Medical Research Archives

European Society of Medicine

OPEN ACCESS

Published: November 30, 2023

Citation: Cobar, O. & Cobar, S., 2023. SARS-CoV-2 Orf1ab Genome Mutations, the Driving Force for Virus Pathogenicity. Medical Research Archives, [online] 11(11). https://doi.org/10.18103/mra.v11i11.4750

DOI: https://doi.org/10.18103/mra.v11i11.4750

ISSN: 2375-1924

RESEARCH ARTICLE

SARS-CoV-2 Orf1ab Genome Mutations, the Driving Force for Virus Pathogenicity.

Cóbar, Oscar[1,2*] & Cóbar, Stella[2]

[1]Pharmacogenomics and Nutrigenomics Research Group, School of Chemical Sciences and Pharmacy, University of San Carlos, Guatemala.
[2]School of Chemical Sciences and Pharmacy, University of San Carlos, Guatemala.

[*]ocobar@gmail.com

ABSTRACT

Background: The proteins codified in the Open Reading Frame 1ab -Orf1ab- region of the SARS-CoV-2 genome are the main responsible of the virus transcription, replication, and translation processes inside the human cell.
Once inside the cell, the viral mRNA encode structural and nonstructural proteins, that direct virus assembly, transcription, replication and host control, and the accessory proteins whose function has not been determined.
The largest gene, Orf1ab, contains overlapping open reading frames that encode polyproteins PP1ab and PP1a.
These polyproteins are cleaved to yield 16 nonstructural proteins, NSP1-16.
Production of the longer (PP1ab) or shorter protein (PP1a) depends on a ribosomal frameshifting event.
The proteins, include the papain-like proteinase (NSP3), 3C-like proteinase (NSP5), RNA-dependent RNA polymerase (NSP12, RdRp), helicase (NSP13, HEL), endoRNAse (NSP15), 2'-O-Ribose-Methyltransferase (NSP16) and other nonstructural proteins.
The SARS-CoV-2 nonstructural proteins are responsible for viral transcription, replication, proteolytic processing, suppression of host immune responses, and suppression of host gene expression.
The purpose of the manuscript is to present a systematic review as of September 30, 2023, on the Orf1ab region mutations of the SARS-CoV-2 genome as of September 30, 2023, with the aim to predict, through the mutations profile on that region, the severity of an infection for a new SARS-CoV-2 variant that could emerge in the near future.
Material and Methods: Original scientific articles published in Medline, Pubmed, Science Direct, Web of Science, Scopus, EBSCO and BioMed Central databases, official health organizations (WHO, CDC, ECDEC, NIH) electronic publications, and specialized media in the subject, were electronically searched to accomplish the aim of the study.
Articles published in any language were included from 2020 to present using a variety of keywords in combination.
The studies relevant to our review were analysed and compared.
Results and discussion: The NIH "National Human Genome Research Institute" define the Open Reading Frames (ORFs) as a portion of a DNA sequence that does not include a stop codon.
The Open Reading Frames encode accessory proteins transcribed from the 3' one-third of the genome to form a set of subgenomic mRNAs (sg mRNAs).
Following entry on the human cell, the viral particle release the genomic RNA molecule that is translated on two large open reading frames, ORF1a and ORF1b.
The resulting polyproteins pp1a and pp1ab are co-translationally and post-translationally processed into the individual non-structural proteins (NSPs) that form the viral replication and transcription complex (RTC).
Translated structural proteins translocate into endoplasmic reticulum (ER) membranes and transit through the ER-to-Golgi intermediate compartment (ERGIC), where interaction with N-encapsidated, newly produced genomic RNA results in budding into the lumen of secretory vesicular compartments.
Two viral proteases, Plpro -NSP3- and 3Cl-pro -NSP5-, process the polyproteins and generate the nonstructural proteins NSP1-NSP16 that directs the transcription, replication and construction of new virions that are secreted by exocytosis from the infected cell.
The evolution of key proteins in viral transcription and replication is clearly observed by carefully studying the structure, function, and evolution of RdRp, Mpro or 3Clpro, and NSP13 proteins directed by the Orf1a and Orf1ab genome mutations.
Conclusions: ORF1ab is cleaved into 16 non-structural proteins involved in SARS-CoV-2 transcription and genome replication.
P323L, P227L, G671S, V776L and A185S are the first five frequent mutations of RdRp (NSP12), the mutations P227L and G671S might have functional consequences in the viral transcription and replication.
Mutations in residues D499 to L514, K545, R555, T611 to M626, G678 to T710, S759 to D761 are directly implicated with the transcription-replication capability of the virus by RdRp.
In Mpro (NSP5) the mutation of residues H41, P132, C145, S145, L226, T234, R298, S301, F305, and Q306 may increase the efficiency of proteolytic cleavage of proteins such as NEMO, thereby improving the ability of the omicron series of viruses to suppress the immune system and accelerate the viral replication.
In Helicase (NSP13) the mutations of residues E261, K218, K288, S289, H290, D374, E375, Q404, K460, R567, and A598 are involved in the separation of the double-stranded RNA or DNA with a 5'→3' polarity as well as 5' mRNA capping activity in the virus transcription-replication process.
In the Orf1ab gene, ORF1b:V2354F mutation, corresponding to NSP15:V303F, may induce a conformational change and result in a disruption to a flanking beta-sheet structure.
The premature stop codon ORF7a:Q94*, truncates the transmembrane protein and cytosolic tail used to mediate protein transport, may affect protein localization to the ER-Golgi.
The analyses of Orf1ab genome mutations, allows us to predict, through the mutations profile on that region, the severity of an infection for a new SARS-CoV-2 variant that could emerge in the near future.
Keywords: SARS-CoV-2 pathogenicity, Open Reading Frames, Orf1ab, Non-Structural Proteins, RNA-dependent RNA polymerase (RdRp), Main Protease (MPro), Helicase.

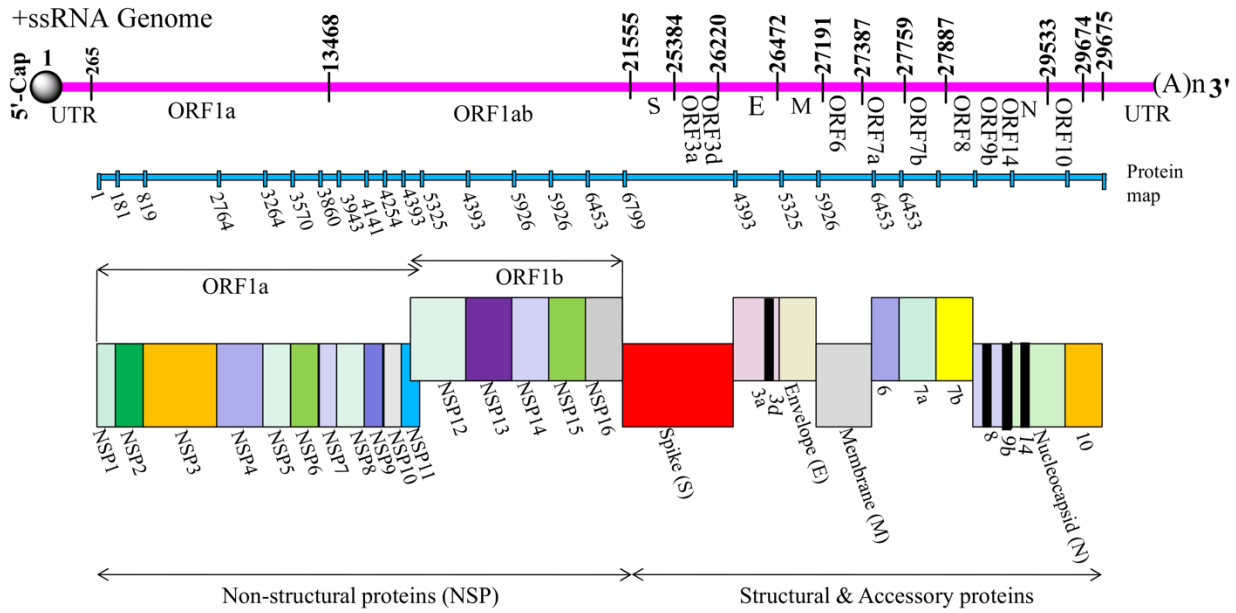
The Genome and SARS-CoV-2 Life Cycle.

SARS-CoV-2 possess remarkably large RNA genomes flanked by 5′ and 3′ untranslated regions that contain *cis*-acting secondary RNA structures essential for RNA synthesis.

At the 5′end, the genomic RNA features two large open reading frames (ORFs; ORF1a and ORF1b) that occupy the two-thirds of the virus genome.

ORF1a and ORF1b encode 15-16 non-structural proteins (NSP), of which 15 constitute the viral replication and transcription complex (RTC) (Figure 1).

Figure 1. Genomic arrangement of SARS-CoV-2.



The genomic organization of the SARS-CoV-2 shows sequential arrangement of various non-structural, structural and accessory genes as follow: 5′-cap-leader-UTR-replicase-S (Spike)–E (Envelope)-M (Membrane)-N (Nucleocapsid)-3UTR-poly (A) tail with accessory genes such as 3a, 3d, 6, 7a,7b, 8, 9b, 14, and 10 interspersed among the structural genes preceding 3′end of the viral RNA genome[1].

The NIH "National Human Genome Research Institute" define the Open Reading Frames (ORFs) as spans of DNA sequence between start and stop codons[1].

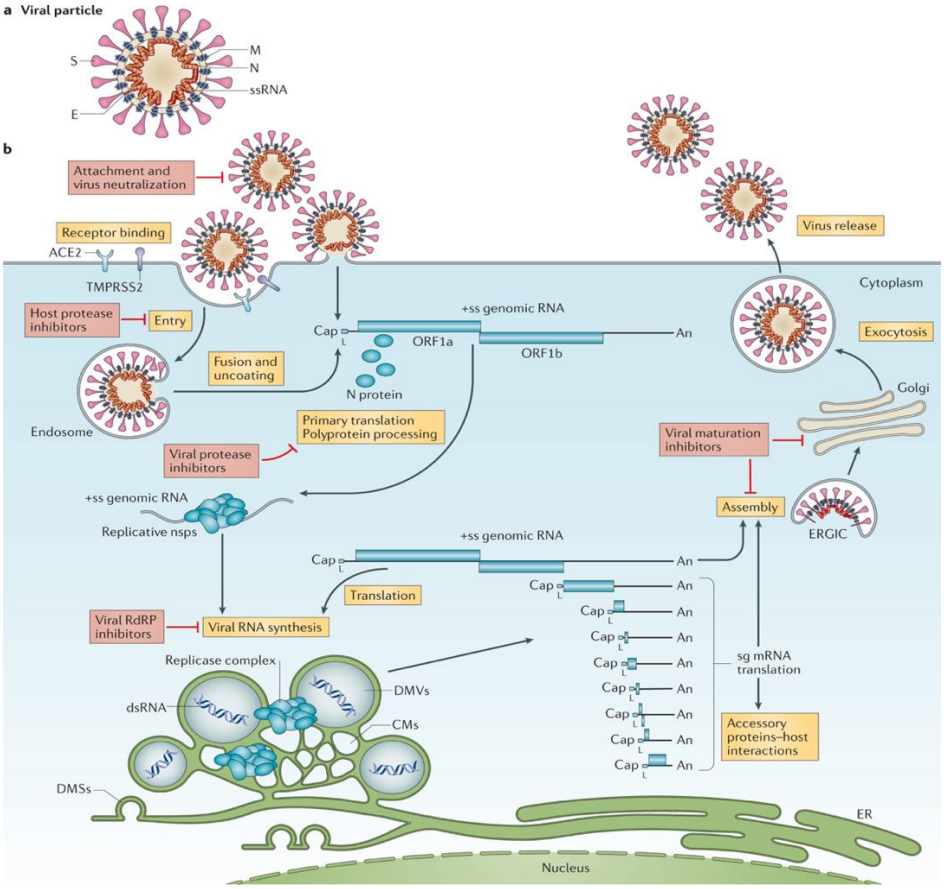
Is a portion of a DNA sequence that does not include a stop codon (which functions as a stop signal).

The Open Reading Frames encode accessory proteins that are transcribed from the 3′ one-third of the genome to form a set of subgenomic mRNAs (sg mRNAs).

Coronavirus accessory proteins are highly variable sets of virus-specific proteins that essentially contribute to modulating host responses to infection and are determinants of viral pathogenicity.

The initial steps of coronavirus infection involve the specific binding of the coronavirus S-protein to the human cellular entry receptor, the Angiotensin-Converting Enzyme 2 -ACE2- (Figure 2).

Figure 2. The coronavirus virion and life cycle.



a Coronavirus virion consists of structural proteins, namely spike (S), envelope (E), membrane (M), nucleocapsid (N) and, for some betacoronaviruses, haemagglutinin-esterase (not shown). The positive-sense, single-stranded RNA genome (+ssRNA) is encapsidated by N, whereas M and E ensure its incorporation in the viral particle during the assembly process. S trimers protrude from the host-derived viral envelope and provide specificity for cellular entry receptors.b Coronavirus particles bind to cellular attachment factors and specific S interactions with the cellular receptors (such as angiotensin-converting enzyme 2 (ACE2)), together with host factors (such as the cell surface serine protease TMPRSS2), promote viral uptake and fusion at the cellular or endosomal membrane[2].

During the intracellular life cycle the virus express and replicate their genomic RNA to produce full-length copies that are incorporated into newly produced viral particles.

Following entry, the incoming genomic RNA is translated of two large open reading frames, ORF1a and ORF1b.

The resulting polyproteins pp1a and pp1ab are co-translationally and post-translationally processed into the individual non-structural proteins (NSPs) that form the viral replication and transcription complex (RTC).

Concordant with the expression of NSPs, the biogenesis of viral replication organelles consisting of characteristic perinuclear double-membrane vesicles (DMVs), convoluted membranes (CMs) and small open double-membrane spherules (DMSs).

These organelles create a protective microenvironment for viral genomic RNA replication and transcription of subgenomic

mRNAs (sg mRNAs) comprising the characteristic nested set of coronavirus mRNAs.

Translated structural proteins translocate into endoplasmic reticulum (ER) membranes and transit through the ER-to-Golgi intermediate compartment (ERGIC), where interaction with N-encapsidated, produced genomic RNA results in budding into the lumen of secretory vesicular compartments.

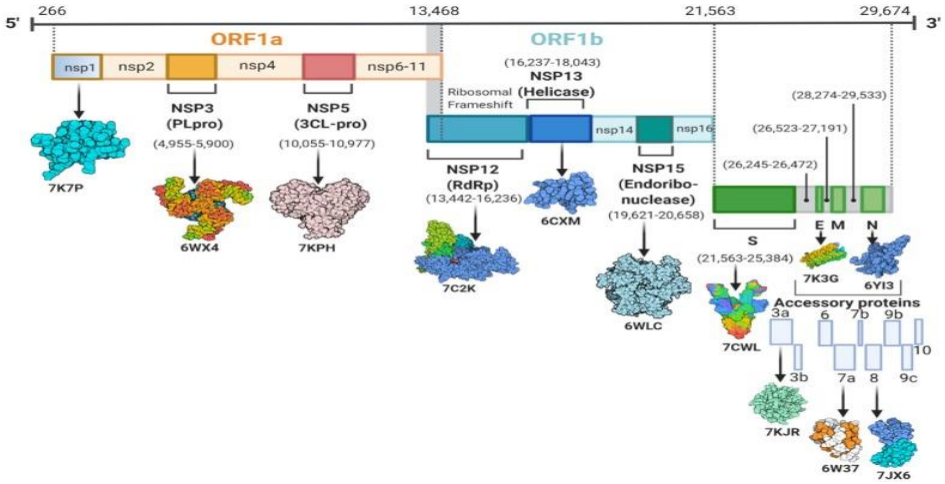
Finally, virions are secreted from the infected cell by exocytosis.

The Structure of "Non-structural Polyproteins" -NSP-

Genomic representation of SARS-CoV-2 consisting of open reading frames that encode structural, non-structural, and accessory proteins.

The genome organization and corresponding structural and non-structural proteins is depicted in the Figure 3[3,4].

Figure 3. Genomic representation of SARS-CoV-2 consisting of open reading frames that encode structural, non-structural, and accessory proteins.



The genome contains four structural proteins such as spike (S) protein, Envelope (E) protein, Membrane (M) protein, and Nucleocapsid (N) protein, along with 14 ORFs that encode 27 proteins[4].

SARS-CoV-2 has fourteen open reading frames (ORFs) in its genome, separated into two portions.

Cellular ribosomes directly translate ORF1a and ORF1ab into two polyproteins (pp1a and pp1ab), found in the first two-thirds of the viral genome from the 5'end.

Table 1 shows the function and properties or SARS-CoV-2 Structural and No-Structural Proteins and key Open Reading Frames.

Table 1. Functions and properties of SARS-CoV-2 structural (S, E, M, N) and non-structural proteins (Nsp1–16) as well as ORFs 3a & b, 6a, 7a & b, 8a & b, and 9b.

Protein	Function	Protein	Function
Nsp1	Likely induces cell arrest in the G_0/G_1 phase and interferes with type I IFNs.	Nsp15	Endoribonuclease activity; Immune evasion; degrades viral polyuridine sequences to prevent host antiviral detection.
Nsp2	Potential nonessential role in pathogenesis.	Nsp16	2′O-methyl-transferase activity
Nsp3	Membrane rearrangements for replication organelle formation. Viral proteolytic activity. Membrane anchoring of other viral proteins to perinuclear membranes; potential role in genome packaging. IFN antagonism.	S	Spike protein for binding/fusion/entry. Role in ER stress and syncytium formation.
Nsp4	Membrane rearrangements for replication organelle formation.	E	Envelope protein essential for virion formation and exit. Viroporin and membrane rearranging activity. Role in NLRP3 inflammasome activation.
Nsp5	Viral proteolytic activity.	M	Membrane/Matrix protein essential for virion formation. Binds to S, E and lipids to form the virial envelope-protein capsule.
Nsp6	Essential for membrane rearrangements for SARS-CoV. May induce autophagy.	N	Nucleocapsid protein, binds to viral RNAs. Necessary for packaging genome and protection from host RNAases.
Nsp7	Cofactor for RdRp complex.	(ORF)3a	Viroporin activity similar to E. Interacts with S, E, and M. Activates NLRP3 inflammasomes.
Nsp8	RNA binding, RNA polymerase activity and essential RdRp complex cofactor protein.	(ORF)3b	Putative function in upregulating cytokine secretion.
Nsp9	Novel ssRNA binding protein. May participate in RNA processing.	(ORF)6a	Colocalizes with nsp3, nsp8 and RO-associated membranes. Upstream and downstream Type I IFN antagonist.
Nsp10	Replicative cofactor to nsp14.	(ORF)7a	Interacts with structural proteins M, E and S. May form a complex with 3a. Possibly essential for viral replication.
Nsp11	Unknown.	(ORF)7b	Possibly essential for viral replication.
Nsp12	RNA-dependent RNA polymerase (RdRp).	(ORF)8a	Viroporin activity similar to 3a and E and activates NLRP3 inflammasome.
Nsp13	Cofactor for the RdRp complex. Viral helicase. Unwinding duplex RNA.	(ORF)8b	Contributor to lysosomal stress, autophagy and inflammation; activation of NLRP3 inflammasome.
Nsp14	S-adenosyl methionine-dependent (N7-guanine)-methyl transferase, assembling cap1 structure at 5′ end of viral mRNA to promote translation and avoid antiviral detection. Proofreading of viral RNA transcripts.	(ORF)9b	Suppresses innate immunity by usurping poly-C-binding protein 2 and HECT domain E3 ligase AIP4 to degrade MAVS/TRAF3/TRAF6 signalosome.

Source: Wong, N. & Saier, M. Jr. The SARS-Coronavirus Infection Cycle: A Survey of Viral Membrane Proteins, Their Functional Interactions and Pathogenesis. Int. J. Mol. Sci. 2021, 22, 1308. https://doi.org/10.3390/ijms22031308

Two viral proteases, papain-like protease (Plpro -NSP3-) and main protease (Mpro or 3Cl-pro -NSP5-), then process the polyproteins and generate sixteen nonstructural proteins, NSP1-NSP16[5,6].

Furthermore, eight accessory proteins such as ORF3a, ORF3b, p6, ORF7a, ORF7b, ORF8b, ORF9b, and ORF14 are located at the 3′ end.

More than 1,500 structures of SARS-CoV-2 proteins have been submitted to the Protein Data Bank (PDB) database[7] since January 2020.

In the case of SARS-CoV-2 structural proteins, ~747 spike proteins, an envelope protein, and ~25 nucleocapsid protein structures have been deposited into the PDB database.

Among the non-structural proteins of SARS-CoV-2, 443 structures have been resolved for NSP5, followed by NSP3, NSP13, NSP15, NSP7 & NSP8, NSP10, NSP12, NSP16, NSP9, and NSP14, respectively.

In addition, several structures of SARS-CoV-2 accessory proteins were deposited into the PDB database, whereas only four structures for ORF8 and two structures for each of the ORF3a, ORF7a, and ORF9b proteins.

Brief description of various structural, non-structural, and accessory proteins of SARS-CoV-2[7].

NSP7

Length 83 amino acids, forms a complex with NSP8 for viral replication and participate as a cofactor for NSP12.

The catalytic residues are S4, D5, K7, C8, H36, L40, N37, V33 S15, L14, V11, A30, W29, E23; Replicase Domain (1-83) R21, K43 and D44 (Wilamowski *et al.* 2021)

NSP8

Length 198 amino acids, makes heterodimer with NSP7 and NSP12.

The catalytic residues are P183, Y149, V131, M129, P133, A125, K127, V130, P121, L122, A110, L128, N118, I119, T123, K79, L117, I106, N109, P116, V115, M94, D112, C114, D99, L95, N104, L91, V83, L98, F92, A162, T84, R80, Q88, M90, I185, M87, A86; Shaft Domain (6–104), Head Domain (105-196) and Drug Binding Sites A102, A150, R190, A194 (Wilamowski *et al.* 2020).

NSP9

Length 198 amino acids, RNA-binding proteinresidues which may participate in viral replication N33, G100, M101, V102, L103, G104, S105; Single Domain Protein (1–109), Motif GxxxG (100–104), Drug Binding Sites M12, S13, N33, T35, F40, L42, L94, N98 (Khan *et al.* 2021b, Littler *et al.* 2021a).

NSP10

Length 139 amino acids, forms heterodimer complex with NSP14 and NSP16, acting as a cofactor for both and stimulates ExoN (viral exoribonuclease) and 2-O-methyltransferase activity, the catalytic residues are N3, V4, T5, F8, K9, D10, P20, T21, Q22, P24, T25, H26, L27, L38, C39, D41, F60, K61, M62, N63, Y64, V66, Y69, T127, N129, N130, T131, K196, K200, I201; Single Domain Protein (1–139); and Drug Binding Sites V21, D22, A26, G35, Q36, P37, I38, GLY52, Q65, R78, P107, V108 (Halder, 2021, Lin *et al.* 2021).

NSP12 (RdRp)

Length 932 amino acids, function Replication and Methylation; the catalytic residues are Y420, F415, F441, F440, F442, N552, A443, P412, G413, D445, Q444, T409, N447, R392, D390, L391, L389, N403, V405, L388, T402, L387, N386, A379, P323, L270, F396, F326, V398, L271, L514, P328, M666, V330, Y273, T324, T344, L329, P339, M380, R331, V338, K332, Y374, F340, A383, D336, S384, V341, S518, F407, L371, F368, D523, W509, S759, D760, D761; the Domains and Motifs are NiRAN (1–250), C-terminal RdRp (398–932), Motif G (499–511), Motif F (544–560), Motif A (612–626), Motif B (678–710), Motif C (753–767), Motif D (771–796), Motif E (810–820); and Drug Binding Sites M542, K545, S549, K551, R553, R555, V557, D618, C622, ASP623, S682, S759, D760, D761, R836 (Ahmed *et al.* 2020, Khan *et al.* 2021b, Zhang *et al.* 2020c)

NSP13 (Helicase)

Length 596 amino acids, a helicase core domain participates in binding interaction

with ATP. Zn-binding domain is involving in replication and transcription; the catalytic residues are R178, H230, N361, S468, T532, D534; the Domains and Motifs are ZBD (1–100), SD (101–150), 1B domain (151–261), 1A domain (262-442), 2A domain (443-601); and Drug Binding Sites V45, Y70, F90, P283, G285, T286, G287, K288, H290, R443, E540 (Chen *et al.* 2020, Malone *et al.* 2021, Yan *et al.* 2021).

NSP14 (HexoN)

Length 527 amino acids, acting on both ssRNA and dsRNA in a 3′ to 5′ direction and a N7-guanine methyltransferase activity; the catalytic residues are D90, E92, E191, H268, D273; the Domains and Motifs are Flanking region (1–50), ExoN (51–287), N7-MTase (288–527), DEDD motif; and Drug Binding Sites W385, N386, Y420, F426, F506 (Devkota *et al.* 2021, Tahir, 2021).

The Structure of RNA-Dependent RNA-Polymerase -RdRp-[8].

The ORF1a/b encodes 16 non-structural proteins (NSP1-NSP16) that participate in the transcription and replication machinery, among these, RNA-dependent RNA polymerase (RdRp, NSP12) is the core enzyme involved in this process[9].

Mutation in SARS-CoV-2 RNA polymerase machinery might considerably impact the virus's virulence and mortality rate caused by the virus[10].

NSP12 acts as the main component of replication machinery of SARS-CoV-2 and comprises RNA dependent RNA polymerase (RdRp), interface, and nidovirus RdRp-associated nucleotidyltransferase -NiRAN-[11].

The biochemical investigation demonstrated that NSP12 is linked to NSP7/8 by subunit to perform efficient RNA synthesis.

Moreover, NSP14 associates with NSP12/7/8 complex by binding to a subunit involved in the final mRNA cap synthesis and proofreading.

Additionally, NSP12 directly interacts with helicase, which results in enhancing the helicase activity[12].

The limited proofreading activity of SARS-CoV-2 RdRp introduces an actual mutation rate (approximately 4 x 104) in the process of virus replication[13].

The location of NSP12 AASs (Amino acid Analog-Sensitive) mutations in the protein structure and their frequency were investigated between January 2020 and June 2021.

The Table 2 describes the first five frequent mutations regardless of geographical distribution.

Table 2. First Five Frequent Mutations of NSP12 regardless of geographical distribution.

RANK	*RESIDUE*	*FREQUENCY*	*TOTAL FREQUENCY*
Top 1	P(323)L	0.98366	1,731,040
Top 2	P(227)L	0.061411	108,071
Top 3	G(671)S	0.028901	50,860
Top 4	V(776)L	0.018056	31,775
Top 5	A(185)S	0.017245	30,348

Interestingly, not all of these mutations are present among the continents as the top five mutations.

Among these mutations, the P323 mutation was present in all continents (North America (0.9862 frequency), South America (0.9896 frequency), Europe (0.9877 frequency), Asia (0.9491 frequency), Oceania (0.9098 frequency), and Africa (0.9394 frequency) as the mutation with the highest incidence rate.

P227 mutation has been observed as one of the top mutations in North America (0.1292 Frequency), Europe (0.0327 Frequency), South America (0.0193 Frequency), and Africa (0.0331 Frequency).

G671 mutation has been remarked in North America (0.008 Frequency), Asia (0.0454 Frequency), and Europe (0.0396 Frequency) among the top five mutations.

Finally, V776 mutation in North America (0.0085 Frequency), Europe (0.0254 Frequency), and Africa (0.0228 Frequency); A185 mutation in Europe (0.0262 Frequency) and America (0.0258 Frequency) are observed among the top five mutations.

The Table 3 shows the incidence of the global NSP12 top-five mutations.

The Table is based on the continents.

Table 3. Incidence of the global NSP12 top-five mutations on the world continents.

Residue	North America	South America	Europe	Asia	Oceania	Africa
P(323)L	0.9862	0.9896	0.9877	0.9491	0.9098	0.9394
P(227)L	0.1292	0.0193	0.0327	0.1280	0.0055	0.0331
G(671)S	0.0080	0.0009	0.0396	0.0454	0.0091	0.0048
V(776)L	0.0085	0.0038	0.0254	0.0013	0.0026	0.0228
A(185)S	0.0042	0.0046	0.0262	0.0018	0.0022	0.0258

Altogether, the data strongly suggest that SARS-CoV-2 is acquiring mutations as it is spreading to new locations.

Most likely, these mutations are helping SARS-CoV-2 to adapt better inside hosts and in new geographical areas.

The mutations P227L and G671S might have functional consequences that need to be addressed in future studies.

The structure of complex between SARS-CoV-2 RdRp and RNA, and the protein-RNA interactions are displayed in the Figures 5 and 6 below[14].

The RdRp of SARS-CoV-2 is composed of a catalytic subunit known as NSP12 as well as two accessory subunits, NSP8 and NSP7.

The structure of this RdRp has recently been reported; it is highly similar to the RdRp of SARS-CoV, a zoonotic coronavirus that spread into the human population in 2002.

The NSP12 subunit contains an N-terminal nidovirus RdRp-associated nucleotidyl transferase (NiRAN) domain, an interface domain and a C-terminal RdRp domain.

The RdRp domain resembles a right hand, comprising the fingers, palm and thumb subdomains that are found in all single-subunit polymerases.

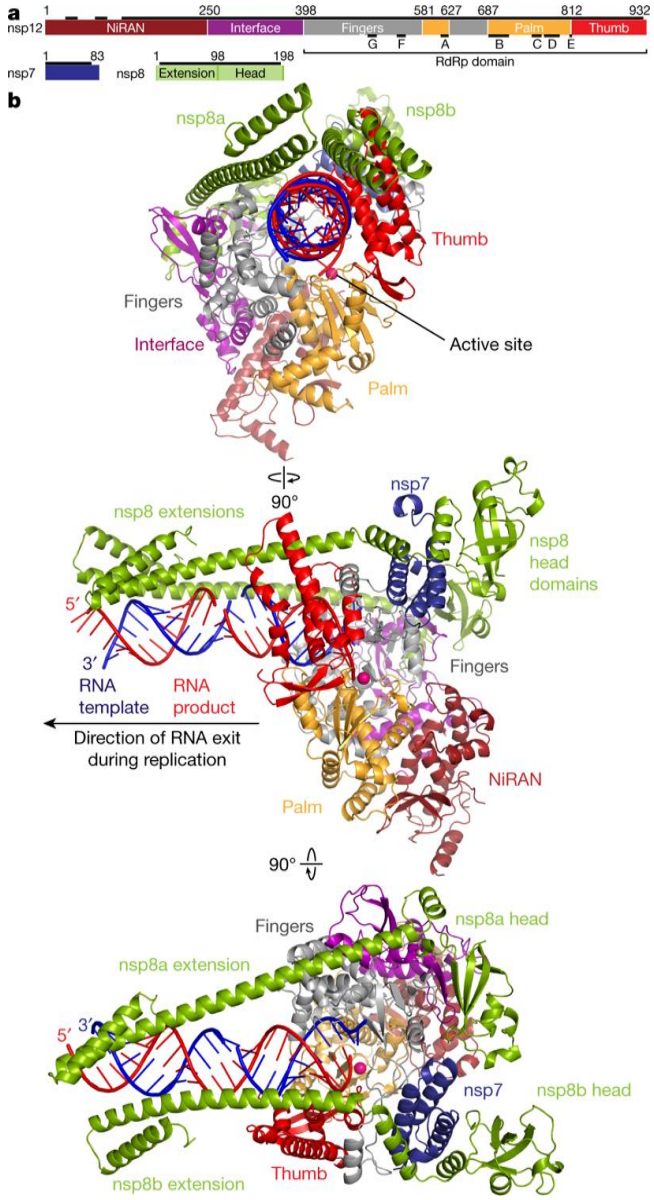
Subunits NSP7 and NSP8 bind to the thumb, and an additional copy of NSP8 binds to the fingers domain.

RdRp in the active site, possess seven conserved motifs (motifs A-G) involved in template binding (motif G), NTP binding (motif F), and polymerization (motifs A-E)[15].

The NSP12 RNA-dependent RNA polymerase possesses some minimal activity on its own, but the addition of the NSP7 and NSP8 co-factors greatly stimulates polymerase activity[16].

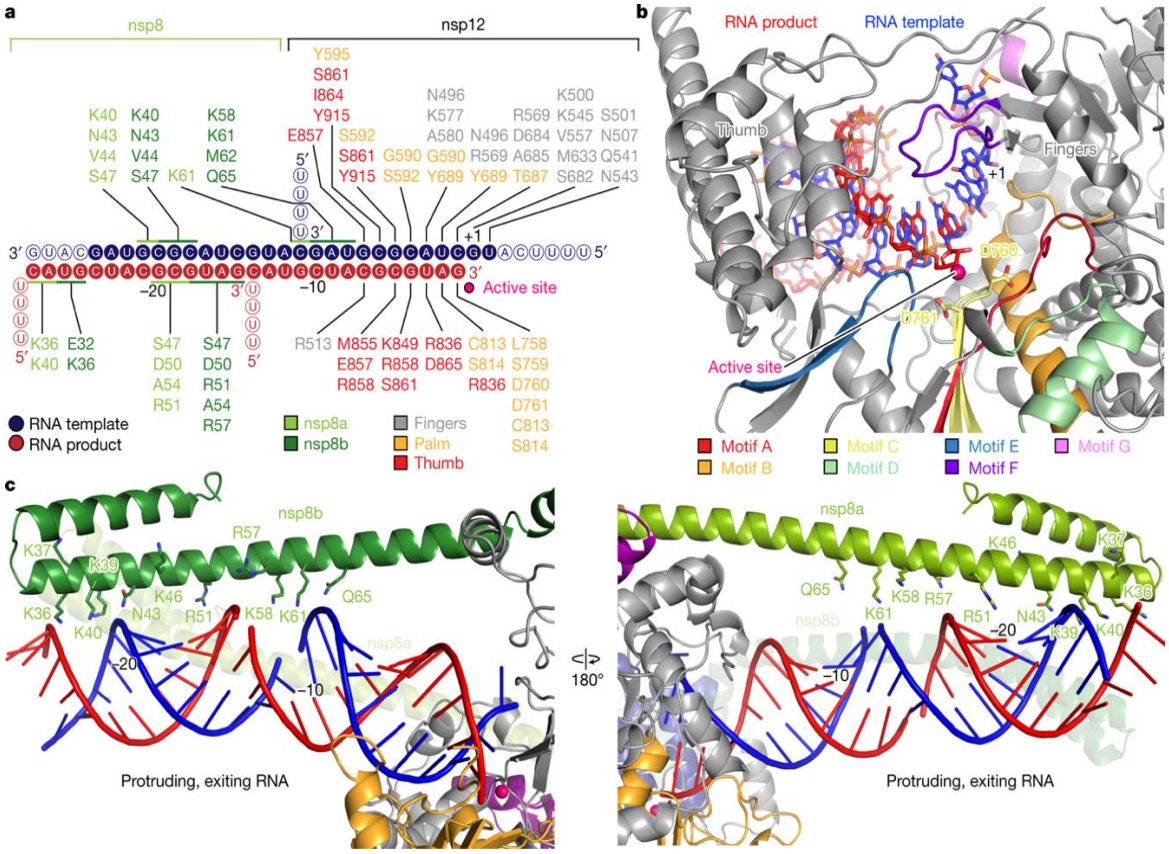
Though additional viral NSP subunits are likely necessary to carry out the full repertoire of replication and transcription activities, the NSP12-NSP7-NSP8 complex so far represents the minimal complex required for nucleotide polymerization.

Figure 5. Structure of complex between SARS-CoV-2 RdRp and RNA.



a, Domain structure of NSP12, NSP8, and NSP7 subunits of RdRp. In NSP12, the conserved sequence motifs A-G are depicted. Regions included in the structure are indicated with black bars.b, Three views of the structure, related by 90° rotations (top, back view; middle, side view; bottom, top view). Colour code for NSP12 (NiRAN, interface, fingers, palm and thumb), NSP8, NSP7, RNA template (blue) and RNA product (red) used throughout. The magenta sphere depicts a modelled metal ion in the active site.

Figure 6. RdRp–RNA interactions.



a, Schematic of protein-RNA interactions. Solid and hollow circles show nucleotides that were included in the structure or invisible, respectively. RNA is assembled from overlapping oligonucleotides. RdRp residues in NSP12 within 4Å of RNA are indicated and contacts are depicted with lines. NSP8 residues and their putative RNA contact regions (horizontal lines, within 5Å) are indicated even if side chain density is absent.b, Interactions of the RdRp active site with the first turn of RNA. Subunit NSP12 is in grey, and conserved motifs A-G are coloured. Active-site residues D760 and D761 shown as sticks. The magenta sphere depicts a modelled metal ion.c, Charged NSP8 residues that may interact with proximal RNA regions. Left, side view; right, inverted side view.

In addition, many other CoV NSPs possess enzyme activities related to RNA modification.

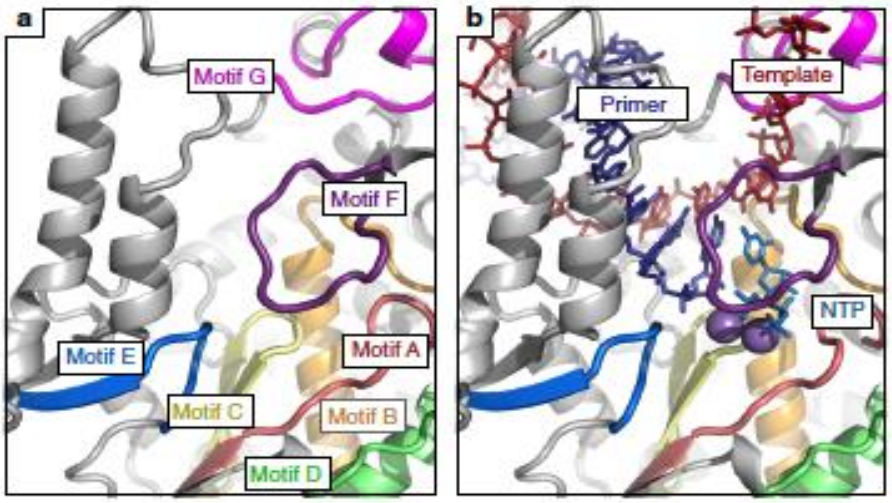
The N-terminal portion of NSP12 contains a β hairpin (residues V31 to K50) and a nidovirus-specific extension domain (NiRAN, residues S115 to A250).

The β-hairpin is sandwiched by the palm subdomain in RdRp core and NiRAN, a configuration not observed in the SARS-CoV RdRp structure.

The C-terminal catalytic domain of NSP12 (residues L366 to F920) connects to NiRAN through an interface subdomain (residues L251 to R365).

The Figure 7 shows the Architecture of the NSP12 polymerase-active site.

Figure 7. Conserved Motif Regions (A–G) involved in template and nucleotide binding and catalysis.



a, RdRp conserved motifs (motifs A–G) involved in template binding (motif G), NTP binding (motif F), and polymerization (motifs A-E). b, Superposition of elongation complexes from poliovirus (3OL6.pdb) [10.2210/ pdb3OL6/pdb] and norovirus (3H5Y.pdb) [10.2210/pdb3H5Y/pdb] polymerases give the approximate positions of RNA template (red), primer (dark blue), incoming NTP (blue), and bound catalytic metal ions (purple).

Recently, an RdRp mutation 14408C>T was identified in Europe, which is associated with an increased mutation rate compared to viral genomes from Asia through an unknown mechanism[17].

The resulting mutation P323L located at the interface domain of RdRp is far from the active catalytic site

The mutation may exert its effects through altered interaction with other components of the replication-transcription complex or with the RNA template[18].

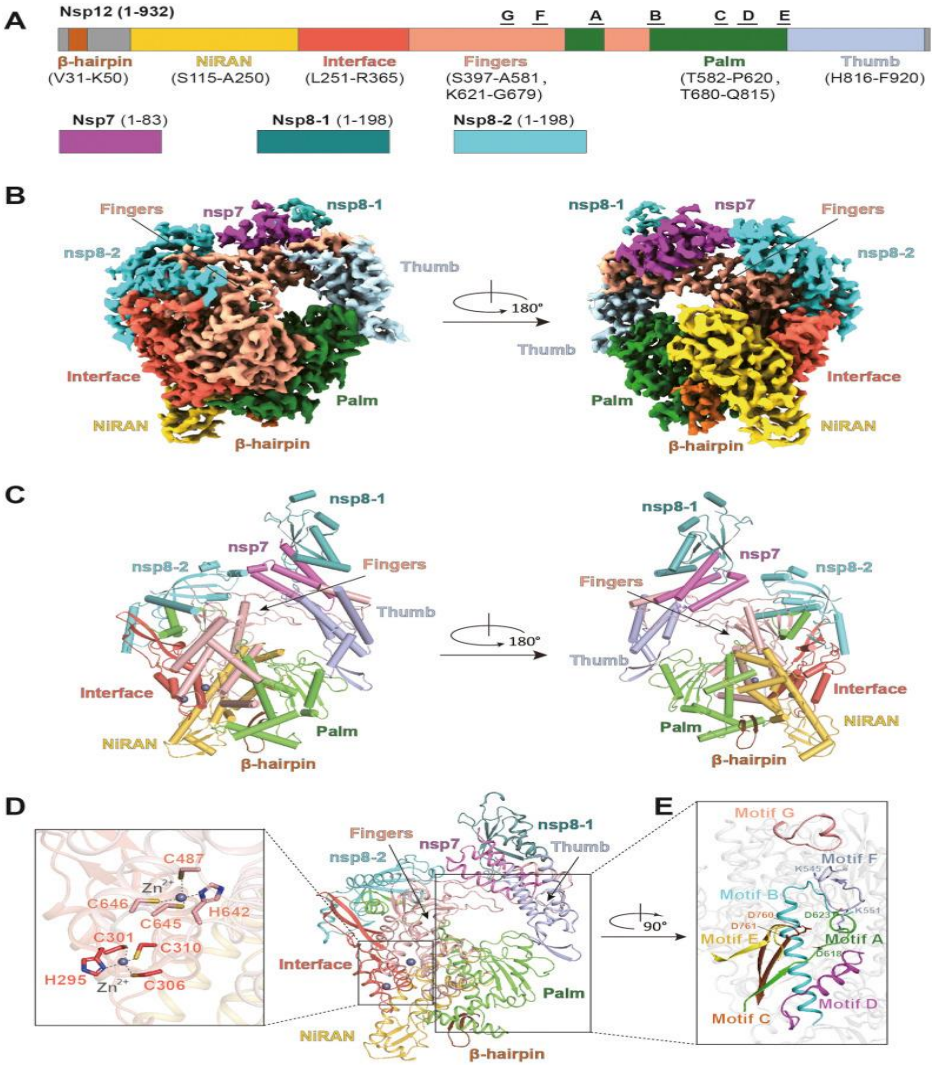
The C-terminal catalytic domain of NSP12 adopts a canonical cupped right-handed configuration of all viral RdRp, composed of the finger, palm, and thumb subdomains.

The finger subdomain (residues S397 to A581 and K621 to G679) forms a closed-ring structure with the thumb subdomain (residues H816 to F920) (see Figure 2A, B and C).

Additionally, two Zinc ions bind to the conserved metal binding motifs constituted by H295-C301-C306-C310 and C487-H642-C645-C646 respectively.

The Figure 8 shows the Structure of the SARS-CoV-2 RdRp complex.

Figure 8: The structural basis for RNA recognition by SARS-CoV-2 RdRp.



a, The schematic diagram for the domain organization of the RdRp complex, containing NSP12, NSP7, and two copies of NSP8 (NSP8-1 and NSP8-2). The polymerase motifs A to G in the catalytic site are highlighted. The β-hairpin is indicated. b-c, Two views of the cryo-EM map (B) and structure (C) of SARS-CoV-2 apo RdRp complex (PDB code: 7BV1). d, The conserved zinc-binding motifs. The zinc-binding residues are shown as sticks. The subdomains and components of the RdRp complex are colored as follows: β-hairpin, chocolate; NiRAN, gold; Interface, tomato; Fingers, salmon; Palm, green; Thumb, light blue; NSP7, magenta; NSP8-1, dark cyan; NSP8-2, cyan. e, The active site of the RdRp complex. The conserved seven motifs A-G are highlighted as indicated colors. The conserved residues, K545 and R555 in motif F, D618 and D623 in motif A, as well as D760 and D761 in motif C, are shown as sticks. (For interpretation of the references to color in this figure legend.)

The active site of the SARS-CoV-2 RdRp is formed by seven conserved catalytic motifs, motifs A to G.

Five of these motifs (A-E) are located within the palm subdomain, while the other two (F and G) reside in the finger subdomain.

Motif A (residues T611 to M626) houses the catalytic motif $DX_{2-4}D$, in which the first aspartic acid D618 is invariant in most viral polymerases, including the hepatitis C virus and poliovirus.

The flexible loop in Motif B (residues G678 to T710) serves as a hinge to undergo conformation arrangement associate with template RNA and substrate binding.

Motif C (residues F753 to N767) contains the catalytic motif SDD (residues S759 to D761), which is essential for binding the metal ion. The recent structures of RdRp also confirm that the D760 and D761 are involved in the coordination of two magnesium ions at the catalytic center[19].

These conserved aspartic acids from catalytic Motif $DX_{2-4}D$ and SDD motifs are involved in the regulation of catalytic activity.

Motif F (residues L544 to V557) interacts with the phosphate group of incoming NTP.

Structurally, the side chains of K545 and R555 contact with the +1 base to direct the incoming NTP to the correct position for catalysis.

Motif G (residues D499 to L514) interacts with the template strand and may direct the RNA template to the active catalytic site (see Figure 2E).

The Structure of Mpro or 3Cl-pro[20].

The genome of coronaviruses, including SARS-CoV-2, encodes for two proteases, a papain like (PL^{pro}) protease and the so-called main protease (M^{pro}), a chymotrypsin-like cysteine protease, also named $3CL^{pro}$ or non-structural protein 5 (NSP5).

M^{pro} is activated by autoproteolysis and is the main protease responsible for cutting the viral polyprotein into functional units.

Aside from this, it is described that M^{pro} proteases are also capable of processing host proteins, including those involved in the host innate immune response.

SARS-CoV-2 Mpro had 391 cleavage events as compared to 130 events for its SARS-CoV equivalent.

SARS-CoV-2 Mpro could also cleave a wider spread of amino acids at the P2 region. Compared with SARS-CoV Mpro, both share highly similar amino acid sequences (96% similarity) and catalytic efficiency.

M^{pro} is responsible for at least ten cleavages along the viral polyprotein.

Works hydrolysing the peptide bonds C-terminal to glutamine residues within the sequence motif (small amino acid)-X-(L/F/M)-Q↓(G/A/S)-X (where X is any amino acid; ↓cleavage site).

The cleavage of the viral polyprotein produces the mature non-structural proteins (NSP), which then form the replication/transcription complex.

M^{pro} it has 640 high-confidence cleavage sites corresponding to 434 unique accession numbers, 418 M^{pro}-derived peptides (from 318 unique proteins), had a positive $\log_2$ change and also contained a glutamine residue at position P1.

Besides, possess 75 N-termini with a negative $\log_2$ fold change, which may represent natural N-termini that are degraded upon the addition of M^{pro}.

Therefore, high-confidence peptides that are found to be degraded upon M^{pro} addition are also of interest, potentially representing protease substrates.

The overlap of the high-confidence N-terminal peptides identified upon addition with the different M^{pro} was very low, in particular for the main proteases of SARS-CoV and SARS-CoV-2, which shows 96% similarity in terms of their amino acid sequence and catalytic efficiency.

SARS-CoV-2 can not only hydrolyse the peptide bond C-terminal to Gln residues but also His residues, albeit at a lower frequency. M^{pro} from SARS-CoV-2 were able to cleave optineurin (OPTN) at two different sites (LQ151.152AE and LQ165.166LK).

OPTN plays a role in the activation of innate immune response during viral infection.

X-ray co-crystal structures of (A) 13b-K (PDB entry: 8A4T) and (B) 13b-H (PDB entry: 8A4Q) in the substrate-binding cleft of the M^{pro}.
The orientation of the view was rotated ~45° in the case of (B) for clarity.

The carbon atoms of the compounds are depicted in yellow and those of the protein in cyan, with the exception of the Ser1 of the other monomer (green).

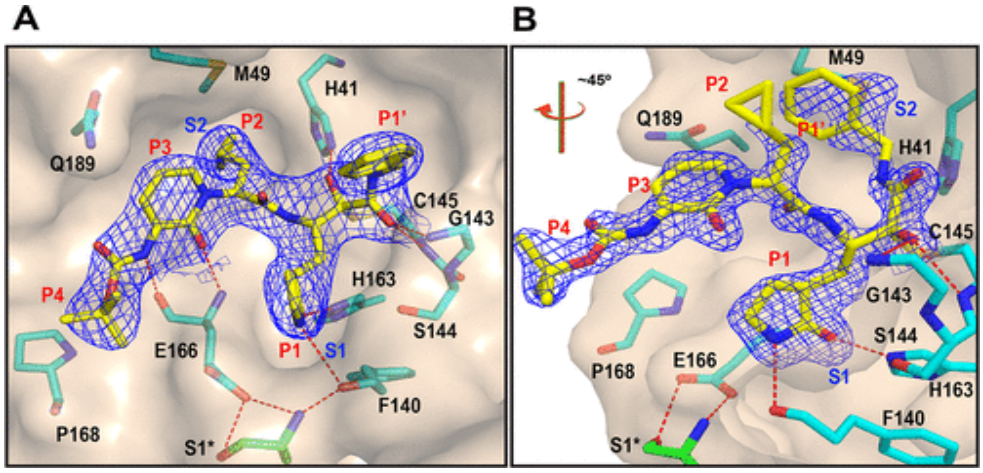
Blue symbols (S1 and S2) indicate the canonical binding pockets for moieties P1-P4 (red symbols) of the inhibitors.
The $2F_o$-F_c map carved around the inhibitor is depicted as a blue mesh (contoured at 1.0).

Polar contacts are represented as red dashed lines.

The antiviral activity of 13b-K against SARS-CoV-2 infection was investigated in different cellular infection models.

Treatment with 13b-K of human bronchial epithelial cells (Calu 3 cells) infected with SARS-CoV-2 revealed a dose-dependent antiviral effect on the release of infectious viral particles, with an EC_{50} of 2.4 ± 0.7 µM (see Figure 9A for details).

Figure 9. X-ray co-crystal structures of (A) 13b-K (PDB entry: 8A4T) and (B) 13b-H (PDB entry: 8A4Q) in the substrate-binding cleft of the M^{pro}.



The reported EC_{50} for the diastereomeric mixture 13b in Calu 3 cells infected with SARS-CoV-2 is 4-5 µM, which is consistent with our findings that only one of the diastereomers is active[21].

13b-K was also tested for its antiviral activity using a fully replication-competent SARS-CoV-2 strain in A549-ACE2-TMPRSS2, Huh7, and VeroE6 cell cultures.

Since VeroE6 cells show a high efflux of some chemotypes, this latter antiviral assay was performed in the presence of the MDR1-inhibitor CP-100356 (0.5 µM).

The dose–response curves in these model systems are shown in Figure 9B, and nonlinear fitting results in EC_{50}values (95% CI) of 1.3 µM (1.16-1.55), 3.4 µM (1.58-5.03), and 0.84 µM (0.64-1.08) in the VeroE6 + CP/SARS2,

Huh7/SARS2, and A549-ACE2-TMPRSS2 systems, respectively, and CC_{50} values >100 μM[22].

Soichi Wakatsuky and coworkers from SLAC National Accelerator Laboratory, Stanford Synchrotron Radiation Lightsource, Structural Molecular Biology, Menlo Park, CA, USA, found that SARS-CoV-2 Mpro binds with itself, forming a dimer with a C-terminal tail when interacting with NEMO.

One of the C terminal tails nestles between beta strands Gly109-Tyr118 and Ser121-Arg131 in the Mpro neighboring dimer[23].

NF-kappa-B essential modulator (NEMO) also known as inhibitor of nuclear factor kappa-B kinase subunit gamma (IKK-γ) is a protein that in humans is encoded by the *IKBKG* gene.

NEMO is a subunit of the IκB kinase complex that activates NF-κB[24].

The Figure 10 shows the comparison of 3CLpro interactions with its N-terminal sequence and the $hNEMO_{226–234}$ peptide.

P132H mutation, present in all Omicron sub-variants, substitutes an uncharged proline, which would constrict the protein structure, with a positive histidine.

The different charge likely results in a major structural change near what appears to be a critical region for SARS-CoV-2 Mpro.

This mutation could affect the ability of ORF9b to suppress the innate immune system through interacting with NEMO and its effects should be explored further[25].

The human gene for IKBKG is located on chromosome Xq28.

Multiple transcript variants encoding different isoforms have been found for this gene.

We speculate that this change may increase the efficiency of proteolytic cleavage of proteins such as NEMO, thereby improving the ability of the omicron series of viruses to suppress the immune system and accelerate their fitness.

In the conclusions of their research, Koudelka *et al.* bring additional specificity and understanding to how Mpro contributes to viral replication.

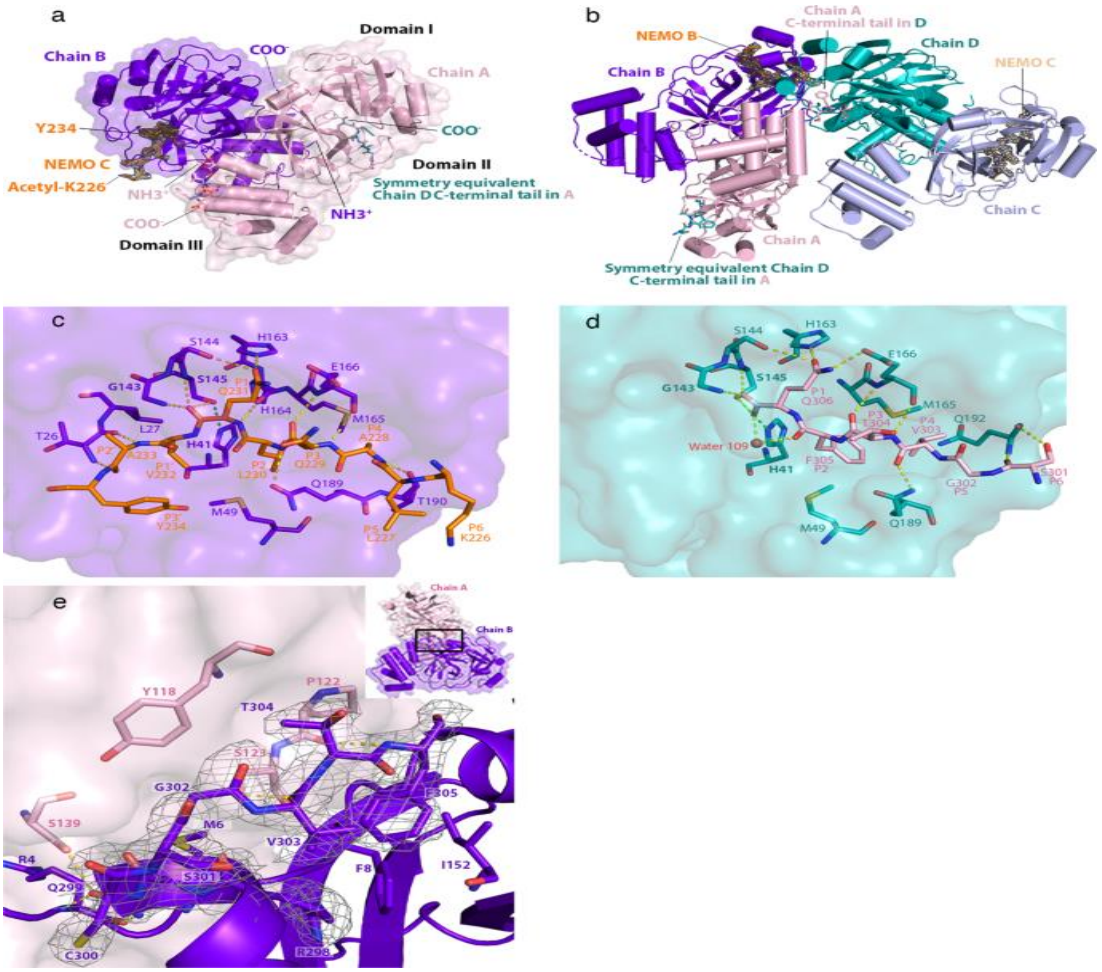
The typical role of Mpro as a cysteine protease is to mature other SARS-CoV-2 proteins and aid processing of translated viral proteins.

However, similarly to N protein, it seems Mpro does double duty by also cleaving cellular proteins essential to innate immune signaling.

Alongside studies on SARS-CoV-2 spike and nucleocapsid, additional investigation could further elucidate the full scope of Mpro's immune suppressive qualities.

As well as identify whether mutations in Mpro significantly contribute to SARS-CoV-2 viral fitness should be considered.

Figure 10. Comparison of 3CLpro Interactions with its N-terminal Sequence and the hNEMO$_{226–234}$ peptide.



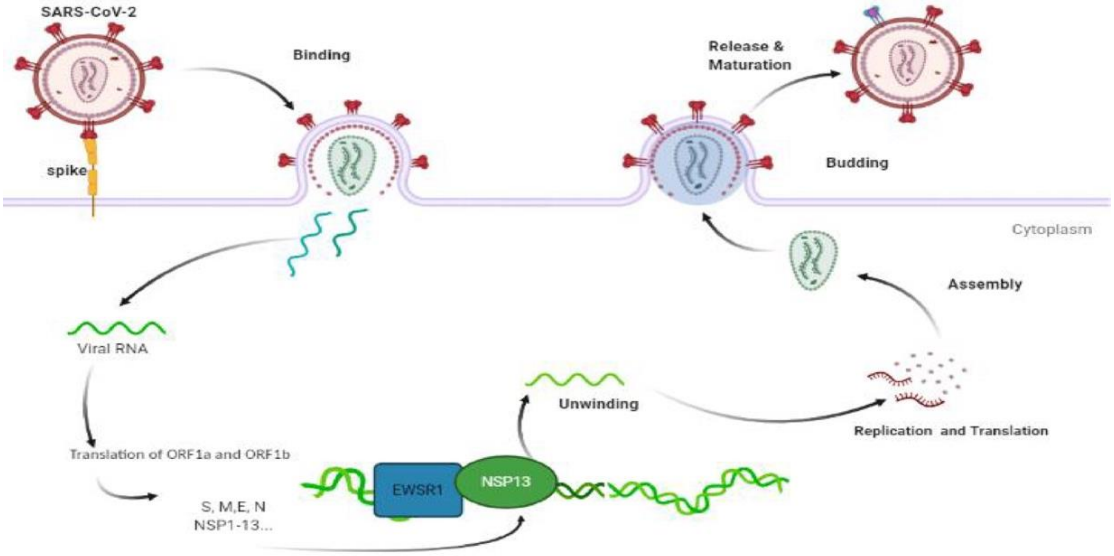
a, The hNEMO$_{226–234}$-bound 3CLpro dimer. N- (NH3+) and C-termini (COO-) are labeled. hNEMO$_{226–234}$ (NEMO B - orange) binds into chain B (purple) and is surrounded by an omit map of electron density (1.0 σ contour level and 1.9Å carving radius). Acetylated Lys226 and Tyr234 at the N- and C-termini of the hNEMO$_{226–234}$ peptide, respectively, are labeled. The C-terminal tail of chain D (teal) from a crystallographic symmetry equivalent binds into the substrate-binding site of chain A (light pink). Domains I, II and III are labeled. b, Structure of the hNEMO-bound C145S variant asymmetric unit. The C-terminal tail of chain A (pink) binds into the substrate-binding site of chain D (teal). hNEMO peptides bound into chain B (purple) (NEMO B - orange) and chain C (light blue) (NEMO C - wheat) are superimposed with an omit map (1.0 σ contour level and 1.9Å carving radius) of their electron density. c, Interactions of hNEMO$_{226–234}$ with the substrate-binding groove of 3CLpro C145S. hNEMO$_{226–234}$ is colored orange. Residues Lys226 to Tyr234 are fully labeled. The surface of chain B (purple) is shown. Substrate-binding residues in chain B are portrayed as sticks and labeled. Catalytically relevant residues are labeled in bold. Hydrogen bonds are depicted as dashed yellow lines. The hydrogen bond predicted to form between Cys145 (in WT) or Ser145 (in C145S) and His41 is depicted as a dashed green line. d, Interactions of the C-terminal tail of chain A with the substrate-binding groove of 3CLpro C145S. The C-terminal tail of chain A is colored light pink. Residues Ser301 to Gln306 are fully labeled. The surface of chain D (teal) is shown. Interacting residues and their interactions are portrayed as in c. The oxygen atom of water 109 is depicted (red sphere). e, The C-terminal tail at the 3CLpro dimer interface. The C-terminal tail of chain B (purple) is depicted as sticks, juxtaposed over its cartoon representation. Density from the omit map (1.0 σ contour level and 1.9Å carving radius) is shown around the C-terminal tail. Arg298 to Phe305 and residues that interact with the C-terminal tail of chain B are labeled. Inset, schematic of chains A and B, showing the position of the C-terminal tail at the 3CLpro dimer interface.

The Structure of the Viral Helicase -NSP13-[26,27]

Helicase -NSP13- it is expressed by the orf1b SARS-CoV-2 genome domain, along with NSP12 (RdRp), NSP14 and NSP16[4].

NSP13 is essential for viral replication SP13 (Figure 11), possesses helicase activity for unwinding double stranded RNA (dsRNA) during the replication process[28].

Figure 11. The working model demonstrating the association of EWSR1 with NSP13 and promotion of SARA-CoV-2 unwinding and replication.



The SARS-CoV-2 NSP13 consisting of 596 amino acids, has a triangular pyramid shape comprising five domains: two "RecA-like" domains (1A and 2A), 1B domain, N-terminal Zinc binding domain (ZBD), and stalk domain, which connects ZBD and 1B domain, 1A, 2A, and 1B domains form the base of the pyramid, whereas ZBD and stalk domains form the apex.

NSP13 is a superfamily 1B (SF1B) helicase that can unwind DNA or RNA substrates with a 5′→3′ directionality[29].

The SARS-CoV-2 NSP13 structure reported similar NTPase active site residues conservation as present in SARS-CoV NSP13, which contains the residues K288, S289, D374, E375, Q404, and R567.

All of these residues were found to be concentrated in the cleft between domain 1A and 2B located at the base Twelve Cys/His conserved residues capable of binding at least three Zn^{2+} ions are present in the Zn^{2+} binding domain.

The helicase domain has six conserved motifs among which two motifs are called the Walker A motif (GXXXXGK(T/S) containing a conserved Lys residue and the Walker B motif (R/K)XXXXGXXXXLhhhhDE) containing a Asp and a Glu residue.

Lys in Walker A and Asp/Glu in Walker B participate in DNA/RNA helicases based ATP-hydrolysis.

The NSP13 helicase separates double-stranded RNA or DNA with a 5′→3′ polarity.

SARS-CoV RdRp, NSP12, enhances the catalytic efficiency of NSP13 twofold by increasing the step size of nucleic acid unwinding.

Besides the helicase activity on double-stranded DNA and RNA, it is also capable of unwinding RNA/DNA duplex.

Moreover, it has NTPase activity as well as 5′ mRNA capping activity.

An analysis of the mutation rate of NSP13 amino acid residues during the 2020 pandemic revealed that five residues, E261, K218, H290, K460, and A598, showed mutation rate greater than 0.01.

NSP13 contains three additional domains: an N-terminal zinc-binding domain (unique to nidoviral helicases), a stalk and a 1B domain.

Previous studies established that the NSP13 helicase forms a stable complex with the RTC, and single-particle cryogenic-electron microscopy (cryo-EM) structures of an $NSP13_2$-RTC (the RTC with two NSP13 protomers bound) have been determined.

In the $NSP13_2$-RTC structure, two protomers of NSP13 sit on top of the RTC with each NSP13-zinc-binding domain interacting with one of the two N-terminal helical extensions of NSP8.

The two NSP13 protomers were originally named NSP13.1 and NSP13.2, but the nomenclature proposed by Hillen, $NSP13_T$ (binds on the 'Thumb' side of the RdRp) and $NSP13_F$ (binds on the 'Fingers' side of the RdRp) is used now.

This overall architecture (see Figure 12 below) places the $NSP13_T$ active site directly in the path of the downstream template-RNA (t-RNA).

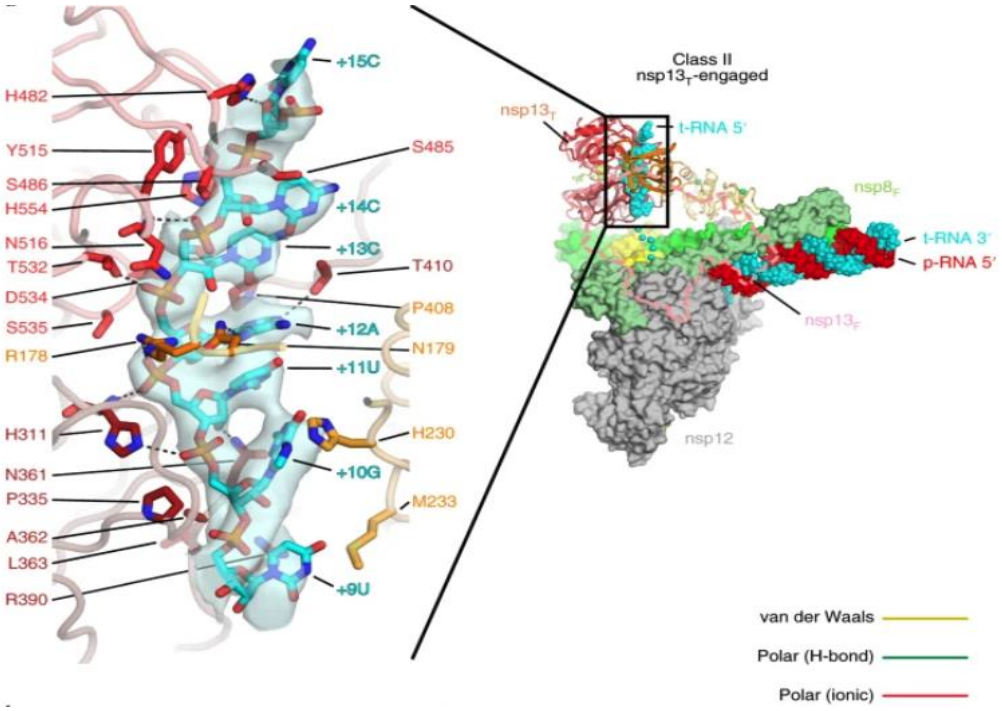
The cryo-EM maps showed that the 5′single-stranded overhang of the t-RNA passed through the nucleic acid binding channel of $NSP13_T$, but the low resolution of the map due to structural heterogeneity precluded detailed modeling.

NSP13 may facilitate processive elongation by the RdRp on the highly structured RNA genome, but may also generate backtracked RTCs for proofreading, template-switching during subgenomic RNA transcription or both.

How NSP13 directs these diverse processes that regulate RdRp function remains less well understood.

For instance, the structures indicate that NSP13 translocates on the t-RNA strand in the 5′ → 3′ direction, while the RdRp would translocate on the same strand in the opposite direction.

Figure 12. In class II (nsp13$_T$-engaged), the nsp13$_T$ RecA domains and 1B domain clamp onto the 5′single-stranded t-RNA.



Right, overall view of the nsp13$_T$-engaged structure. Proteins are shown as molecular surfaces except nsp13$_T$ is shown as a backbone ribbon, and nsp13$_F$ is removed and shown only as a dashed outline. The RNA is shown as atomic spheres. The boxed region is magnified on the left. Left, nsp13$_T$ is shown as a backbone worm but with side chains that interact with the t-RNA shown. Cryo-EM density for the downstream 5′ t-RNA segment is shown (transparent blue surface) with the t-RNA model superimposed. The pattern of purines/pyrimidines in the RNA density was clear and unique, allowing the identification of the sequence register for the nsp13$_T$-bound RNA.

Structural analyses revealed the basis for coupling of the essential NSP13 helicase with the RNA-dependent RNA polymerase (RdRp) where the holo-RdRp and RNA substrate (the replication–transcription complex or RTC) associated with two copies of NSP13 (NSP13$_2$-RTC).

One copy of NSP13 interacts with the template-RNA in an opposing polarity to the RdRp and is envisaged to drive the RdRp backward on the RNA template (backtracking), prompting questions as to how the RdRp can efficiently synthesize RNA in the presence of NSP13.

Here we use cryogenic-electron microscopy and molecular dynamics simulations to analyze the NSP13$_2$-RTC, revealing four distinct conformational states of the helicases.

The results indicate a mechanism for the NSP13$_2$-RTC to turn backtracking on and off, using an allosteric mechanism to switch between RNA synthesis or backtracking in response to stimuli at the RdRp active site.

The NSPs 10, 13, 14, and 16, play significant roles in mRNA capping[30].

The mRNA cap has many significant biological roles in the virus as it is critical for the stability, translation, and evasion of the host immune response to mRNAs.

Uncapped RNA molecules in cytoplasmic granular compartments are degraded as they can trigger innate immune response.

Initially, 5'γphosphate of the nascent RNA chains (pppN-RNA) is hydrolyzed by the NSP13 helicase, an RNA 5'-triphosphatase.

Then, GpppN-RNA is formed after a yet unidentified GTase transfers a GMP molecule to the 5'-diphosphate of the RNA chains (ppN-RNA).

Furthermore, the domain NSP14C-terminal N7-MTase methylates the cap structure at the guanosine's N7 position, generating cap-0 (m7GpppN-RNA), using SAM (S-Adenosyl Methionine) as a methyl donor.

Finally, NSP16 (SAM)-dependent 2'-O-MTase activity promotes the methylation of the ribose 2'-O position to form cap-1 (m7GpppNm-RNA) in the first transcribed nucleotide.

In the final steps, NSP10 serves as an allosteric activator.

The viral helicase, NSP13 is an important enzyme involved in SARS-CoV-2 replication and transcription.

Hypothetical SARS-CoV-2 entry and replication inside human host cell is depicted in the Figure 13 below.

Without the stalk domain, NSP13 loses its dsRNA unwinding ability due to the lack of ATPase activity.

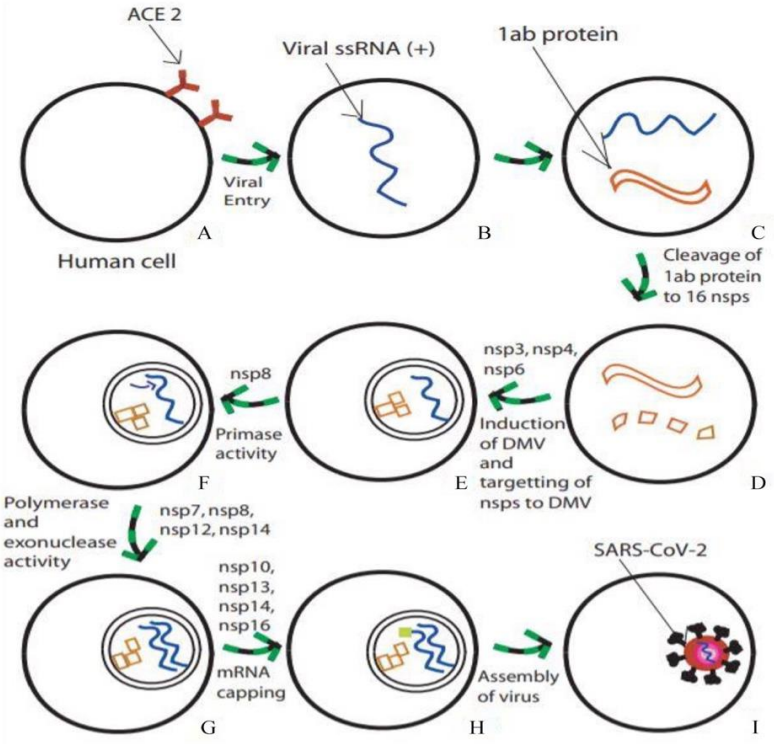
The stalk domain of NSP13 also provides a rigid connection between the ZBD and helicase domain.

We found that the tight connection between the stalk and helicase is necessary for NSP13-mediated dsRNA unwinding.

When a short flexible linker was inserted between the stalk and helicase domains, the helicase activity of NSP13 was impaired, although its ATPase activity remained intact.

The linker insertion between the stalk and helicase domains attenuated the RNA binding ability and affected the thermal stability of NSP13.

Figure 13. Hypothetical SARS-CoV-2 entry and replication inside human host cell.



a, Human host cell with angiotensin-converting enzyme 2 (ACE 2) receptors, which attaches the virus and aids its entry into the cell. b, Positive-sense single-stranded viral RNA within the cell. c, Translation of the RNA into 1ab proteins (pp1a and pp1b). d, Cleavage of pp1a and pp1b into 16 nonstructural proteins by virally encoded chymotrypsin-like protease and two papain-like proteases. e, Induction of double-membraned vesicles (DMV) and localization of cleaved nsps with the help of NSP3, NSP4 and NSP6. f, NSP8 acts as a primase for RNA replication. g, NSP7, NSP8, NSP12, and NSP14 assist polymerase and exonuclease activities. h, mRNA capping is assisted by NSP10, NSP13, NSP14 and NSP16. i, Finally, replicated RNA and other translated viral proteins assembled into a new virus.

SARS-CoV-2 Orf1ab and Orf1a; Key Mutations

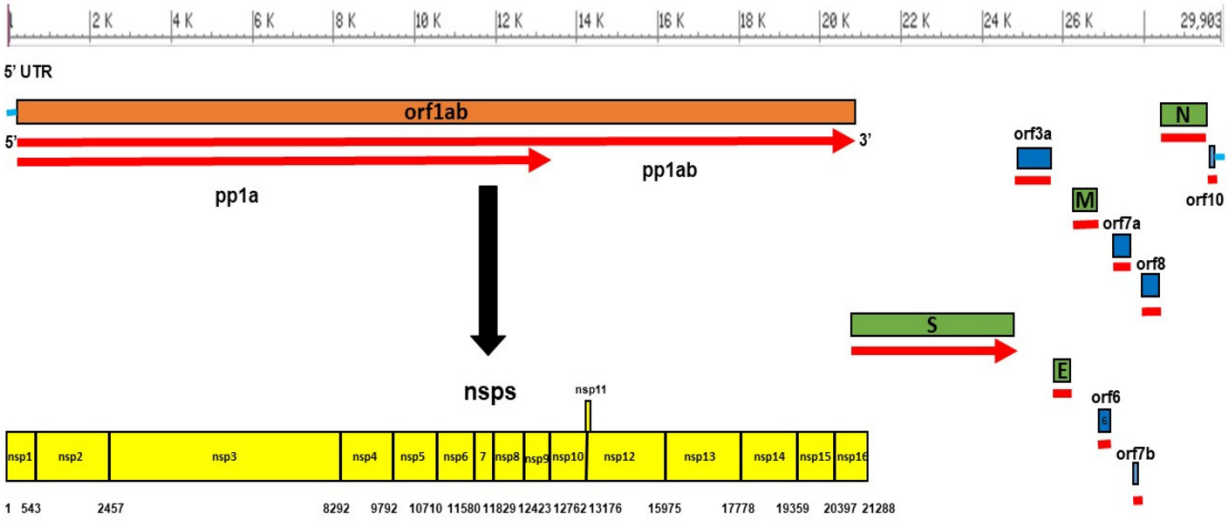
SARS-CoV-2 is an enveloped, positive-sense, single-stranded RNA virus that causes COVID-19.

Virus particles include the RNA genetic material and structural proteins needed for invasion of host cells.

ORF1ab, the largest SARS-CoV-2 gene, contains overlapping open reading frames that encode polyproteins PP1ab and PP1a.

Orf1ab is 21,290 nucleotides and encodes either replicase proteins pp1a of 4,405 amino acids (aa) (NSP1-NSP11) or pp1ab of 7,096 aa (NSP1-NSP16), according to ribosomal frameshift (Figure 14).

Figure 14. Genome Organization of SARS CoV-2 Orf1ab Gene and its Encoded Proteins.



The orf1ab gene constitutes two-thirds of the genome, encodes a total of 16 non-structural proteins (NSPs) within the pp1ab gene, as shown in yellow, which are NSP1 (180 aa), NSP2 (638 aa), NSP3 (1945 aa), NSP4 (500 aa), NSP5 (306 aa), NSP6 (290 aa), NSP7 (83 aa), NSP8 (198 aa), NSP9 (113 aa), NSP10 (139 aa), NSP11 (13 aa), NSP12 (932 aa), NSP13 (601 aa), NSP14 (527 aa), NSP15 (346 aa), and NSP16 (298 aa). The other third of SARS CoV-2 includes four genes (in green) that encode four structural proteins (S, M, E, N), and six accessory genes (in blue) that encode six accessory proteins (orf3a, orf6, orf7a, orf7b, orf8, and orf10)[31].

The polyproteins are cleaved to yield 16 nonstructural proteins, NSP1-16.

Production of the longer (PP1ab) or shorter protein (PP1a) depends on a -1 ribosomal frameshifting event.

The proteins, based on similarity to other coronaviruses, include the papain-like proteinase protein (NSP3), 3C-like proteinase (NSP5), RNA-dependent RNA polymerase (NSP12, RdRp), helicase (NSP13, HEL), endoRNAse (NSP15), 2'-O-Ribose-Methyltransferase (NSP16) and other nonstructural proteins.

SARS-CoV-2 nonstructural proteins are responsible for viral transcription, replication, proteolytic processing, suppression of host immune responses and suppression of host gene expression.

Based on the whole-genome sequence alignment, SARS-CoV-2 shares 89% identity with bat SARS-like CoVZXC21, 82% with SARS-CoV, and 96.3% with bat CoV RaTG13.

Alignment of the predicted protein sequences of SARS-CoV-2 to those of SARS-CoV or SARS-like coronaviruses revealed a total of 380 amino acid substitutions between these viruses.

These amino acid substitutions were distributed as follows: 348 mutations in nonstructural proteins (ORF1ab, 3a, 3b, 7a, 7b, 9b, and ORF14), 27 in S protein, and 5 in N protein.

No amino acid substitutions were detected in E or M proteins, indicating that E and M proteins are highly conserved among these viruses.

The analysis of the receptor-binding domains (RBD) of SARS-CoV and SARS-CoV-2 S protein revealed similar binding affinities.

Aiping Wu and colleagues from Center for Systems Medicine, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences, Beijing, China[32] found a total of 27 amino acid substitutions in the S protein but not in the receptor-binding motif (RBM) that directly interacts with hACE2, which may affect host tropism.

These 27 substituted residues were distributed as follows: 17 in the S1 subunit [6 in the RBD and 6 in the subdomain (SD)] and 10 in the S2 subunit.

The Figure 15 shows the phylogenetic analysis of the Omicron variant by Ultrafast Sample placement of Existing tRee (UShER)[33].

The phylogenetic tree (unrooted tree) was constructed with a total of 2012 SARS-CoV-2 strains involving 12 strains of the Omicron variant (red color) and 2,000 strains from 25 different clades.

SARS-CoV-2 ORF3b may play a role in viral pathogenicity in addition to its inhibitory effects on interferon (IFN) expression and signaling[34].

Recently, a novel short putative protein was identified in ORF3b of SARS-CoV-2[35] however, the function of this novel protein is still not known.

SARS-CoV-2 ORF8 is closer to those of bat SARS CoV ZXC21 and ZC45 and distant from that of human SARS-CoV.

Because of its adaptability, the Orf1ab gene expresses a significant number of mutations in its regions, resulting in alterations in this protein and the structural and nonstructural proteins NSP6, NSP11, and NSP13[36].

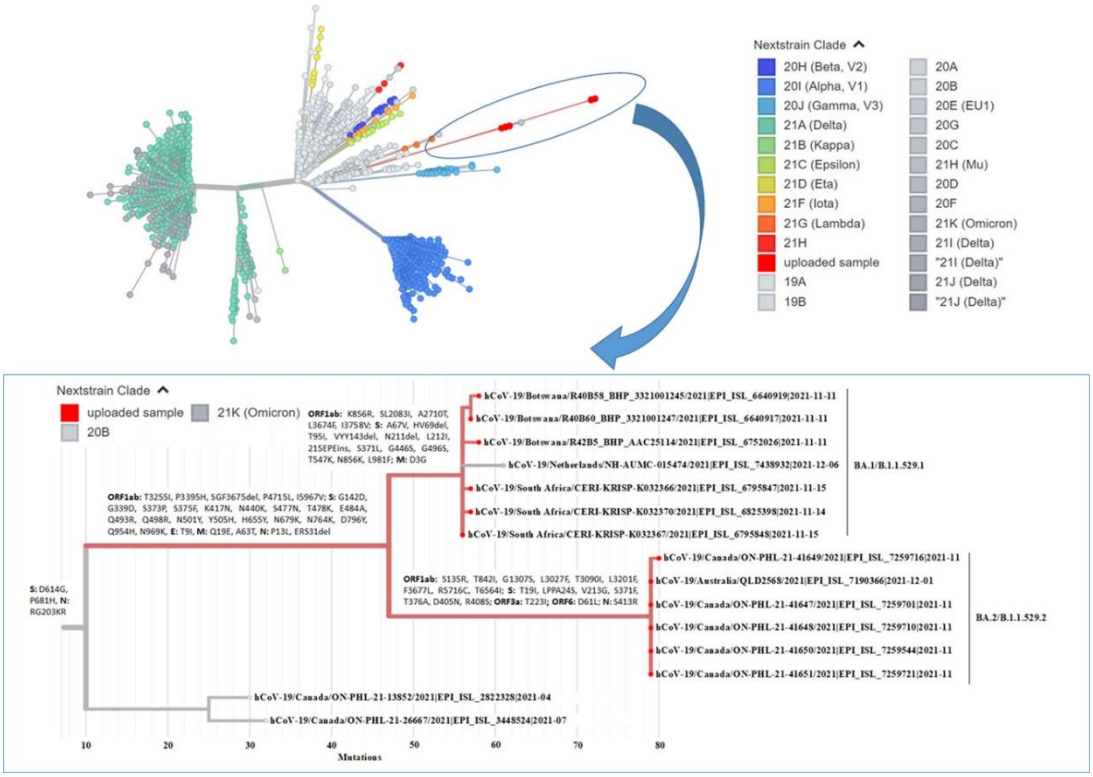
Evolutionary genomic analysis revealed two novel mutations, ORF1b:V2354F and a premature stop codon ORF7a:Q94*, identified in a cluster of SARS-CoV-2 Delta isolates collected from vaccinated individuals in Colorado[37].

The ORF1b:V2354F mutation, corresponding to NSP15:V303F, may induce a conformational change and result in a disruption to a flanking beta-sheet structure.

The premature stop codon ORF7a:Q94*, truncates the transmembrane protein and cytosolic tail used to mediate protein transport.

This may affect protein localization to the ER-Golgi.

Figure 15. Phylogenetic Analysis of the Omicron Variant by Ultrafast Sample Placement of Existing tRee (UShER).



Each color in the tree is representing a different clade/lineage.

Additionally, the vaccinated cluster of isolates contains non-synonymous mutations within ORF8 and NSPs which further distinguish this cluster from the respective ancestral Delta variant.

ORF8 is highly variable, which may be a viral strategy to counter immune pressure from the host[38].

Given the importance of the residues between 117-120 on dimer stabilization, the deletion of residues 119-120 in the Delta lineage may indicate a reduced functional role for ORF8.

Despite numerous and occasionally large deletions, such as a 382-nucleotide deletion in 45 of 191 diagnostic samples reported by Su et al. (2021), there is evidence that the SARS-CoV-2 virus containing ORF8 mutations is able to replicate and cause disease.

During SARS-CoV-2 infection, polyprotein ORF1ab is proteolytically cleaved into sixteen NSPs (NSP1-16) that collectively form the viral replication-transcriptional complex -RTC-[39].

ORF1a comprising NSP1-10 plays an important role in coping with cellular stresses and maintaining the functional integrity of the cellular components along with the pivotal roles in viral replication.

ORF1b encodes viral RNA-dependent RNA polymerase (NSP12), helicase (NSP13), exonuclease (NSP14), a poly-U-specific endonuclease (NSP15), and methyltransferase -NSP16-[40].

NSP4 is essential for viral replication-transcriptional complex -RTC- formation and in coordination with NSP3 and NSP6 anchor the RTC to modified ER membranes.

The large luminal loop formed between the first and second transmembrane regions is a crucial component of NSP4 for ER membrane rearrangement induced by its interaction with NSP3.

The disruption of this region impairs viral replication.

The V167L and T492I NSP4 mutations located in the large luminal loop and C-terminal end, respectively, are characteristic of the Delta variant.

NSP13 comprised a five-domain, triangular pyramid structure and is a highly conserved multifunctional protein possessing both NTPase and RNA helicase activities.

The N-terminal zinc-binding domain contains three zinc-finger motifs and a bridging stalk, responsible for NSP12 interaction.

The 1B domains and C-terminal RecA-like helicase domains, RecA1 and RecA2, provide the NTP and nucleic acid binding activities.

Characteristic of the Delta variant and conserved among all the CO isolates sequenced in this study is a P77L mutation located within the N-terminal zinc-binding domain required for catalytic activity and interaction with NSP12[41].

Additionally, three CO isolates co-harbor a V226L mutation located in the 1B domain involved in substrate binding.

NSP14 is a bifunctional replicase consisting of both an N-terminal exoribonuclease (ExoN) and C-terminal (guanine-N7) methyl transferase (N7-MTase) domains[42].

ExoN is important for proofreading and N7-MTase functions in mRNA capping.

Conserved among our CO isolates and canonical to the Delta variant, is a A394V mutation located in the N7-MTase domain.

Additionally, a single CO isolate co-harbors a novel non-codon-aligned insertion in NSP14 (19,486-GGT-19487) resulting in lysine and serine residues replacing valine at position 483.

This mutation occurs in the N7-MTase domain and adjacent to Cys484, a residue part of the zinc-finger 3 motif.

This mutation could impact protein–protein interactions, but the overall effect is unknown given it´s distant location from the catalytic site.

Mutations in two of the motif residues only marginally affected MTase activity in other studies.

Two additional isolates belonging to the S:S112L sub-lineage, but outgrouping from the breakthrough cluster contain substitutions in the NSP14 ExoN domain.

Both isolates were obtained and sequenced during the same April-June 2021 time period.

However, one near neighbor isolate was obtained from a vaccinated individual collected in the same location, and the second more distantly related isolate was obtained from an individual with unknown vaccination status in a different location.

The isolate obtained from a vaccinated individual contains tandem asparagine and arginine substitutions for isoleucine and glycine at NSP14 positions 201 and 202 respectively.

While the functional implication of these polar replacements are unknown, they could impact NSP10/NSP14-ExoN complex stability[42].

The remaining isolate collected in a different CO location contains a conservative A220V mutation.

Last, NSP15 is a uridine-specific endoribonuclease composed of an N-terminal domain, a middle domain and a C-terminal endonuclease domain[43].

NSP15 is involved with the degradation of viral RNA to evade the host defense system.

Located within the poly-U-specific endonuclease domain (EndoU) of NSP15, V303F introduces a potential conformational change to one of five α-helices which flank the two antiparallel β-sheets comprising the catalytic domain of NSP15, a region widely conserved among the *Coronaviridae*.

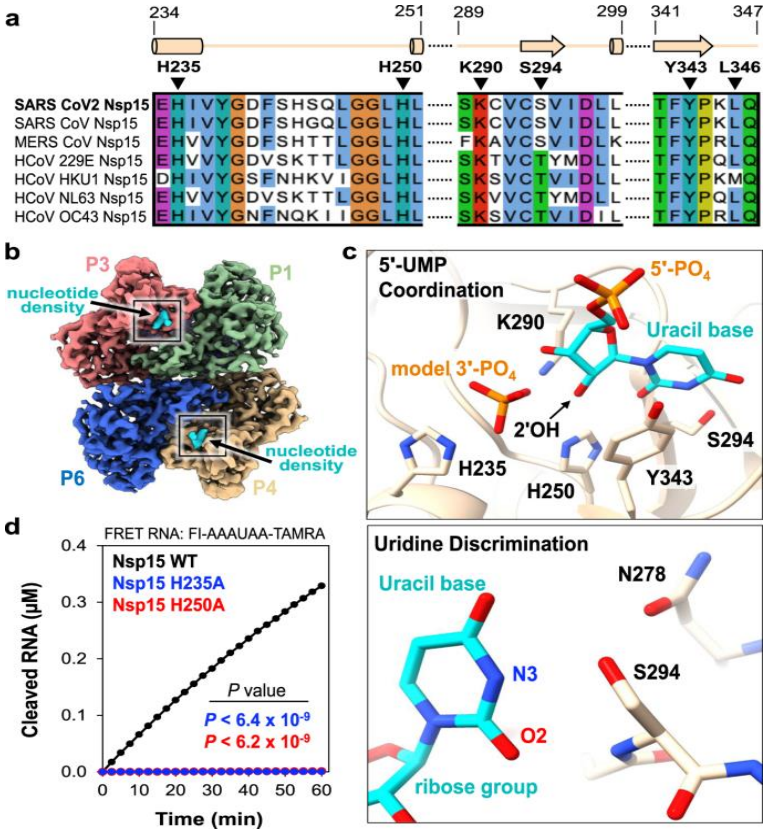
While the impact of this substitution on NSP15 structure and function is unknown, the introduction of an additional aromatic residue replacing the highly conserved valine at position 303 may disorder the nearby beta-sheet and suggests a potentially beneficial influence.

Some of these isolates may be evolving under positive selection due to their potential fitness benefit (e.g. S:S112L, NSP15:V303F and ORF7a:Q94*) resulting in their persistence and subsequent emergence of novel lineages.

On December 27, 2022, Cov-lineages reports[44] than BF.7 detected in China, possess ORF7a:H47Y, ORF1b:L238F, S:C1243F, ORF1a:V274L, and C29632T mutations.

The Figure 16 shows the 5′-UMP coordination by the Nsp15 endoU active site.

Figure 16. 2: 5′-UMP Coordination by the Nsp15 endoU Active Site[43].



a, Amino Acid sequence alignment of endoU active site residues from NSP15 homologs. Secondary structure motifs observed in the nucleotide-bound NSP15 cryo-EM structure are shown above with their corresponding amino acid residue boundaries.b, Nucleotide-bound NSP15 cryo-EM map reconstruction with protomers. Excess UTP was added to the sample resulting in additional density within all six endoU active sites. The nucleotide density is colored in cyan and the black box demarcates the endoU active site.c, NSP15 coordination of 5'-UMP ligands. Due to the poor density of the UTP β- and γ-phosphates, 5'-UMP was modeled into the active site. Cartoon model of the 5'-UMP-bound NSP15 endoU active site where 5'-UMP (cyan) and individual residues H235, H250, K290, S294, and Y343 are shown as sticks (top). Model of uracil base discrimination shown as sticks (bottom). d, RNA cleavage activity of NSP15 variants (2.5 nM) incubated with FRET RNA substrate (0.8 μM) over time. RNA cleavage was quantified from three technical replicates. The mean and standard deviation are plotted and P values of wt-NSP15 compared to H235A (blue; $P < 6.4 \times 10^{-9}$) and H250A (red; $P < 6.2 \times 10^{-9}$) are reported from two-tailed Student's t tests.

The Nsp15 active site contains several residues that are well conserved amongst Nsp15 homologues and important for nuclease activity and specificity (Figure 16a).

The active site within each individual Nsp15 protomer lies near the interface with a neighboring endoU domain (Figure 16b).

This close positioning of the active site to the neighboring protomer hints at a possible mechanism of allosteric communication between the Nsp15 protomers.

The active site residues within the endoU domain have well defined side-chain density in the cryo-EM reconstruction along with additional density for a ligand (Figure 16b).

While we used an excess of UTP when vitrifying Nsp15, we modeled uridine 5′-monophosphate (5′-UMP) into the active site because there was no observable density to account for the β- and γ-phosphates (Figure 16c).

We observed additional ambiguous density next to 5′-UMP, which was modeled as a phosphate ion (Figure 16c; model 3′-PO_4).

The ribose sugar of uridine forms van der Waals interactions with Y343, suggesting that this residue is critical for orienting the ribose within the active site for cleavage (Figure 16c; top).

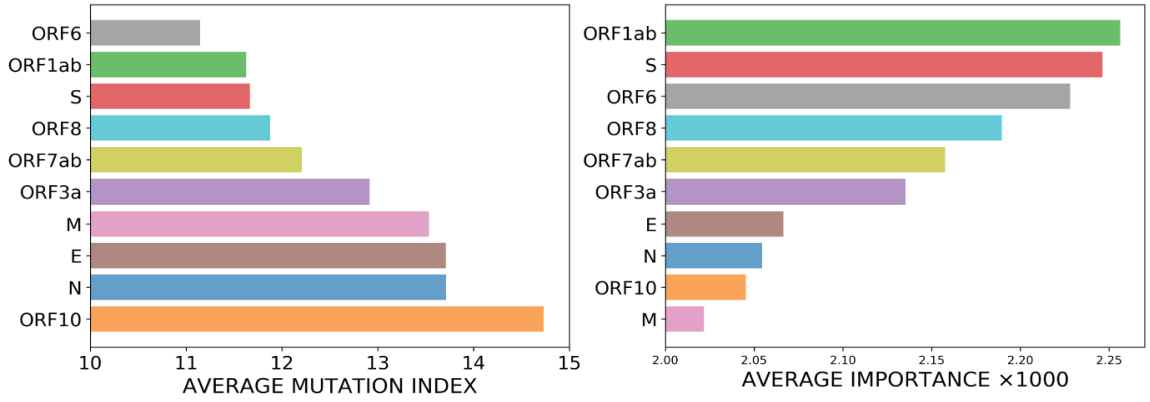
Mutation of the equivalent tyrosine to alanine in SARS-CoV-1 and MERS (Middle East Respiratory Syndrome) NSP15 leads to an almost complete loss in nuclease activity, underscoring the importance of this tyrosine residue.

Amir Darooneh and colleagues from Department of Applied Mathematics, University of Waterloo, ON, Canada, in the paper "A novel statistical method predicts mutability of the genomic segments of the SARS-CoV-2 virus", perform a large-scale analysis of the viral mutations in nearly 80,000 publicly available SARS-CoV-2 predecessor whole genome sequences.

They show that these results are highly correlated with the segments predicted by the statistical method used for keyword detection[45].

Building on these results (Figure 17), they ranked the SARS-CoV-2 genes according to the average of the relative density and importance of their constituent codons.

Figure 17. Average Relative Density and Average Importance of SARS-CoV-2 viral genes.



CONCLUSIONS

Subsequent to the exhaustive and in-depth analysis of recent scientific publications describing the structure and function of proteins expressed by the orf1ab region of the SARS-Cov-2 genome moreover their mutations during the evolution of the SARS-CoV-2 virus in the last 3 years, we can conclude:

1. The most frequent mutations occur in the S protein of the virus, making its interaction with human cell receptors ACE2 and TMPRSS2 increasingly efficient, favoring its entry into the human cell, which increases its transmissibility.

2. The analyses of Orf1ab genome mutations, allows us to predict, through the mutations profile on that region, the severity of an infection for a new SARS-CoV-2 variant that could emerge in the near future.

3. The evolution of key proteins in viral transcription and replication is clearly observed by carefully studying the structure, function, and evolution of RNA-dependent RNA-polymerase (RdRp), Main Protease (Mpro or 3CLpro) and Helicase (NSP13) proteins.

4. P323L, P227L, G671S, V776L and A185S are the first five frequent mutations of RdRp (NSP12).

5. P323L was present in all continents, P227L in five continents (absent in Asia), G671S in three continents (absent in Africa, Oceania and South América), V776L was present in all continents and, A185S only is observed in Europe and America.

6. The mutations P227L and G671S might have functional consequences in the viral transcription and replication.

7. P323L located at the interface domain of RdRp, may exert its effects through altered interaction with other components of the replication-transcription complex or with the RNA template.

8. Mutations in residues D499 to L514, K545, R555, T611 to M626, G678 to T710, S759 to D761 require attention as they are directly implicated with the transcription-replication capability of the virus by RdRp.

9. In Mpro (NSP5) the mutation of residues H41, P132, C145, S145, L226, T234, R298, S301, F305, and Q306 may increase the efficiency of proteolytic cleavage of proteins such as NEMO, thereby improving the ability of the omicron series of viruses to suppress the immune system and accelerate the viral replication.

10. In Helicase (NSP13) the mutations of residues E261, K218, K288, S289, H290, D374, E375, Q404, K460, R567, and A598 are involved in the separation of the double-stranded RNA or DNA with a 5′→3′ polarity as well as 5′ mRNA capping activity in the virus transcription-replication process.

11. In the Orf1ab gene, ORF1b:V2354F mutation, corresponding to NSP15:V303F, may induce a conformational change and result in a disruption to a flanking beta-sheet structure.

12. The premature stop codon ORF7a:Q94*, truncates the transmembrane protein and cytosolic tail used to mediate protein transport, may affect protein localization to the ER-Golgi.

13. P77L mutation located within the N-terminal zinc-binding domain required for catalytic activity and interaction with NSP12.

14. V226L mutation located in the 1B domain is involved in substrate binding.

15. A220, A394V and C484 residue mutations could impact protein–protein interactions.

16. V303F mutation introduces a potential conformational change to one of five α-helices which flank the two antiparallel β-sheets comprising the catalytic domain of NSP15.

17. ORF1b:L238F and ORF1a:V274L mutations along with Orf7a:H47Y, S:C1243F and S:29632T are the finger prints of BF.7 variant.

Conflict of Interest Statement:

None

Funding Statement:

None

Acknowledgement Statement:

None

References:

1. Yadav R, Chaudhary JK, Jain N, Chaudhary PK, Khanra S, *et al.* Role of Structural and Non-Structural Proteins and Therapeutic Targets of SARS-CoV-2 for COVID-19. *Cells.* 2021; 10(4): 821. https://doi.org/10.3390/cells10040821

2. V'kovski Ph, Kratzel A, Steiner S, Stalder H, Thiel V. Coronavirus biology and replication: implications for SARS-CoV-2. *Nature Reviews Microbiology*. 2021. 19:155-170. https://doi.org/10.1038/s41579-020-00468-6

3. Huang J, Song W, Huang H, Sun Q. Pharmacological Therapeutics Targeting RNA-Dependent RNA Polymerase, Proteinase and Spike Protein: From Mechanistic Studies to Clinical Trials for COVID-19. *Journal of Clinical Medicine*. 2020; 9(4):1131. https://doi.org/10.3390/jcm9041131

4. Islam J, Islam NN, Alom S, Kabir M, Halim MA. A review on structural, non-structural, and accessory proteins of SARS-CoV-2: Highlighting drug target sites. Immunobiology. 2023; 228(1):152302. https://doi.org/10.1016/j.imbio.2022.152302

5. Chen J, Malone B, Llewellyn E, Grasso M, Shelton PMM, *et al.* Structural Basis for Helicase-Polymerase Coupling in the SARS-CoV-2 Replication-Transcription Complex. *Cell*. 2020; 182:1560-1573. https://doi.org/10.1016/J.CELL.2020.07.033

6. Lu R, Zhao X, Li J, Niu P, Yang B, *et al.* Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet*. 2020; *395*:565-574.

7. Protein Data Bank data base: https://rcsb.org/covid19

8. Mazhari S, Alavifard H, Rahimian K, Karimi Z, Mahmanzar M, *et al.* SARS-CoV-2 NSP-12 mutations survey during the pandemic in the world. *Research Square*. 2021. https://doi.org/10.21203/rs.3.rs-877078/v1

9. Gao Y, Yan L, Huang Y, Liu F, Zhao Y, *et al.* Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science*. 2020; 368:779-782. http://doi.org/10.1126/science.abb 7498

10. Becerra-Flores M, Cardozo T. SARS-CoV-2 viral spike G614 mutation exhibits higher case fatality rate. *International journal of clinical practice*. 2020; 74:e13525. https://doi.org/10.1111/ijcp.13525

11. Ilmjärv S, Abdul F, Acosta-Gutiérrez S, Estarellas C, Galdadas L, *et al.* concurrent mutations in RNA-dependent RNA polymerase and spike protein emerged as the epidemiologically most successful SARS-CoV-2 variant. *Scientific Reports*. 2021; 11:13705. https://doi.org/10.1038/s41598-021-91662-w

12. Wang Q, Wu J, Wang H, Gao Y, Liu Q, *et al.* Structural Basis for RNA Replication by the SARS-CoV-2 Polymerase. *Cell*. 2020; 182:417-428.https://doi.org/10.1016/j.cell.2020.05.034

13. Venkataraman S., Prasad BVLS, *Selvarajan R. RNA dependent RNA*

polymerases: insights from structure. function and evolution. *Viruses*. 2018; 10(2):76. https://doi.org/10.3390%2Fv10020076

14. Hillen HS, Kokic G, Farnung L, Dienemann Ch, Tegunov D, *et al*. Structure of Replicating SARS-CoV-2 Polymerase. *Nature*. 2020; 584:154-156. https://doi.org/10.1038/s41586-020-2368-8

15. Kirchdoerfer RN, Ward AB. Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nature Communications*. 2019; 10:2342. https://doi.org/10.1038/s41467-019-10280-3

16. Jiang, Y, Yin W, Xu H. RNA-dependent RNA polymerase: Structure, mechanism, and drug discovery for COVID-19. *Biochemical and Biophysical Research Communications*. 2021; 538:47-53. https://doi.org/10.1016%2Fj.bbrc.2020.08.116

17. Pachetti M, Marini B, Benedetti F, Giudici F, Mauro E, et al. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *Journal of Translational Medicine*. 2020; 18:179. https://doi.org/10.1186/s12967-020-02344-6

18. Eskier D, Karakülah G, Suner A, Oktay Y. RdRp Mutations are Associated with SARS-CoV-2 Genome Evolution. *PeerJ Microbiology*. 2020; 8:e9857. https://doi.org/10.7717/peerj.9587

19. Yin W, Mao Ch, Luan X, Shen DD, Shen Q, *et al*. Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science*. 2020; 368: 1499-1504. https://doi.org/10.1126/science.abc1560

20. Koudelka T, Boger J, Henkel A, Schönherr R, Krantz R, *et al*. N-Terminomics for the Identification of *In Vitro* Substrates and Cleavage Site Specificity of the SARS-CoV-2 Main Protease. *Proteomics*. 2020; 21(2): 20246.https://doi.org/10.1002/pmic.202000246

21. Zhang L, Lin D, Sun X, Curth U, Drosten Ch, *et al*. Crystal Structure of SARS-CoV-2 Main Protease Provides a Basis for Design of Improved a-Ketoamide Inhibitors. *Science*. 2020; 368:409-412. https://doi.org/10.1126/science.abb3405

22. Cooper MS, Zhang L, Ibrahim M, Zhang X, Sun X, *et al*. Diastereomeric Resolution Yields Highly Potent Inhibitor of SARS-CoV-2 Main Protease. *Journal of Medicinal Chemistry*. 2022; 65(19):13328-13342. https://doi.org/10.1021/acs.jmedchem.2c01131

23. Hameedi MA, Prates ET, Garvin MR, Mathews II, Amos BK, *et al*. Structural and functional characterization of NEMO cleavage by SARS-CoV-2 3CLpro. *Nature Communications*. 2022; 13:5285. https://doi.org/10.1038/s41467-022-32922-9

24. Rothwarf DM, Zandi E, Natoli G, Karin M, *et al*. IKK-gamma is an essential regulatory subunit of the IkappaB kinase complex. *Nature*. 1998; 395(6699):297-300. https://doi.org/10.1038/26261

25. Orgera, J. SARSNTdb database: Factors affecting SARS-CoV-2 sequence

conservation. *Frontiers Virology*. 2022; 2:1028335. https://doi.org/10.3389/fviro.2022.1028335

26. Yue K, Yao B, Shi Y, Yang Y, Qian Zh, *et al*. The stalk domain of SARS-CoV-2 NSP13 is essential for its helicase activity. *Biochemical and Biophysical Research Communications*. 2022; 601:129-136. https://doi.org/10.1016/j.bbrc.2022.02.068

27. Gorkhali R, Koirala P, Rijal S, Mainali A, Baral A, *et al*. Structure and Function of Major SARS-CoV-2 and SARS-CoV Proteins. *Bioinformatics and Biology Insights*. 2021; 15. https://doi.org/10.1177/11779322211025876

28. Zeng H, Gao X, Xu G, Zhang Sh, Cheng L, *et al*. SARS-CoV-2 helicase NSP13 hijacks the host protein EWSR1 to promote viral replication by enhancing RNA unwinding activity. *Infectious Medicine*. 2022; 1(1):7-16. https://doi.org/10.1016/j.imj.2021.12.004

29. Chen J, Wang Qi, Malone B, Llewellyn E, Pechersky Y, *et al*. Ensemble cryo-EM reveals conformational states of the nsp13 helicase in the SARS-CoV-2 helicase replication–transcription complex. *Nature Structural & Molecular Biology*. 2022; 29:250-260. https://doi.org/10.1038/s41594-022-00734-6

30. Khalili Yazdi A, Pakarian P, Perveen S, Hajian T, Santhakumar V, *et al*. Kinetic Characterization of SARS-CoV-2 nsp13 ATPase Activity and Discovery of Small-Molecule Inhibitors. *ACS Infectious Diseases*. 2022; 8(8):1533-1542. https://doi.org/10.1021/acsinfecdis.2c00165

31. Helmy YA, Fawzy M, Elaswad A, Sobieh A, Kenney SP, Shehata AA. The COVID-19 Pandemic: A Comprehensive Review of Taxonomy, Genetics, Epidemiology, Diagnosis, Treatment, and Control. *Journal of Clinical Medicine*. 2020; 9:1225. https://doi.org/10.3390/jcm9041225

32. Wu A, Peng Y, Huang B, Ding X, Wang X, *et al*. Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. *Cell Host Microbe*. 2020; 27:325-328.https://doi.org/10.1016/j.chom.2020.02.001

33. Majumdar S, Sarkar R. Mutational and phylogenetic analyses of the two lineages of the Omicron variant. *Journal of Medical Virology*. 2021; 94(5):1777-1779. https://doi.org/10.1002/jmv.27558

34. Kopecky-Bromberg SA, Martínez-Sobrido L, Frieman M, Baric RA, Palese P. Severe acute respiratory syndrome coronavirus open reading frame (ORF) 3b, ORF 6, and nucleocapsid proteins function as interferon antagonists. *Journal of Virology*. 2007; 81:548-557. https://doi.org/10.1128/jvi.01782-06

35. Chan JF, Kok KH, Zhu Zh, Chu H, To KK,. *et al*. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerging Microbes & Infections*. 2020; 9(1):221-236. https://doi.org/10.1080/22221751.2020.1719902

36. Abulsoud AI, El-Husseiny HM, El-Husseiny AA, El-Mahdy HA, Ismail A, et al. Mutations in SARS-CoV-2: Insights on structure, variants, vaccines, and biomedical interventions. *Biomedicine & Pharmacotherapy*. 2023; 157:113977. https://doi.org/10.1016/j.biopha.2022.113977

37. Brinkac L, Diepold S, Mitchell S, Sarnese S, Kolakowski LF, *et al*. SARS-CoV-2 Delta variant isolates from vaccinated individuals. *BMC Genomics*. 2022; 23:417. https://doi.org/10.1186/s12864-022-08652-z

38. Su Y, Anderson DE, Young BE, Linster M, Zhu F, *et al*. Discovery and Genomic Characterization of a 382-Nucleotide Deletion in ORF7b and ORF8 during the Early Evolution of SARS-CoV-2. *mBio*. 2021; 11(4): 01610-20. https://doi.org/10.1128/mbio.01610-20

39. Gorkhali R, Koirala P, Rijal S, Mainali A, Baral A, Bhattarai HK, *et al*. Structure and function of major SARS-CoV-2 and SARS-CoV proteins. *Bioinformatics and Biology Insights*. 2021; 15. https://doi.org/10.1177/11779322211025876

40. Romano M, Ruggiero A, Squeglia F, Maga G, Berisio R. A structural view of SARS-CoV-2 RNA replication machinery: RNA synthesis, proofreading and final capping. *Cells*. 2020; 9(5):1267. https://doi.org/10.3390/cells9051267

41. Peacock T, Penrice-Randal R, Hiscox J, Barclay W. *et al*. SARS-CoV-2 one year on: evidence for ongoing viral adaptation. *Journal of General Virology*. 2021;102(4). https://doi.org/10.1099/jgv.0.001584

42. Lin, Sh.; Chen, H.; Chen, Z.; Yang, F.; Ye, Fe. *et al*. Crystal structure of SARS-CoV-2 nsp10 bound to nsp14-ExoN domain reveals an exoribonuclease with both structural and functional integrity. *Nucleic Acids Research*. 2021;. 49(9):5382-92. https://doi.org/10.1093/nar/gkab320

43. Pillon M, Frazier M, Dillard L, Williams J, Kocaman S. *et al*. Cryo-EM structures of the SARS-CoV-2 endoribonuclease Nsp15 reveal insight into nuclease specificity and dynamics. *Nature Communications*. 2021; 12(1):636. https://doi.org/10.1038/s41467-020-20608-z

44. https://github.com/cov-lineages/pango-designation/issues/1470

45. Darooneh A, Przedborski M, & Kohandel M. A novel statistical method predicts mutability of the genomic segments of the SARS-CoV-2 virus. *QRB Discovery*. 2022; 3:E1. https://doi:10.1017/qrd.2021.13