

RESEARCH ARTICLE

Mutations in membrane-fusion subunit of spike glycoprotein play crucial role in the recent outbreak of COVID-19

Soumita Podder¹  | Avishek Ghosh² | Tapash Ghosh^{1,3}¹Department of Microbiology, Raiganj University, Uttar Dinajpur, West Bengal, India²Department of Microbiology, Maulana Azad College, Kolkata, West Bengal, India³Department of Bioinformatics, Bose Institute, Kolkata, West Bengal, India**Correspondence**

Soumita Podder, Department of Microbiology, Raiganj University, Uttar Dinajpur-733134, W.B, India.

Email: s.podder@raiganjuniversity.ac.in and piu2007.soumita@gmail.com**Abstract**

Coronavirus disease-2019 (COVID-19), the ongoing pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a major threat to the entire human race. It is reported that SARS-CoV-2 seems to have relatively low pathogenicity and higher transmissibility than previously outbreak SARS-CoV. To explore the reason of the increased transmissibility of SARS-CoV-2 compared with SARS-CoV, we have performed a comparative analysis on the structural proteins (spike, envelope, membrane, and nucleoprotein) of two viruses. Our analysis revealed that extensive substitutions of hydrophobic to polar and charged amino acids in spike glycoproteins of SARS-CoV2 creates an intrinsically disordered region (IDR) at the beginning of membrane-fusion subunit and intrinsically disordered residues in fusion peptide. IDR provides a potential site for proteolysis by furin and enriched disordered residues facilitate prompt fusion of the SARS-CoV2 with host membrane by recruiting molecular recognition features. Here, we have hypothesized that mutation-driven accumulation of intrinsically disordered residues in spike glycoproteins play dual role in enhancing viral transmissibility than previous SARS-coronavirus. These analyses may help in epidemic surveillance and preventive measures against COVID-19.

KEYWORDS

intrinsically disordered region, molecular recognition feature, SARS-CoV2, spike glycoprotein

1 | INTRODUCTION

Novel coronavirus (2019-nCoV or SARS-CoV-2) has caused ongoing global epidemics with high morbidity and mortality. Coronaviruses (order Nidovirales, family Coronaviridae, and subfamily Coronavirinae) are primarily known to cause enzootic infections in birds and mammals; however, in the last few decades; they have crossed the animal-human species barrier.^{1,2} The outbreak of severe acute respiratory syndrome (SARS) in 2002–2003 and, more recently, Middle-East respiratory syndrome (MERS) in 2012 has confirmed the lethality of CoVs when they crossed the species barrier and started to infect humans. The consequent outbreak of SARS in 2003, 8096 cases and 774 deaths reported worldwide, resulting in a fatality rate of 9.6%.³ Whereas the outbreak of MERS in April 2012 up until

October 2018, 2229 cases, and 791 associated deaths have been confirmed globally, resulting in a case-fatality rate of 35.5%.⁴ The novel coronavirus has reported to share about 79% sequence similarity with the SARS-coronavirus, about 50% with the MERS-coronavirus.⁵ SARS-CoV-2 is associated with an ongoing outbreak of atypical pneumonia (coronavirus disease-2019 [COVID-2019]) that has affected 4,425,485 people and killed 302,059 of those affected in more than 60 countries as on May 16, 2020.⁶ On January 30, 2020, the World Health Organization declared the SARS-CoV-2 as a pandemic.

Coronaviruses are enveloped viruses with a positive sense, single-stranded RNA genome. The viral genome encodes four major structural proteins: the spike (S) protein, nucleocapsid (N) protein, membrane (M) protein, and the envelope (E) protein, all of which are

crucial to produce a structurally complete viral particle.⁷ Coronavirus enter into host cells by using transmembrane spike (S) glycoprotein that forms homotrimers extended from the viral envelope.⁸ S encompasses two functional subunits-S1, responsible for binding to the host cell receptor and S2, involved in fusion of the viral envelope and host cellular membranes. For many CoVs, S protein is cleaved at the boundary between the S1 and S2 subunits, which remain as a single polypeptide in the prefusion conformation.⁹ The distal S1 subunit comprises the receptor-binding domain (RBD) and facilitates the stabilization of the prefusion state of the membrane-anchored S2 subunit containing the fusion machinery.¹⁰ The cleavage at S1/S2 boundary has been anticipated to stimulate the protein by irreversible conformational changes for membrane fusion.¹¹ The host proteases for S protein cleavage differ among different coronaviruses, which play crucial roles in determining the epidemiological and pathological features of virus, including host range, tissue tropism, transmissibility, and mortality. For example, a variety of human proteases, such as trypsin, trypsin-like protease, and transmembrane protease serine 2 are reported to cleave and activate the S protein of SARS-CoV.^{12,13} Depending on the viral species, coronaviruses recognize a variety of entry receptors to infect the host. SARS-CoV and several SARS-related coronaviruses (SARSr-CoV) interact directly with angiotensin-converting enzyme 2 via S protein to enter into the target cells.¹⁴ Recently, it is reported that mutation in the RBD in SARS-CoV-2 renders more efficient human-human transmission.¹⁵ Scientists have found that SARS-CoV-2 S glycoprotein possesses a furin cleavage site at the boundary between the S₁/S₂ subunits which helps in activating the fusion machinery of the virus.^{16,17} These two distinctive features in SARS-CoV2 could partially explain the efficient transmission of SARS-CoV-2 in humans.

A recent study by Zhao et al.¹⁸ has estimated basic reproduction number (R_0) for 2019-nCoV in the early phase of the outbreak and revealed that mean R_0 for SARS-CoV2 is ranging from 3.3 to 5.5 which is higher than those of SARS-CoV (R_0 : 2–5). The higher transmissibility of this virus turns the outbreak into a pandemic. Thus, it is of the prime interests of the researchers to untangle all the uniqueness of this newly emerged coronavirus by comparing with the previous human infecting SARS-CoV for designing protective measures against it.

We have studied an in depth mutational spectra and evolutionary dynamics of these four structural proteins by comparing SARS-CoV2 and human infecting SARS-CoV. Analyzing the impact of a mutation in proteins we have found that an intrinsically disordered region is acquired at the beginning of fusion protein (S2) which offers furin cleavage site in SARS-CoV2. Moreover, higher predisposition of intrinsically disordered residues in S2 observed to contain three molecular recognition features (MoRFs). We here hypothesized a unique fusion mechanism favored by the MoRFs present in the fusion peptide of the novel coronavirus. Thus, our study provides new insight into the genomic feature responsible for the rapid transmission of SARS-CoV2 as well as it could help in designing preventives against COVID-19.

2 | MATERIALS AND METHODS

2.1 | Sequence retrieval

Up to date, 1590 genome sequences of SARS-CoV2 are deposited in ViPR database (<https://www.viprbrc.org>).¹⁹ Complete genome is available for 1017 isolates from different geographical regions. We have retrieved 1017 coding sequences of SARS-CoV2 structural proteins (S, E, M, and N) and human infecting SARS-CoV three isolates from China, Germany, and USA. Genome accession numbers for all genome studied here are provided in Table S1.

2.2 | Calculation of evolutionary rate

Coding sequences of S, E, M, and N of SARS-CoV and SARS-CoV2 were aligned by CLUSTAL W, then Pairwise synonymous (ds) and nonsynonymous (dn) distances between the orthologous genes were calculated using the Phylogenetic Analysis Using Maximum Likelihood²⁰ package (PAML, yn00) to identify regions and sites under evolutionary selection.²¹

Pairwise amino acid residues of all proteins between SARS-CoV and SARS-CoV2 and within different strains of SARS-CoV2 were aligned by CLUSTAL omega and amino acid substitution were calculated by house-build Perl program.

Mutation analysis was done by PROVEAN (<http://provean.jcvi.org/index.php>) and SIFT (<https://sift.bii.a-star.edu.sg/>). PROVEAN is very much useful for filtering sequence variants to identify non-synonymous or indel variants that are predicted to be functionally important²² and in SIFT algorithm for sorting Intolerant from Tolerant could efficiently predict whether an amino acid substitution affects protein function or not.²³

2.2.1 | Prediction of IDR and MoRF

Coordinates of S1 and S2 subunits of S proteins in SARS-CoV and SARS-CoV-2 were retrieved from Pfam (<https://pfam.xfam.org/>).

Intrinsically disordered region (IDRs) of all four proteins of SARS-CoV and CoV2 were predicted by PONDR® VLXT (<http://www.pondr.com/>), predictors of natural disorder region. PONDR® VLXT applies three different feedforward neural networks (VL1, XN, and XC). XN and XC [22] for N- terminal and C-terminal region, respectively and VL1²⁴ for the internal region of the sequence. This method is frequently used for disorder calculation in virus.^{25–27}

MoRF was predicted by MoRFchibi (<https://morf.msl.ubc.ca/index.xhtml>). This tool is used for its high accuracy predictions and it provides more than double the precision of other predictors.²⁸

2.3 | Statistical test

All the statistical tests were performed using the SPSS package.

3 | RESULTS

3.1 | Mutational spectrum in the structural protein-coding genes in SARS-CoV2 arisen during evolution

Bayesian evolutionary rate and divergence date estimates were shown that nonsynonymous-to-synonymous substitution rate ratio is decreasing from SARS (1.41) to MERS (0.35) and MERS to HCoV-OC43 (0.133).²⁹ Here, we have measured evolutionary distance between 1017 strain of SARS-CoV2 spreading throughout the world with human infecting SARS-CoV predominantly had spread in three different geographical regions (USA, China, and Germany) in 2003–2004. For this, we have calculated nonsynonymous substitution per nonsynonymous site (dn) and synonymous substitution per synonymous site (ds) among the four structural protein-coding sequences (S, M, E, and N) of orthologous genes present in SARS-CoV and SARS-CoV-2. We have noticed significantly ($p = .001$) higher nonsynonymous substitution rates and synonymous substitution rates in S protein compared with the other three proteins (Table 1). A similar observation is also documented in the recent study of Tang et al.³⁰ Pairwise alignment by Needleman–Wunsch algorithm between the four structural proteins of SARS-CoV and SARS-CoV2, have shown that average percentages of amino acid substitutions occurred in S, M, N, and E proteins are 21.9, 9.5, 8.76, and 4, respectively. These results imply that significantly ($p = .001$) higher amino acid substitutions have occurred in S proteins than the other three proteins during evolution which is reflected in the higher dN value of S proteins. Analyzing the effects of amino acids substitution in proteins by PROVEAN and SIFT, we revealed five deleterious mutations, that is, the mutations may cause protein structural destabilization (C19T, L54S, L286T, P335A, and Y1070H) have occurred in S protein whereas no such deleterious mutations were observed in other three proteins. S protein is very much crucial for virus entry since it interacts with receptor and fuse with host membrane. It was evident from different studies that genes which crucial for the survival of the organisms are remained conserved (low dN/dS) over the evolutionary time scale.³⁰ Thus, according to the neutral theory of evolution³¹ higher nonsynonymous substitution rate compelled S proteins to experience significantly ($p = .001$) higher synonymous substitution rates than other proteins to retain overall conservation of the S proteins (Table 1).

Though, it was well evidenced that RNA viruses accumulate more mutation rates than DNA viruses due to lack of proofreading

TABLE 1 Comparison on average nonsynonymous (dN) and synonymous (dS) substitution rates of four structural proteins in SARS-CoV-2

Protein name	Average dN ($n = 1017$)	Average dS ($n = 1017$)
S	0.156	1.262
M	0.067	0.551
N	0.057	0.368
E	0.031	0.136

Abbreviations: E, envelope; M, membrane; N, nucleocapsid; S, spike, SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

activity in RNA polymerase they have encoded.³² However, it would be interesting to investigate whether the accumulation of non-synonymous mutation preferentially in S proteins than others offer any benefits to the virus for enhancing their potency of infectivity.

3.2 | Effects of mutations on the protein structural features in SARS-CoV2

We have analyzed the properties of substituted amino acids in the four structural proteins during evolution from SARS-CoV to SARS-CoV2 to investigate whether the amino acids substituted with a similar group of amino acids or not. We have detected extensive hydrophobic (Hy) to polar (P) and charged (C) amino acids exchange in S proteins than other three structural proteins (S: Hy-P/C = [51/91] = 56.1% and P/C-Hy = [40/91] = 43.9% [p -value = .02]; M: Hy-P/C = [4/7] = 57.2%, P/C-Hy = [3/7] = 42.8% [p -value = .59]; N: Hy-P/C = [7/15] = 46.6%, P/C-Hy = [8/15] = 53.3% [p -value = .71]). In E proteins, no substitution of amino acid with the different groups has found. The details of amino acid mutations of S, M, and N proteins which took place during evolution from SARS-CoV to SARS-CoV2 have delineated in Table 2. Several high throughput studies on protein structure have evidenced that regions in a protein-enriched with polar and charged amino acids have a tendency to conform IDR.³³ Moreover, IDRs in virus endure several structural features associated with viral pathogenicity.³⁴ Thus, we have predicted IDR in all structural proteins in SARS-CoV2 by PONDR-VLXT and compared the predisposition of IDR with the corresponding proteins present in SARS-CoV. Here, we have found that M and E proteins of both SARS-CoV and SARS-CoV2 do not contain any IDR (consecutive disordered residues > 30 amino acids) in their proteins (Table 3). N proteins contain three IDRs and percentages of intrinsically disordered residues in their proteins are remarkably high. However, the enrichment of IDR in N proteins is similar for both viruses (Table 3). Interestingly, we have revealed an IDR (671–708) in S proteins of SARS-CoV-2 but no IDR is found in their previous orthologous SARS-CoV (Figure 1A, Table 3). Moreover, percentages of disordered residues (PID) are significantly ($p = .035$) increased in S proteins of SARS-CoV2 compared with SARS-CoV which implies more disorder residues become enriched in S proteins after evolution (Table 3). Genomic analysis of S genes deduced that out of 1017 genome of SARS-CoV2, 491 viral strains are 100% similar and 526 strains have differed from each other. Multiple alignments of Spike protein sequences from 526 different isolates with one of the similar isolates have revealed a total of 31 Single Amino Acid Polymorphisms (SAPs) but none of them has occurred in the predicted IDR (671–708) which indicates the region is conserved among all of them (Figure 2A). We also noticed D614G mutation is predominant in 504 isolates (Figure 2A; Table S1). Next, comparing the IDR between SARS-CoV/CoV2, we have traced an insertion mutation which allows to incorporate three new disorder promoting amino acids serine, proline, and arginine in S proteins of SARS-CoV2 (Figure 2A). Along with this, the substitution of order promoting to disorder promoting amino

TABLE 2 Position wise amino acid substitution between S, M, and N proteins of SARS-CoV and SARS-CoV2

Protein	Position of hydrophobic to polar/charged amino acids substitution	Position of polar/charged to hydrophobic amino acids substitution
S	L16C; F22R; M37T; L54S; G77D; A91S; A126K; P143K; M144S; M151R; F153Y; F157N; F193Y; G199K; F232Q; A233T; A237Y; I244S; G246T; F253Y; A284S; L286T; V308Q; G311E; F360S; V404K; M417T; A430S; G446S; V458E; F460Y; P461Q; G464S; P470E; L515K; P540E; F558T; A590T; A618T; V663Q; L665N; A714T; A732S; L792S; A854Q; A857S; A866S; A1037S; F1079S; F1092Y; A1229C	T9P; T11V; D17V; C19L; N29A; S36F; T51V; Y63F; T71A; H74G; S120A; T146M; T150F; E174G; K190I; T215A; T247A; S248G; S279A; Q280L; T359A; S432V; T433G; K439L; Y442L; D463G; T485P; Q546L; S556A; S607P; T608V; T669V; S670A; T775P; N827A; S861A; Q904L; S914G; S924A; S1052A
M	A29T; M32C; A210S; G211S	C27F; S39A; T188G
N	G32E; P38S; G80S; G193N; G214N; A380T; P391Q	D26G; S121G; N153A; T158I; S213G; T218A; Q268A; T377A

Abbreviations: M, membrane; N, nucleocapsid; S, spike, SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

TABLE 3 Comparison on the intrinsic disorder content of four proteins between SARS-CoV and SARS-CoV-2

Disorder features	SARS-CoV-S	SARS-CoV2-S
Total disorder residues	65	98
No of disorder region (>30a.a)	NIL	1
PID	5.18	7.70
	SARS-CoV-M	SARS-CoV2-M
Total disorder residues	14	13
No of disorder region (>30a.a)	NIL	NIL
PID	6.28	5.86
	SARS-CoV-E	SARS-CoV2-E
Total disorder residues	10	9
No of disorder region (>30a.a)	NIL	NIL
PID		
	SARS-CoV-N	SARS-CoV2-N
Total disorder residues	212	208
No of disorder region (>30a.a)	3	3
PID	50.74	49.6

Note: PID indicates the proportion of disorder residues to the total length of amino acids predicted by PONDR-VLXT.

Abbreviations: E, envelope; M, membrane; N, nucleocapsid; PID, percentages of disordered residues; S, spike, SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

acids in five positions (H661Q, V663Q, L665N, L666A, and D684E) help to create the new IDR in SARS-CoV2 (Figure 2B). These results imply that the novel coronavirus acquired a new intrinsically disordered region in their spike glycoprotein which is crucial for entry into the host. Thus, it is imperative to explore the connection between IDR and elevated transmissibility in SARS-CoV2.

3.3 | Role of IDR in S protein in the rapid transmission of SARS-CoV-2

S proteins contain two subunits-S1 and S2. Pfam prediction on S proteins of two viruses depicted two domains: (i) Receptor binding domain (321–556 in SARS-CoV and 330–583 in SARS-CoV-2); (ii) S2 domain (635–1240 in SARS-CoV and 671–1270 in SARS-CoV-2). Thus, it is clearly observed that the IDR (671–708) has enriched in membrane fusion domain (S2) of spike glycoproteins in SARS-CoV-2 (Figure 1A). Recently, two research groups have reported that fusion protein acquired a new furin cleavage site (682–685) at the upstream of fusion peptide (S2').^{16,17} We have found that this cleavage site actually resides in the IDR. Since, intrinsically disordered proteins (IDPs)/IDPRs lack stable well-folded three-dimensional structures, the structural instability renders exceptional sensitivity to proteolysis.³⁵ Thus, the new IDR offers the cleavage site of furin protease as well as assists in efficient proteolytic cleavage of S proteins to activate the fusion peptide in SARS-CoV-2.

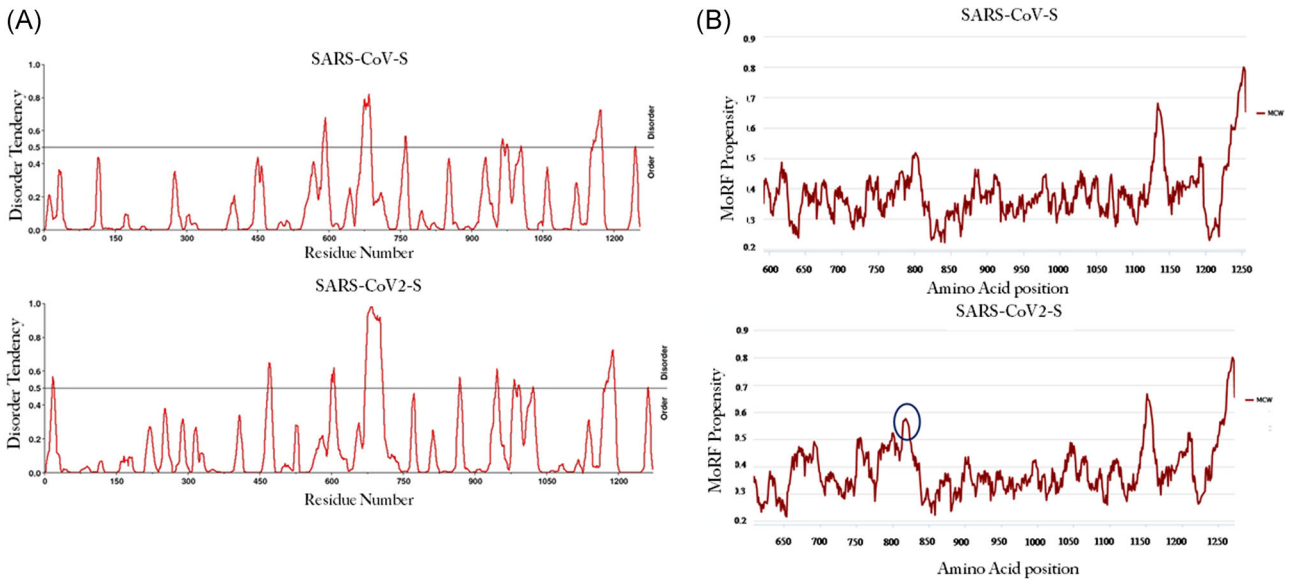


FIGURE 1 Comparison on IDR and MoRF between SARS-CoV and SARS-CoV2. (A) Comparison on the intrinsic disorder tendency of the amino acid residues in spike glycoproteins of SARS-CoV and SARS-CoV2. Disorder score above 0.5 is considered as cut-off value. (B) Comparison between MoRF content in SARS-CoV and SARS-CoV2. MoRF propensity score 0.5 was considered as cutoff. Circle shows the enrichment of additional MoRF in SARS-CoV2. IDR, intrinsically disordered region; MoRF, molecular recognition feature; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2

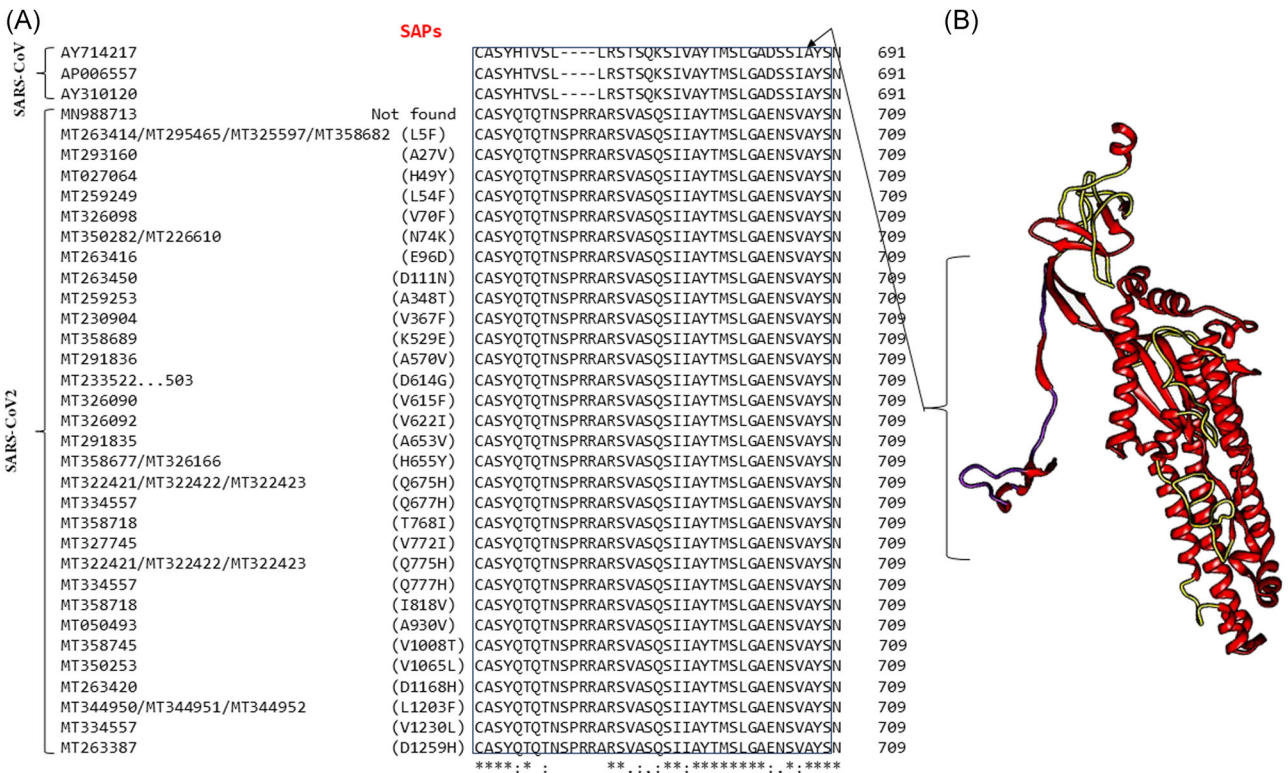


FIGURE 2 Conservation of IDR in all isolates of SARS-CoV2. (A) Pairwise alignment of IDR of S proteins in three strains of SARS-CoV and different strains of SARS-CoV-2 having change in amino acid level shows the amino acid substitutions between the two viruses as well as the conservation of this region (indicated by black box) in all SARS-CoV-2 isolates. (B) IDR in S2 subunits of spike glycoprotein in SARS-CoV-2 (PDB Id: 6VSB) represents in purple color and other disorder residues in green color. IDR, intrinsically disordered region; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2

Moreover, analyzing the PID separately in S2 domains of SARS-CoV and SARS-CoV-2, we have found that PID of S2 is significantly ($p = .025$) higher in SARS-CoV-2 (13.83) than SARS-CoV (8.75). Earlier it was reported that the enormous flexibility of intrinsically disordered regions in the membrane proteins imposes the potentiality to involve in the membrane remodeling process.^{36,37} The membrane remodeling is essential for efficient fusion of the enveloped virus with host cellular membrane. Subsequently, it was also described that IDRs mediated remodeling of the membrane depends on the presence of MoRFs and posttranslational protein modifications.³⁸ The MoRF is a short peptide (10–70 residues) present in the disordered region and the flexibility of this facilitates membrane curvature. MoRF prediction by MoRFchibi has revealed the three MoRFs (804–823, 1147–1159, and 1249–1272) in the fusion peptide of SARS-CoV-2, whereas in SARS-CoV it was two (1129–1142 and 1232–1255; Figure 1B). In an earlier study,³⁸ it was elucidated that the membrane curvature increases with two factors: the size of the inserted MoRF and surface density of the disordered protein. Thus, the acquisition of one additional MoRF in SARS-CoV-2 escalated MoRF and disorder residue density on the viral protein which could be able to trigger more rapid fusion with host membrane than the SARS-CoV. Together these results have deduced that preponderance of intrinsically disordered residues in S2 domain offers protease-sensitive region for prompt activation of fusion peptide and enrichment of MoRFs for efficient fusion with host membrane. So, it could be treated as a novel feature observed exclusively in 2019-nCoV distinguishing this virus from SARS-CoV.

4 | DISCUSSION

The protein expressed on the surface of a pathogen is supposed to be more accessible to surveillance by the immune system than one within the interior of a pathogen.³⁹ Thus, more genetic variations in surface proteins are the signatures of host-pathogen coevolution. In this study, we have found that amongst the four structural proteins, an extensive higher rate of nonsynonymous substitution is occurred in spike glycoproteins of SARS-CoV-2 when compared with the human infecting SARS-CoV strain. Along with the amino acid substitution having neutral effects on virus fitness, S proteins also experienced five deleterious mutations that may cause destabilization of viral structure. The neutral theory of molecular evolution suggested that the mutations decreasing the carrier's fitness tend to disappear from populations through the process of negative or purifying selection ($dN/dS < 1$).⁴⁰ Thus, S protein has also experienced higher synonymous substitution rate to balance overall selection pressure on it. Now, it was also depicted that slightly deleterious and slightly advantageous mutations are engulfed by neutral mutations. Thus, the ratio of dN and dS is frequently used to study positive Darwinian selection operating at highly variable genetic loci, but it could not able to detect adaptively important codons offering benefits to the organism for adaptation.⁴¹ Thus, we have extensively studied amino acid changes in all the structural proteins of

SARS-CoV-2 occurred during evolution from SARS-CoV to search out the mutation posing advantages to the novel virus for their systematic infection in the human body.

We have revealed that mutations in the four structural proteins of SARS-CoV-2 prompt a significant hydrophobic to polar and charged amino acids exchange in S proteins compared with E, M, and N proteins (Table 2). This trend of amino acid exchange in S proteins is observed to generate an intrinsically disordered region (38 residues) at the upstream of fusion peptide in S2 domain which is embedded inside the envelope of SARS-CoV-2. However, amino acid substitution in M, E, and N proteins did not show any enrichment of new IDR in SARS-CoV-2 compared with SARS-CoV. Though, it was earlier reported that N proteins of SARS-CoV extensively enriched with intrinsically disordered residues.⁴² We found that the propensity of disordered residues in N proteins of SARS-CoV-2 (49.6%) is nearly similar with SARS-CoV (50.7%). The enrichment of disordered residues in N proteins has suggested as a crucial phenomenon for their transmission in respiratory routes.⁴² Whereas, lower content of disordered residues in shell proteins (E and M) of SARS-CoV as well as SARS-CoV-2 eliminate the chances of transmission via oral–fecal routes.⁴² Since, the intrinsic disorder content in E, M, and N proteins already reported to regulate the behavior of viral transmission, it is a prerequisite to illustrate the impact of IDR in S proteins. Viral entry is mediated by S proteins containing RBD and fusion domain (S2). The IDR which is exclusively found in SARS-CoV-2 is located in S2 domain. The IDR in this domain offers the furin cleavage site (682–685). Furin protease is ubiquitously expressed in a wide range of organs and tissues, including the brain, lung, gastrointestinal tract, liver, pancreas, and reproductive tissues. The structural flexibility of IDR stimulates this region sensitive to proteolysis.^{35,43} Similar kind of observation was also reported in Zika virus where the prefusion protein prM contains IDR with a protease cleavage site.⁴⁴ Thus, the acquisition of a disordered region imposes efficient proteolytic cleavage of S2 which in turn activates fusion peptide to fuse with the host membrane. Moreover, in comparison with the structured proteins, disordered proteins of similar length have large volumes and flexibility, so that they are able to furnish different coupled binding-folding reactions.⁴⁵ The highly flexible nature of IDR is frequently exploited by eukaryotic cells including viruses for modulating membrane properties during membrane trafficking.³⁷ Viral envelopes are made up of lipid bilayer where a number of spike proteins with considerable disordered regions are observed to be anchored. They are free to diffuse in the lipid leaflets. According to the hypothesis described in Fakhree et al.,³⁹ we could explicate our observation in a way that the free movement of intrinsically disordered residues containing in the fusion peptide will result in collisions with other membrane-anchored macromolecules. The collisions generate a lateral pressure on the membrane. The presence of a large fraction of charged and polar amino acids in disordered proteins makes them more efficient in generating lateral pressure. This pressure is consequently used to incite membrane curvature. It has been seen that many IDRs induce membrane curvature by recruiting MoRFs.⁴⁶ MoRFs are relatively short (10–70 residues) and typically possessing higher numbers of hydrophilic amino acids and prolines.^{47,48} Thus they could play a vital role in protein–protein interactions, metal binding, and

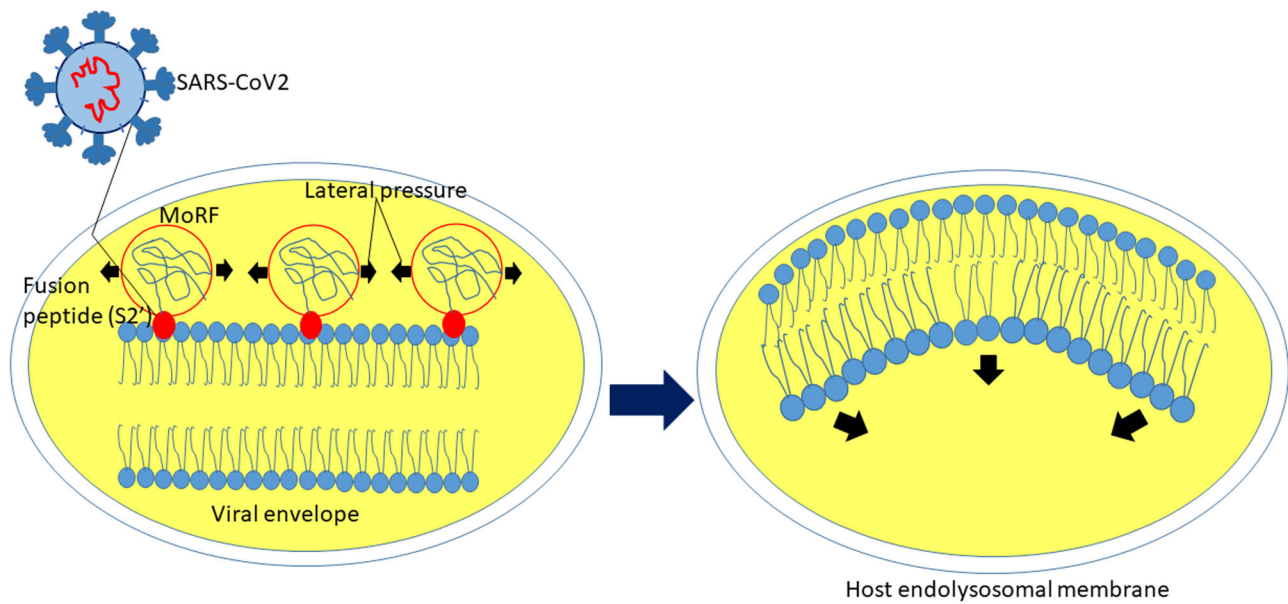


FIGURE 3 Mechanism of viral lipid bilayer curvature by MoRF which helps prompt fusion of SARS-CoV-2 with host endolysosomal membrane. MoRF, molecular recognition feature; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2

cellular communications.⁴⁹ Several roles of MoRFs are also documented in Chikungunya virus.⁵⁰ We have noticed that an abundance of disordered residues in SARS-CoV-2 generates three MoRFs in the fusion peptide. It was explained in an earlier review³⁸ that the presence of MoRF result in a unilateral increase in the surface area of the membrane. This changes the ratio between the outer and inner surface area of the membrane and subsequently to adopt topography with this new ratio, the membrane remodels by increasing its curvature.^{38,51} Thus, it could be interpreted that MoRFs help for curving the lipid bilayer of the virus and initiate efficient fusion with the host cell membrane (Figure 3). Such ability of intrinsically disordered domains to create steric pressure on membrane surfaces to drive its curvature was depicted in endocytic adapter proteins, Epsin1 and AP180 by a combination of in vitro biophysical studies and quantitative experiments in live cells.⁵² Another couple of examples of membrane curvature by the MoRF motif of was established in an IDP ArfGAP1 (ADP-ribosylation factor GTPase-activating protein 1)⁵³ and α -synuclein, by site-directed mutagenesis, limited proteolysis, circular dichroism experiments,⁵⁴ and FRET microscopy study in live cell,⁵⁵ respectively. Hence, we have hypothesized that the acquisition of disordered residues in SARS-CoV-2 makes them highly competent for systematic infection in human. Nowadays, IDPs are becoming attractive candidates for therapeutic intervention by small drug-like molecules. Thus, this study will help in epidemic surveillance and designing drug targets to battle COVID-19.

In summary, these analyses provide insights into the mutational effects in originating intrinsically disordered residues in the S2 subunits of spike glycoprotein present in SARS-CoV-2. We have also hypothesized a unique fusion mechanism of the viral envelope and host membrane by MoRF. However, these propositions are mainly based on our sequence studies and experimental evidence in other

organisms, thus further experimental validations are required to confirm this mechanism in coronavirus.

ACKNOWLEDGMENTS

We are thankful to Mr. Sanjib Kumar Gupta, Senior technical assistant, Bioinformatics Center, Bose institute for his kind help. We are also thankful to two anonymous referees for their valuable suggestions.

CONFLICT OF INTERESTS

All the authors declare that there are no conflict of interests.

AUTHOR CONTRIBUTIONS

Soumita Podder designed, executed experiments, and wrote the manuscript. Avishek Ghosh performed some parts of the experiment. Tapash Ghosh helped in manuscript preparation.

DATA AVAILABILITY STATEMENT

All data will be available in the Supplementary Table.

ORCID

Soumita Podder  <http://orcid.org/0000-0002-8140-508X>

REFERENCES

- Chan JFW, To KKW, Tse H, Jin DY, Yuen KY. Interspecies transmission and emergence of novel viruses: Lessons from bats and birds. *Trends Microbiol.* 2013;21(10):544–555. <https://doi.org/10.1016/j.tim.2013.05.005>
- Lu G, Wang Q, Gao GF. Bat-to-human: spike features determining 'host jump' of coronaviruses SARS-CoV, MERS-CoV, and beyond. *Trends Microbiol.* 2015;23(8):468–478. <https://doi.org/10.1016/j.tim.2015.06.003>

3. World Health Organization WHO. Summary of probable SARS cases with onset of illness from 1 November 2002 to 31 July 2003. http://www.who.int/csr/sars/country/table2004_04_21/en/index.html Accessed August 9, 2020.
4. World Health Organization. WHO MERS-CoV Global Summary and Assessment of Risk, August 2018 (WHO/MERS/RA/August18) 2018. http://www.who.int/csr/disease/coronavirus_infections/riskassessment-august-2018.pdf Accessed June 6, 2019.
5. Lu R, Zhao X, Li J, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet*. 2020;395(10224):565–574. [https://doi.org/10.1016/S0140-6736\(20\)30251-8](https://doi.org/10.1016/S0140-6736(20)30251-8)
6. World Health Organization (2020) Coronavirus disease (COVID-19) outbreak. Situation report – 117, 16th May 2020. <https://www.who.int>. Accessed May 17, 2020.
7. Masters PS. The molecular biology of coronaviruses. *Adv Virus Res*. 2006;66:193–292. [https://doi.org/10.1016/S0065-3527\(06\)66005-3](https://doi.org/10.1016/S0065-3527(06)66005-3)
8. Tortorici MA, Vesler D. Structural insights into coronavirus entry. *Adv Virus Res*. 2019;105:93–116. <https://doi.org/10.1016/bs.aivir.2019.08.0028>
9. Belouzard S, Chu VC, Whittaker GR. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. *Proc Natl Acad Sci USA*. 2009;106:5871–5876. <https://doi.org/10.1073/pnas.0809524106>
10. Gui M, Song W, Zhou H, et al. Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding. *Cell Res*. 2017;27:119–129. <https://doi.org/10.1038/cr.2016.152>
11. Millet JK, Whittaker GR. Host cell proteases: critical determinants of coronavirus tropism and pathogenesis. *Virus Res*. 2015;202:120–134. <https://doi.org/10.1016/j.virusres.2014.11.021>
12. Bosch BJ, Bartelink W, Rottier PJ. Cathepsin L functionally cleaves the severe acute respiratory syndrome coronavirus class I fusion protein upstream of rather than adjacent to the fusion peptide. *J Virol*. 2008;82:8887–8890. <https://doi.org/10.1128/JVI.00415-08>
13. Bertram S, Glowacka I, Muller MA, et al. Cleavage and activation of the severe acute respiratory syndrome coronavirus spike protein by human airway trypsin-like protease. *J Virol*. 2011;85:13363–13372. <https://doi.org/10.1128/JVI.05300-11>
14. Kirchdoerfer RN, Wang N, Pallesen J, et al. Ward Stabilized coronavirus spikes are resistant to conformational changes induced by receptor recognition or proteolysis. *Sci Rep*. 2018;8:15701. <https://doi.org/10.1038/s41598-018-34171-7>
15. Wan Y, Shang J, Graham R, Baric RS, Li R. Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus. *J Virol*. 2020;94(7):1–9. <https://doi.org/10.1128/JVI.00127-20>
16. Wang Q, Qiu Y, Li JY, Zhou ZH, Liao CH, Ge XY. A unique protease cleavage site predicted in the spike protein of the novel pneumonia coronavirus (2019-nCoV) potentially related to viral transmissibility. *Virologica Sinica*. 2020;35:337–339. <https://doi.org/10.1007/s12250-020-00212-7>
17. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Vesler D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell*. 2020;180:1–12. <https://doi.org/10.1016/j.cell.2020.02.058>
18. Zhao S, Lin Q, Ran J, et al. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: a data-driven analysis in the early phase of the outbreak. *Int J Infect Dis*. 2020;92:214–217. <https://doi.org/10.1016/j.ijid.2020.01.050>
19. Pickett BE, Sadat EL, Zhang Y, et al. ViPR: an open bioinformatics database and analysis resource for virology research. *Nucleic Acids Res*. 2012;40:D593–D598. <https://doi.org/10.1093/nar/gkr859>
20. Yang Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol Biol Evol*. 2007;24(8):1586–1591. <https://doi.org/10.1093/molbev/msm088>
21. Nei M, Gojbori T. Simple methods for estimating the number of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol*. 1986;3:418–426. <https://doi.org/10.1093/oxfordjournals.molbev.a040410>
22. Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*. 2015;31(16):2745–2747. <https://doi.org/10.1093/bioinformatics/btv195>
23. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009;4(7):1073–1081. <https://doi.org/10.1038/nprot.2009.86>
24. Li X, Romero P, Rani M, Dunker AK, Obradovic Z. Predicting protein disorder for N-, C-, and internal regions. *Genome Inform Ser Workshop Genome Inform*. 1999;10:30–40.
25. Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK. Sequence complexity of disordered protein. *Proteins*. 2001;42:38–48. <https://doi.org/10.1002/1097-0134>
26. Charon J, Barra A, Walter J, et al. First experimental assessment of protein intrinsic disorder involvement in an RNA virus natural adaptive. *Process Mol. Biol. and Evol*. 2018;35(1):38–49. <https://doi.org/10.1093/molbev/msx249>
27. Singh A, Kumar A, Yadav R, Uversky VN, Giri R. Deciphering the dark proteome of Chikungunya virus. *Sci Rep*. 2018;8:5822. <https://doi.org/10.1038/s41598-018-23969-0>
28. Redwan EM, AlJaddawi AA, Uversky VN. Structural disorder in the proteome and interactome of Alkhurma virus (ALKV). *Cell Mol Life Sci*. 2019;76:577–608. <https://doi.org/10.1007/s00018-018-2968-8>
29. Malhis N, Jacobson M, Gsponer J. MoRFchibi SYSTEM: software tools for the identification of MoRFs in protein sequences. *Nucleic Acids Res*. 2016;44:W488–W493. <https://doi.org/10.1093/nar/gkw409>
30. Tang X, Wu C, Li X, et al. On the origin and continuing evolution of SARS-CoV-2 [published online ahead of print March 03, 2020]. *Microbiology*. 2020. <https://doi.org/10.1093/nsr/nwaa036>
31. Podder S, Ghosh TC. Exploring the differences in evolutionary rates between monogenic and polygenic disease genes in human. *Mol Biol Evol*. 2010;27:934–941. <https://doi.org/10.1093/molbev/msp297>
32. Kimura M. Evolutionary rate at the molecular level. *Nature*. 1968;217(5129):624–626. <https://doi.org/10.1038/217624a0>
33. Peck KM, Lauring AS. Complexities of viral mutation rates. *J Virol*. 2018;92(14):e01031-17. <https://doi.org/10.1128/JVI.01031-17>
34. Uversky VN, Gillespie JR, Fink AL. Why are “natively unfolded” proteins unstructured under physiologic conditions? *Proteins: Struct Funct Genet*. 2000;41:415–427. <https://doi.org/10.1002/1097-0134>
35. Saha D, Podder S, Ghosh TC. Overlapping regions in HIV-1 genome act as potential sites for host–virus interaction. *Front Microbiol*. 2016;7:1735.
36. Uversky VN. Intrinsically disordered proteins and their “mysterious” meta physics. *Front Phys*. 2019;7(10). <https://doi.org/10.3389/fphy.2019.00010>
37. Snead WT, Hayden CC, Gadok AK, Zhao C, Lafer EM, Rangamani P. Membrane fission by protein crowding. *Proc Natl Acad Sci USA*. 2017;114:E3258–E3267. <https://doi.org/10.1073/pnas.1616199114>
38. Snead D, Eliezer D. Intrinsically disordered proteins in synaptic vesicle trafficking and release. *J Biol Chem*. 2019;294(10):3325–3342. <https://doi.org/10.1074/jbc.REV118.006493>
39. Fakhree MAA, Blum C, Claessens MMAE. Shaping membranes with disordered proteins. *Arch Biochem Biophys* 2019;677:108163. <https://doi.org/10.1016/j.abb.2019.108163>
40. Davies MN, Flower DR. Harnessing bioinformatics to discover new vaccines. *Drug Discov Today*. 2007;12:389–395. <https://doi.org/10.1016/j.drudis.2007.03.010>
41. Duret L. Neutral theory: the null hypothesis of molecular evolution. *Nature Education*. 2008;1(1):218.
42. Nei M. Selectionism and neutralism in molecular evolution. *Mol Biol Evol*. 2005;22(12):2318–2342. <https://doi.org/10.1093/molbev/msi242>
43. Goh GKM, Dunker AK, Uversky VN. Understanding viral transmission behavior via protein intrinsic disorder prediction: coronaviruses. *J Pathol*. 2012;1–13. <https://doi.org/10.1155/2012/738590>

44. Johnson DE, Xue B, Sickmeier MD, Meng J, Cortese MS, Oldfield CJ. High-throughput characterization of intrinsic disorder in proteins from the protein structure initiative. *J Struct Biol.* 2012;180:201–215. <https://doi.org/10.1016/j.jsb.2012.05.013>
45. Giri R, Kumar D, Sharma N, Uversky VN. Intrinsically disordered side of the zika virus proteome. *Front Cell Infect Microbiol.* 2016;6(144):1–12. <https://doi.org/10.3389/fcimb.2016.00144>
46. Miller M. The importance of being flexible: the case of basic region leucine zipper transcriptional regulators. *Curr Protein Pept Sci.* 2009;10(3):244–269. <https://doi.org/10.2174/138920309788452164>
47. Cheng Y, Oldfield CJ, Meng J, Romero P, Uversky VN, Dunker AK. Mining α -helix-forming molecular recognition features with cross species sequence alignments. *Biochemistry.* 2007;46(47):13468–13477. <https://doi.org/10.1021/bi7012273>
48. Vacic V, Oldfield CJ, Mohan A, et al. Characterization of molecular recognition features, MoRFs, and their binding partners. *J Proteome Res.* 2007;6(6):2351–2366. <https://doi.org/10.1021/pr0701411>
49. Katuwawala P, Zhenling Y, Lukasz JK. Computational prediction of MoRFs, short disorder-to-order transitioning protein binding regions. *Comput Struct Biotechnol J.* 2019;17:454–462. <https://doi.org/10.1016/j.csbj.2019.03.013>
50. Kotta-Loizou TGN, Hamodrakas SJ. Analysis of molecular recognition features (MoRFs) in membrane proteins. *Biochim Biophys Acta.* 2013;1834(4):798–807. <https://doi.org/10.1016/j.bbapap.2013.01.006>
51. Singh AK, Uversky VN, Giri R. Understanding the interactability of chikungunya virus proteins via molecular recognition feature analysis. *RSC Adv.* 2018;8:27293–27303. <https://doi.org/10.1039/C8RA04760J>
52. Braun AR, Sevcsik E, Chin P, Rhoades E, Tristram-Nagle S, Sachs JN. Alphasynuclein induces both positive mean curvature and negative Gaussian curvature in membranes. *J Am Chem Soc.* 2012;134:2613–2620. <https://doi.org/10.1021/ja208316h>
53. Bigay J, Casella JF, Drin G, Mesmin B, Antony B. ArfGAP1 responds to membrane curvature through the folding of a lipid packing sensor motif. *EMBO J.* 2005;24:2244–2253. <https://doi.org/10.1038/sj.emboj.7600714>
54. Busch DJ, Houser JR, Hayden CC, Sherman MB, Lafer EM, Stachowiak JC. Intrinsically disordered proteins drive membrane curvature. *Nat Commun.* 2015;6:7875. <https://doi.org/10.1038/ncomms8875>
55. Fakhree MAA, Zijlstra N, Raiss CC, et al. The number of α -synuclein proteins per vesicle gives insights into its physiological function. *Sci Rep.* 2016;6:30658. <https://doi.org/10.1038/srep30658>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Podder S, Ghosh A, Ghosh T.

Mutations in membrane-fusion subunit of spike glycoprotein play crucial role in the recent outbreak of COVID-19. *J Med Virol.* 2021;93:2790–2798. <https://doi.org/10.1002/jmv.26598>