Ord et al., 2020). The extension of the S2 segment from pre-fusion to fusion intermediate states – sometimes referred to as a "spring-loaded mechanism" (Carr and Kim, 1993; Carr et al., 1997) – allows the FP region (residues 816–828 in S2a) to bind the host cell membrane.

To go from a pre-fusion state to a post-fusion state, the S2 segment of the S protein has to go through a complex, large-scale structural transition. Conceptually, we can divide the pre-fusion to post-fusion transition of the S2 segment into two steps: (1) S2 adopting an extended conformation that can physically connect the viral and the host cell membranes, and (2) bending of the HR2 helices toward the HR1 triple-helix domain to form a 6-helix bundle (Walls et al., 2017; Xia et al., 2020b), bringing the viral membrane into the close proximity with the host cell membrane and initiating membrane fusion.

In spite of copious structural data for the S protein of SARS-CoV-2 (Pinto et al., 2020; Schoof et al., 2020; Wang et al., 2020; Wrapp et al., 2020), many of the atomic-scale structures of the S protein in different stages of the fusion pathway remain unknown, leaving large swaths of drug-target space unexplored. The atomistic structure of the complete S2 segment, including fusion peptide regions, and the complete fusion intermediate are also unknown and represent significant gaps in our knowledge of the infectious pathway of SAR-CoV-2 (Gur et al., 2020).

The focus of this study is on filling the knowledge gaps of S protein structure and developing accurate, atomic-scale models of this crucial viral protein. Additionally, we seek to address the functional implications of the structural transitions from pre-fusion to post-fusion states and identify ways to target drugs to specific conformational states. Leveraging high performance computing resources and coding methodologies developed at Los Alamos National Laboratory with the wealth of structures related to the SARS-CoV-2 S protein available in the Protein Data Bank (PDB), we model the different transition states of SARS-CoV-2 S protein and help elucidate the large-scale protein conformational changes involved in the process of viral-host fusion for the SARS-CoV-2 S protein. Additionally, we identify interactions between these conformations and multiple binding agents, including nanobodies and antiviral drugs. This demonstrates the utility of these models for identifying

potential therapeutics for arresting these structural transitions in the viral life cycle.

2. Results

Based on current structural information, we can divide the fusion pathway for the S protein of SARS-Cov-2 into four different states: (1) pre-fusion 1, (2) pre-fusion 2, (3) fusion intermediate, and (4) post-fusion state. While at least two other pre-fusion states have been identified (Peng et al., 2021), as a first step, we consider two pre-fusion states. Transitions between states (1)–(4) are designated as transition 1, transition 2 and transition 3, respectively, and are described in the caption of Fig. 2. Large portions of the S protein in the pre-fusion-1, pre-fusion-2 and post-fusion states are known and available in the Protein Data Bank (PDB ID: 6VXX, 6VSB, 6M17, 6LZG, 6XRA) and are shown as the shaded areas of Fig. 2. The structural modeling of the remaining portions of the S protein in different stages of the transition (structures in unshaded regions of Fig. 2) as well as functional implications of the transitions from the pre-fusion to post-fusion states will be described in the following sections.

2.1. S protein in the pre-fusion state

Coronavirus S proteins have been rigorously studied since the first outbreak in 2003 with hundreds of structures being deposited in the PDB. These S protein structures share a similar fold and function as trimers. When arranged in a trimeric formation, the S1b domain of the S1 segment can adopt two different conformations ("up" and "down") (Schoof et al., 2020; Henderson et al., 2020). When all S1b domains of the trimeric arrangement are modeled in the "up" conformation, steric clashes prevent any individual S1b domain from binding to the ACE2 receptor. These clashes are avoided when one S1b is in the "up" conformation, and this RBD (S1b) in "up" conformation binds to the host ACE2 receptor. Here, we designate the pre-fusion 1 state as the trimer arrangement where all three S1b domains are in the "down" conformation, the pre-fusion 2 state as the trimer arrangement with a single S1b in the "up" conformation, and the transition from pre-fusion 1 to pre-fusion 2 as transition 1. The known structures of SARS-CoV-2 S protein include *apo* forms (e.g., PDB ID: 6VXX, 6VSB), receptor-bound complexes (PDB ID: 6LZG, 6VW1) and antibody-bound complexes (PDB ID: 6YLA, 6XDG, 6WP, 7BZ5). In either *apo* or complexed forms, arrangements of the individual domains, S1a and S1b, and S2, remain unchanged.

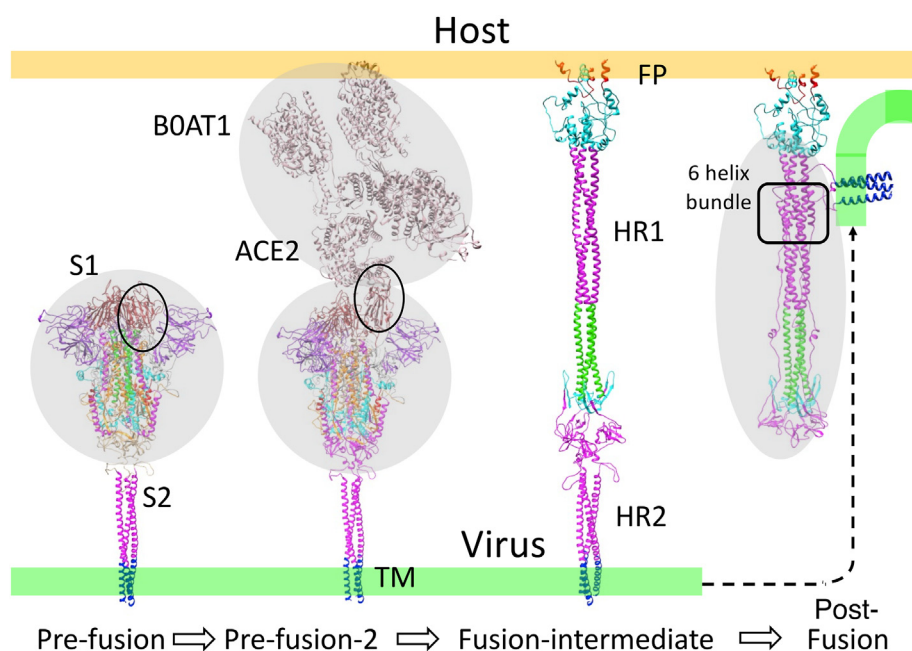### 2.2. Transition 1: pre-fusion 1 state to pre-fusion 2 state

Using structures of S protein from the Protein Databank (PDB_ID: 6VXX), we model a trimer structure comprising three SARS-CoV-2 S proteins (A, B and C), each with the S1b domain in the 'down' position. We find that the mammalian/human cell receptor ACE2 is unable to bind to the S protein trimer in this arrangement due to the steric hindrance that each of the three S1b domains exerts on its neighbors. In order to bind ACE2, one S1b domain from monomer A in the trimer must transition from a "down" to an "up" conformation. A small cleft between the S1b from monomer C and S1a from monomer B of the S protein trimer is just wide enough (~4.5 nm) to allow this transition to occur (Fig. 3a). Transition intermediates were generated through structural intercalation using a linear transition between the two states.

Since the structural transition of S1b from "down" to "up" is crucial to host cell receptor binding, blocking the transition would strongly interfere with the viral life cycle. Sites in the S protein trimer that are directly involved in the transition can serve as drug targets. Using this concept, nanobodies have been developed (PDB_ID: 7KKK) (Schoof et al., 2020) to bridge across the trimer S1b domains in the "down" conformation (Fig. 3b). By locking the S1b in the inactive conformation, this type of nanobody is able to block the receptor-binding of the S protein. With the transition being a necessary step before the receptor binding, we propose that sites along the transition pathway could also serve as potential targets for drug development.
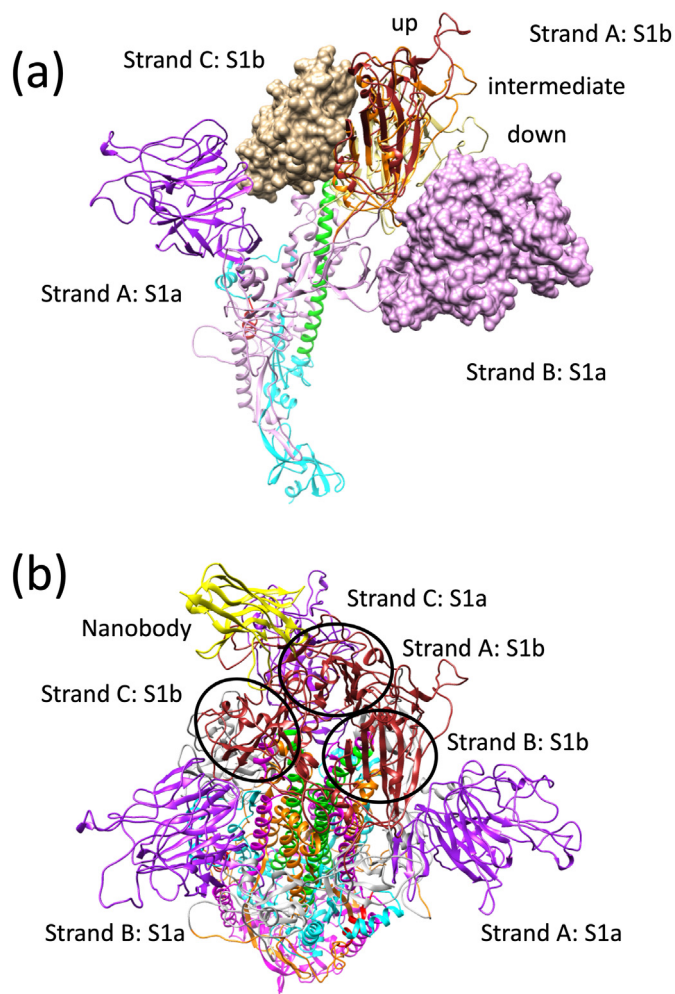
2.3. Transition 2: pre-fusion 2 state to fusion intermediate state

After binding the cell receptor, the S protein goes through a second, large conformational change (Fig. 2). During this transition, the S2 domain adopts an extended conformation to allow the fusion peptide (FP) – which was previously buried inside the trimer – to bind the host cell membrane. After cell proteases cleave S protein at the S1/S2 and S2' sites and remove the covalent linkage between the two functional subunits, the S1 crown is shed, and the FP is exposed (Carr et al., 1997).

A large portion of the S protein S2 segment in this extended state was solved using cryo-EM (PDB_ID: 6XRA) (Cai et al., 2020). Specifically, this

Fig. 2. Four different states of the S protein along the pre-fusion to post-fusion transition pathway. Grey shaded regions are known protein structures from the PDB; unshaded regions are not. Transition 1 (from the pre-fusion to the pre-fusion-2 state) involves the receptor-binding domain (RBD) going from a "down" to an "up" conformation (black ovals) in order to bind the host cell receptor ACE2 (B0AT1). Transition 2 (from the pre-fusion-2 state to the fusion-intermediate state) involves formation a long triple-helix with most of the S protein, as well as two heptad repeat domains (HR1 and HR2 in magenta). In this 'fusion-intermediate' state, the fusion peptide region (FP) of the S protein can insert into the host cell membrane. Transition 3 (from the fusion-intermediate state to the post-fusion state) involves the 6-helix bundle (black box) between the HR2 domain and the HR1 domain, which brings the viral membrane into close proximity with the host cell membrane and allows the initiation of membrane fusion.

Fig. 3. Transition 1 of S protein trimer from pre-fusion-1 to pre-fusion-2 states. (a) The S1b domain of strand A moves from a "down" (khaki/yellow) to an "intermediate" (orange), to an "up" (brown) conformation, transitioning through a cleft between S1a (magenta surface) and S1b (gold surface) from strands B and C respectively. (b) A nanobody (yellow) binds to two S1b domains (brown, strands A and C) and locks the two S1b domains in the "down" conformation, preventing the transition to pre-fusion-2 state. Color coding of the spike protein is following the definition shown in Fig. 1. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

data describes the core structure of the S2 in the transition intermediate state. However, both the N-terminal (residues 816–911) and C-terminal (residues 1198–1273) regions are not present in the cryo-EM model due to lack of strong density in the cryo-EM reconstruction. For the sake of completeness, we constructed atomistic models of these regions.

2.3.1. Structure of the N-terminal end of S2b for the fusion intermediate state

The fusion peptide region of the S protein is made of hydrophobic amino acids that insert into the cell membrane to induce viral host membrane fusion and subsequent entry of the viral genome into the cell (Granet, 2021). Using a computational approach, Sainz Jr et al. (Sainz et al., 2005) were able to identify a putative fusion peptide of the SARS-CoV S protein. Using sequence similarity, this putative fusion peptide is mapped to residues 770–788 of the SARS-CoV S protein, indicating that the fusion peptide maps to residues 788–806 in the SARS-CoV-2 S protein. This information can be found in functional maps of S protein (Xia et al., 2020b).

Millet. et al. argued that the fusion peptide of SARS-CoV corresponds

to region immediately following the S2' cleavage site (R-869/S-870) of the S protein (residues 870–896) (Miller and Whittaker, 2015). This information is consistent with the fusion peptide of the SARS-CoV-2 S protein residing between residues 816–842.

Based on these findings, the putative SARS-CoV-2 fusion peptide is located in the 788–842 region of the S protein. In the pre-fusion state structure (e.g., 6VSB), this region is located close to the base of the trimer. Additionally, residues 812–815 and 829–842 are missing in the solved structure (i.e. the deposited model corresponding to the cryo-EM map). However, structures of the fusion peptide from either murine hepatitis virus or influenza are available from the PDB (PDB_IDs: 3JCL and 1IBN, respectively), and in both structures, the fusion peptide takes the form a kinked helix. This type of arrangement was shown to be the functional structure of the fusion peptide (Lai et al., 2006).

We align the putative SARS-CoV-2 fusion peptide sequence from 6VSB with known structures of fusion peptides (PDB IDs: 3JCL and 1IBN) in Fig. 4 and use the structure of the fusion peptide from influenza (3JCL) as a template to model structure of SARS-CoV-2 fusion peptide (residues 886–896). In the pre-fusion structure, residues 788–806 are in a loop conformation while residues 816–828 are in a helical conformation.

2.3.2. Structure of the C-terminal end of S2b

The C-terminal region of S2, which includes the HR2 (residues 1197 to 1212) and TM domains (residues 1213 to 1237), is absent in all structures of the coronavirus S protein. Both the HR2 and TM domains are recognized for their ability to adopt a coil-coil conformation (Parry, 1982; Lupas et al., 1991). The solution structure of a SARS-CoV HR2 domain (PDB ID: 2FXP) shows a triple-helix arrangement. This sequence shows 97% homology with the corresponding HR2 domain (residues 1157 to 1201) of SARS-CoV-2. The pre-transmembrane sub-domain
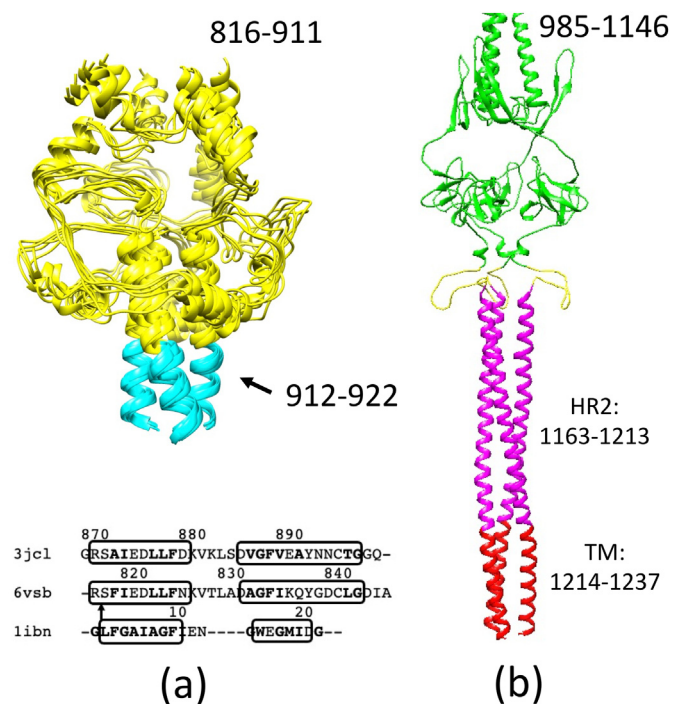

Fig. 4. Modeling of the N-terminal and C-terminal domains of the S2 region of the S protein of the fusion intermediate state. (a) Ensemble of structures (upper left) produced by flexible fitting of cryo-EM reconstructions of the post-fusion state of the N-terminal domain trimer including fusion peptide (FP) and fusion peptide proximal region (FPPR), and sequence alignment (lower left). Yellow, FP and FPPR; cyan, portions of HR1. (b) The HR2 (magenta) and TM (red) domains triple-helix structure are extended from the known structure of S2 (colored in green, PDB:6VSB). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

(residues 1189–1206) of a different SARS-CoV S protein adopts helical (residues 1191 to 1196) and a helical-like (residues 1198–1203) conformations. This domain maps to residues 1203–1220 of SARS-CoV-2 S protein and covers a region including the HR2/TM boundary. A BLAST search of the SARS-CoV-2 TM sequence (residues 1220–1232) shows 62% homology to a helical region (residues 100–112) of thermophilic rhodopsin (6KFQ) (Hayashi et al., 2020). Taken together, these structures and sequence sources indicate that the HR2-TM regions of the SARS-CoV-2 S protein adopt a triple-helix conformation, which we use to model this region of the S protein.

2.4. Transition 3: fusion intermediate state to post-fusion state

The transition from the fusion intermediate to post-fusion state requires another large conformational change (transition 3). This transition brings the HR2 helices into close proximity with the HR1 triple-helix, forming a six-helix bundle: a critical step in the fusion process (Walls et al., 2017; Xia et al., 2020b). By adopting a six-helix bundle, the TM region of the trimer is moved closer to the fusion-peptide (Fig. 2) and the viral membrane is brought closer to the host cell membrane to initiate host-virus fusion.

While known structures of the six-helix bundle with the core region of S2 segment have been characterized via cryo-EM (PDB ID: 6XRA) (Cai et al., 2020), the feasibility of the connections between the six-helix bundle and the TM triple-helix remains an open question. Coordinates were not deposited in the PDB for this region of the complex due to the weak cryo-EM density observed in regions corresponding to FP and FPPR. As a feasibility test, we model the loops that connect the six-helix bundle region to the TM region (Fig. 5). We also use phenix.cryo_fit to perform flexible fitting molecular simulations, obtaining structures and simulated cryo-EM maps of the full post-fusion complex ('post-fusion state'), including the FP and FPPR regions. The simulated cryo-EM maps are highly consistent with both strong and weak cryo-EM density experimentally measured by Cai et al. (Cai et al., 2020) (see Fig. 6).
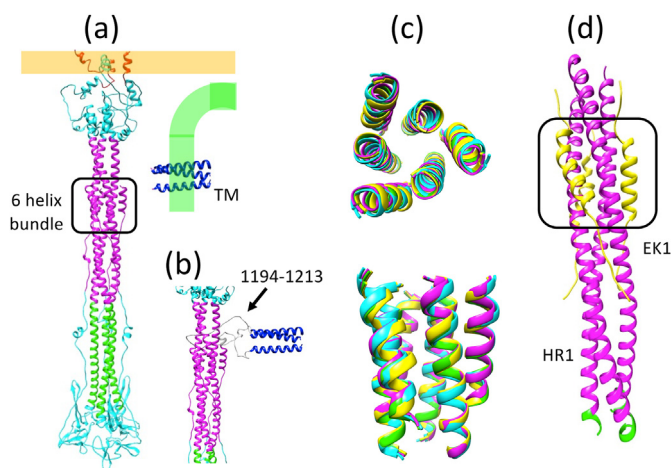
Fig. 5. Structure of the S protein trimer in the post-fusion state. (a) A 20 amino acid peptide loop (residues 1194–1213, shown explicitly in (b)) that connects the tail of HR2 in the six-helix bundle (black box) to the TM has limited the placement of the TM group (blue). As a result, it brings the cell membrane (yellow/orange rectangle) and viral membrane (thick light green curve) to close proximity. Magenta, HR1 and HR2; Cyan colored portion in upper part of image, FPPR; red, FP; green and cyan colored portions in lower part of image, region connecting HR1 and HR2. (b) The 20 aa peptide loop can easily bridge across the TM and six-helix bundle segments of the S protein. (c) The superposition of the six-helix bundle from SARS-CoV-2 (HR1/HR2), SARS (HR1/EK1), MERS (HR1/EK1) and 229E (HR1/EK2) shown in yellow, cyan, magenta and green, respectively. (d) A model of EK1 bound S protein HR1 six-helix bundle. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)
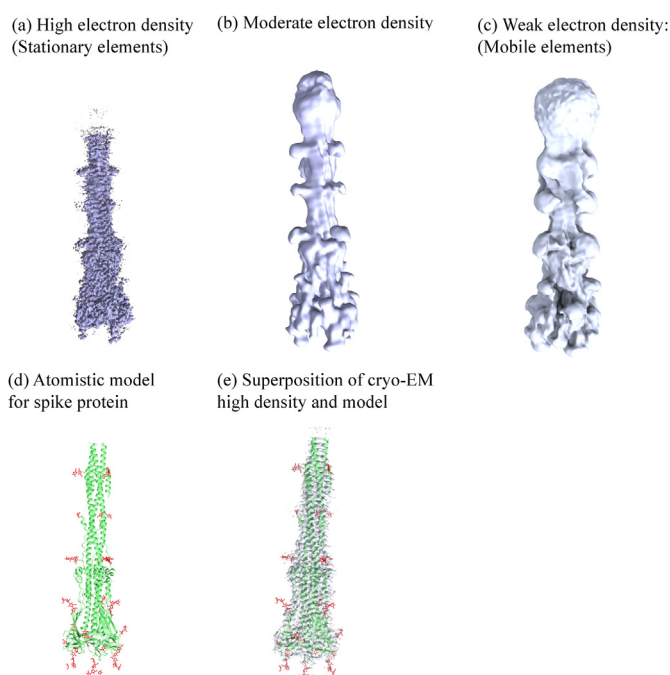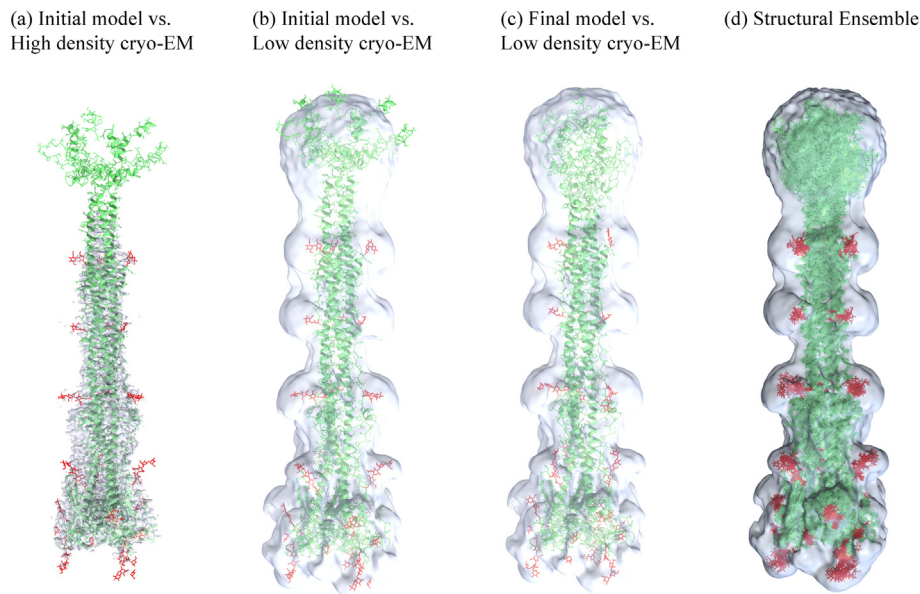

Fig. 6. Cryo-EM reconstruction from Cai et al. (Cai et al., 2020) shown at different threshold levels reveal different features of the complex. (a) Higher threshold level (0.00843) showing stronger density, generated by stationary elements of complex. (b) Moderate threshold level (0.00329) showing moderate density. (c) Lower threshold level (0.00256), showing weaker density, some of which is generated by presumably mobile elements of the complex (i.e., fusion peptide region). (d) Structural model of stationary elements. (e) Structural model of stationary elements superposed with stronger density cryo-EM (0.0121). In (b) and (c), map was filtered with a 2 Å Gaussian filter using Chimera. Grey/blue, cryo-EM density; green, protein models; red, glycan models. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Models for the post-fusion state are shown in Fig. 5. We see that the 20-residue loop (residues 1194 to 1213) provides the requisite flexibility to connect the six-helix and the triple-helix regions. The formation of a six-helix bundle motif between HR1 and HR2 domains is crucial to the S protein adopting the post-fusion state structure, making it a viable target to block the fusion transition of the S protein. Studies have shown that EK1, an inhibitory peptide, is a potent pan-coronavirus fusion inhibitor (Xia et al., 2020a, 2020b). Structures of the HR1/EK1 complexes from SARS (PDB: 5ZVM), MERS (PDB: 5ZVK) and 229E (PDB: 5ZUV) were solved previously using x-ray crystallography (Xia et al., 2019). These six-helix bundle structures are completely superimposable with the HR1/HR2 six-helix bundle from SARS-CoV-2 (PDB: 6LXT) as shown in Fig. 5c.

To further compare with experimental data (i.e., cryo-EM reconstruction from Cai et al. (Cai et al., 2020)), we performed flexible fitting using molecular simulations with phenix.cryo_fit. Here, the molecular simulations are performed in the presence of a cryo-EM correlation term, where, throughout the simulation, a correlation function between a simulated map, based on the model structure at the current step in the simulation, and the experimentally determined cryo-EM map is computed. Once the maximum correlation is achieved, the simulation is continued in the presence of the cryo-EM term, producing a large ensemble of configurations consistent with the cryo-EM map.

As a first step, we compared our final model to the experimentally determined cryo-EM structure using a threshold value representing strong cryo-EM density, as is commonly done in the cryo-EM field. Our final model shows reasonable agreement between the model and the cryo-EM map throughout the complex, including the six-helix bundle (Fig. 7).
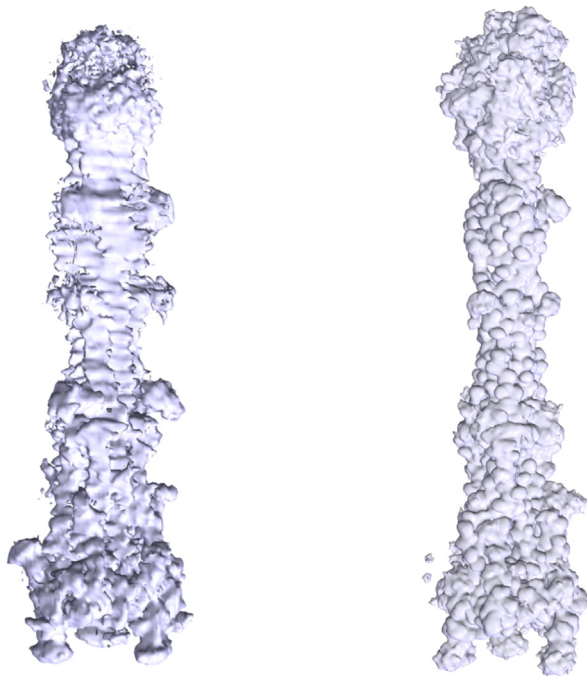
(a) Initial model vs. High density cryo-EM

(b) Initial model vs. Low density cryo-EM

(c) Final model vs. Low density cryo-EM

(d) Structural Ensemble

Fig. 7. Structural ensemble of models generated by combining homology modeling with cryo_fit is consistent with experimentally determined low density cryo-EM map. (a) Initial model of full S protein complex in post-fusion state, including fusion peptide regions (FP and FPPR), superposed with stronger density cryo-EM (threshold = 0.0121). (b) Initial model of full S protein complex, including fusion peptide regions, superposed with weaker density cryo-EM (threshold = 0.0016). (c) Final model of full S protein complex, including fusion peptide regions, generated from molecular dynamics fitting by cryo_fit, superposed with weaker density cryo-EM (threshold = 0.0016). (d) Superposition of 11 configurations of S protein complex, including fusion peptide regions, generated from molecular dynamics fitting by cryo_fit, superposed with weaker density cryo-EM (threshold = 0.0016). In (b), (c) and (d), map was filtered with a 2 Å Gaussian filter using Chimera. Grey/blue, cryo-EM density; green, protein models; red, glycan models. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

The 3D coordinates for the FP and FPPR regions were not reported in the cryo-EM study. This was presumably due to high mobility of these regions, resulting in weak cryo-EM density. Using our homology model of these regions (Fig. 7) as a starting structure, we continued our simulations of the full S protein complex to obtain agreement between our model of the FP and FPPR regions with weaker cryo-EM density in this region (Fig. 8). To closely compare simulation and experiment, we calculated simulated cryo-EM maps for 21 configurations (Fig. 8) occurring after maximum convergence between the model and cryo-EM map was obtained. The 21 maps were averaged, and the averaged simulated map compared with the experimentally determined cryo-EM. The bulb toward the top of the experimentally determined cryo-EM at weak density is recapitulated in the averaged simulated map, along with some of the horizontal strata corresponding to the positions of glycan molecules. The observed qualitative agreement between our model and the experimentally determined cryo-EM map further supports our homology model (Fig. 8). We note that differences between the experimentally determined cryo-EM at weak density and averaged simulated map could be attributed to detergent micelles. Interestingly, the majority of the surface of the FP/FPPR bulb region contains exposed hydrophobic residues that can interact with the cell membrane.

2.4.1. Arbidol (Umifenovir) as a prophylactic drug against SARS-CoV-2

As a proof of concept for predicting drug-binding to our post-fusion state model we tested a known antiviral drug against our model. Arbidol is an antiviral medication for the treatment of influenza infection (Xia et al., 2019) and has been shown to inhibit viral entry into the targeted cell by interfering with the fusion process (Leneva et al., 2009). Arbidol has been considered as a potential drug for treatment of SARS-CoV-2 (Kadam and Wilson, 2017; Yang et al., 2020), and a study with small sample size (164 subjects) has shown that prophylactic oral Arbidol was associated with a lower incidence of SARS-CoV-2 infections (Kadam and Wilson, 2017). Simulations employing molecular dynamics and docking have supported hypotheses that Arbidol can target the SARS-CoV-2 S protein and impede trimerization (Yang et al., 2020).

The structure of *Influenza* HA/Arbidol complex has been solved using x-ray crystallography (5T6N) (Leneva et al., 2009). The interactions between the drug and the HA protein include both a hydrophobic interaction and a network of salt bridges. These strong interactions function as molecular glue that stabilizes the pre-fusion conformation of the HA trimer (Leneva et al., 2009) and blocks the pre-fusion to post-fusion transition. Simple docking of Arbidol to the S protein trimer provided a weak argument of this drug being able to impede S protein trimerization (Yang et al., 2020), so instead we use a homology modeling approach to investigate whether Arbidol interacts with the S protein in a binding mode similar to *Influenza* HA in order to block the transition from pre-fusion to post-fusion intermediate states.

The structure of HA proteins in the vicinity of Arbidol-binding site show some structural similarities to a compatible region of three basic

(a) Experiment

(b) Simulation

Fig. 8. Comparison between experiment and simulation. (a) Cryo-EM reconstruction of full S protein complex shown for moderate density levels (threshold = 0.0041). (b) Simulated cryo-EM map, averaged over simulated maps from 21 models, generated from molecular dynamics fitting by cryo_fit (threshold = 0.0167). In (a), map was filtered with a 1 Å Gaussian filter using Chimera.

residues in the S protein (K-947, R-1014, R-1019). Using this target site, we docked Arbidol into the S protein trimer showing that indeed Arbidol can bind to the S2 domain of the S protein in a manner similar to HA (Fig. 9a). In this binding mode, Arbidol fits between HR1 and HR2 instead of between two HR1 helices (the rectangle box highlighted in Fig. 9a) as proposed by the aforementioned docking study (Yang et al., 2020; Vankadari, 2020). In our model, salt-bridges can form between Arbidol and three basic residues (K-947, R-1014, R-1019) from two different S proteins in the trimer (see Fig. 9b). These strong interactions lock the HR1 (K-947) and HR2 (R-1014, R-1019) helices in the pre-fusion structure and block the transition from pre-fusion state to fusion intermediate state.

While Arbidol was proposed as a prophylactic drug against SARS-CoV-2 infection and proven to be an efficient inhibitor of SARS-CoV-2 *in vitro* (Wang et al., 2020), the low relative Selectivity Index (SI = 7.73) may have impeded the effectiveness of Arbidol as an antiviral drug for COVID-19 (Wang et al., 2020).

To improve the effectiveness of Arbidol as an antiviral drug, we looked for ways to modify Arbidol and improve binding to the S2 protein trimer. With residues E−773 and K-776 of S2 in close proximity to Arbidol, we propose to modify C31–N32(C33)C34 portion of Arbidol into C31–N32–C34–O35, as highlighted in Fig. 9c. As a result, two additional salt bridges can form between the drug and the S2 trimer. The approach described here can be used to inform syntheses for new drug compounds and drive development of new therapeutics.

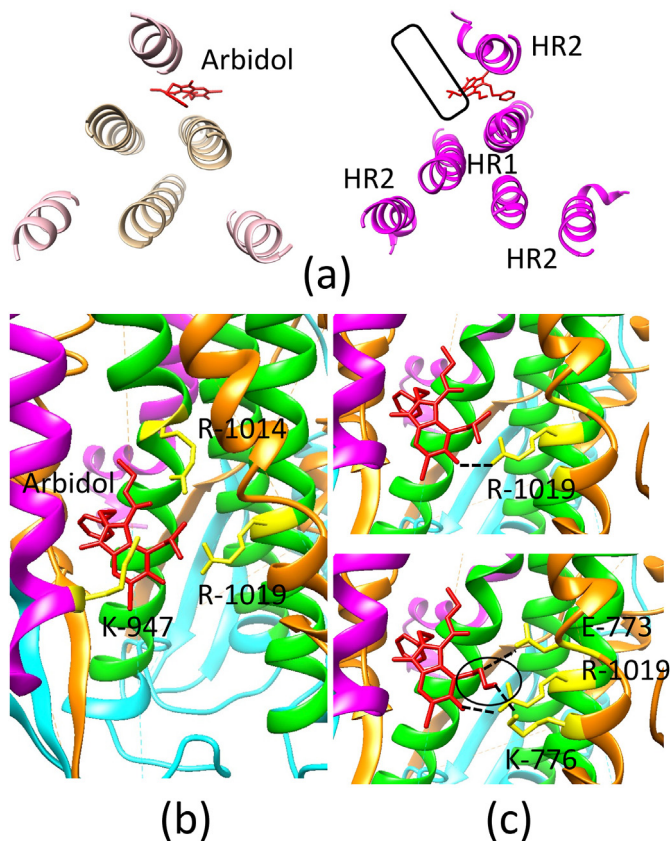The delta variant has been found to be more contagious and more virulent than other variants of SARS-CoV-2. Of the mutations associated with the delta variant (Table 1), P-681-R, is the only mutation distal from the receptor binding domain of the S protein. This residue is mutated to histidine (P-681-H) in the alpha and omicron variants (both BA.1 and BA.2 subvariants). While residue 681 is not responsible for the receptor binding, it is adjacent to the furin cleavage site (685/686) of the S protein. Cleavage by furin primes the S protein for additional cleavage by TMPRSS2, a necessary step for liberation of the fusion peptide region and membrane fusion (Papa et al., 2021; Hoffman et al., 2020).

The furin cleavage site resides near the interface between the S1 and S2 domains (magenta region connecting S1ab between S1b and S2a in Fig. 1). To assess the impact of the P-681-R mutation on the function of the S protein, we developed a feasibility model of the S protein/furin complex using a homology modeling approach. The mouse furin-inhibitor peptide complex (PDB accession code: 1P8J) was chosen as the template. The mouse furin structure in the complex was substituted

Fig. 9. Arbidol-binding to the S protein trimer in the fusion intermediate state. (a) Binding of Arbidol to hemagglutinin (left image, PDB: 5T6N) and to the S protein (right image) with the previously-proposed Arbidol binding site marked by a rectangular box. (b) Detailed interactions of the S-Arbidol complex show multiple interactions with basic side-chain residues (K-947, R-1014, and R-1019). (c) Arbidol (top) and a modified Arbidol compound (bottom, modification site is highlighted by an oval box) both shown binding to the S protein trimer.

Table 1
Mutations of the S protein in major variants of SARS-CoV-2 (causes COVID-19).

Type	Lineage	Country	Mutations	Location (virus/host)
Alpha	B.1.1.7	UK	N501Y, E484K P681H	S1b/ACE2 S1–S2/Furin
Beta	B.1.351	South Africa	N501Y, K417N, E484K	S1b/ACE2
Delta (plus)	B.1.617.2	India	T478K, L452R, (K417N) P681R	S1b/ACE2 S1–S2/Furin
			D614G	S1ab
Gamma	P.1, B.1.1.28	Brazil	E484K, N501Y, K417T	S1b/ACE2
Epsilon	B.1.427 B.1.429	US	L452R	S1b/ACE2
Lambda	C.37	Peru	G75V, T76I, L452Q, F490S	S1a S1b/ACE2
			D614G	S1ab
			T859N	FPPR
Omicron	B.1.1.529, clade 21K, a.k.a. BA.1	South Africa	A67V, T95I, G145D, L212I	S1a
			G339D, S371L, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, Q496S, Q498R, N501Y, Y505H, T547K, D614G, H655Y, N679K, P681H	S1b
			N764K,	S1b
			D796Y,	S1b
			N856K,	S1b
			Q954H, N969K, L981F	S1b S1ab S1ab S2a S2a FPPR HR1
Omicron	B.1.1.529, clade 21L, a.k.a. BA.2		T19I, A27S, G142D, V213G	S1a
			G339D, S371F, S373P, S375F, D405N, R408S, K417N, N440K, S477N, T478K, E484A, Q493R, Q498R, N501Y, Y505H, D614G, H655Y, N679K, P681H,	S1b S1b S1b
			N764K,	S1b
			D796Y,	S1b
			Q954H, N969K	S1ab S1ab S2a S2a HR1

by a human furin (PDB accession code: 4Z2A). The inhibitor peptide (RVKR) is remodeled using the S protein proteolytic site sequence (682-RRAR-685), and furin was docked to the S protein such that the connecting loops between the proteolytic site sequence and the main body of the S protein are sterically feasible. The two proteins are not overlapping. The resulting structure of the complex is shown in Fig. 10.

Residue 681 is one amino acid upstream of the proteolytic site (682–685) and close to the E−230 of furin (Fig. 10). The P-681-R mutation (delta variant) makes it possible to add one hydrogen bond between the two proteins, which stabilizes the interactions between the two proteins and is positioned to increase the furin cleavage efficiency of the S protein. Both the P-681-R (delta) and P-681-H (alpha, omicron BA.1 and BA.2) mutations enhance the likelihood of salt bridge formation. We note that increased cleavage efficiency would promote the completion of the pre-fusion to post-fusion transition and improve the fusion process of the delta and omicron (both BA.1 and BA.2) variants of the virus, consistent with the observed increase in infectivity. Interestingly, the omicron variants (BA.1 and BA.2) contain mutations located along the stalk of the spike protein in regions (HR1) involved in large conformational changes during viral entry (i.e., during the pre-fusion to post-fusion transitions) (Fig. 11). The omicron variant (BA.1) and lambda variant contain mutations in the fusion region (FPPR), important for anchoring the virus to the host during viral entry.

It is an interesting issue to know whether the TMs can maintain their triple helix conformation inside the cell membrane instead of each TM helix moving independently during the transition process, as proposed by Dodero-Rojas et al., an exciting study where extensive molecular simulations were performed to determine the energy landscape of the pre-fusion to post-fusion transition (Dodero-Rojas et al., 2021). Our model, consistent with the schematic rendition of a potential transition pathway, is shown in Fig. 12. In this transition pathway, the HR2 triple helix dissociates into three individual helices with a large portion of the domain changes to a non-helical flexible loop conformation to accommodate the large structural rearrangement. The long loops are flexible enough to allow the TM triple helix to move together with the cell membrane to reposition such that the small helical region from HR2 can form the six-helix bundle motif with helices from HR1.

3. Discussion

Unlike many other proteins that perform their functions without having to significantly alter their structural folds (e.g., enzymes), viral envelope proteins (e.g. Influenza hemagglutinin, HIV-gp160, EBOV-GP, Zika-E) have to go through large conformational changes to carry out their functions. Thus, the native fold of viral envelope proteins (e.g., pre-fusion structure) provides only partial information about protein function. The S protein of coronavirus undergoes several conformational changes to initiate receptor-binding and membrane fusion. To better understand how SARS-CoV-2 S proteins function, elucidating the structures of the S protein at different stages of the viral fusion process is important. A clear picture of the S protein at different structural transition stages informs strategies to block the transitions and prevent the membrane fusion process.

The transition from the fusion-intermediate state to the post-fusion state is a complex folding process. By using LANL-developed software, we have filled gaps of the experimentally-determined S protein structures and constructed models of both the N-terminal domain (FP) and C-terminal domain (HR2/TM) of the S2 fragment of the S protein. All of this information provides a better view of the changes in S protein conformations during the fusion process. In a feasibility study of the fusion intermediate 2 structure, with modeled connections between the TM (as a triple-helix) and the HR2 (in six-helix bundle) domains, we show that the loops between the two domains are sufficiently flexible to make the connection between both domains, in turn, connecting the host and the virus (Figs. 2 and 12). We also find that the TM can remain in a triple helix conformation while going through transition 3 to initiate the fusion process.
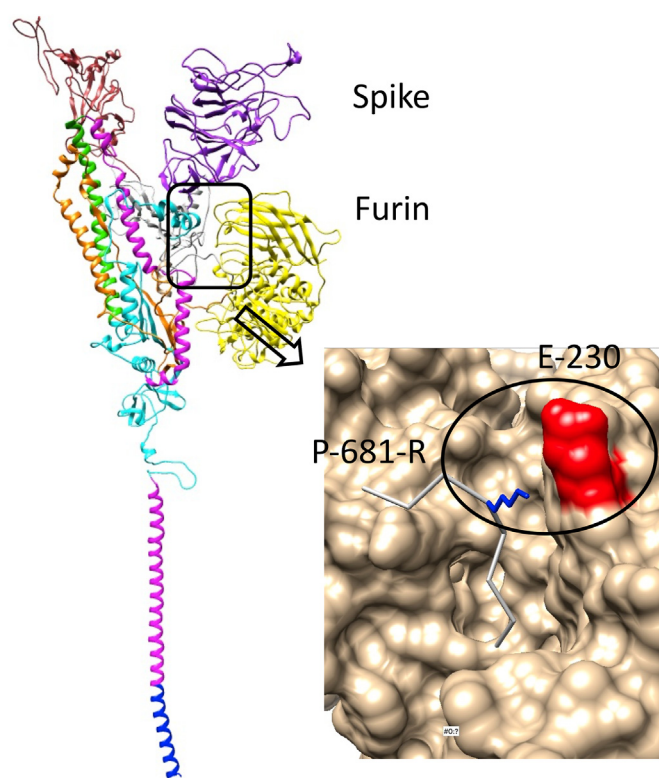

Fig. 10. S protein/furin complex. Left, model of the S protein furin complex. For the S protein, colors are as in Fig. 1. Furin is depicted in yellow. The square highlights the binding regions between the S protein and furin. Right, region in the square is enlarged to show detailed interactions between the S protein (white) and furin (goldish brown). The P681R mutation (blue) makes it possible to add one hydrogen bond between the S protein and the E230 residue (red) from furin, stabilizing the interaction between the two proteins and potentially increasing the cleavage efficiency of the S protein. The increased cleavage efficiency may promote completion of the pre-fusion to post-fusion transition and improve the fusion process of the delta variant of the virus. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Transition 1 of S protein (from pre-fusion-1 to pre-fusion-2) involves moving the S1b domain from a "down" to an "up" conformation. It was shown that S1b can be locked in the "down" conformation by nanobody NB6 (PDB_ID: 7KKK) or bound and blocked from transition by antibody S209 (PDB_ID: 6WPS) (Schoof et al., 2020). In both cases, binding to cell receptor ACE2 is prevented.

Between the known pre-fusion-2 and post-fusion states, we propose that the S2 fragment of the S protein goes through a fusion intermediate state. After the binding of S1b to the cell receptor protein (pre-fusion-2 state in Fig. 2), the fusion peptide is still distal from the host cell membrane. In order to bridge the viral and host cell membranes, the S protein has to go through a two-step conformational change: (1) S1 fragment cleavage to expose the fusion peptide (FP), and (2) S2 fragment undergoing a conformational change (as shown in Fig. 2). In this fully extended conformation, the FP on the N-terminal end of S2 can insert into the host cell membrane while the TM remains in the viral membrane. In this extended conformation, S2 can directly bridge the viral and the host cell membranes.

(RELOCATED FROM BELOW) While vaccines development has focused on the receptor binding domain (RBD) of S1b of the S protein, the mutation rates in RBD are higher than those in the S2 domain. We demonstrate that several sites on the S2 domain of the S protein can be used for developing drugs against the pre-fusion to post-fusion transitions. Drugs targeting S2 domain may likely be less affected by viral mutations.
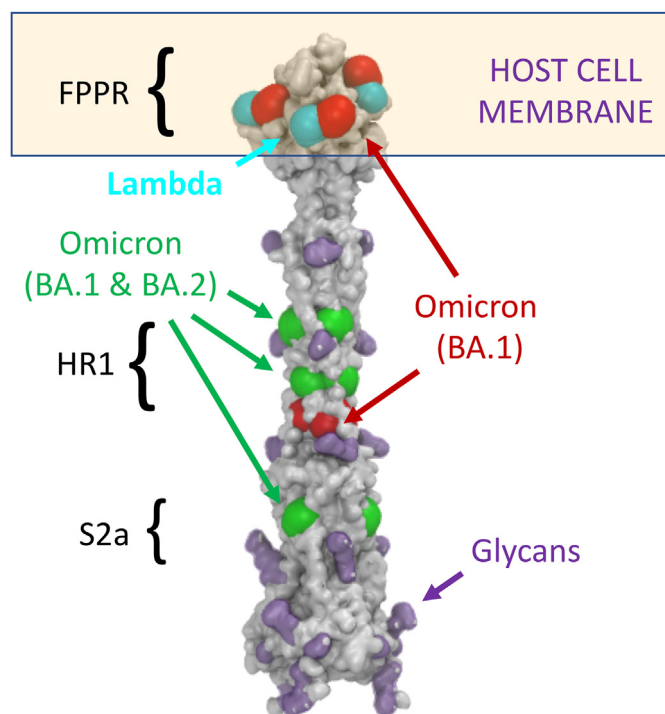
Fig. 11. Mutations of the S protein post-fusion complex present in major variants of SARS-CoV-2 (causes COVID-19). Cyan, lambda variant; red, omicron variant (subvariant BA.1); green, omicron variant (occurring in both subvariant BA.1 and BA.2); purple, glycan molecules; grey, spike protein complex. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

We also note that later variants such as omicron (BA.1 and BA.2 subvariants) possess mutations in regions important for the conformational changes occurring during the fusion process (Fig. 11). In addition, the BA.1 subvariant and the lambda variant contain mutations in the
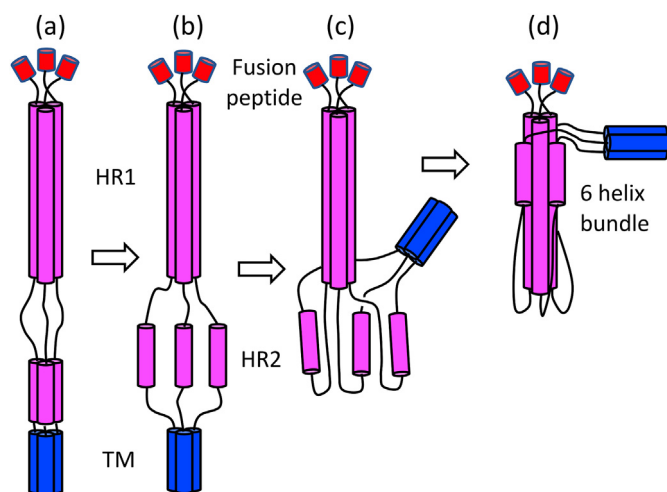
Fig. 12. Schematic of the fusion-intermediate to post-fusion transition (transition no. 3). A 20 aa peptide loop that connects HR2 in the six-helix bundle to the TM (blue) has limited the placement of the TM group and helps bring the cell and viral membrane to close proximity. (a) Fusion-intermediate state. HR2 helices in triple-helix bundle (magenta). (b) HR2 helices loosen. (c) Conformational change occurs, allowing HR2 helices to form 6-helix bundle with HR1 helices. (d) Resulting position of HR1 helices constraints position of TM, helping to bring host cell membrane and viral membrane closer. Red, fusion peptide (FP). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

fusion region (FP/FPPR), critical for anchoring the virus to the host during fusion and viral entry. It may be that the virus exhausted mutations related to host receptor recognition and then moved on to mutations related to downstream events in the process of infection, such as fusion and viral entry. Overall, understanding S2 and its conformational changes during fusion may play an important role in understanding the future evolution of SARS-CoV-2, vaccine design (e.g., the Pfizer and Moderna vaccines are based on the S protein) and new antiviral targets.

To demonstrate how drug-binding assays that can be done with these models, we propose an Arbidol binding site on the S trimer. By binding to this potential site, Arbidol can act as molecular glue and interfere with the transition of the S protein from the pre-fusion-2 state to the fusion-intermediate state. A relatively low selectivity index value makes Arbidol a weak antiviral drug for COVID-19, prompting us to explore the possibility of modifying Arbidol to enhance the inhibition of the transition. With a small modification, additional salt bridges can form, along with a stronger imputed binding, between the modified Arbidol and the S protein trimer.

On the whole, by combining known structures of complete and partial S proteins under various conditions using a variety of computational modeling approaches, we have developed complete structural models of the S protein along the pre-fusion to post-fusion transition pathway. This modeled transition pathway yields detailed predictive models of the structural changes in atomic detail. While the model predictions need to be experimentally, they provide a basis for designing drugs for blocking the pre-fusion to post-fusion transitions to prevent the viral entry of the host cells. Our modeling approach is generalizable and can be used to probe pre-fusion to post-fusion transitions of envelop proteins from other viruses such as HIV, influenza and Ebola.

4. Methods

4.1. Homology modeling using a motif-matching fragment assembly method (MMFA)

The MMFA, described in our earlier work (Tung et al., 2002; Tung and Sanbonmatsu, 2004), was used to conduct homology modeling of the ribonucleoprotein (RNP) complex and its polymerase (POL). This method is general. It can be used to assemble structures of protein complexes (e.g., the spike trimer at different stages of structural transition) using partial structures of the protein at different resolutions from crystallography and cryo-EM data sources.

4.2. Structural refinement

To resolve any bad contacts in our S protein trimer model we perform energy minimization and microcanonical ensemble (constant number, N, volume, V and energy, E) simulations using the GROMACS software package (ver. 4.5.5) (Pronk et al., 2013).

4.3. Graphics

Molecular graphics images are produced using the UCSF Chimera package (Meng et al., 2006) from the Resource for Bio-computing, Visualization and Informatics at the University of California, San Francisco and VMD (Humphrey et al., 1996).

4.4. Cryo-EM flexible fitting of the post-fusion state

The homology model for the full post-fusion state complex including fusion peptide region (Fig. 7) was combined with the glycans from the deposited cryo-EM model 6XRA (Cai et al., 2020) and used as a starting structure. The Smog native contact potential was used to generate GROMACS topology files (Whitford et al., 2009). The deposited cryo-EM map EMD-22293 (Cai et al., 2020) was filtered at 4 Å using Chimera. The starting structure was minimized for 5000 steps using GROMACS with

the Smog native contact potential. MDFIT (Ratje et al., 2010; Whitford et al., 2011) and phenix.cryo_fit (Kim et al., 2019) were used to perform flexible fitting into the cryo-EM map EMD-22293. A correlation function between the simulated and experimentally determined cryo-EM maps is employed to evolve the simulation towards a structure whose density volume map that best matches the cryo-EM density volume map. Native contacts involving linker regions within the fusion peptide region were turned off. Simulations were run for 300,000 steps with a cutoff of 4 Å. The correlation coefficient between the simulated maps and experimentally determined cryo-EM map was re-evaluated every 200 steps.

Comparison between simulated and experimentally determined cryo-EM reconstructions. After the correlation coefficient between the simulated maps and experimentally determined cryo-EM map approached its maximum (100,000 steps), simulated maps were collected every 10,000 steps for a total of 21 simulated cryo-EM maps. The 21 maps were averaged using Situs (Wriggers et al., 1999). The averaged map was compared with the experimentally determined map (Fig. 8).

Analysis of SARS-CoV-2 variants. Variants observed previously (Emma, 2021) were mapped onto the post-fusion complex 3D structure (Fig. 11).

CRediT authorship contribution statement

Jacob C. Miner: Formal analysis, Investigation, Methodology, Validation, Visualization. **Paul W. Fenimore:** Investigation. **William M. Fischer:** Investigation. **Benjamin H. McMahon:** Investigation. **Karissa Y. Sanbonmatsu:** Formal analysis, Funding acquisition, Investigation, Methodology, Software, Validation, Visualization. **Chang-Shung Tung:** Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

References

Belouzard, S., Chu, V.C., Whittaker, G.R., 2009. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. Proc. Natl. Acad. Sci. USA 106, 5871–5876.

Belouzard, S., Millet, J.K., Licitra, B.N., Whittaker, G.R., 2012. Mechanisms of coronavirus cell entry mediated by the viral spike protein. Viruses 4, 1011–1033.

Bertram, S., Glowacka, I., Muller, M.A., Lavender, H., Gnirss, K., Nehlmeier, I., Niemeyer, D., He, Y., Simmons, G., Drosten, C., Soilleux, E.J., Jahn, O., Steffen, I., Pohlmann, S., 2011. Cleavage and activation of the severe acute respiratory syndrome coronavirus spike protein by human airway trypsin-like protease. J. Virol. 85, 13363–13372.

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

Carr, C.M., Kim, P.S., 1993. A spring-loaded mechanism for the conformational change of influenza hemagglutinin. Cell 73, 823–832.

Carr, C.M., Chaudhry, C., Kim, P.S., 1997. Influenza hemagluttinin is spring-loaded by a metastable native conformation. Proc. Natl. Acad. Sci. USA 94, 14306–14313.

Dodero-Rojas, E., Onuchic, J.N., Whitford, P.C., 2021. Sterically-confined rearrangements of SARS-CoV-2 spike protein control cell invasion. Elife 10, e70362. https://doi.org/10.7554/eLife.70362, 2021.

Emma, B., Hodcroft, 2021. CoVariants: SARS-CoV-2 mutations and variants of interest. https://covariants.org/.

Fantini, J., Scala, C.D., Chahinian, H., Yahi, N., 2020. Structural and molecular modelling studies reveal a new mechanism of action of chloroquine and hydroxychloroquine against SARS-CoV-2 infection. Int. J. Antimicrob. Agents 55, 105960.

Gorbalenya, A.E., Baker, S.C., Baric, R.S., de Groot, R.J., Drosten, C., Gulyaeva, A.A., Haagmans, B.L., Lauber, C., Leontovich, A.M., Neuman, B.W., Penzar, D., Perlman, S., Poon, L.L.M., Samborskiy, D.V., Sidorov, I.A., Sola, I., Ziebuhr, J., 2020. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat. Microbiol. 5, 536–544.

Granet, R., August 10, 2021. CAS COVID-19 resources | CAS. https://www.cas.org/resources/covid19.

Gur, M., Taka, E., Yilmaz, S.Z., Kilinc, C., Aktas, U., Golcuk, M., 2020. Conformational transition of SARS-CoV-2 spike glycoprotein between its closed and open states. J. Chem. Phys. 153, 075101.

Hayashi, T., Yasuda, S., Suzuki, K., Akiyama, T., Kanehara, K., Kojima, K., Tanabe, M., Kato, R., Senda, T., Sudo, Y., Murata, T., Kinoshita, M., 2020. How does a microbial rhodopsin RxR realize its exceptionally high thermostability with the proton-pumping function being retained? J. Phys. Chem. B 124, 990–1000.

Henderson, R., Edwards, R.J., Mansour, K., Janowska, K., Stalls, V., Gobeil, M.C., Kopp, M., Li, D., Parks, R., Hsu, A.L., Borgnia, M.J., Haynes, B.F., Acharya, P., 2020. Controlling the SARS-CoV-2 spike glycoprotein conformation. Nat. Struct. Mol. Biol. 27, 925–933.

Hoffman, M., Klein-Weber, H., Pohlmann, S.A., 2020. Multibasic cleavage site in the spike protein of SARS-CoV-2 is essential for infection of human lung cells. Mol. Cell. 78, 779–784.

Hoffmann, M., Kleine-Weber, H., Pohlmann, S., 2020. A multibasic cleavage site in the spike protein of SARS-CoV-2 is essential for infection of human lung cells. Mol. Cell. 78, 779–784.

Humphrey, W., Dalke, A., Shulten, K., 1996. VMD – visual molecular dynamics. J. Mol. Graph. 14, 33–38.

Jaimes, J.A., Millet, J.K., Whittaker, G.R., 2020. Proteolytic cleavage of the SARS-CoV-2 spike protein and the role of the novel S1/S2 site. iScience 23, 101212.

Kadam, R.U., Wilson, I.A., 2017. Structural basis of influenza virus fusion inhibition by the antiviral drug Arbidol. Proc. Natl. Acad. Sci. USA 114, 206–214.

Kim, D.N., Moriarty, N.W., Kirmizialtin, S., Afonine, P.V., Poon, B., Sobolev, O.V., Adams, P.D., Sanbonmatsu, K.Y., 2019. Cryo_fit: democratization of flexible fitting for cryo-EM. J. Struct. Biol. 208, 1–6.

Kim, D., Lee, J.Y., Yang, J.S., Kim, J.W., Kim, V.N., Chang, H., 2020. The architecture of SARS-CoV-2 transcriptome. Cell 181, 914–921.

Lai, A.L., Park, H., White, J.M., Tamm, L.K., 2006. Fusion peptide of influenza hemagglutinin requires a fixed angle boomerang structure for activity. J. Biol. Chem. 281, 5760–5770.

Leneva, I.A., Russell, R.J., Boriskin, Y.S., Hay, A.J., 2009. Characteristics of arbidol-resistant mutants of influenza virus: implications for the mechanism of anti-influenza action of arbidol. Antivir. Res. 81, 132–140.

Li, F., 2015. Receptor recognition mechanisms of coronaviruses: a decade of structural studies. J. Virol. 89, 1954–1964.

Li, F., Li, W.H., Farzan, M., Harrison, S.C., 2005. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. Science 309, 1864–1868.

Lu, G., Hu, Y., Wang, Q., Qi, J., Gao, F., Li, Y., Zhang, Y., Zhang, W., Yuan, Y., Bao, J., Zhang, B., Shi, Y., Yan, J., Gao, G.F., 2013. Molecular basis of binding between novel human coronavirus MERS-CoV and its receptor CD26. Nature 500, 227–231.

Lupas, A., Van Dyke, M., Stock, J., 1991. Predicting coiled coils from protein sequences. Science 252, 1162–1164.

Meng, E.C., Pettersen, E.F., Couch, G.S., Huang, C.C., Ferrin, T.E., 2006. Tools for integrated sequence-structure analysis with UCSF Chimera. BMC Bioinf. 7, 339.

Milanetti, E., Miotto, M., Rienzo, L.D., Monti, M., Gosti, G., Rocco, G., 2020. In-Silico evidence for two receptors based strategy of SARS-CoV-2. bioRxiv. https://doi.org/10.1101/2020.03.24.006197 [Preprint].

Miller, J.K., Whittaker, G.R., 2015. Host cell proteases: critical determinants of coronavirus tropism and pathogenesis. Virus Res. 202, 120–134.

Ord, M., Faustova, I., Loog, M., 2020. The sequence at spike S1/S2 site enables cleavage by furin and phosphor-regulation in SARS-CoV-2 but not in SARS-CoV-1 or MERS-Cov. Sci. Rep. 10, 16944.

Papa, G., Mallery, D.L., Albecka, A., Welch, L.G., Cattin-Ortola, J., Luptak, J., Paul, D., McMahon, H.T., Goodfellow, I.G., Carter, A., Munro, S., James, L.C., 2021. Furin cleavage of SARS-CoV-2 Spike promotes but is not essential for infection and cell-cell fusion. PLOS Path 17, e1009246.

Parry, D.A., 1982. Coiled-coils in alpha-helix-containing proteins: analysis of the residue types within the heptad repeat and use of these data in the prediction of coiled-coils in other proteins. Biosci. Rep. 2, 1017–1024.

Peng, G., Sun, D., Rajashankar, K.R., Qian, Z., Holmes, K.V., Li, F., 2011. Crystal structure of mouse coronavirus receptor-binding domain complexed with its murine receptor. Proc. Natl. Acad. Sci. USA 108, 10696–10701.

Peng, R., Wu, L.-A., Wang, Q., Qi, J., Gao, G.F., 2021. Cell entry by SARS-CoV-2. Trends Biochem. Sci. Article 0968-0004.

Pike, Y.-J., Walls, A.C., Wang, Z., Sauer, M.M., Li, W., Tortorici, M.A., Bosch, B.-J., DiMaio, F., Veesler, D., 2019. Structures of MERS-CoV spike glycoprotein in complex with sialoside attachment receptors. Nat. Struct. Mol. 26, 1151–1157.

Pinto, D., Pike, Y.-J., Beltramello, M., Walls, A.C., Tortorici, M.A., Bianchi, S., Jaconi, S., Culap, K., Zatta, F., De Marco, A., Peter, A., Guarino, B., Spreafico, R., Cameroni, E., Case, J.B., Chen, R.E., Havenar-Daughton, C., Snell, G., Telenti, A., Virgin, H.W., Lanzavecchia, A., Diamond, M.S., Fink, K., Veesler, D., Corti, D., 2020. Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. Nature 583, 290–295.

Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M.R., Smith, J.C., Kasson, P.M., van der Spoel, D., Hess, B., Lindahl, E., 2013. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics 29, 845–854.

Ratje, A.H., Loerke, J., Mikolajka, A., Brunner, M., Hildbrand, P.W., Starosta, A.L., Donhofer, A., Connell, S.R., Fucini, P., Mielke, T., Whitford, P.C., Onuchic, J.N., Yu, Y., Sanbonmatsu, K.Y., Hartmann, R.K., Penczek, P.A., Wilson, D.N., Spahn, C.M.T., 2010. Head swivel on the ribosome facilitates translocation by means of intra-subunit tRNA hybrid states. Nature 468, 713–716.

Reguera, J., Santiago, C., Mudgal, G., Ordono, D., Enjuanes, L., Casasnovas, J.M., 2012. Structural bases of coronavirus attachment to host aminopeptidase N and its inhibition by neutralizing antibodies. PLoS Pathog. 8, e1002859.

Sainz Jr., B., Rausch, J.M., Gallaher, W.R., Garry, R.F., Wimley, W.C., 2005. Identification and characterization of the putative fusion peptide of the severe acute respiratory syndrome-associated coronavirus spike protein. J. Virol. 79, 7195–7206.

Schoof, M., Faust, B., Saunders, R.A., Sangwan, S., Rezelj, V., Hoppe, N., Boone, M., Bilesbølle, C.B., Purchades, C., Azumaya, C.M., Kratochvil, H.T., Nguyen, H.C., Chio, C.M., Merz, G.E., Thompson, M.C., Diwanji, D., Schaefer, K., Anand, A.A., Dobzinski, Niv, Zha, B.S., Simoneau, C.R., Leon, K., White, K.M., Chio, U.S., Gupta, M., Jin, M., Li, F., Liu, Y., Zhang, K., Bulkley, D., Sun, M., Smith, A.M., Rizo, A.N., Moss, F., Brilot, A.F., Pourmal, S., Trenker, R., Pospiech, T., Gupta, S., Barsi-Rhyne, B., Belyy, V., Barile-Hill, A.W., Nock, S., Liu, Y., Krogan, N.J., Ralston, C.Y., Swaney, D.L., Garcia-Sastre, Adolfo, Ott, M., Vignuzzi, M., QCRG Structural Biology Consortium, Walter, P., Manglik, A., 2020. An ultrapotent synthetic nanobody neutralizes SARS-CoV-2 by stabilizing inactive Spike. Science 370, 1473–1479.

Sharif-Askari, N.S., et al., 2021. Enhanced expression of immune checkpoint receptors during SARS-CoV-2 viral infection. Mol. Ther. - Methods & Clin. Dev. 20, 109–121.

Song, W., Gui, M., Wang, X., Xiang, Y., 2018. Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. PLoS Pathog. 14, e1007236.

Tung, C.S., Sanbonmatsu, K.Y., 2004. Atomic model of the Thermus thermophilus 70S ribosome developed in silico. Biophys. J. 87 (4), 2714–2722.

Tung, C.S., Joseph, S., Sanbonmatsu, K.Y., 2002. All-atom homology model of the Escherichia coli 30S ribosomal subunit. Nat. Struct. Biol. 9, 750–755.

Vankadari, N., 2020. Arbidol: a potential antiviral drug for the treatment of SARS-CoV-2 by blocking trimerization of the spike glycoprotein. Int. J. Antimicrob. Agents 56, 105998.

Walls, A.C., Tortorici, M.A., Snijder, J., Xiong, X., Bosch, B.-J., Rey, F.A., Veesler, D., 2017. Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion. Proc. Natl. Acad. Sci. USA 114, 11157–11162.

Wang, N., Shi, X., Jiang, L., Zhang, S., Wang, D., Tong, P., Guo, D., Fu, L., Cui, Y., Liu, X., Arledge, K.C., Chen, Y.H., Zhang, L., Wang, X., 2013. Structure of MERS-CoV spike receptor-binding domain complexed with human receptor DPP4. Cell Res. 23, 986–993.

Wang, X., Cao, R., Zhang, H., Liu, J., Xu, M., Hu, H., Li, Y., Zhao, L., Li, W., Sun, X., Yang, X., Shi, Z., Deng, F., Hu, Z., Zhong, W., Wang, M., 2020. The anti-influenza virus drug, Arbidol is an efficient inhibitor of SARS-CoV-2 in vitro. Cell Discov 6, 28.

Whitford, P.C., Noel, J.K., Gosavi, S., Schug, A., Sanbonmatsu, & K.Y., Onuchic, J.N., 2009. An all-atom structure-based potential for proteins: bridging minimal models with all-atom empirical forcefields. Proteins 75, 430–441.

Whitford, P.C., Ahmed, A., Yu, Y., Hennelly, S.P., Tama, F., Spahn, C.M.T., Onuchic, J.N., Sanbonmatsu, K.Y., 2011. Excited states of ribosome translocation revealed through integrative molecular modeling. Proc. Natl. Acad. Sci. USA 108, 18943–18948.

Wrapp, D., Wang, N., Corbett, K.S., Goldsmith, J.A., Hsieh, C.L., Ablona, O., Graham, B.S., McLellan, J.S., 2020. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science 367, 1260–1263.

Wriggers, W., Milligan, R.A., McCammon, J.A., 1999. Situs: a package for docking crystal structures into low- resolution maps from electron microscopy. J. Struct. Biol. 125, 185–195.

Wu, K., Li, W.B., Peng, G., Li, F., 2009. Crystal structure of NL63 respiratory coronavirus receptor-binding complexed with its human receptor. Proc. Natl. Acad. Sci. USA 106, 19970–19974.

Xia, S., Yan, L., Xu, W., Agrawal, A.S., Algaissi, A., Tseng, C.-T.K., Wang, Q., Du, L., Tan, W., Wilson, I.A., Jiang, S., Yang, B., Lu, L., 2019. A pan-coronavirus fusion inhibitor targeting the HR1 domain of human coronavirus spike. Sci. Adv. 5, eaav4580.

Xia, S., Zhu, Y., Liu, M., Lan, Q., Xu, W., Wu, Y., Ying, T., Liu, S., Shi, Z., Jiang, S., Lu, L., 2020a. Fusion mechanism of 2019-nCoV and fusion inhibitors targeting HR1 domain in spike protein. Cell. Mol. Immunol. 765–767.

Xia, S., Liu, M., Wang, C., Xu, W., Lan, Q., Feng, S., Qi, F., Bao, L., Du, L., Liu, S., Qin, C., Sun, F., Shi, Z., Zhu, Y., Jiang, S., Lu, L., 2020b. Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion. Cell Res. 30, 343–355.

Yang, C., Ke, C., Yue, D., Li, W., Hu, Z., Liu, W., Hu, S., Wang, S., Liu, J., 2020. Effectiveness of arbidol for COVID-19 prevention in health professionals. Ront. Public Health 8, 249.