ISBN: 978-93-90847-59-4

4. Molecular Biology of SARS-CoV-2

Chandrasekaran Sankaranarayanan, Perumal Subramanian

Department of Biochemistry & Biotechnology,
Annamalai University, Annamalainagar,
Tamil Nadu, India.

4.1 Introduction:

Coronaviruses (CoV), belong to the Nidovirales order has been associated with a broad spectrum of diseases in animals and humans. First recognized by scientists in 1968, coronaviruses (CoVs) are among the largest family of viruses currently known. These viruses are enveloped and possess (+)-single stranded RNA as its genetic material. Human coronaviruses (HCoVs) are known to cause enteric/respiratory infections. The main HCoV is severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), responsible for COVID-19 pandemic, activate an excessive immune response called as a cytokine storm, cause acute respiratory distress syndrome (ARDS), multi-organ failure, and death with 2% mortality rate. Observations in people with COVID-19 suggest that the virus can be carried in the blood and infect various organs or tissues.

COVID-19 first emerged in December 2019, when a cluster of patients with pneumonia of unknown cause was recognized in Wuhan, China. As of July 15, 2021, in India, the number of identified cases of SARS-CoV-2 is 3,10,26,829 with 4,12,531 confirmed deaths. The outbreak of SARS- CoV-2 pandemic has made its extraordinary global, societal and economic troublemaking impact on human population.

4.2 Structure:

CoVs are the largest known RNA viruses, with a genomic size of about 30 kb. Their genome is monopartite, ss RNA of positive polarity, modified with a 5'-cap, and 3'-poly (A) tail that can be translated upon entry into the host cell.

The genes that code for the five important structural proteins (the spike (S) protein, the nucleocapsid (N) protein), the hemagglutinin-esterase (HE, present in some β-coro-naviruses) the membrane (M) protein and the envelope (E) protein) occupy less about 30 % of the genome at the 3'-end. The remaining 60 -70 % of the genome is occupied by a single gene that codes for the viral replicase/transcriptase proteins.

The virions are spherical or pleomorphic enveloped particles with a diameter of 80-120 nm, formed by the structural proteins. The S glycoprotein (150 kD with 1300 amino acid residues) with a club-like structure is a large type I transmembrane polypeptide having a N-terminal exodomain and a C-terminal endodomain that assembles into trimeric spikes that project from the virus and this whole structure is involved in receptor binding and membrane fusion.

Another major protein of the viral envelope is M (about 30 kD with approximately 235 amino acid residues) which is a type III transmembrane glycoprotein which is crucial for the shape of the virus, N-terminal and C-terminal are ecto and endodomains respectively. The endodmain binds with RNA of the virus. This protein shows structural resemblance with M protein of MERS and SARS and with prokaryotic sugar transport protein (Braakman and Van Anken 2000). The antigenic role of this protein is attributed to N-linked and O-linked glycosylation (Braakman and Van Anken, 2000).

The E protein (about 10 kD with approximately 100 amino acid residues) is a small transmembrane, hydrophobic protein present in lower copy numbers acts a key structural component of viral envelope. Further, this protein functions as ion channel, viroporin activity and virus assembling (Schoeman and Fielding 2019). The proline residue present in C-terminal (CT) region is highly conserved and is essential for the maturation of the protein in the Golgi of the host cell (Tseng et al., 2014). The important region in CT is PDZ-binding motif (PBM) that is crucial for cell signalling and pathogenicity. The PBM is highly conserved among SARS-CoV-2 variants and is responsible for stimulating cytokine storm, inflammasome and oedema in lungs. Key post-translational changes that occur in Eprotein are glycosylation, palmitoylation, myristoylation, and ubiquitination, (Schoeman and Fielding, 2019). Because of the functional diversity, the inhibitors of this E- protein are suitable agents for SARS-CoV-2 treatment (Alam et al. 2020).

The viral envelope surrounds a nucleocapsid with helical symmetry, which is not characteristic for positive-strand RNA viruses but rather typical of negative-strand RNA viruses. The helical nucleocapsid is formed by the N phosphoprotein, which associates with the RNA genome in a beads-on-a-string fashion. The HE (hemagglutinin-esterase) is the fourth protein constituent of the viral envelope. It forms shorter dimeric projections with C-endo and N-exo terminals. It has acetylesterase and haemagglutinating properties and may participate during cell entry and egress. The mature protein is stabilized by disulfide bonds (Liu et al., 2014), promotes reversible attachment of the virus to the host cell and acts as receptor-destroying enzyme (RDE) and as lectin.

The N-protein (43-48 kD) is ribonucleo protein rich in positively charged amino acids (lysine and arginine) is tightly bound with negatively charged RNA via N-terminal domain in a beads-on-a-string manner. N protein dimer possesses the shape of a rectangular slab in which the four-stranded β-sheet is on one side of the face and the α-helices form the opposite face of the slab (McBride et al., 2014). The protein is dimerized via C-terminal which is abundant in serine and arginine that undergo phosphorylation (Kumar et al. 2020). The reason for phosphorylation is not clearly known till date. N protein plays important role in regulation of transcription and packaging of viral genome in the formation of viable virion. The protein contains three important regions, viz. C-terminal domain (CTD), the linker region (LKR) and N-terminal domain (NTD).

Among the three, LKR can directly interact with RNA under in vitro conditions. The LKR can also bind with M protein, heterogenous nucler ribonucleoprotein A1 (hnRNPA1) and helps with the tight binding of N protein with RNA genome (McBride et al. 2014). N protein could arrest the cell cycle of host cell in G2/M phase (Wurm et al. 2001) and as it shows less variation than other envelope proteins it is considered as a vaccine candidate for SARS-CoV-2 (Ahmed et al. 2020).

A. Non-Structural Proteins:

The ribosomes of the host cell translate the viral genome in a cap-dependent mechanism produce two large polyproteins, pp1a and pp1ab. Cotranslational and autoproteolytic processing of pp1a and pp1ab result in the production of 15 or 16 nonstructural proteins (nsp) (Table 1). From pp1a, Nsp1 to nsp11 and from pp1ab nsp12 to nsp16 are produced. These sixteen nsps contribute to transcription and replication.

Table 4.1: Characteristics of non structural proteins

Name of the protein	Structural features	Functions
nsp1	180 amino acids; subjected to numerous point mutations causing variants	Degradation of host mRNA and 40s ribosomal subunit resulting in inhibition of host gene expression and cell division (Kamitani et al., 2009).
nsp2	61 amino acids; 65 kD	Interaction with nsp8; essential for the structure of double membrane vesicle (DMV)-anchored replication-transcription complexes (RTCs); promotion of cell death pathways and inhibition of cell differentiation; coding gene is essential role for RT-PCR assay
nsp3	~200 kD; ubiquitin-like domains 1 and 2 (Ubl1 and 2), Glu-rich acidic domain, macrodomain, papain-likeprotease2 (PL2pro), nsp3 ecto domain, Y1 domain and two conserved transmembrane (TM1 and TM2) domains (Lei et al., 2018)	Disruption of the host cell cycle (Ma-Lauer et al., 2016); critical role in the formation of Replication-Transcription Complex (RTC) and suppression of host immune response
nsp4	Contains four TMDs (TM1–TM4)	Key roles in the development, arrangement, and function of viral replication complexes; necessary for the structure of RTCs anchored with double membrane vesicles (DMV)
nsp5	306 amino acids	Acts as main protease and plays important role in individual viral protein synthesis; inhibition of interferon I signalling via NF-kB and STAT 1 transcription factor
nsp6	Forms complex nsp 3 and nsp4	Induction of cell autophagosis; nsp3 and nsp4, can close the DMVs
nsp7	10 kD	Co-factor component; forms heterodimer with nsp8; stabilization of RNA binding site in nsp12

Name of the protein	Structural features	Functions
nsp8	22 kD; N-terminal interacts with nsp7	Possesses RNA dependent RNA polymerase (RdRp) replicase; essential for nsp12 polymerization activity
nsp9	198 amino acids	Essential for RNA replication; requires nsp8 for its function
nsp10	148 amino acids containing two zinc finger domains	Replication regulator; interacts with nsp 14 and nsp16
nsp11	13 amino acids	endo-ribonuclease activity; inhibitory effects on TNF-α development and IL1 signaling (He et al., 2015) main roles in DNA replication, cell cycle and signalling, modification of histones, proteasomal degradation
nsp12	932 amino acids	Possesses primer-dependent and primer-independent RdRp and nucleotidyltransferase activities; enhances the helicase activity of nsp13
nsp13	Has several domains, such as NT Cys/His-rich domain (CH) with three zinc atoms and β-barrel domain	Helicase activity; hydrolyzes all forms of NTPs
Nsp14	527 amino acids	Exonuclease (3'→5') activity; interacts with nsp 10; involved in synthesis of 5' cap
Nsp15	345 amino acids	Endoribonuclease activity specific at 3'U residues; can delay interferon signalling
Nsp16	298 amino acids	nsp16 requires an interaction with nsp10 to perform powerful methyltransferase, which can transform cap-0 (7MeGpppN) specifically into the cap-1 structure; affects the immune system and activates the antiviral response in the body; interferes with interferon β and Interferon Stimulated Response Element (ISRE) signalling

4.3 Multiplication Cycle:

Attachment and entry of SARS-CoV-2 into host cell:

The binding and entry of SARS-CoV-2 into the host-cell depend on the viral S glycoprotein, comprised of two domains S1 and S2. The vastly variable S1-RBD domain (receptor binding domain) recognizes and binds with angiotensin-converting enzyme II (ACE2) receptor, brings conformational changes and promotes the fusion of S2 domain with host cell membrane (Sun et al., 2014a). This binding stimulates signalling pathway via phosphorylation of the receptor by casein kinase 2, which leads to activation of AP1 (activator protein-1) and ERK1/2 (extracellular signal-regulated kinase), CCL2 (chemokine (C-C motif) ligand 2) expression and eventually resulting in pulmonary fibrosis.

Further, the proteolytic cleavage of S protein is crucial to allow fusion. This is achieved by the cell-surface serine protease, TMPRSS2 (transmembrane protease serine 2) which is essential for priming and entry of SARS-CoV-2. The endosomal cathepsins B and L (cat B and catL) facilitate this entry process.

TMPRSS2 is expressed in the human respiratory tract and therefore robustly contributes to both SARS- CoV2 spread and pathogenesis. Particularly, as SARS- CoV-2 entry relies mainly on TMPRSS2, its inhibition is considered as a powerful therapeutic approach. Subsequently, through clathrin-dependent or clathrin-independent ways SARS-CoV-2 enters the host cell via endocytosis (Algarroba et al., 2020). The host cell lysosomal enzymes such as catL and trypsin facilitate the release of viral genome into the host cell cytosol.

Viral RNA synthesis and translation:

After release, the virus alters the host transcription process to its favour through the onset of complex events of viral gene expression. Consequently, ribosomes translate the viral RNA via a cap-dependent mechanism to produce pp1a and pp1ab (two large polyproteins) which are further processed to produce 16 non-structural proteins (nsps).

The RTC assembly on the modified ER membrane is associated with viral RNA synthesis. The establishment of the viral RTC is vital for virus replication and therefore a promising target for antiviral agents against SARS- CoV-2. The RTC is produced in DMVs synthesize a cluster, including subgenomic RNAs in a discontinuous transcription manner.

Using (-) strand intermediate RNA, a series of genomic and subgenomic RNAs are produced. The new genomic RNA molecules are encapsidated with structural proteins M, N, S and accessory proteins and incorporated into progeny virions on membranes of the endoplasmic reticulum-Golgi intermediate compartment (ERGIC). Ultimately, the viral particle loaded vesicles are attached with the plasma membrane to release the virus.

4.5 Variants of Coronavirus:

Genetic variants of SARS-CoV-2 have been emerging as a result constant mutation and they are of serious concern globally. WHO classifies SARS-CoV-2 variants as Variants of Concern (VOC) and Variants of Interest (VOI). Besides, Variants of High Consequence (VOHC) are also included. A variant has one or more mutations that distinguish it from other SARS-CoV-2 strains. The documented variants so far are: α, β, γ, δ and δ^{+}.

4.6 Conclusion Remarks:

High transmissibility and pathogenicity of SARS-CoV-2 is due to a high mutation rate leading to the synthesis of different protein structures such as S protein compared with SARS-CoV and MERS-CoV (Wang et al., 2020a). In addition, CoVs may adapt swiftly to varying ecological niches due to high recombination frequencies that originate from the complex mechanisms.

4.7 References:

1. Ahmed, S.F., Quadeer, A.A., and McKay, M.R.J.V. (2020). Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. Viruses 12, 254.
2. Alam I, Kamau A, Kulmanov M, Arold ST, Pain A, Gojobori T et al. (2020) Functional pangenome analysis suggests inhibition of the protein E as a readily available therapy for COVID-2019. BioRxiv [Preprint]. https://doi.org/ 10.1101/2020.02.17.952895.
3. Algarroba, G.N., Rekawek, P., Vahanian, S.A., Khullar, P., Palaia, T., Peltier, M.R., et al. (2020). Visualization of SARSCoV- 2 virus invading the human placenta using electron microscopy. Am J Obstet Gynecol [Epub ahead of print]; DOI: 10.1016/j.ajog.2020.05.023.
4. Braakman I, Van Anken EJT (2000) Folding of viral envelope glycoproteins in the endoplasmic reticulum. Traffic 1, 533–539.
5. He, Q., Li, Y., Zhou, L., Ge, X., Guo, X., and Yang, HJVr (2015). Both Nsp1b and Nsp11 are responsible for differential TNF-a production induced by porcine reproductive and respiratory syndrome virus strains with different pathogenicity in vitro. Virus Res 201, 32–40.
6. Kamitani, W., Huang, C., Narayanan, K., Lokugamage, K.G., Makino, S.J.Ns., and Biology, M. (2009). A two-pronged strategy to suppress host protein synthesis by SARS coronavirus Nsp1 Protein 16, 1134.
7. Kumar, A., Parveen, A., Kumar, N., Bairy, S., Kaushik, V., Chandola, C., et al. (2020). Characterization of nucleocapsid (N) protein from novel coronavirus SARS-CoV-2. Preprints 2020, 2020050413.
8. Liu L. Fields virology. Clin Infect Dis. (2014); 59:613.
9. Ma-Lauer, Y., Carbajo-Lozoya, J., Hein, M.Y., Mu¨ ller, M.A., Deng, W., Lei, J., et al. (2016). p53 down-regulates SARS coronavirus replication and is targeted by the SARS-unique domain and PLpro via E3 ubiquitin ligase RCHY1. PNAS 113, E5192–E201.
10. McBride, R., van Zyl, M., Fielding, B.C. (2014). The coronavirus nucleocapsid is a multifunctional protein. Viruses 6, 2991–3018. https://www.ncbi.nlm.nih.gov/pmc /articles/PMC4147684/.
11. Schoeman D, Fielding BCJV (2019) Coronavirus envelope protein: current knowledge. Virol J 16, 69.
12. Sun, Y., Li, D., Giri, S., Prasanth, S.G., and Yoo, DJBri. (2014a). Differential host cell gene expression and regulation of cell cycle progression by nonstructural protein 11 of porcine reproductive and respiratory syndrome virus. Biomed Res Int 2014, 430508.
13. Tseng Y.-T, Wang, S.-M, Huang K.-J, Wang C.-T (2014) SARS-CoV envelope protein palmitoylation or nucleocapid association is not required for promoting virus-like particle production. J Biomed Sci 21, 34.
14. Wang, Q., Qiu, Y., Li, J.-Y., Zhou, Z.-J., Liao, C.-H., and Ge, X.-Y. (2020a). A unique protease cleavage site predicted in the spike protein of the novel pneumonia coronavirus (2019- nCoV) potentially related to viral transmissibility. Virol Sin 20, 1–3.
15. Wurm, T., Chen, H., Hodgson, T., Britton, P., Brooks, G., and Hiscox, J.A. (2001). Localization to the nucleolus is a common feature of coronavirus nucleoproteins, and the protein may disrupt host cell division. J Virol 75, 9345–9356.