# Spike protein fusion loop controls SARS-CoV-2 fusogenicity and infectivity

Debnath Pal

Indian Institute of Science, Bengaluru 560012, India.

* Debnath Pal

**Email:** dpal@iisc.ac.in

ORCID: 0000-0002-3591-5978

**Classification**

Computational Biology, Bioinformatics, Virology

**Keywords**

coronavirus, COVID-19, MHV-A59, two consecutive prolines, structural rigidity

**Author Contributions**

DP did all the work and wrote the paper.

**This PDF file includes:**

Main Text
Figures 1 to 4

**Abstract**

Compared to the other human coronaviruses, SARS-CoV-2 has a higher reproductive number that is driving the COVID-19 pandemic. The high transmission of SARS-CoV-2 has been attributed to environmental, immunological, and molecular factors. The Spike protein is the foremost molecular factor responsible for virus fusion, entry and spread in the host, and thus holds clues for the rapid viral spread. The dense glycosylation of Spike, its high affinity of binding to the human ACE2 receptor, and the efficient priming by cleavage have already been proposed for driving efficient virus-host entry, but these do not explain its unusually high transmission rate. I have investigated the Spike from six β-coronaviruses, including the SARS-CoV-2, and find that their surface-exposed fusion peptides constituting the defined fusion loop are spatially organized contiguous to each other to work synergistically for triggering the virus-host membrane fusion process. The architecture of the Spike quaternary structure ensures the participation of the fusion peptides in the initiation of the host membrane contact for the virus fusion process. The SARS-CoV-2 fusion peptides have unique physicochemical properties, accrued in part from the presence of consecutive prolines that impart backbone rigidity which aids the virus fusogenicity. The specific contribution of these prolines shows significantly diminished fusogenicity *in vitro* and associated pathogenesis *in vivo*, inferred from comparative studies of their deletion-mutant in a fellow murine β-coronavirus MHV-A59. The priming of the Spike by its cleavage and subsequent fusogenic conformational transition steered by the fusion loop may be critical for the SARS-CoV-2 spread.

**Significance Statement**

The three proximal fusion peptides constituting the fusion loop in Spike protein are the membranotropic segments most suitable for engaging the host membrane surface for its disruption. Spike's unique quaternary structure architecture drives the fusion peptides to initiate the protein host membrane contact. The SARS-CoV-2 Spike trimer surface is relatively more hydrophobic among other human coronavirus Spikes, including the fusion peptides that are structurally more rigid owing to the presence of consecutive prolines, aromatic/hydrophobic clusters, a stretch of consecutive β-branched amino acids, and the hydrogen bonds. The synergy accrued from the

location of the fusion peptides, their physicochemical features, and the fusogenic conformational transition appears to drive the virus fusion process and may explain the high spread of the SARS-CoV-2.

## Main Text

### Introduction

The novel Coronavirus Disease 2019 (COVID-19) caused by SARS-CoV-2, Severe Acute Respiratory Syndrome (SARS) by SARS-CoV, and Middle Eastern Respiratory Syndrome (MERS) by MERS-CoV, induce severe acute respiratory distress in patients. Though these diseases share similar clinical and pathological features, COVID-19 differs in overlapping yet distinct phases of infection (1, 2). The degree of infectivity is significantly high in SARS-CoV-2, and far more aggressive, as evidenced by the current global pandemic. This can be quantified by the preliminary reproductive number ($R_0$) of COVID-19 (2.0-2.5), which is higher than the $R_0$ of SARS (1.7-1.9) and far higher than that of MERS (<1) (3). The significant difference in $R_0$ may accrue due to environmental, immunological or molecular reasons. COVID-19 transmission has been attributed to the long life of SARS-CoV-2 outside the host as it increases the chances of infection through cross-contamination by contact in the population (4). The large distance distribution of the SARS-CoV-2 particles from the infected person due to activities like sneezing and coughing (5), and the tiny size of the virus droplets may be more efficient in penetrating deeply into the pulmonary system to allow rapid spread of the disease (6). However, the SARS-CoV has high genomic similarity with the SARS-CoV-2, and one would have expected it to have similar transmission behavior and $R_0$, which is evidently not the case. For that matter, the environmental spread of other viruses should have been far more widespread than coronaviruses, given that the coronaviruses have the largest RNA viral genomes and therefore the largest particle size and consequently higher aerosol size compared to many other viruses. This is again not that we observe in practice. Another possibility of high viral spread may accrue from intense viral shedding, where SAR-CoV-2 has succeeded early on in rapid viral replication and cell-to-cell spread before the onset of acute inflammatory

response. Here the extent of the viral replication is dependent on the immune containment-response, but given that MERS has shown 34% case fatality compared to 9.5% for SARS-CoV and only 2.3% for SARS-CoV-2 thus far (3), one can argue that the $R_0$ values should also have been in the same order, which is evidently not the case. This suggests that neither environmental nor immunological response suitably explains the higher $R_0$ of SARS-CoV-2, suggesting that the key reason may be molecular. The Spike glycoprotein that protrudes ~150 Å out of the 500-2000 Å diameter coronavirus envelope is the most suitable molecule for making the first contact with the host cell, and is, therefore, a key molecular factor that determines virus fusion, entry and spread in the host, and thus holds clues for the rapid spread of SARS-CoV-2.

Spike (S) is a key well-studied gene among all coronaviruses that determines virus infectivity due to its established role in initiating virus-host attachment, fusion, entry, and cell-to-cell fusion post replication (7, 8). The expressed protein executes its function through a trimeric quaternary structure, with each monomer segmented into the S1 and S2 domains, where S1 consists of the N-terminal domain (NTD) and the C-terminal domain (CTD), while the S2 (also known as the fusion domain) consists of the fusion peptides (FPs), central helix (CH), heptad repeat regions HR1 and HR2, and a transmembrane domain (TM) followed by the cytoplasmic tail (CP) (Fig. 1A). The NTD or CTD embeds the receptor-binding domain (RBD) depending on the individual virus. In SARS-CoV-2 the CTD embeds the RBD. The RBD recognizes the host-cell and mediates the virus-host attachment through interaction with a receptor, while the S2 domain drives the viral entry into the host cell and cell-to-cell fusion by catalyzing the virus-host membrane fusion process. The TM anchors the protein in the viral envelope, while the FPs contribute to the trigger that drives the virus-host membrane fusion process. A transition pathway has been proposed to proceed through a pre-hairpin intermediate structure followed by pre-bundle hemifusion membrane-associated structure, bundle structure, and the eventual post-fusion structure (9). It is remarkable that despite their similarity in function, β-coronavirus Spike proteins have widely diverged, with the RBD diverging further (21%-61% pairwise sequence identity), explained in part by the diversity of its host receptor targets like the human ACE2 for SARS-CoV and SARS-CoV-2, DPP4 for MERS-

CoV, and 9-O-Ac-Sia for HCoV-OC43 and HCoV-HKU1. The S2 domain has also diverged, but comparatively less with >33% overall pairwise sequence identity – the heptad repeat regions being the most conserved (Fig. 1B). It is pertinent to ask what features of Spike sequence and structure determine the virus fusogenicity and infectivity.

Densely glycosylated Spike protein has been suggested as the prime reason for the high SARS-CoV-2 infectivity (7). This extensive glycosylation of the S2 domain is driven by an intracellular N-terminal signaling peptide for transport and retention in the endoplasmic reticulum (see Fig. 1B Multiple Sequence Alignment bottom panel). However, this signal sequence is absent in a fellow murine β-coronavirus MHV-A59 Spike protein (10), indicating limited glycosylation. Yet, MHV-A59 aggressively infects the mouse liver and brain. Upon intracranial inoculation in mice, it can cause acute stage meningoencephalitis and myelitis, chronic stage demyelination, and axonal loss (11-13). It infects the neurons profusely and can spread from neuron to neuron. Its propagation from grey matter neuron to white matter and release at the nerve ends to infect the oligodendrocytes by cell-to-cell fusion (12) are robust mechanisms to evade immune responses and induce chronic stage progressive neuroinflammatory demyelination concurrent with axonal loss in the absence of functional virions (12, 14). Therefore, high infectivity of the MHV-A59 Spike does not appear to be contingent on glycosylation and one can argue the same for SARS-CoV-2 Spike, where its glycosylation may only marginally raise the basal fusion efficiency. Surface glycosylation thus may not be a contributing factor to host cell binding, although the successive virus-to-cell and cell-to-cell fusion may all-together play an important role in higher virus infectivity.

It has been suggested that enhanced virus-to-cell infection can be propelled by the increased number of hydrogen-bonded contacts between SARS-CoV-2 RBD and ACE2 receptor leading to higher affinity and improved host targeting compared to the SARS-CoV (7, 15). However, a significantly higher affinity between ACE2 and SARS-CoV-2 has not been experimentally corroborated (16). Besides, such a proposition is weak because the RBDs in all HCoVs are diverse, including SARS-CoV, where the minimal RBD (318 to 510 residues) (17) shares only 74% sequence identity with SARS-CoV-2 (Fig. S1). Also, the SARS-CoV-2 Spike may interact with other

receptors such as DC-SIGN and DC-SIGNR as in SARS-CoV to increase tropism (18) and viral spread. Therefore, there is no direct consequence of ACE2 recognition with infectivity unless a virus entry can be realized; however, when RBDs interact with ACE2 in large numbers during the acute stage of the infection, they may modulate the host immune response by downregulating hydrolysis of the pro-inflammatory angiotensin II to anti-inflammatory angiotensin 1-7 in the renin-angiotensin signaling pathway (19). This can alter the immune response and increase infectivity. But such effects can manifest only beyond the early stage of the infection, and for that to happen the efficiency of viral entry is the rate-limiting step.

The cleavage of the Spike protein is said to prime it for the efficient virus-host membrane fusion process. How essential is this for virus fusogenicity and infectivity is an important consideration. The Spike cleavage potentially removes any *in situ* covalent and noncovalent constraints that the S1 domain may impose on the S2 domain impeding its conformational transition that facilitates the virus entry. It has been proposed that SARS-CoV-2 Spike is preactivated by cleavage at the S1/S2 site when it is packaged inside the host, and the S2' site is cleaved when the Spike gets attached to the host receptor, which makes the priming process very efficient (20). However, a comparison of the S1/S2 cleavage signal sequence …RXXR… shows that SARS-CoV-2 "…RRARS…" Furin recognition site is similar to MHV-A59 Spike's "…RRAHR…", and others like MERS, HCoV-OC43 and HCoV-HKU1 Spike have a conserved motif sequence as well (Fig. 1B). The cleavage site signal at S2' embedding a single Arginine is highly conserved across all HCoVs. Therefore, the efficient priming advantage available to SAR-CoV-2 Spike is equally present for MHV-A59, MERS, HCoV-OC43, and HCoV-HKU1 Spike. In contrast, the canonical S1/S2 cleavage recognition sequence is missing in SARS-CoV with only a single Arginine present there. A regular cleavage at this site has not been reported, and cleavage by trypsin has been shown to activate the virus independent of the pH due to the presence of a single Arginine. The importance of this region has been aptly corroborated by S2' site cleavage studies in SARS-CoV (21). Besides, it is also possible for Spike to be activated by the low pH environment through protonation of residues if it internalizes in the endosome post interaction with the host receptor. In contrast, fusion

processes are known to happen in MHV-A59 Spike without cleavage as well (22). Therefore, based on the similarity SARS-CoV-2 Spike with others, one can argue that it is competent to access multiple pathways for priming that enhances its infection capability, including the possibility that it can infect without a cleavage as well – though these advantages are not unique.

Among all the components of the fusion apparatus in the Spike S2 domain, the FPs are the least studied although they have been suggested to contribute to the trigger that drives the virus-host fusion process by initiating the protein-host membrane contacts. Limited experimental information available shows mutation in FP of SARS-CoV Spike can significantly perturb the fusion efficiency (23) as much as >70% (24). Most studies of the FP regions have used synthetic peptides in a fusion assay system to understand their membrane perturbing capabilities (25-28), and how $Ca^{+2}$ ions may interact with these peptides to modulate fusion (9). Interestingly, the FPs also contains a central proline (29) in several viruses such as the Avian Sarcoma/Leucosis virus (30), Ebola virus (31), Vesicular Stomatitis virus (32), and Hepatitis C virus (33), where its important role has been investigated through mutation studies. The location of coronavirus FPs proximal to the N-terminal of the S2 domain is reminiscent of FPs from HIV-1, influenza virus, and paramyxoviruses. They have been suggested to be located at the head a pre-hairpin intermediate structure (34) predicted for the current model of class I viral fusion proteins.

Although the FPs are believed to be the early initiators of protein and host membrane contact, there is still no consensus on their location. For example, the fusion peptide for SARS-CoV-2 Spike has been cited at 788-806 position by Xia et al. (8) compared to 816-833 by Wrap et al. (7). When inferred from alignment to SARS-CoV Spike, two additional FP segments at 875-902 position and 1203-1220 can be proposed based on experimental studies by Ou et al. (35) and Guillén et al. (26), respectively. In reality, all four Spike fusion domain segments (FP-I to FP-IV; Fig. 1B) mentioned can be identified by a simple window-based analysis using interfacial hydrophobicity scales such as from Wimley and White (36). Given that FP-1 to FP-III are contiguous to each other in the sequence, the whole segment spanning the beginning of FP-1 to the end of FP-III can be termed together as the “fusion loop” (Fig. 1A). But how these FPs in the loop can act

synergistically to rapidly trigger the membrane fusion process is what we explore further in this study.

## Results

**Spike receptor binding and the fusion loop.** To understand the synergy of the FPs in the trigger process, it is important to understand the structure of the Spike protein in the proper context. The FP-I to FP-III are surface exposed and are contiguous to each other in space as seen from the three-dimensional structures of the Spike fusion domain (Fig. 2, cartoon diagrams). While FP-I to FP-III are always surface exposed in the full-length, the FP-IV is deeply buried and interface the virus membrane. (The location of FP-IV although not available from the three-dimensional structure can be inferred from the primary structure and the location of the transmembrane domain (Fig. 1B)). Therefore, while FP-I to FP-III are always available for early contacts with the host membrane, FP-IV can participate in the process post the conformational transition which may expose it for interaction with the host membrane. To understand what guarantees the FP surfaces to make the initial protein-host membrane contact, one must look at the possible modes of virus-host attachment mediated by the Spike. For this, we propose a new contact initiation model, where there is no requirement of a fusion peptide to be at the N-terminal of the conformationally transformed pre-hairpin intermediate (9). To understand the model, let us consider the different options for host receptor binding. For example, if all three RBDs in the trimeric structure find the host receptors, it can attain a tripod binding mode (Fig. 3A). A recent structure of the trimeric Spike complexed with a host receptor reveals the precise geometry of CEACAM1 binding the RBD of Spike from MHV-A59 (37). However, a tripod binding requires receptor molecules on the host membrane to be pre-available in a specific arrangement. High expression of receptor molecules on the host cell surface is expected to increase the probability of tripod binding, but there is no existing information whether such a precise arrangement is present on the host surface suitable for interaction with the trimeric RBD. Moreover, the membrane bilayer structure does not contain any feature that can direct such regular host receptor arrangement. It may be noted that only in a tripod arrangement, the N-terminal

segment of the S2 domain will interact early during the protein-membrane contact post intermediate structure formation, and in such a case the pre-hairpin/pre-bundle helices of the fusion domain are expected to interact head-on with the host membrane (9). However, FP-I and FP-II are located at the middle of the cylinder-like Spike S2 structure (Fig. 3A) and an intermediate formation through conformational transition is needed to place them near the head of the cylinder. Here, FP1 is expected to make the early host-membrane contact if the cleavage is at the S1/S2 site and FP-II if the cleavage is at the S2' site (see Fig. 1A). The FP-III region has limited scope to make any early contact due to its farthest position from the host membrane surface. In tripod binding mode the virus membrane is still ~150Å away, and the three HR2 regions need to fold back and bind to the hydrophobic grooves of the HR1 trimer in an antiparallel manner to bridge this gap and form a hemifusion structure with the host membrane (9).

**The Contact Initiation Model - Spike fusion peptide trigger.** If only one or two RBDs bind the receptor, the vertical anchoring of the Spike fusion domain relative to the host surface lacks the third anchor rendering the vertical orientation unstable and unfeasible. Also, a recent trimeric structure of SARS-CoV-2 Spike has shown a single RBD to be in the open conformation (7), where it is swiveled away from its core structure originally interacting with the NTD and the fusion domain. In such a state, the interaction of the open RBD with the host (Fig. 3B) is not expected to stabilize the Spike anchoring in any specific orientation relative to the host due to the weak interaction with the fusion domain. Here a post-cleaved S2 domain is expected to interact side-on to the membrane surface through a "belly" landing to trigger the fusion process (Fig. 3C). Such a process would be sterically facile if there are no other Spike in the vicinity on the virus surface. It is to be noted that the shape of the trimeric S2 domain is not a proper cylinder, but with a bulge in the mid-segment which we call a "belly". The FP-II and FP-I surfaces are located at the crest of this bulge, such that it is able the make the initial contact with the host membrane. The structural constraints that guarantee the "belly" landing can be understood from the overall geometry of the Spike (Fig. 3D). It is to be noted that the most stable and eventual landing posture of an object having an uneven surface will be the one that guarantees the largest surface area of contact with the host landing

surface. The Spike architecture is such where the relative location of the NTDs approximately form three vertices of a triangle, while the FP-I to F-III are located midpoint of the sides, forming the vertices of an inner triangle. Based on the maximum landing surface premise, if we consider the receptor attachment to be in two RBD locations, the S2 landing will be close to the midpoint of the two vertices of the outer triangle coincident to the FP surfaces. Even for one-legged attachment, the contact must always be directed towards the midpoint of the two NTDs because that allows maximum contact surface to be formed where the Spike can stably rest on the host surface. During the contact, the membranotropic segment of the fusion loop spanning the FPs is expected to engage the host membrane during the fusogenic conformational change. In this case, the FP sites are proximal to the virus membrane surface such that hemifusion membrane structures can be initiated early in comparison to the tripod-binding mode which requires an intermediate pre-hairpin structure to be formed. It is also to be noted that weak RBD binding to S2 domain or disintegration of the Spike trimer post tripod binding can mimic the one- or two-legged binding mode. For additional lucidity, the structural basis of the Contact Initiation model is further explained pictorially in a lay manner in Fig. S2.

**SARS-CoV-2 fusion peptide distinctions.** The physicochemical property of the fusion loop and the synergy of the FPs therein is critical to the rapid initiation and transition to the hemifusion stage of the membrane fusion process. The proposed Contact Initiation Model ensures that the S2 domain lies on the belly contacting the host surface during the conformational transition. Since the exerted force during the conformation transition is tangential to the host surface, the orthogonal frictional forces may allow an efficient scything action based on the physiochemical nature of the contact engaged by the fusion loop. The nature of the initial surfaces of the fusion loop can be obtained from the electrostatic potential of the FP-I to FP-III surface patches (38) (Fig. 2, surface diagrams). Among the six Spike proteins considered in this study, SARS-CoV-2 has the most hydrophobic/neutral electrostatic FP patches (WHITE colored) most suitable for membrane disruption. It also has the least amount of highly negative electrostatic surface (RED colored patches) which repel the attachment of the S2 domain to the host membrane due to the repelling

hydrophilic surface of the outer membrane composed of negatively charged fatty acid groups. On the other hand, a positive electrostatic surface (BLUE colored patches) may allow protein surface to tightly attach to the membrane exterior through charge attraction. The presence of an interfacial ion like the $Ca^{+2}$ can cap the negative charge at the FP surface to assist fusion trigger - consistent with the membrane charge compatibility requirements. The disruption efficiency of surface patches is, therefore, likely to be highest for the hydrophobic/neutral, followed by the positive and negative electrostatic potential, unless modulated by an interfacial cation like the $Ca^{2+}$. One may argue that large patches of negative electrostatic surface potential (RED color) in the HCoV-HKU1 and HCoV-OC43 Spike fusion domain may explain the mild nature of those viruses.

Aside from the electrostatics, the physical rigidity of the fusion loop in Spike is of prime importance for the virus fusogenicity. This is because the Spike fusion domain being metastable, any local alteration of rigidity has global implications for the molecule. This has been alluded to by the mutation studies on the fusion peptide central prolines (30-33), but its criticality was recently revealed from our comprehensive studies on centrally located consecutive prolines in MHV-A59 Spike fusion peptide (39). Proline being an imino-acid with unique structure, has restricted torsional freedom, which in turn restricts the torsional freedom of the protein backbone where it is located (40). When two consecutive prolines are located, the rigidity of the protein backbone is further enhanced. Among the six Spike proteins in our study, three have a single central proline in FP-I, and two centrally located consecutive prolines occur in SARS-CoV-2 Spike FP-1. There are no central prolines in FP-II and a single conserved central proline in FP-IV, while consecutive prolines exist at a central location in HCoV-HKU1, HCoV-OC43, and MHV-A59 FP-III Spikes. To understand the intrinsic flexibility/rigidity of the surface exposed FPs, we set up 500 ns molecular dynamics simulations (Fig. 4). The starting structures of all the FP-I fragments are devoid of stable secondary structures such as helices or sheets and is dominated by turns, and irregular structures indicating that the region prefers to be in a loop conformation. This is true even for the MHV-A59 Spike, which is flexible as experimentally evidenced by the lack of coordinate from the electron density map for complete FP-I in the PDB file: 3JCL. In comparison, FP-IIs in all cases are dominated by helical

conformation, with the N-terminal FP segments always in a helical state. This is true for FP-III region fragments as well. Given an irregular structure, the effect of proline is more dramatic on the FP-I region in contrast to FPs from FP-II and FP-III which are already stabilized by hydrogen bonds in helices. For all FPs in the FP-I region, although local secondary structures are induced during simulation, the lowest root-mean-square fluctuation (RMSF) for a residue is achieved only by the SARS-CoV-2 consecutive proline containing FP. This can also be confirmed from the energy landscape plot created from the molecular dynamics simulation trajectory (Fig. S3) where it has the most compact single conformational well among all the FP-Is. If we look at the FP-III regions where other consecutive proline containing FPs exist, HCoV-OC43 has the lowest residue RMSF at the double prolines; SARS-CoV-2 FP-III segment does not have any central proline, but it contains a unique “Thr-Ile-Thr” segment constituted of three consecutive β-branched side-chain residues that can impart substantial rigidity based on steric considerations (40) if not as much as the consecutive prolines. This can be again confirmed from the energy landscape plot which shows a single compact conformational well in contrast to the other FP-IIIs (Fig. S3). The observations are consistent with our previous comparative molecular dynamics studies (39) on MHV-A59 FP-III and single central proline containing FP-III from another parental non-neurotropic strain of murine hepatitis virus (MHV), MHV-2, where we found the FP-III from the former became more rigid than the latter in methanolic conditions compared to water. Additionally, NMR studies on the MHV-A59 FP-III fragment revealed its unique ability to form *cis*-peptide at the central P-P peptide bond (39). A *cis*-trans isomerization during the membrane fusion process has the potential to expose the hydrophobic residues efficiently around the isomerized-peptide neighborhood, enhancing the fusion trigger potential. While the structural role of proline cannot be discounted, other stabilizing interactions also dominate the FPs, among which the formation of the aromatic/hydrophobic clusters is of relevance to the fusion process (see Fig. S3 for examples). As observed from the packing of the loops, a combination of aromatic and Val/Ile/Leu side chains pack tightly to exclude water. But when these regions become exposed during the conformational transition, they would enhance the hydrophobic interaction of the protein surface with the membrane. In all S2 domains,

however, FP-III regions are in part masked by the FP-I segment, as evident from the relative solvent accessible surface area (SASA) values indicated in Fig. 4. The FP-III region can get fully exposed post cleavage at the S2' site which dislodges the sheathing FP-I segment, or when the site gets progressively exposed during the conformational transition amidst the fusion process.

**Experimental corroboration.** Additional evidence of "rigidity" imparted by proline being critical for virus fusogenicity, infectivity and pathogenicity can be obtained from our previous in vitro and in vivo studies (39). We generated the Spike containing two consecutive prolines at 938-939 (S-MHV-A59(PP)) in the FP-III and its proline deletion (Δ938) mutant S-MHV-A59(P) from isogenic recombinant strain of MHV-A59, RSA59 (39) Spike protein. The proline deleted mutant only Spike gene construct S-MHV-A59(P) show slower trafficking to the cell surface, and significantly less fusogenicity. Established on differential properties of proline deletion in Spike construct, we generated proline deleted targeted recombinant mutant strain RSA59(P-) which contains S-MHV-A59(P) and compared it with two consecutive proline-containing parental isogenic strain RSA59 containing S-MHV-A59(PP). Proline deleted mutant RSA59(P-) infection in neuronal cell line demonstrated less aggressive and fewer syncytia formation, and one order lower viral titers post-infection *in vitro*. Based on these *in vitro* differential properties of proline deletion in Spike construct we generated proline deleted targeted recombinant mutant strain of RSA59(P-) which contains S-MHV-A59(P) and compared it with two consecutive proline containing parental isogenic strain RSA59 containing S-MHV-A59(PP). The *in vivo* studies in mice parallels the *in vitro* studies demonstrating significantly reduced viral replication and consecutive disease pathologies, like less severity in meningitis, encephalitis, and demyelination, and inability to infect the retina nor induce loss of retinal ganglion cells (41). The non-neurotropic strain MHV-2 sharing 91% genome identity with MHV-A59, and 83% pairwise Spike sequence identity, with a single central proline in FP-III, causes only meningitis and is unable to invade the brain parenchyma. Computational studies of S2 fusion domains of S-MHV-A59(PP) and MHV-2 Spike involving molecular dynamics confirmed the former to be more rigid and containing more residues in the regular secondary structure (39).

**Target for therapy.** Our study brings out the importance of the fusion loop region which could be a legitimate target for the design of vaccine or synthetic agents for therapy against COVID-19. The S2 domain serves as a better therapeutic target than S1 due to higher evolutionary conservation, but the few attempts made have mainly focussed attention on the heptad repeat regions (9). Two important features that also make the fusion loop an attractive therapeutic target are its accessibility due to its surface exposure in the full-length Spike, and the relatively high conservation of residues in the fusion loop, especially around the FP-II region which increases its scope as a pan-coronavirus target that can cater to future pandemic threats as well. Mimetic peptides can be designed to bind to the fusion loop to inhibit the fusogenic conformational transition of the S2 domain. Impairment of the fusion trigger would have a direct bearing on the fusogenicity of the virus and contribute to the reduction of lung invasion and damage that clinically results in acute pneumonia. Systematic studies can identify the minimal motif in the fusion loop serving as the fusogenic determinant to improve our selection of a potential therapeutic target to prevent cell-to-cell fusion and subsequent pathogenesis.

## Discussion

The interplay of the outlined physicochemical features determines the virus-entry process to become more efficient. The local and global stability of the S2 domain is important. Since the S2 domain undergoes a conformational transition, local stability means reasonably rigid moving parts, and global stability means a well-defined conformational transition pathway from the metastable to the stable state. This local and global stability requirement could be attributed to the physicochemical efficiency needed in disturbing the host membrane. Secondly, the electrostatic potential of the fusion peptide derived surface patches must be neutral or positive to be able to engage the host membrane. The presence of glycosylation sites adds to the hydrophobicity of the S2 domain surface, but it is a small fraction of the available surface for interaction with the host membrane. The fusogenic conformational transition requires optimal synergy between the physical and chemical properties of the fusion loop to allow a concurrent scything action to rapidly facilitate transition to the host-virus hemifusion membrane state. The free-energy available from the

conformational transition of S2 to a more relaxed helical bundle is available to disrupt the host membrane and overcome the kinetic barrier to bring the host and virus membrane lipid bilayers together. The Contact Initiation Model ensures that the virus and host membrane are in close proximity for the formation of a hemifusion structure. The pre-hairpin S2 intermediate as suggested to exist by many researchers may be one of the many conformational states interacting with the host membrane. Priming by Spike cleavage is important for facilitating the rapid fusion process and therefore a part of the synergy at play. However, the open conformation of RBD seen in PDB ID: 6VSB for SARS-CoV-2 suggests that flexible linker segments loosely connect the RBD back to the fusion domain leaving it relatively free for unfettered conformational transition. Therefore, multifarious options to prime and trigger appear to be available to SARS-CoV-2 for viral entry, which contributes to its increased infectivity. Preventing the trigger by inhibiting the fusion loop is therefore a suitable target for therapy. Given the importance, a more extensive study of SARS-CoV-2 Spike protein and the mechanistic hypothesis described here is therefore warranted.

**Materials and Methods**

The sequences used in this study were downloaded from the NCBI database (URL:http://www.ncbi.nlm.nih.gov). The multiple sequence alignments were performed using the T-coffee webserver (http://tcoffee.crg.cat/). The default parameter values for alignment available in the server were used. The server combines several methods to come up with an optimal multiple sequence alignment (42).

All protein three-dimensional structures were downloaded from the Protein Data Bank (http:/www.rcsb.org). The PDB IDs for the downloaded structures are HCoV-HKU1: 5I08, MHV-A59: 3JCL, 6VSJ, HCoV-OC43: 6NZK, MERS-CoV: 6Q04, SARS-CoV-2: 6VXX, 6VSB, and SARS-CoV: 5XLR. Structures with the highest resolution was preferred when more than one model was available. Coordinates from these files were extracted for obtaining starting models of the FP-I, FP-II, FP-III, used in our molecular dynamics simulations. Whenever there were missing coordinates, they were modeled as an extended structure in the FP. The SASA was calculated by the program

NACESS (http://wolf.bms.umist.ac.uk/naccess/); residues for which no atom coordinates are present in the PDB file was considered as fully exposed to solvent while calculating the relative SASA values. The secondary structure was calculated by the SECSTR program from the PROCHECK suite (https://www.ebi.ac.uk/thornton-srv/software/PROCHECK/). The electrostatic surface potential was calculated using the APBS plugin (38) inside the PyMol software (http://pymol.org). The default parameters were used for the calculations. All cartoon diagrams and surfaces were rendered by the PyMol software.

The molecular dynamics simulations were performed with GROMACS software (43) (http://www.gromacs.org). The simulations were performed using the CHARMM 27 forcefield with cmap (44). Each FP was placed in a cubic box solvated with water (SPC model). Solvent molecules were randomly replaced with $Na^+$ and $Cl^-$ to neutralize the system and bring the final concentration of NaCl to 0.1 M. Periodic boundary conditions were enforced in all three directions. The system was minimized until the maximum force in the system reached below 1000 kJ $mol^{-1}$ $nm^{-1}$. The systems were equilibrated for 2 ns under the NVT ensemble. This was followed by equilibration for 2 ns in the NPT ensemble at 1 atm and 300 K. The production run was executed for 500 ns saving the output every 100 ps yielding 5,000 frames for analysis. The analysis of the trajectory was performed using GROMACS utilities. The diagrams were created using our in-house software.

**Acknowledgments**

DP thanks the Department of Biotechnology, New Delhi for supporting the computational facilities, and Prof. Jayasri Das Sarma, Department of Biological Sciences, Indian Institute of Science Education and Research Kolkata for her generous support.

**References**

1. Shi Y*, et al.* (2020) COVID-19 infection: the perspectives on immune responses. *Cell Death & Differentiation* 27:1451-1452.
2. Zheng J (2020) SARS-CoV-2: an Emerging Coronavirus that Causes a Global Threat. *Int J Biol Sci* 16(10):1678-1685.
3. Petrosillo N, Viceconte G, Ergonul O, Ippolito G, & Petersen E (2020) COVID-19, SARS and MERS: are they closely related? *Clin Microbiol Infect* 26(6):729-734.

4. van Doremalen N*, et al.* (2020) Aerosol and Surface Stability of SARS-CoV-2 as Compared with SARS-CoV-1. *The New England journal of medicine* 382(16):1564-1567.
5. Morawska L & Cao J (2020) Airborne transmission of SARS-CoV-2: The world should face the reality. *Environment International*:105730.
6. Pedersen SF & Ho Y-C (2020) SARS-CoV-2: a storm is raging. *The Journal of clinical investigation* 130(5).
7. Wrapp D*, et al.* (2020) Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* 367(6483):1260-1263.
8. Xia S*, et al.* (2020) 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 research*:1-13.
9. Tang T, Bidon M, Jaimes JA, Whittaker GR, & Daniel S (2020) Coronavirus membrane fusion mechanism offers as a potential target for antiviral development. *Antiviral Research*:104792.
10. Sadasivan J, Singh M, & Sarma JD (2017) Cytoplasmic tail of coronavirus spike protein has intracellular targeting signals. *J Biosci* 42(2):231-244.
11. Das Sarma J*, et al.* (2008) Demyelinating and nondemyelinating strains of mouse hepatitis virus differ in their neural cell tropism. *Journal of virology* 82(11):5519-5526.
12. Das Sarma J, Kenyon LC, Hingley ST, & Shindler KS (2009) Mechanisms of primary axonal damage in a viral model of multiple sclerosis. *J Neurosci* 29(33):10272-10280.
13. Das Sarma J, Scheen E, Seo SH, Koval M, & Weiss SR (2002) Enhanced green fluorescent protein expression may be used to monitor murine coronavirus spread in vitro and in the mouse central nervous system. *Journal of neurovirology* 8(5):381-391.
14. Kenyon LC*, et al.* (2015) Gliopathy of Demyelinating and Non-Demyelinating Strains of Mouse Hepatitis Virus. *Front Cell Neurosci* 9:488.
15. Tai W*, et al.* (2020) Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. *Cellular & Molecular Immunology*.
16. Walls AC*, et al.* (2020) Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* 181(2):281-292.
17. Hofmann H & Pöhlmann S (2004) Cellular entry of the SARS coronavirus. *Trends Microbiol* 12(10):466-472.
18. Marzi A*, et al.* (2004) DC-SIGN and DC-SIGNR interact with the glycoprotein of Marburg virus and the S protein of severe acute respiratory syndrome coronavirus. *Journal of virology* 78(21):12090-12095.
19. Vaduganathan M*, et al.* (2020) Renin–Angiotensin–Aldosterone System Inhibitors in Patients with Covid-19. *New England Journal of Medicine*.
20. Hoffmann M*, et al.* (2020) SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell*.
21. Belouzard S, Madu I, & Whittaker GR (2010) Elastase-mediated activation of the severe acute respiratory syndrome coronavirus spike protein at discrete sites within the S2 domain. *The Journal of biological chemistry* 285(30):22758-22763.
22. Hingley ST, Leparc-Goffart I, Seo SH, Tsai JC, & Weiss SR (2002) The virulence of mouse hepatitis virus strain A59 is not dependent on efficient spike protein cleavage and cell-to-cell fusion. *Journal of neurovirology* 8(5):400-410.
23. Broer R, Boson B, Spaan W, Cosset F-L, & Corver J (2006) Important role for the transmembrane domain of severe acute respiratory syndrome coronavirus spike protein during entry. *Journal of virology* 80(3):1302-1310.
24. Petit CM*, et al.* (2005) Genetic analysis of the SARS-coronavirus spike glycoprotein functional domains involved in cell-surface expression and cell-to-cell fusion. *Virology* 341(2):215-230.
25. Alsaadi EAJ, Neuman BW, & Jones IM (2019) A Fusion Peptide in the Spike Protein of MERS Coronavirus. *Viruses* 11(9).

26. Guillen J, Kinnunen PK, & Villalain J (2008) Membrane insertion of the three main membranotropic sequences from SARS-CoV S2 glycoprotein. *Biochim Biophys Acta* 1778(12):2765-2774.
27. Guillen J, Perez-Berna AJ, Moreno MR, & Villalain J (2008) A second SARS-CoV S2 glycoprotein internal membrane-active peptide. Biophysical characterization and membrane interaction. *Biochemistry* 47(31):8214-8224.
28. Sainz B, Jr., Rausch JM, Gallaher WR, Garry RF, & Wimley WC (2005) Identification and characterization of the putative fusion peptide of the severe acute respiratory syndrome-associated coronavirus spike protein. *Journal of virology* 79(11):7195-7206.
29. White JM (1990) Viral and cellular membrane fusion proteins. *Annu Rev Physiol* 52:675-697.
30. Delos S, Gilbert J, & White J (2000) The central proline of an internal viral fusion peptide serves two important roles. *Journal of virology* 74(4):1686-1693.
31. Gómara MaJ, Mora P, Mingarro I, & Nieva JL (2004) Roles of a conserved proline in the internal fusion peptide of Ebola glycoprotein. *FEBS letters* 569(1-3):261-266.
32. Fredericksen BL & Whitt MA (1995) Vesicular stomatitis virus glycoprotein mutations that affect membrane fusion activity and abolish virus infectivity. *Journal of virology* 69(3):1435-1443.
33. Drummer HE & Poumbourios P (2004) Hepatitis C virus glycoprotein E2 contains a membrane-proximal heptad repeat sequence that is essential for E1E2 glycoprotein heterodimerization and viral entry. *Journal of Biological Chemistry* 279(29):30066-30072.
34. Harrison SC (2008) Viral membrane fusion. *Nature structural & molecular biology* 15(7):690-698.
35. Ou X*, et al.* (2016) Identification of the fusion peptide-containing region in betacoronavirus spike glycoproteins. *Journal of virology* 90(12):5586-5600.
36. Wimley WC & White SH (1996) Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nat Struct Biol* 3(10):842-848.
37. Shang J*, et al.* (2020) Structure of mouse coronavirus spike protein complexed with receptor reveals mechanism for viral entry. *PLoS pathogens* 16(3):e1008392.
38. Jurrus E*, et al.* (2018) Improvements to the APBS biomolecular solvation software suite. *Protein Science* 27(1):112-128.
39. Singh M*, et al.* (2019) A proline insertion-deletion in the spike glycoprotein fusion peptide of mouse hepatitis virus strongly alters neuropathology. *J Biol Chem* 294(20):8064-8087.
40. Chakrabarti P & Pal D (2001) The interrelationships of side-chain and main-chain conformations in proteins. *Progress in biophysics and molecular biology* 76(1-2):1-102.
41. Rout SS, Singh M, Shindler KS, & Das Sarma J (2020) One proline deletion in the fusion peptide of neurotropic mouse hepatitis virus (MHV) restricts retrograde axonal transport and neurodegeneration. *J Biol Chem*.
42. Di Tommaso P*, et al.* (2011) T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic acids research* 39(suppl_2):W13-W17.
43. Abraham MJ*, et al.* (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1:19-25.
44. Bjelkmar P, Larsson P, Cuendet MA, Hess B, & Lindahl E (2010) Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects from Correction Maps, Virtual Interaction Sites, and Water Models. *Journal of Chemical Theory and Computation* 6:459–466.

**Figures**

Cleavage site
S1/S2↓
816/797
S2'↓
Fusion Loop
S2 Domain
(A)
FL
HR1
CH
CD
HR2
TM
CP
SARS2 686 788 902 987 1035 76 141 163 212 237 1273
SARS 667 770 884 969 1017 58 123 145 194 219 1255
SARS2 788-806
SARS 770-788
816-833
798-815
875-902
857-884
Surface exposed Fusion peptides: FP-I, FP-II, FP-III
(B) Multiple Sequence Alignment
BAD AVG GOOD
FC:Furin Cleavage Site (S1/S2 domain boundary)[1]:RXXR↓
SARS2|QIK02964. PTWRVYSTGSNVFQTR--AGCLIGAEHVNN--SYECDIPIGAGICASYQTQT-NSPRRARSVA 688
SARS|NP_828851. PAWRIYSTGNNVFQTQ--AGCLIGAEHVDT--SYECDIPIGAGICASYHTVS-L-LR---STS 670
MERS|QBM11748.1 SMLKRRDSTYGPLQTP--VGCVLGLVNSSL-FVEDCKLPLGQSLCALPDTPSTLTPRSVRSVP 754
MHV-A59|AAU0635 --------PLNYFDSY--LGCVVNADNRTDEALPNCDLRMGAGLCVDYSKS----RRAHRSVS 720
OC43|YP_0095552 --------PINYFDSY--LGCVVNAYNSTAISVQTCDLTVGSGYCVDYSKN----RRSRGAIT 761
HKU1|Q14EB0.1 --------P-FYFDSY--LGCVLNAVNLTSYSVSSCDLRMGSGFCIDYALPS-S-RRKRRGIS 759
NL63|YP_003767. --------PNFYYVSNGGNNCTTA----------VMTYSNFGICADGSLIP-VRPRNSS--- 751
228E|NP_073551. --------PKFFYASNGTYNCTDA----------VLTYSSFGVCADGSIIA-VQPRNVS--- 570
Consensus
SARS2|QIK02964. SQSI----IAYTMSLGAENSVAY-SNNSIAIPTNFTISVTTEILPVSMTKTSVDCTMYICGDS 746
SARS|NP_828851. QKSI----VAYTMSLGADSSIAY-SNNTIAIPTNFSISITTEVMPVSMAKTSVDCNMYICGDS 728
MERS|QBM11748.1 GEMRLAS-IAFNHPIQVDQLNS--SYFKLSIPTNFSFGVTQEYIQTTIQKVTVDCKQYVCNGF 814
MHV-A59|AAU0635 TGYRLTTFEPYTPMLVNDSVQSVDGLYEMQIPTNFTIGHHEEFIQTRSPKVTIDCAAFVCGDN 783
OC43|YP_0095552 TGYRFTNFEPFTVNSVNDSLEPVGGLYEIQIPSEFTIGNMVEFIQTSSPKVTIDCAAFVCGDY 824
HKU1|Q14EB0.1 SPYRFVTFEPFNVSFVNDSVETVGGLFEIQIPTNFTIAGHEEFIQTSSPKVTIDCSAFVCSNY 822
NL63|YP_003767. ----------DNGISAI-ITANLSIPSNWTTSVQVEYLQITSTPIVVDCATYVCNGN 797
228E|NP_073551. ----------YDSVSAI-VTANLSIPSNWTTSVQVEYLQITSTPIVVDCSTYVCNGN 616
Consensus
← Fusion Peptide I[3]
SARS2|QIK02964. TECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKT-PPI-----KDFG-GFNF 802
SARS|NP_828851. TECANLLLQYGSFCTQLNRALSGIAAEQDRNTREVFAQVKQMYKT-PTL-----KYFG-GFNF 784
MERS|QBM11748.1 QKCEQLLREYGQFCSKINQALHGANLRQDDSVRNLFESVKSSQSS-PII-----PGFGGDFNL 871
MHV-A59|AAU0635 TACRQQLVEYGSFCVNVNAILNEVNNLLDNMQLQVASALMQGVTISSRLPDGISGPID-DINF 845
OC43|YP_0095552 AACKSQLVEYGSFCDNINAILTEVNELLDTTQLQVANSLMNGVTLSTKLKDGVNFNVD-DINF 886
HKU1|Q14EB0.1 AACHDLLSEYGTFCDNINSILNEVNDLLDITQLQVANALMQGVTLSSNLNTNLHSDVD-NIDF 884
NL63|YP_003767. PRCKNLLKQYTSACKTIEDALRLSAHLETNDVSSMLTFDSNAFSL-ANV-----TSFG-DYNL 853
228E|NP_073551. VRCVELLKQYTSACKTIEDALRNSARLESADVSEMLTFDKKAFTL-ANV-----SSFG-DYNL 672
Consensus
→ S2'↓← Fusion Peptide II[1] →
SARS2|QIK02964. SQILPD----------PSKPSKRSFIEDLLFNKVTLADAGFIKQ-YGDCL--GDIAARDLIC 851
SARS|NP_828851. SQILPD----------PLKPTKRSFIEDLLFNKVTLADAGFMKQ-YGECL--GDINARDLIC 833
MERS|QBM11748.1 TLLEPVSIST-------GS-RSARSAIEDLLFDKVTIADPGYMQG-YDDCMQQGPASARDLIC 925
MHV-A59|AAU0635 SPLLGCIGSTCAEDGNGPSAIRGRSAIEDLLFDKVKLSDVGFVEA-YNNCT--GGQEVRDLLC 905
OC43|YP_0095552 SPVLGCLGSEC-------SKASSRSAIEDLLFDKVKLSDVGFVEA-YNNCT--GGAEIRDLIC 939
HKU1|Q14EB0.1 KSLLGCLGSQC--------GSSSRSLLEDLLFNKVKLSDVGFVEA-YNNCT--GGSEIRDLLC 936
NL63|YP_003767. SSVLPQRNIR-------SSRIAGRSALEDLLFSKVVTSGLGTVDVDYKSCT--KGLSIADLAC 907
228E|NP_073551. SSVIPSLPTS-------GSRVAGRSAIEDILFSKLVTSGLGTVDADYKKCT--KGLSIADLAC 726
Consensus
← Fusion Peptide III[4] → ←Heptad
SARS2|QIK02964. AQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQN 914
SARS|NP_828851. AQKFNGLTVLPPLLTDDMIAAYTAALVSGTATAGWTFGAGAALQIPFAMQMAYRFNGIGVTQN 896
MERS|QBM11748.1 AQYVAGYKVLPPLMDVNMEAAYTSSLLGSIAGVGWTAGLSSFAAIPFAQSIFYRLNGVGITQQ 988
MHV-A59|AAU0635 VQSFNGIKVLPPVLSESQISGYTTGATAAAMFPPWS----AAAGVPFSLSVQYRINGLGVTMN 964
OC43|YP_0095552 VQSYKGIKVLPPLLSENQISGYTLAATSASLFPPWT----AAAGVPFYLNVQYRINGLGVTMD 998
HKU1|Q14EB0.1 VQSFNGIKVLPPILSETQISGYTTAATVAAMFPPWS----AAAGVPFSLNVQYRINGLGVTMD 995
NL63|YP_003767. AQYYNGIMVLPGVADAERMAMYTGSLIGGMVLGGLT----SAAAIPFSLALQARLNYVALQTD 966
228E|NP_073551. AQYYNGIMVLPGVADAERMAMYTGSLIGGIALGGLT----SAVSIPFSLAIQARLNYVALQTD 785
Consensus

```
                    Repeat Region 1 Begin[1]
SARS2|QIK02964.     VLYENQKLIANQFNSAIGKIQDSLSSTA--------------SALGKLQDVVNQNAQALNTLV 963
SARS|NP_828851.     VLYENQKQIANQFNKAISQIQESLTTTS--------------TALGKLQDVVNQNAQALNTLV 945
MERS|QBM11748.1     VLSENQKLIANKFNQALGAMQTGFTTTN--------------EAFQKVQDAVNNNAQALSKLA 1037
MHV-A59|AAU0635     VLSENQKMIASAFNNALGAIQDGFDATN--------------SALGKIQSVVNANAEALNNLL 1013
OC43|YP_0095552     VLSQNQKLIANAFNNALYAIQEGFDATN--------------SALVKIQAVVNANAEALNNLL 1047
HKU1|Q14EB0.1       VLNKNQKLIANAFNKALLSIQNGFTATN--------------SALAKIQSVVNANAQALNSLL 1044
NL63|YP_003767.     VLQENQKILAASFNKAINNIVASFSSVNDAITQTAEAIHTVTIALNKIQDVVNQQGSALNHLT 1029
228E|NP_073551.     VLQENQKILAASFNKAMTNIVDAFTGVNDAITQTSQALQTVATALNKIQDVVNQQGNSLNHLT 848
Consensus           ** :*** :*  **.*:  :  .:  .                 *: *:* .** :..:*. *
                    Heptad Repeat Region 1 End→←Central Helix Begin[1]                Central
SARS2|QIK02964.     KQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAA 1026
SARS|NP_828851.     KQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAA 1008
MERS|QBM11748.1     SELSNTFGAISASIGDIIQRLDVLEQDAQIDRLINGRLTTLNAFVAQQLVRSESAALSAQLAK 1100
MHV-A59|AAU0635     NQLSNRFGAISASLQEILTRLEAVEAKAQIDRLINGRLTALNAYISKQLSDSTLIKVSAAQAI 1076
OC43|YP_0095552     QQLSNRFGAISASLQEILSRLDALEAEAQIDRLINGRLTALNAYVSQQLSDSTLVKFSAAQAM 1110
HKU1|Q14EB0.1       QQLFNKFGAISSSLQEILSRLDNLEAQVQIDRLINGRLTALNAYVSQQLSDITLIKAGASRAI 1107
NL63|YP_003767.     SQLRHNFQAISNSIQAIYDRLDSIQADQQVDRLITGRLAALNAFVSQVLNKYTEVRGSRRLAQ 1092
228E|NP_073551.     SQLRQNFQAISSSIQAIYDRLDTIQADQQVDRLITGRLAALNVFVSHTLTKYTEVRASRQLAQ 911
Consensus           .:*   * ***  :  *  **: :: . *:****.*** :*:.:::: *        .   *
                    Helix End→                                          ←Connector
SARS2|QIK02964.     TKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHDGK--- 1086
SARS|NP_828851.     TKMSECVLGQSKRVDFCGKGYHLMSFPQAAPHGVVFLHVTYVPSQERNFTTAPAICHEGK--- 1068
MERS|QBM11748.1     DKVNECVKAQSKRSGFCGQGTHIVSFVVNAPNGLYFMHVGYYPSNHIEVVSAYGLCDSANPTN 1163
MHV-A59|AAU0635     EKVNECVKSQTTRINFCGNGNHILSLVQNAPYGLYFIHFSYVPISFTTANVSPGLCISGDR-- 1137
OC43|YP_0095552     EKVNECVKSQSSRINFCGNGNHIISLVQNAPYGLYFIHFSYVPTKYVTARVSPGLCIAGDR-- 1171
HKU1|Q14EB0.1       EKVNECVKSQSPRINFCGNGNHILSLVQNAPYGLLFIHFSYKPTSFKTVLVSPGLCLSGDR-- 1168
NL63|YP_003767.     QKINECVKSQSNRYGFCGNGTHIFSIVNSAPDGLLFLHTVLLPTDYKNVKAWSGICVDGIY-- 1153
228E|NP_073551.     QKVNECVKSQSKRYGFCGNGTHIFSIVNAAPEGLVFLHTVLLPTQYKDVEAWSGLCVDGTN-- 972
Consensus           *:.*** .*: * .***:* *:.*:   ** *: *:*    * .        .:*  .
                     Domain Begin[1]                                Connector Domain End→
SARS2|QIK02964.     AHFPREGVFV--SNGT-----HWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQ 1142
SARS|NP_828851.     AYFPREGVFV--FNGT-----SWFITQRNFFSPQIITTDNTFVSGNCDVVIGIINNTVYDPLQ 1124
MERS|QBM11748.1     CIAPVNGYFI--KTNNTRIVDEWSYTGSSFYAPEPITSLNTKYV-APQVTYQNISTNLPPPLL 1223
MHV-A59|AAU0635     GLAPKAGYFV--QDDG-----EWKFTGSSYYYPEPITDKNSVIMSSCAVNYTKAPEVFLNTSI 1193
OC43|YP_0095552     GIAPKSGYFV--NVNN-----TWMYTGSGYYYPEPITENNVVVMSTCAVNYTKAPYVMLNTSI 1227
NL63|YP_003767.     GYVLRQPNLVLYSDNG-----VFRVTSRVMFQPRLPVLSDFVQIYNCNVTFVNISRVELHTVI 1211
228E|NP_073551.     GYVLRQPNLALYKEGN-----YYRITSRIMFEPRIPTMADFVQIENCNVTFVNISRSELQTIV 1030
Consensus                  :    .  ------  :  *    : *.     :       *
                                          ←Heptad Repeat Region 2[3]
SARS2|QIK02964.     PE--LDSFKEELDKY-FKNHT-SPDVDLG-DISGINASVVNIQKEID--------------RL 1186
SARS|NP_828851.     PE--LDSFKEELDKY-FKNHT-SPDVDLG-DISGINASVVNIQKEID--------------RL 1168
MERS|QBM11748.1     GNSTGIDFQDELDEF-FKNVS-TSIPNFG-SLTQINTTLLDLTYEML--------------SL 1269
MHV-A59|AAU0635     PN--PPDFKEELDKW-FKNQT-SIAPDLSLDFEKLNVTLLDLTYEMN--------------RI 1238
OC43|YP_0095552     PN--LPDFKEELDQW-FKNQT-SVAPDLS--LDYINVTFLDLQVEMN--------------RL 1270
HKU1|Q14EB0.1       PN--LSDFEAEFSLW-FKNHT-SIAPNLTF-NSHINATFLDLYYEMN--------------VI 1268
NL63|YP_003767.     PD--YVDVNKTLQEF-AQNLPKYVKPNF--DLTPFNLTYLNLSSELKQLEAKTASLFQTTVEL 1269
228E|NP_073551.     PE--YIDVNKTLQELSYKL-PNYTVPDL--VVEQYNQTILNLTSEISTLENKSAELNYTVQKL 1088
Consensus           :    ..:  :.   :  .     ::      * : :::  *:                  :
                     Heptad Repeat Region 2 End→←Transmembrane Domain[3]→←  Cytoplasmic tail
SARS2|QIK02964.     NEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSC---LKG 1246
SARS|NP_828851.     NEVAKNLNESLIDLQELGKYEQYIKWPWYVWLGFIAGLIAIVMVTILLCCMTSCCSC---LKG 1228
MERS|QBM11748.1     QQVVKALNESYIDLKELGNYTYYNKWPWYIWLSFIAGLVALALCVFFILCCTGCGTN--CMG- 1329
MHV-A59|AAU0635     QDAIKKLNESYINLKEVGTYEMYVKWPWYVWLLIGLAGVAVCVLLFFICCCTGCGSC--CFK- 1298
OC43|YP_0095552     QEAIKVLNQSYINLKDIGTYEYYVKWPWYVWLLICLAGVAMLVLLFFICCCTGCGTS--CFK- 1330
HKU1|Q14EB0.1       QESIKSLNSSFINLKEIGTYEMYVKWPWYIWLLIVILFIIFLMILFFICCCTGCGSA--CFS- 1328
NL63|YP_003767.     QGLIDQINSTYVDLKLLNRFENYIKWPWWVWLIISVVFVVLLSLLVFCCLSTGCCGCCNCLT- 1331
228E|NP_073551.     QTLIDNINSTLVDLKWLNRVETYIKWPWWVWLCISVVLIFVVSMLLLCCCSTGCCGFFSCFA- 1150
Consensus           :    . :*.: ::*: :.    * ****::** :    : .    .:    *.*    :
                     ER or ERGIC localization signal[5]:←  →
SARS2|QIK02964.     CCSCGSCCKFD--EDDSEP-VLKGVKLHYT 1273
SARS|NP_828851.     ACSCGSCCKFD--EDDSEP-VLKGVKLHYT 1255
MERS|QBM11748.1     KLKCNRCCDRY--EEYDLE--PH--KVHVH 1353
MHV-A59|AAU0635     --KCGNCCDEY--GGHQDSIVIHNISSHED 1324
OC43|YP_0095552     --KCGGCCDDY--TGYQE-LVIK--TSHDD 1353
HKU1|Q14EB0.1       --KCHNCCDEY--GGHND-FVIK--ASHDD 1351
NL63|YP_003767.     -SSMRGCCDCGSTKLPYYE--FE--KVHVQ 1356
228E|NP_073551.     -SSIRGCCEST--KLPYYD--VE--KIHIQ 1173
Consensus           .   **.                .
```

**Figure 1. The S2 domain organization of SARS-CoV-2 Spike glycoprotein and its sequence comparison against seven other coronaviruses.** (A) The denomination of the various S2 domain segments and their positions. The labels are as follows, FL: Fusion Loop, HR1: Hepatad Repeat 1, CH: Central Helix, CD: Connector Domain, HR2: Heptad Repeat 2, TM: Transmembrane Domain, CP: Cytoplasmic tail. The span of the segments can be estimated from the position markers indicated for SAR-CoV-2 (SARS2) and SARS-CoV (SARS). (B) Multiple sequence alignment (MSA) of the S2 domain of Spike protein from eight coronaviruses. The segments of the S2 domain are also marked by position markers. Residues underlined in the MSA have no coordinates in the structure file and are deemed flexible regions. The PDB files used to mark the same are SARS-CoV-2: 6VXX, SARS-CoV:5XLR, MERS-CoV: 6Q04, MHV-A59: 3JCL, HCoV-OC43: 6NZK, HCoV-HKU1: 5I08. The sequences for HCoV-NL63 and HCoV-2289E are not marked because they belong to α-coronaviruses; whereas, all others are from β-coronaviruses. The RED marked region spanning Heptad repeat region 2 C-terminal and trans-membrane domain N-terminal region is also called the aromatic domain or the FP-IV. The MSA of the full-length protein and further annotation details can be found in Fig. S1.

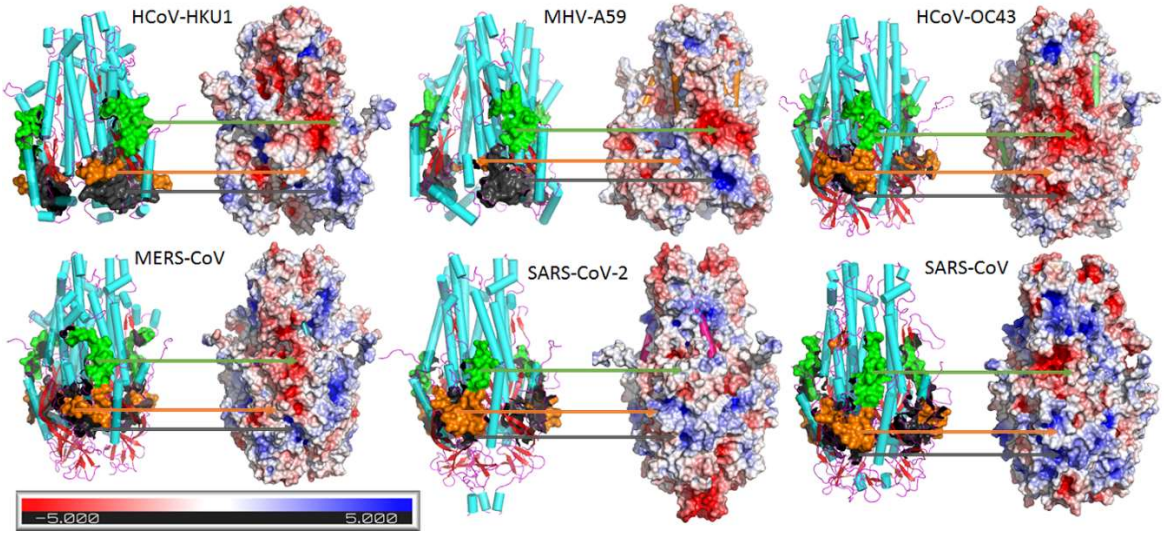



**Figure 2. Cartoon diagram showing the fusion peptide regions FP-I (orange), FP-II (green), and FP-III (dark grey) on the trimeric fusion domain structure of Spike from six viruses.** The electrostatic surface of the fusion domain is shown adjacent to each structure and the corresponding fusion peptide regions are marked by an arrow. The orientation of the cartoon structure and the electrostatic surface are aligned among themselves. Note that one face of the electrostatic surface that is visible is repeated on the other side due to the symmetry arising out of the trimeric quaternary structure. The fusion domain structures are shown in cyan for helices, strands in red, and loops in magenta. Please refer to Fig. 1 for the sequence alignment and its legend for the PDB ID of files used to draw the structures. Note that a part of the FP-I surface is absent for MHV-A59 and HCoV-HKU1 due to unavailable atom coordinates in the PDB file. The same is true for a small section of FP-II from SARS-CoV-2. For residue details, please refer to Fig. 4. All FP-I regions have at least one Asn-linked site, only SARS-CoV-2, SARS-CoV, and HCoV-HKU1 has a site in FP-II, and HCoV-OC43 and HCoV-HKU1 in FP-III. Contiguous to the FP-III are the heptad repeat regions where the glycosylation and sequence conservation among Spike is the highest. Spike from MHV-A59 shares a 64-66% overall sequence identity with HCoV-HKU1 and OC43, 71-76% identity with the corresponding S2 domains. Similar match with SARS-CoV, MERS-CoV, and SARS-CoV-2 are at <30% pairwise sequence identity.

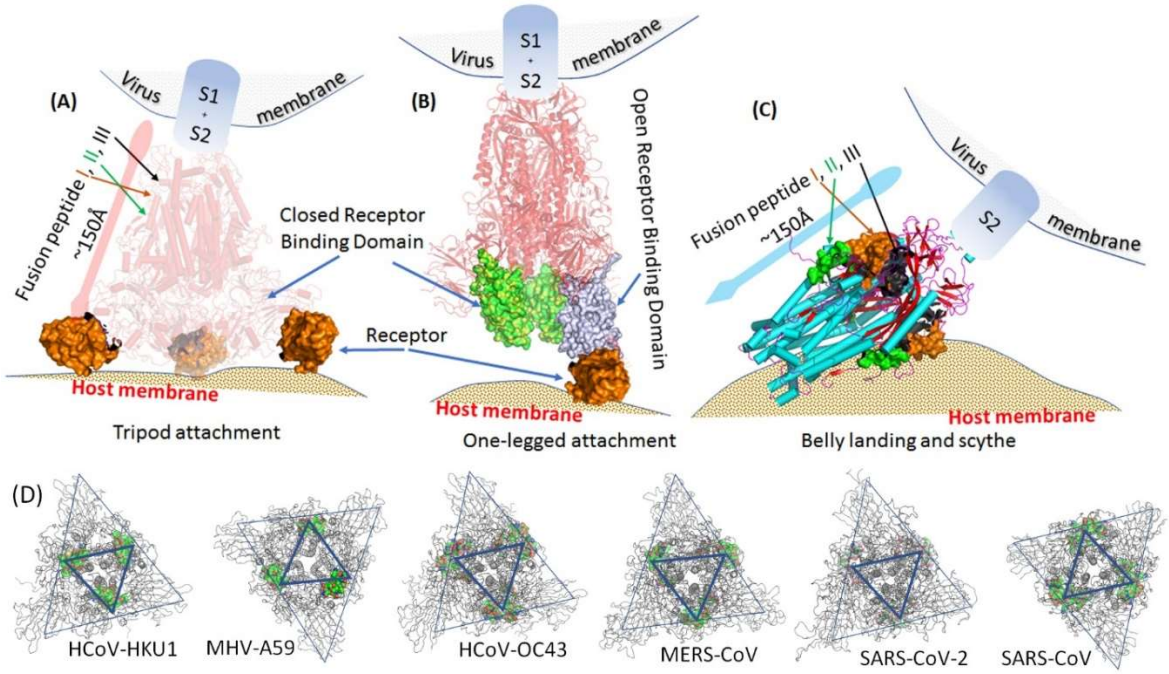



**Fig. 3**. **Schematic diagrams explaining the mode of binding of Spike protein to the host receptor and its putative orientation relative to the host membrane surface during the virus-host attachment.** (A). A cartoon diagram created using PDB ID: 6VSJ where the Spike protein from MHV-A59 is in complex with CEACAM1 receptor from mouse. Since all three subunits from the Spike RBD attach to the receptor, it allows the virus to anchor in a tripod mode. The approximate length of the Spike in the longer dimension is indicated along with the FP locations. (B). A cartoon diagram created using PDB ID: 6VSB, where one of the RBDs are shown in an open conformation. A receptor molecule ACE2 has been drawn to show the putative attachment in one-legged mode. Two-legged mode may also be possible similarly. (C). The S2 domain of SARS-CoV-2 Spike protein is shown in a belly-landing orientation. The approximate length of the Spike in the longer dimension is indicated along with the marked FP locations. The S1 domain is likely to be loosely bound and not shown for clarity. (D) A ribbon diagram of the trimeric Spike proteins from the six coronaviruses used in this study. Two triangles are marked on each structure, where the relative locations of the protruding NTDs appear near the outer triangle vertices, and the FP-I, FP-II, and FP-III colored surfaces are located near the inner triangle vertices. The triangles are marked to

bring out the relative positions of the NTDs and the FP surfaces. The Spike structure is oriented such that the RBD appears closest to the eye, followed by the NTD and then the FPs.

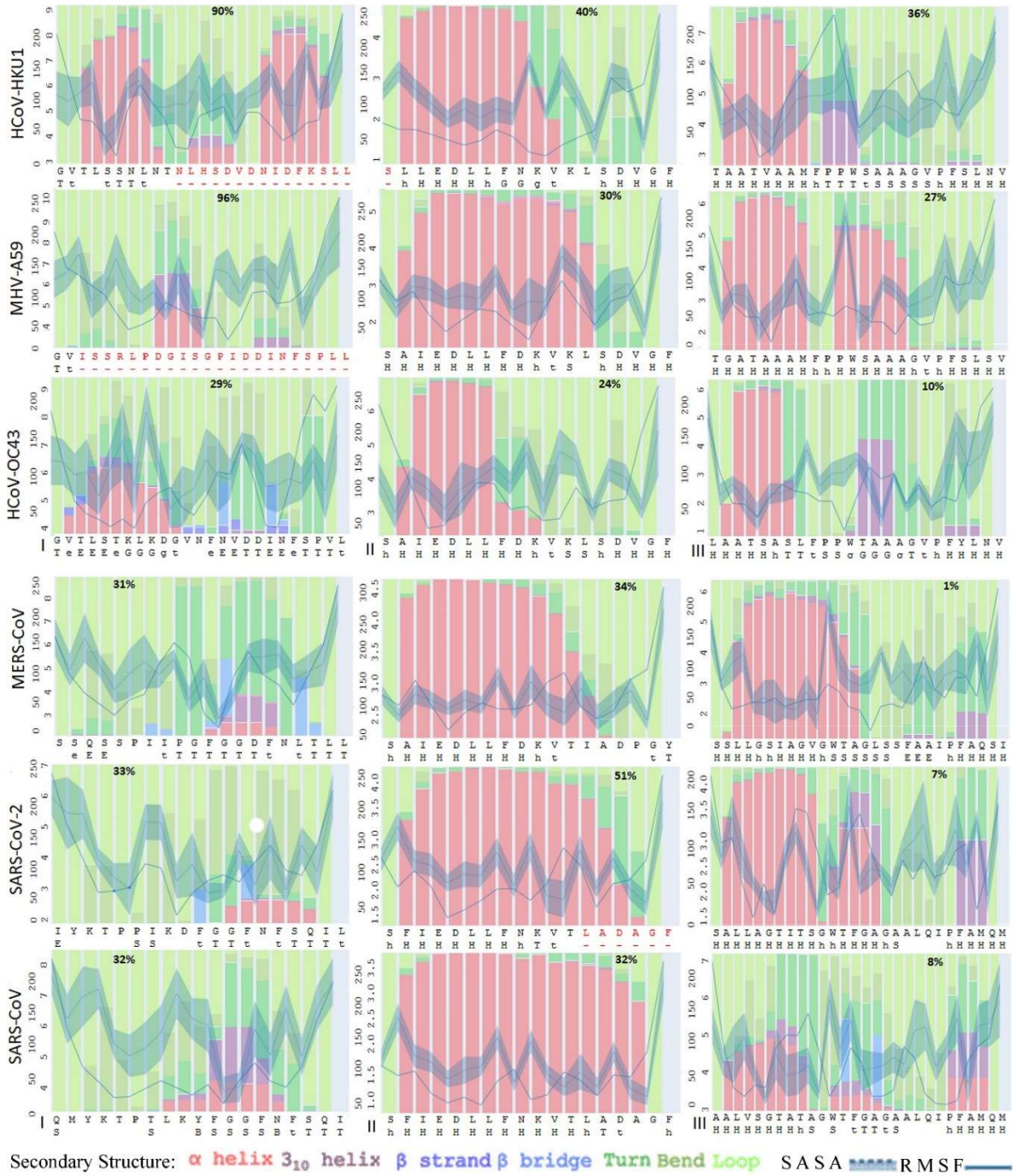



**Figure 4**. **Graphs showing various parameters calculated from the simulation trajectory of the isolated fragments from FP-I, FP-II, and FP-III of six coronaviruses.** The bar diagrams are drawn showing the secondary structure prevalence in 500 ns simulation, overlaid by the curves for Solvent Accessible Solvent Area (SASA; Å$^2$), and Root-Mean-Square-Fluctuation (RMSF; Å). The

Y-axis scales provide values for SASA and RMSF side-by-side. The standard deviation of the SASA in the simulation is indicated by a shaded background around the SASA curve. The X-axis indicates the residues for each FP fragment. The central proline residues are marked in BOLD and those residues for which the atom coordinates are absent in the PDB file are marked in RED. Below the residues, the corresponding secondary structures as indicated as present in the full-length protein. The symbols mean as follows, H:α-helix, h: α-helix termini, G: $3_{10}$-helix, g: $3_{10}$-helix termini, B: β-bridge, E: β-strand, e: β-strand termini, T: hydrogen-bonded turn, t: hydrogen-bonded turn termini, S: Bend, and <space>: irregular secondary structure. The first, second, and third columns are indicated by Labels I, II, and III corresponding to FP-I, FP-II, and FP-III in each Spike protein from a given virus, respectively. The relative SASA value of the whole FP fragment in the S2 domain expressed as a percentage is indicated in BOLD on the top part of each plot.

**Supplementary Material**

# Spike protein fusion loop controls SARS-CoV-2 fusogenicity and infectivity

**Debnath Pal**

Department of Computational and Data Sciences, Indian Institute of Science, Bengaluru, Karnataka-560012, India

## Supplementary Material

**Fig. S1. Multiple sequence alignment of Spike glycoproteins from eight coronaviruses.**

```
BAD AVG GOOD

                                  ←N-terminal Domain Begin[1]
SARS2|QIK02964.    MF--VFLVL-LPLVSSQCV-----NLTT--------RTQL--PPAY---------------- 28
SARS|NP_828851.    MF--IFLLF-LTLTSGSDL-----DRCT-TF-----DDVQ--APNY---------------- 30
MERS|QBM11748.1    MIHSVFLLM-FLLTPTESYVDVGPDSAKSACIEVDIQQTF--FDKTWPR-------------- 46
MHV-A59|AAU0635    ML-FVFILF-LPSCLGYIGD----FRCI-QLVNSNGANVS--APSISTE-------------- 40
OC43|YP_0095552    MF--LILLISLPTAFAVIGD----LKCT-SD-NINDKDTG--PPPISTD-------------- 39
HKU1|Q14EB0.1      MF--LIIFI-LPTTLAVIGD----FNCT-NS-FINDYNKT--IPRISED-------------- 38
NL63|YP_003767.    MK--LFLIL-LVLPLASCF-----FTCN-SNANLSMLQLG--VPDNSSTIVTGLLPTHWFCAN 52
228E|NP_073551.    MF--VLLVA-YALLHIA-G-----CQTT-NGLN-TSYSVCNGCVGYSEN-------------- 38
Consensus          *  ::::.

                                  Locations interacting with host receptor
SARS2|QIK02964.    ---TNSFTRGVYYPDK-VFRS-----------------------SVLHSTQDLFLPFFSNVT 63
SARS|NP_828851.    -TQHTSSMRGVYYPDE-IFRS-----------------------DTLYLTQDLFLPFYSNVT 67
MERS|QBM11748.1    -PIDVSKADGIIYPQGRTYSN-----------------------ITITY-QGLF-PYQGDHG 82
MHV-A59|AAU0635    -TVEVSQGLGTYYVLDRVYLN-----------------------ATLLL-TGYY-PVDGSKF 76
OC43|YP_0095552    -TVDVTNGLGTYYVLDRVYLN-----------------------TTLFL-NGYY-PTSGSTY 75
HKU1|Q14EB0.1      -VVDVSLGLGTYYVLNRVYLN-----------------------TTLLF-TGYF-PKSGANF 74
NL63|YP_003767.    QSTSVYSANGFFYIDVGNHRS-----------------------AFALH-TGYY-DANQYYI 89
228E|NP_073551.    -VFAVE--SGGYIPSDFAFNNWFLLTNTSSVVDGVVRSFQPLLLNCLWSV-SGLR-FTTGFVY 96
Consensus                *        . .                                .

SARS2|QIK02964.    ---WFHAIHVSGTNGTKRFDNPV---L------------PFNDGVYFAS-------------- 94
SARS|NP_828851.    ---GFHTINH-------TFGNPV---I------------PFKDGIYFAA-------------- 91
MERS|QBM11748.1    DMYVYSAGHATGTTPQKLFVANYSQDVK-----------QFANGFVVRIGAAANSTGTVIISP 134
MHV-A59|AAU0635    ---RNLALTGTNSVSLSWFQPPY---LS-----------QFNDGIFAKVQNLKTS-------- 114
OC43|YP_0095552    ---RNMALKGSVLLSRLWFKPPF---LS-----------DFINGIFAKVKNTKVI-------- 113
HKU1|Q14EB0.1      ---RDLALKGSKYLSTLWYKPPF---LS-----------DFNNGIFSKVKNTKLY-------- 112
NL63|YP_003767.    YVTNEIGLNASVTLKICKFSRNT---TFDFLSNASSSFDCIVNLLF----------------- 132
228E|NP_073551.    ---FN----GTGRGDCKGFSSDV---LS-----------DVIRYNL----------------- 121
Consensus                                :                .

SARS2|QIK02964.    -------TEKSNIIRGWIFGTTL---------------------------------------- 110
SARS|NP_828851.    -------TEKSNVVRGWVFGSTM---------------------------------------- 107
MERS|QBM11748.1    STSAIIRKIYPAF----MLGSSV---------------------------------------- 153
MHV-A59|AAU0635    -TPSGATAYFPTI----VIGSLF---------------------------------------- 132
OC43|YP_0095552    -KDRVMYSEFPAI----TIGSTF---------------------------------------- 131
HKU1|Q14EB0.1      -VNNTLYSEFSTI----VIGSVF---------------------------------------- 130
NL63|YP_003767.    -------TEQ--------LGAPLGITISGETVRLHLYNVTRTFYVPAAYKLTKLSVKCYFNYS 180
228E|NP_073551.    -------NFEENL----RRGTIL---------------------------------------- 133
Consensus                                *:  .

SARS2|QIK02964.    -----DSKTQSL------------------------LIVNNATN------------VVIKV- 130
SARS|NP_828851.    -----NNKSQSV------------------------IIINNSTN------------VVIRA- 127
MERS|QBM11748.1    -----GNFSYGKMGRFFNHTLVLLPDGCGTLLRAFYCILEPRSGN------------YCPAGN 199
MHV-A59|AAU0635    -----GYTSYTV------------------------VIEPYN------------GVIMASV- 152
OC43|YP_0095552    -----VNTSYSV------------------------VVQPRTINSTQDGDNKLQGLLEVSV- 163
HKU1|Q14EB0.1      -----VNTSYTI------------------------VVQPHN------------GILEITA- 150
NL63|YP_003767.    CVFSVVNATVTV------------------------NVTTHNG----------------RV- 201
228E|NP_073551.    -----FKTSYGV------------------------VVFYCTN------------------- 147
Consensus                :

SARS2|QIK02964.    --CEFQFCNDP---------------------FLGVYYHKNNKSWMES-EFRVY--------- 160
SARS|NP_828851.    --CNFELCDNP---------------------FFAVSKP--MGTQTH---TMIF--------- 153
MERS|QBM11748.1    SYTSFATYHTP---------------------ATDCSDGNYNRNASLN-SFKE---------- 230
MHV-A59|AAU0635    --CQYTICQLP---------------------YTDCKPNT-NGNKLIG-FWHTD--------- 181
OC43|YP_0095552    --CQYNMCEYP---------------------QTICHPNL--GNHRKE-LWHLD--------- 191
HKU1|Q14EB0.1      --CQYTMCEYP---------------------HTVCKSK---GSIRNE-SWHID--------- 177
NL63|YP_003767.    --VNYTVCDDCNGYTDNIFSVQQDGRIPNGFPFNNWFLL-TNGSTLVDGVSRLYQPLRLTCLW 261
228E|NP_073551.    ------------------------------------------------------------- 147
Consensus
```

# Supplementary Material

```
SARS2|QIK02964.    -----------SS------------------ANNCTFEYVSQPFL----MDLEGKQGNFKNLR 190
SARS|NP_828851.    -----------DN------------------AFNCTFEYISDAFS----LDVSEKSGNFKHLR 183
MERS|QBM11748.1    ----------YFN------------------LRNCTF-----MYT----YNITED------EI 250
MHV-A59|AAU0635    -----------VK------------------PPICVLKR---NFT----LNVNAD------AF 202
OC43|YP_0095552    -----------TG------------------VVSCLYKR---NFT----YDVNAD------YL 212
HKU1|Q14EB0.1      -----------SS------------------EPLCLFKK---NFT----YNVSAD------WL 198
NL63|YP_003767.    PVPGLKSSTGFVYFNATGSDVNCNGYQHNSVVDVMRYNL---NFSANSLDNLKSG------VI 315
228E|NP_073551.    ------------------------------------------------------------- 147
Consensus

SARS2|QIK02964.    E--FVFKNI---DGYFKIYSKHTPINLVRDLPQGFSALEPLV-DLPIGINITRFQ-------- 239
SARS|NP_828851.    E--FVFKNK---DGFLYVYKGYQPIDVVRDLPSGFNTLKPIF-KLPLGINITNFR-------- 232
MERS|QBM11748.1    LEWFGITQTAQGVHLFSSRYV--------DLY---GGNMFQFATLPVYDTIKYYS-------- 294
MHV-A59|AAU0635    Y--FHFYQH---GGTFYAYYA--------DKP---SATTFLF-SVYIGDILTQYY-------- 240
OC43|YP_0095552    Y--FHFYQE---GGTFYAYFT--------DTG---VVTKFLF-NVYLGMALSHYY-------- 250
HKU1|Q14EB0.1      Y--FHFYQE---RGVFYAYYA--------DVG---MPTTFLF-SLYLGTILSHYY-------- 236
NL63|YP_003767.    V--FKTLQY---DVLFYCSNS--------SSG------VLD-TTIPFGPSSQPYYCFINSTIN 358
228E|NP_073551.    --------------------N--------TLV------SGD-AHIPFGTVLGNFYCFVNTTIG 175
Consensus

SARS2|QIK02964.    ---------TLLALHRSYLTPGDSSSGWTA--------------------------GAAAYYV 267
SARS|NP_828851.    ---------AILTAF------SPAQDIWGT--------------------------SAAAYFV 254
MERS|QBM11748.1    ---------IIPHSIR------SIQSD-RK--------------------------AWAAFYV 315
MHV-A59|AAU0635    ---------VLPFICN------PTA--GST--------------------------FAPRYWV 260
OC43|YP_0095552    ---------VMPLTCN------S------K--------------------------LTLEYWV 266
HKU1|Q14EB0.1      ---------VMPLTCK------AISSNTDN--------------------------ETLEYWV 258
NL63|YP_003767.    TTHVSTFVGILPPTVR------EIVVA-RTGQFYINGFKYFDLGFIEAVNFNVTTASATDFWT 414
228E|NP_073551.    NETTSAFVGALPKTVR------EFVIS-RTGHFYINGYRYFTLGNVEAVNFNVTTAETTDFCT 231
Consensus

                                   N-terminal Domain End→                  ← Minimal
SARS2|QIK02964.    GYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRVQP-TESIVR- 328
SARS|NP_828851.    GYLKPTTFMLKYDENGTITDAVDCSQNPLAELKCSVKSFEIDKGIYQTSNFRVVP-SGDVVR- 315
MERS|QBM11748.1    YKLQPLTFLLDFSVDGYIRRAIDCGFNDLSQLHCSYESFDVESGVYSVSSFEAKP-SGSVVE- 376
MHV-A59|AAU0635    TPLVKRQYLFNFNQKGVITSAVDCASSYTSEIKCKTQSMLPSTGVYELSGYTVQP-VGVVYRR 322
OC43|YP_0095552    TPLTSRQYLLAFNQDGIIFNAVDCMSDFMSEIKCKTQSIAPPTGVYELNGYTVQP-IADVYRR 328
HKU1|Q14EB0.1      TPLSRRQYLLNFDEHGVITNAVDCSSSFLSEIQCKTQSFAPNTGVYDLSGFTVKP-VATVYRR 320
NL63|YP_003767.    VAFATFVDVLVNVSATNIQNLLYCD-SPFEKLQCEHLQFGLQDGFYSANFLDDNVLPETYVA- 475
228E|NP_073551.    VALASYADVLVNVSQTSIANIIYCN-SVINRLRCDQLSFDVPDGFYSTSPIQSVELPVSIVS- 292
Consensus

                   Receptor Binding Domain Begin[2]
SARS2|QIK02964.    FPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGV--SPTKLND 389
SARS|NP_828851.    FPNITNLCPFGEVFNATKFPSVYAWERKKISNCVADYSVLYNSTFFSTFKCYGV--SATKLND 376
MERS|QBM11748.1    QAEGVE-CDFSPLLFGT-PPQVYNFKRLVFTNCNYNLTKLLSLFSVNDFTCSQI--SPAAIAS 435
MHV-A59|AAU0635    VANLPA-CNIEEWLTARSVPSPLNWERKTFQNCNFNLSSLLRYVQAESLFCNNI--DASKVYG 382
OC43|YP_0095552    KPNLPN-CNIEAWLNDKSVPSPLNWERKTFSNCNFNMSSLMSFIQADSFTCNNI--DAAKIYG 388
HKU1|Q14EB0.1      IPNLPD-CDIDNWLNNVSVPSPLNWERRIFSNCNFNLSTLLRLVHVDSFSCNNL--DKSKIFG 380
NL63|YP_003767.    LP-------------------------IYYQHTDINFT------------------ATASFGG 495
228E|NP_073551.    LP-------------------------VYHKHTFIVLY---------------VDFKPQSGGG 315
Consensus

SARS2|QIK02964.    LCFTNV--------------------YADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGC 432
SARS|NP_828851.    LCFSNV--------------------YADSFVVKGDDVRQIAPGQTGVIADYNYKLPDDFMGC 419
MERS|QBM11748.1    NCYSSL--------------------ILDYFSYPLSMKSDLSVSSAGPISQFNYKQSFSNPTC 478
MHV-A59|AAU0635    RCFGSI--------------------SVDKFAVPRSRQVDLQLGNSGFLQTANYKIDTAATSC 425
OC43|YP_0095552    MCFSSI--------------------TIDKFAIPNGRKVDLQLGNLGYLQSFNYRIDTTATSC 431
HKU1|Q14EB0.1      SCFNSI--------------------TVDKFAIPNRRRDDLQLGSSGFLQSSNYKIDISSSSC 423
NL63|YP_003767.    SCYVCKPHQVNISLNGN----TSVCVRTSHFSIRYI-YNRVKSGSPG---DSSWHIYLKSGTC 550
228E|NP_073551.    KCFNCYPAGVNITLANFNETKGPLCVDTSHFTTKYV-AVYANV--------GRWSASINTGNC 369
Consensus

SARS2|QIK02964.    VIAWNSNNLDSKVG----GNYNYLYRLFRKSNLKPFERDIS--------TEIYQA--GSTPCN 481
SARS|NP_828851.    VLAWNTRNIDATST----GNYNYKYRYLRHGKLRPFERDIS--------NVPFSP--DGKPCT 468
MERS|QBM11748.1    LILATVPHNLTTIT----KPLKYSYINKCSRLLSDD-RTEV--------LQLVNA-NQYSPCV 527
MHV-A59|AAU0635    QLHYTLPKNNVTINNHNPSSWNRRYGFNDAGVFGK------NQHDVVYAQQCFTVRSSYCPCA 482
OC43|YP_0095552    QLYYNLPAANVSVSRFNPSTWNKRFGFIEDSVFKPRPAGVLTNHDVVYAQHCFKAPKNFCPCK 494
```

# Supplementary Material

```
HKU1|Q14EB0.1      QLYYSLPLVNVTINNFNPSSWNRRYGFGS---FNV------SSYDVVYSDHCFSVNSDFCPCA 477
NL63|YP_003767.    PFSF-------------------SK-------LNNF-QKFK--------TICFSTVEVPGSCN 578
228E|NP_073551.    PFSF------------------GK-------VNNF-VKFG--------SVCFSLKDIPGGCA 397
Consensus           :                              .                          *

SARS2|QIK02964.    GVEGFNCYFPLQSYG-----FQPTNGVG--------------YQPYRV-V------------- 511
SARS|NP_828851.    P-PALNCYWPLNDYG-----FYTTTGIG--------------YQPYRV-V------------- 497
MERS|QBM11748.1    S-IVPSTVWEDGDY-------Y-RKQLS--------------PLEGGG-W------------- 553
MHV-A59|AAU0635    Q-P-------DIVSPC------------------------------------------------ 490
OC43|YP_0095552    L-----------NGSCVGSGPGKNNGIGTCPAGTNYLTCDN------------LCTPDPITFT 534
HKU1|Q14EB0.1      D-P-------SVVNSCVKSKPLS----AICPAGTKYRHCDLDTTLYVNNWCRCSCLPDPISTY 528
NL63|YP_003767.    F-P-------LEATW-----HYTSYTIV--------------GALYVT-W------------- 600
228E|NP_073551.    M-P-------IVANW-----AYSKYYTI--------------GSLYVS-W------------- 419
Consensus

                                                                   Minimal Receptor
SARS2|QIK02964.    ----------------------------------------------VLSFELL--------- 518
SARS|NP_828851.    ----------------------------------------------VLSFELL--------- 504
MERS|QBM11748.1    --------LV------------------------------------ASGSTVAMTEQLQMGFG 572
MHV-A59|AAU0635    TTQTKPK-SAFVNVGDHCEGLGVLEDNCGNADPHKGCICANNSFIGWSHDTCLVNDRCQI-FA 551
OC43|YP_0095552    GTYKCPQTKSLVGIGEHCSGLAVKSDYCGGN----SCTCRPQAFLGWSADSCLQGDKCNI-FA 592
HKU1|Q14EB0.1      SPNTCPQKKVVVGIGEHCPGLGINEEKCGTQLNHSSCSCSPDAFLGWSFDSCISNNRCNI-FS 590
NL63|YP_003767.    ----------------------------------------------SEGNSIT--------- 607
228E|NP_073551.    ----------------------------------------------SDGDGIT--------- 426
Consensus                                                            .

Binding Domain End→
SARS2|QIK02964.    ------HAPATVCG---PKKSTNLV--KNKCVNFNFNGLTGTGVLTESNKKFL-PFQQFGRDI 569
SARS|NP_828851.    ------NAPATVCG---PKLSTDLI--KNQCVNFNFNGLTGTGVLTPSSKRFQ-PFQQFGRDV 555
MERS|QBM11748.1    ITVQYGTDTNSVCPKLEFANDTKIVSQLGNCVEYSLYGVSGRGVFQNCTAVGV-PQQRFVYDA 634
MHV-A59|AAU0635    NILLNGINSGTTCSTDLQLPNTEVV--TGICVKYDLYGITGQGVFKEVKADYYNSWQTLLYDV 612
OC43|YP_0095552    NFILHDVNSGLTCSTDLQKANTDII--LGVCVNYDLYGILGQGIFVEVNATYYNSWQNLLYDS 653
HKU1|Q14EB0.1      NFIFNGINSGTTCSNDLLYSNTEVS--TGVCVNYDLYGITGQGIFKEVSAAYYNNWQNLLYDS 651
NL63|YP_003767.    ------GVPYPVSG---IREFSNLV--LNNCTKYNIYDYVGTGIIRSSNQSLA-GGITYV-SN 657
228E|NP_073551.    ------GVPQPVEG---VSSFMNVT--LDKCTKYNIYDVSGVGVIRVSNDTFL-NGITYT-ST 476
Consensus                .  .       .:   . *.::.: .   * *::   .        .

SARS2|QIK02964.    ADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLT-- 630
SARS|NP_828851.    SDFTDSVRDPKTSEILDISPCAFGGVSVITPGTNASSEVAVLYQDVNCTDVSTAIHADQLT-- 616
MERS|QBM11748.1    YQNLVGYYSD-DGNYYCLRACVSVPVSVIYDK--ETKTHATLFGSVACEHISSTMSQYSRSTR 694
MHV-A59|AAU0635    NGNLNGFRDLTTNKTYTIRSCYSGRVSAAFHK--DAPEPALLYRNINCSYVFSNNISREEN-- 671
OC43|YP_0095552    NGNLYGFRDYITNRTFMIRSCYSGRVSAAFHA--NSSEPALLFRNIKCNYVFNNSLTRQLQ-- 712
HKU1|Q14EB0.1      NGNIIGFKDFLTNKTYTILPCYSGRVSAAFYQ--NSSSPALLYRNLKCSYVLNNIS-FISQ-- 709
NL63|YP_003767.    SGNLLGFKNVSTGNIFIVTPCNQPDQVAVY-Q--QSIIGAMTAVNE-SRYGLQNLL--Q-L-- 711
228E|NP_073551.    SGNLLGFKDVTKGTIYSITPCNPPDQLVVY-Q--QAVVGAMLSENF-TSYGFSNVV--E-L-- 530
Consensus              .  .       : .*      .      :   *    .

                                   FC:Furin Cleavage Site (S1/S2 domain boundary)[1]:RXXR↓
SARS2|QIK02964.    PTWRVYSTGSNVFQTR--AGCLIGAEHVNN--SYECDIPIGAGICASYQTQT-NSPRRARSVA 688
SARS|NP_828851.    PAWRIYSTGNNVFQTQ--AGCLIGAEHVDT--SYECDIPIGAGICASYHTVS-L-LR---STS 670
MERS|QBM11748.1    SMLKRRDSTYGPLQTP--VGCVLGLVNSSL-FVEDCKLPLGQSLCALPDTPSTLTPRSVRSVP 754
MHV-A59|AAU0635    --------PLNYFDSY--LGCVVNADNRTDEALPNCDLRMGAGLCVDYSKS----RRAHRSVS 720
OC43|YP_0095552    --------PINYFDSY--LGCVVNAYNSTAISVQTCDLTVGSGYCVDYSKN----RRSRGAIT 761
HKU1|Q14EB0.1      --------P-FYFDSY--LGCVLNAVNLTSYSVSSCDLRMGSGFCIDYALPS-S-RRKRRGIS 759
NL63|YP_003767.    --------PNFYYVSNGGNNCTTA-----------VMTYSNFGICADGSLIP-VRPRNSS--- 751
228E|NP_073551.    --------PKFFYASNGTYNCTDA-----------VLTYSSFGVCADGSIIA-VQPRNVS--- 570
Consensus                     :    .*          .  . *      *

SARS2|QIK02964.    SQSI----IAYTMSLGAENSVAY-SNNSIAIPTNFTISVTTEILPVSMTKTSVDCTMYICGDS 746
SARS|NP_828851.    QKSI----VAYTMSLGADSSIAY-SNNTIAIPTNFSISITTEVMPVSMAKTSVDCNMYICGDS 728
MERS|QBM11748.1    GEMRLAS-IAFNHPIQVDQLNS--SYFKLSIPTNFSFGVTQEYIQTTIQKVTVDCKQYVCNGF 814
MHV-A59|AAU0635    TGYRLTTFEPYTPMLVNDSVQSVDGLYEMQIPTNFTIGHHEEFIQTRSPKVTIDCAAFVCGDN 783
OC43|YP_0095552    TGYRFTNFEPFTVNSVNDSLEPVGGLYEIQIPSEFTIGNMVEFIQTSSPKVTIDCAAFVCGDY 824
HKU1|Q14EB0.1      SPYRFVTFEPFNVSFVNDSVETVGGLFEIQIPTNFTIAGHEEFIQTSSPKVTIDCSAFVCSNY 822
NL63|YP_003767.    ----------------DNGISAI-ITANLSIPSNWTTSVQVEYLQITSTPIVVDCATYVCNGN 797
228E|NP_073551.    ----------------YDSVSAI-VTANLSIPSNWTTSVQVEYLQITSTPIVVDCSTYVCNGN 616
Consensus                     :    .      : **:::: .    * :         :**   ::*..

                                                   ←   Fusion   Peptide I [3]
```

# Supplementary Material

```
SARS2|QIK02964.     TECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKT-PPI-----KDFG-GFNF 802
SARS|NP_828851.     TECANLLLQYGSFCTQLNRALSGIAAEQDRNTREVFAQVKQMYKT-PTL-----KYFG-GFNF 784
MERS|QBM11748.1     QKCEQLLREYGQFCSKINQALHGANLRQDDSVRNLFESVKSSQSS-PII-----PGFGGDFNL 871
MHV-A59|AAU0635     TACRQQLVEYGSFCVNVNAILNEVNNLLDNMQLQVASALMQGVTISSRLPDGISGPID-DINF 845
OC43|YP_0095552     AACKSQLVEYGSFCDNINAILTEVNELLDTTQLQVANSLMNGVTLSTKLKDGVNFNVD-DINF 886
HKU1|Q14EB0.1       AACHDLLSEYGTFCDNINSILNEVNDLLDITQLQVANALMQGVTLSSNLNTNLHSDVD-NIDF 884
NL63|YP_003767.     PRCKNLLKQYTSACKTIEDALRLSAHLETNDVSSMLTFDSNAFSL-ANV-----TSFG-DYNL 853
228E|NP_073551.     VRCVELLKQYTSACKTIEDALRNSARLESADVSEMLTFDKKAFTL-ANV-----SSFG-DYNL 672
Consensus             * . * :*   *  ::  *            .:     .  .  . :       ..  . ::
                      →                    S2'↓← Fusion Peptide II[1] →
SARS2|QIK02964.     SQILPD-----------PSKPSKRSFIEDLLFNKVTLADAGFIKQ-YGDCL--GDIAARDLIC 851
SARS|NP_828851.     SQILPD-----------PLKPTKRSFIEDLLFNKVTLADAGFMKQ-YGECL--GDINARDLIC 833
MERS|QBM11748.1     TLLEPVSIST-------GS-RSARSAIEDLLFDKVTIADPGYMQG-YDDCMQQGPASARDLIC 925
MHV-A59|AAU0635     SPLLGCIGSTCAEDGNGPSAIRGRSAIEDLLFDKVKLSDVGFVEA-YNNCT--GGQEVRDLLC 905
OC43|YP_0095552     SPVLGCLGSEC-------SKASSRSAIEDLLFDKVKLSDVGFVEA-YNNCT--GGAEIRDLIC 939
HKU1|Q14EB0.1       KSLLGCLGSQC--------GSSSRSLLEDLLFNKVKLSDVGFVEA-YNNCT--GGSEIRDLLC 936
NL63|YP_003767.     SSVLPQRNIR-------SSRIAGRSALEDLLFSKVVTSGLGTVDVDYKSCT--KGLSIADLAC 907
228E|NP_073551.     SSVIPSLPTS-------GSRVAGRSAIEDILFSKLVTSGLGTVDADYKKCT--KGLSIADLAC 726
Consensus           . :                  ** :**:**.*:  :. * :.  * .*          ** *
                                         ←        Fusion Peptide III [4]     →      ←Heptad
SARS2|QIK02964.     AQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQN 914
SARS|NP_828851.     AQKFNGLTVLPPLLTDDMIAAYTAALVSGTATAGWTFGAGAALQIPFAMQMAYRFNGIGVTQN 896
MERS|QBM11748.1     AQYVAGYKVLPPLMDVNMEAAYTSSLLGSIAGVGWTAGLSSFAAIPFAQSIFYRLNGVGITQQ 988
MHV-A59|AAU0635     VQSFNGIKVLPPVLSESQISGYTTGATAAAMFPPWS----AAAGVPFSLSVQYRINGLGVTMN 964
OC43|YP_0095552     VQSYKGIKVLPPLLSENQISGYTLAATSASLFPPWT----AAAGVPFYLNVQYRINGLGVTMD 998
HKU1|Q14EB0.1       VQSFNGIKVLPPILSETQISGYTTAATVAAMFPPWS----AAAGVPFSLNVQYRINGLGVTMD 995
NL63|YP_003767.     AQYYNGIMVLPGVADAERMAMYTGSLIGGMVLGGLT----SAAAIPFSLALQARLNYVALQTD 966
228E|NP_073551.     AQYYNGIMVLPGVADAERMAMYTGSLIGGIALGGLT----SAVSIPFSLAIQARLNYVALQTD 785
Consensus           .*   *  *** :      : ** .   .      :    :   :**   :  *:* :.:  :
                    Repeat Region 1 Begin[1]
SARS2|QIK02964.     VLYENQKLIANQFNSAIGKIQDSLSSTA--------------SALGKLQDVVNQNAQALNTLV 963
SARS|NP_828851.     VLYENQKQIANQFNKAISQIQESLTTTS--------------TALGKLQDVVNQNAQALNTLV 945
MERS|QBM11748.1     VLSENQKLIANKFNQALGAMQTGFTTTN--------------EAFQKVQDAVNNNAQALSKLA 1037
MHV-A59|AAU0635     VLSENQKMIASAFNNALGAIQDGFDATN--------------SALGKIQSVVNANAEALNNLL 1013
OC43|YP_0095552     VLSQNQKLIANAFNNALYAIQEGFDATN--------------SALVKIQAVVNANAEALNNLL 1047
HKU1|Q14EB0.1       VLNKNQKLIANAFNKALLSIQNGFTATN--------------SALAKIQSVVNANAQALNSLL 1044
NL63|YP_003767.     VLQENQKILAASFNKAINNIVASFSSVNDAITQTAEAIHTVTIALNKIQDVVNQQGSALNHLT 1029
228E|NP_073551.     VLQENQKILAASFNKAMTNIVDAFTGVNDAITQTSQALQTVATALNKIQDVVNQQGNSLNHLT 848
Consensus           ** :*** :*  **.*:  :  .:  .                 *: *:* .** :..:*. *
                    Heptad Repeat Region 1 End→←Central Helix Begin[1]                Central
SARS2|QIK02964.     KQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAA 1026
SARS|NP_828851.     KQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAA 1008
MERS|QBM11748.1     SELSNTFGAISASIGDIIQRLDVLEQDAQIDRLINGRLTTLNAFVAQQLVRSESAALSAQLAK 1100
MHV-A59|AAU0635     NQLSNRFGAISASLQEILTRLEAVEAKAQIDRLINGRLTALNAYISKQLSDSTLIKVSAAQAI 1076
OC43|YP_0095552     QQLSNRFGAISASLQEILSRLDALEAEAQIDRLINGRLTALNAYVSQQLSDSTLVKFSAAQAM 1110
HKU1|Q14EB0.1       QQLFNKFGAISSSLQEILSRLDNLEAQVQIDRLINGRLTALNAYVSQQLSDITLIKAGASRAI 1107
NL63|YP_003767.     SQLRHNFQAISNSIQAIYDRLDSIQADQQVDRLITGRLAALNAFVSQVLNKYTEVRGSRRLAQ 1092
228E|NP_073551.     SQLRQNFQAISSSIQAIYDRLDTIQADQQVDRLITGRLAALNVFVSHTLTKYTEVRASRQLAQ 911
Consensus           .:*   * ***  :  *  **: :: . *:****.*** :*:.:::: *         .   *
                    Helix End→                                          ←Connector
SARS2|QIK02964.     TKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHDGK--- 1086
SARS|NP_828851.     TKMSECVLGQSKRVDFCGKGYHLMSFPQAAPHGVVFLHVTYVPSQERNFTTAPAICHEGK--- 1068
MERS|QBM11748.1     DKVNECVKAQSKRSGFCGQGTHIVSFVVNAPNGLYFMHVGYYPSNHIEVVSAYGLCDSANPTN 1163
MHV-A59|AAU0635     EKVNECVKSQTTRINFCGNGNHILSLVQNAPYGLYFIHFSYVPISFTTANVSPGLCISGDR-- 1137
OC43|YP_0095552     EKVNECVKSQSSRINFCGNGNHIISLVQNAPYGLYFIHFSYVPTKYVTARVSPGLCIAGDR-- 1171
HKU1|Q14EB0.1       EKVNECVKSQSPRINFCGNGNHILSLVQNAPYGLLFIHFSYKPTSFKTVLVSPGLCLSGDR-- 1168
NL63|YP_003767.     QKINECVKSQSNRYGFCGNGTHIFSIVNSAPDGLLFLHTVLLPTDYKNVKAWSGICVDGIY-- 1153
228E|NP_073551.     QKVNECVKSQSKRYGFCGNGTHIFSIVNAAPEGLVFLHTVLLPTQYKDVEAWSGLCVDGTN-- 972
Consensus            *:.*** .*: * .***:* *:.*:   ** *: *:*    * .        .:*  .
                     Domain Begin[1]                                Connector Domain End→
SARS2|QIK02964.     AHFPREGVFV--SNGT-----HWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQ 1142
SARS|NP_828851.     AYFPREGVFV--FNGT-----SWFITQRNFFSPQIITTDNTFVSGNCDVVIGIINNTVYDPLQ 1124
MERS|QBM11748.1     CIAPVNGYFI--KTNNTRIVDEWSYTGSSFYAPEPITSLNTKYV-APQVTYQNISTNLPPPLL 1223
MHV-A59|AAU0635     GLAPKAGYFV--QDDG-----EWKFTGSSYYYPEPITDKNSVIMSSCAVNYTKAPEVFLNTSI 1193
OC43|YP_0095552     GIAPKSGYFV--NVNN-----TWMYTGSGYYYPEPITENNVVVMSTCAVNYTKAPYVMLNTSI 1227
```

## Supplementary Material

```
HKU1|Q14EB0.1      GIAPKQGYFI--KQND-----SWMFTGSSYYYPEPISDKNVVFMNSCSVNFTKAPFIYLNNSI 1224
NL63|YP_003767.    GYVLRQPNLVLYSDNG-----VFRVTSRVMFQPRLPVLSDFVQIYNCNVTFVNISRVELHTVI 1211
228E|NP_073551.    GYVLRQPNLALYKEGN-----YYRITSRIMFEPRIPTMADFVQIENCNVTFVNISRSELQTIV 1030
Consensus                 :     .      :  *    : *.      :        *
                                          ←Heptad Repeat Region 2[3]
SARS2|QIK02964.    PE--LDSFKEELDKY-FKNHT-SPDVDLG-DISGINASVVNIQKEID--------------RL 1186
SARS|NP_828851.    PE--LDSFKEELDKY-FKNHT-SPDVDLG-DISGINASVVNIQKEID--------------RL 1168
MERS|QBM11748.1    GNSTGIDFQDELDEF-FKNVS-TSIPNFG-SLTQINTTLLDLTYEML--------------SL 1269
MHV-A59|AAU0635    PN--PPDFKEELDKW-FKNQT-SIAPDLSLDFEKLNVTLLDLTYEMN--------------RI 1238
OC43|YP_0095552    PN--LPDFKEELDQW-FKNQT-SVAPDLS--LDYINVTFLDLQVEMN--------------RL 1270
HKU1|Q14EB0.1      PN--LSDFEAEFSLW-FKNHT-SIAPNLTF-NSHINATFLDLYYEMN--------------VI 1268
NL63|YP_003767.    PD--YVDVNKTLQEF-AQNLPKYVKPNF--DLTPFNLTYLNLSSELKQLEAKTASLFQTTVEL 1269
228E|NP_073551.    PE--YIDVNKTLQELSYKL-PNYTVPDL--VVEQYNQTILNLTSEISTLENKSAELNYTVQKL 1088
Consensus           :   ..:  :.   :  .    ::     * : :::  *:                   :
                     Heptad Repeat Region 2 End→←Transmembrane Domain[3]→←  Cytoplasmic tail
SARS2|QIK02964.    NEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSC---LKG 1246
SARS|NP_828851.    NEVAKNLNESLIDLQELGKYEQYIKWPWYVWLGFIAGLIAIVMVTILLCCMTSCCSC---LKG 1228
MERS|QBM11748.1    QQVVKALNESYIDLKELGNYTYYNKWPWYIWLSFIAGLVALALCVFFILCCTGCGTN--CMG- 1329
MHV-A59|AAU0635    QDAIKKLNESYINLKEVGTYEMYVKWPWYVWLLIGLAGVAVCVLLFFICCCTGCGSC--CFK- 1298
OC43|YP_0095552    QEAIKVLNQSYINLKDIGTYEYYVKWPWYVWLLICLAGVAMLVLLFFICCCTGCGTS--CFK- 1330
HKU1|Q14EB0.1      QESIKSLNSSFINLKEIGTYEMYVKWPWYIWLLIVILFIIFLMILFFICCCTGCGSA--CFS- 1328
NL63|YP_003767.    QGLIDQINSTYVDLKLLNRFENYIKWPWWVWLIISVVFVVLLSLLVFCCLSTGCCGCCNCLT- 1331
228E|NP_073551.    QTLIDNINSTLVDLKWLNRVETYIKWPWWVWLCISVVLIFVVSMLLLCCCSTGCCGFFSCFA- 1150
Consensus          :   . :*.: ::*: :.    * ****::** :    : .    .:    *.*     :
                    ER or ERGIC localization signal[5]:←  →
SARS2|QIK02964.    CCSCGSCCKFD--EDDSEP-VLKGVKLHYT 1273
SARS|NP_828851.    ACSCGSCCKFD--EDDSEP-VLKGVKLHYT 1255
MERS|QBM11748.1    KLKCNRCCDRY--EEYDLE--PH--KVHVH 1353
MHV-A59|AAU0635    --KCGNCCDEY--GGHQDSIVIHNISSHED 1324
OC43|YP_0095552    --KCGGCCDDY--TGYQE-LVIK--TSHDD 1353
HKU1|Q14EB0.1      --KCHNCCDEY--GGHND-FVIK--ASHDD 1351
NL63|YP_003767.    -SSMRGCCDCGSTKLPYYE--FE--KVHVQ 1356
228E|NP_073551.    -SSIRGCCEST--KLPYYD--VE--KIHIQ 1173
Consensus           .   **.                 .
```

*Residues underlined have no coordinates in the structure file and are deemed flexible regions. The PDB files used to mark the same are SARS-CoV-2: 6VXX, SARS-CoV:5XLR, 6Q04: MERS-CoV, MHV-A59:3JCL, HCoV-OC43:6NZK, HCoV-HKU1:5I08. The sequences for HCoV-NL63 and HCoV-2289E are not marked because they belong to α-coronaviruses, whereas all others are from β-coronaviruses. The RED marked region spanning Heptad repeat region 2 C-terminal and trans-membrane domain N-terminal region is also called the aromatic domain or the fourth fusion peptide FP-IV.*

**Accession number of Genome/Spike protein sequences used in this alignment**

1: AY700211.1 / AAU06356.1 (Murine hepatitis virus strain A59)
2: KT029139.1 / QBM11748.1 (MERS-CoV/KOR/KNIH/002_05_2015)
3: NC_004718.3 / NP_828851.1 (SARS coronavirus)
4: MN985325.1 / QIK02964.1 (2019-nCoV/USA-WA1/2020)
5. NC_005831.2 /YP_003767.1 (Human Coronavirus NL63,)
6. NC_006213.1 / YP_009555241.1 (Human coronavirus OC43 strain ATCC VR-759)
7. NC_006577.2 / Q14EB0.1 (Human coronavirus HKU1)
8. NC_002645.1 / NP_073551.1 (Human coronavirus 229E)

**References used to annotate the alignment:**


1 Wrapp, D. *et al.* Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260-1263 (2020).

2 Hofmann, H. & Pöhlmann, S. Cellular entry of the SARS coronavirus. *Trends Microbiol* **12**, 466-472 (2004).

3 Xia, S. *et al.* 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 research*, 1-13 (2020).

4 Singh, M. *et al.* A proline insertion-deletion in the spike glycoprotein fusion peptide of mouse hepatitis virus strongly alters neuropathology. *J Biol Chem* 294, 8064-8087, DOI:10.1074/jbc.RA118.004418 (2019).
5 Sadasivan, J., Singh, M. & Sarma, J. D. Cytoplasmic tail of coronavirus spike protein has intracellular targeting signals. J Biosci 42, 231-244, DOI:10.1007/s12038-017-9676-7 (2017).

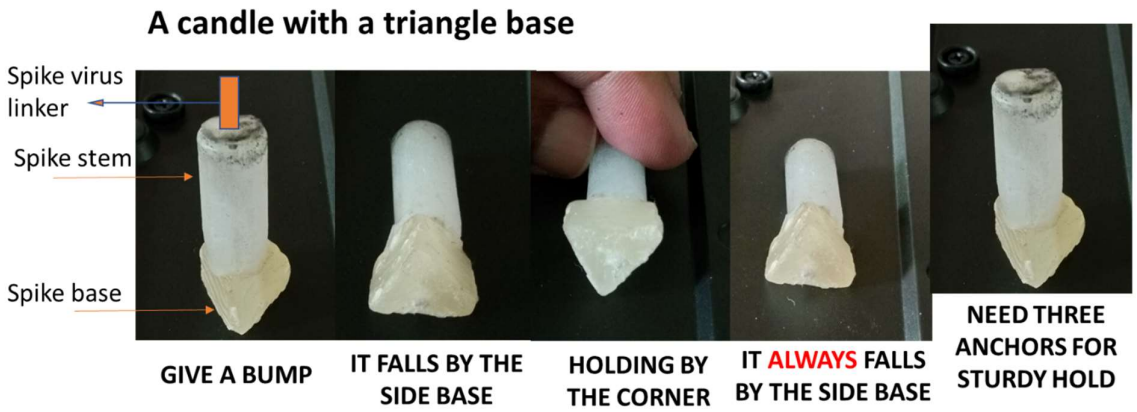
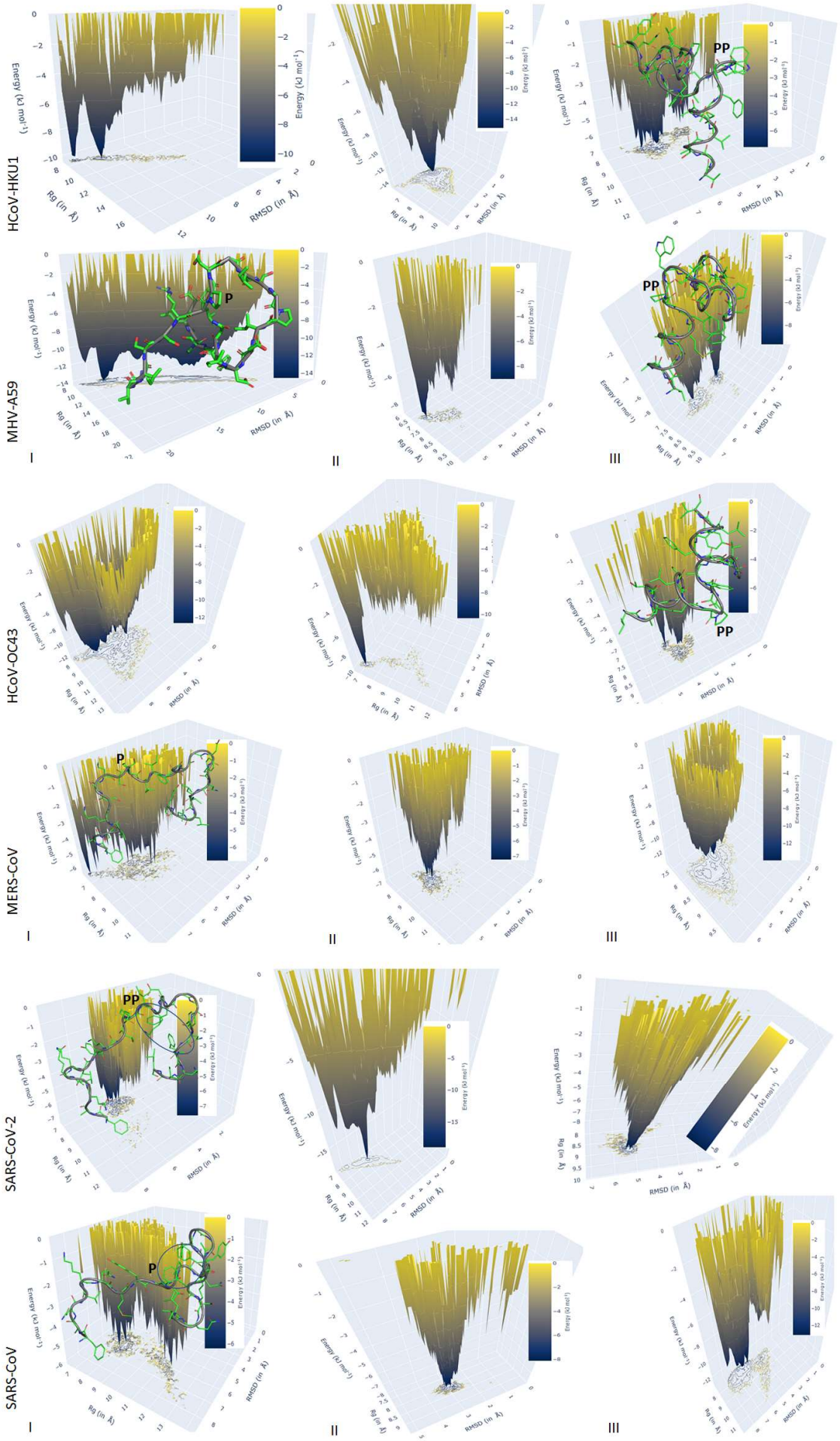



**Fig. S2.** A series of snapshots showing the properties of the Contact Initiation Model. The model is created using candle wax, where the base represents the triangular shape as found for the trimeric spike S1 domain, and the cylindrical stem represents the trimeric S2 domain. The stem is attached to the virus envelope through trimeric flexible linkers at the C-terminal end of the Spike ectodomain. The model on the extreme left shows that an unanchored structure on a triangle base, when bumped from the side, will fall in a position where one of the base sides rests on the surface. This in turn would ensure that the fusion peptide surface on the stem aligned along the middle of the base will make the initial contact and trigger the fusion process (Please check the triangles marked in Fig. 2D alongside). The fact that the orientation holding it by the corner of the base is not stable is emphasized in third and fourth illustrations, and the model always falls such that it rests on the side base. This architecture guarantees the fusion peptides in the S2 domain make the initial contact given that they are located at the crest of the bulge on the S2 domain surface. The last figure emphasizes that three anchors are needed for a stable vertical "tripod" orientation of the spike with respect to the host surface.

# Supplementary Material



**Supplementary Material**

**Fig. S3. Energy landscape plot corresponding to all fusion peptides shown in Fig. 4 (main text)**. The ball and stick model of the starting structures of all the fusion peptides with central proline is shown along with their backbone trace. The central prolines are marked by "P". Carbon atoms are in green, oxygen in red, nitrogen in blue, and sulfur in yellow. The location of the aromatic cluster is marked by oval in SARS-COV-2 and SARS-CoV Spike protein. The spread of the conformation in each FP can be visually estimated from the contours projected onto the X-Y plane.