



E-ISSN: 2706-9575
 P-ISSN: 2706-9567
 IJARM 2020; 2(1): 18-21
 Received: 19-01-2020
 Accepted: 21-02-2020

Vikram S Chauhan
 Department of Biology,
 Government Bangur (PG)
 College, Pali, Rajasthan, India

Meenakshi Parihar
 Government Senior Secondary
 School, Daantiwara, Jodhpur,
 Rajasthan, India

Molecular biology of coronaviruses

Vikram S Chauhan and Meenakshi Parihar

DOI: <https://doi.org/10.22271/27069567.2020.v2.i1a.33>

Abstract

COVID-19 is related to Coronavirus (CoV) group of positive strand single stranded RNA viruses. It replicates within host cytoplasm without entering into nucleus of host cells. Genetics of CoV reveals that many nonstructural protein and structural proteins impacts host cell protein metabolism and gene expression. After entering into various tissues including lungs it up-regulates and down-regulates some critical host genes that are related to cell cycle regulation and apoptosis. It also affects nonspecific immune system which leads to inflammation of various respiratory organs.

Keywords: Capsid, ORF, polyprotein, interferon, serological groups

Introduction

The great pandemic of COVID-19 of present times is proving worst among all epidemics known. Starting from 'Wet market' of city of Wuhan (China), it is spreading in remarkable chain reaction. Because Wuhan viral pneumonia was discovered in end of year 2019, so WHO in January 2020 named it as COVID-19. These human coronaviruses were first discovered by David A.J. Tyrrell during cold epidemics in 1965. Electron micrographs revealed virus particles that are surrounded by an envelope containing embedded proteins, which confer on them the appearance of a solar corona or "halo" (in Latin, corona). COVID-19 is now well established to be another example of animal to human transmission of a viral disease. Scientists have identified three phases of pathogenesis of CoV. First, viral replication, second, immune hyperactivity and third and most fatal is pulmonary destruction. Pulmonary pathology includes alveolar damage, epithelial proliferation, and increased infiltration of macrophages. There is down regulation of genes responsible for fibrinolysis leading to lung fibrosis. COVID-19 represents most of death in elderly people (>60 years) with average mortality of 3.4 percent. It is presently, impossible to differentiate clinically and hence a broad spectrum antiviral will be required to treat this global epidemic. The present review work summarized basic molecular biology and pathogenic impacts of viral infection. This work is aimed to introduce basic mechanism for the sake of awareness among biologists and doctors to understand its genetics at molecular level.

Classification

The coronaviruses are divided into three serological groups, and the two human prototypes are OC43 and 229E. As per Baltimore Classification of viruses, the coronaviruses are subdivided into two subfamilies¹. Their characteristics and host range are summarized in following table.

Table 1: Shows Coronavirinae and Torovirinae

Coronavirinae	<ul style="list-style-type: none"> • Coronaviruses that infect humans as well as many mammals such as ungulates, carnivores and bats have recently been assigned into the genera Alphacoronavirus and Betacoronavirus. • SARS-related coronaviruses has been classified into the Betacoronaviruses isolated from civet cats and bats. • The genera Gammacoronavirus and Deltacoronavirus cause disease in various birds.
Torovirinae	<ul style="list-style-type: none"> • In the subfamily Torovirinae, the white bream virus is a member of the genus Bafinivirus and it is known for pathogenic for fish, whereas other bovine, equine, human and porcine toroviruses belong to the genus Torovirus and they cause gastrointestinal infection in the respective host

Corresponding Author:
Vikram S Chauhan
 Department of Biology,
 Government Bangur (PG)
 College, Pali, Rajasthan, India

Structure of Virus Particle

The enveloped virions have a diameter of 80-180 nm. Their single stranded, positive-sense RNA (+ssRNA) genome is associated with the viral N protein and forms a nucleocapsid in the interior of the particle. The nucleocapsid has a helical shape with a diameter of 10-20 nm. Some specific amino acids of the N protein interact with the carboxy-terminal domain of the M protein that is inserted in the envelope. So, the nucleocapsid is associated through protein interactions with the inner side of the envelope. Apart from the M

protein, an amino-terminally glycosylated protein of 20-30 kDa, two other viral proteins are embedded in the envelope, which are the glycosylated S protein (180-200 kDa) is present in club-shaped trimmer, which protrude about 20 nm from the envelope surface and are responsible for the character “corona appearance” of the virus, and the E protein of 9-12 kDa. Another membrane-associated protein, haemagglutinin esterase (HE) of molecular mass of 65 kDa is present only in members of Betacoronavirus [2].

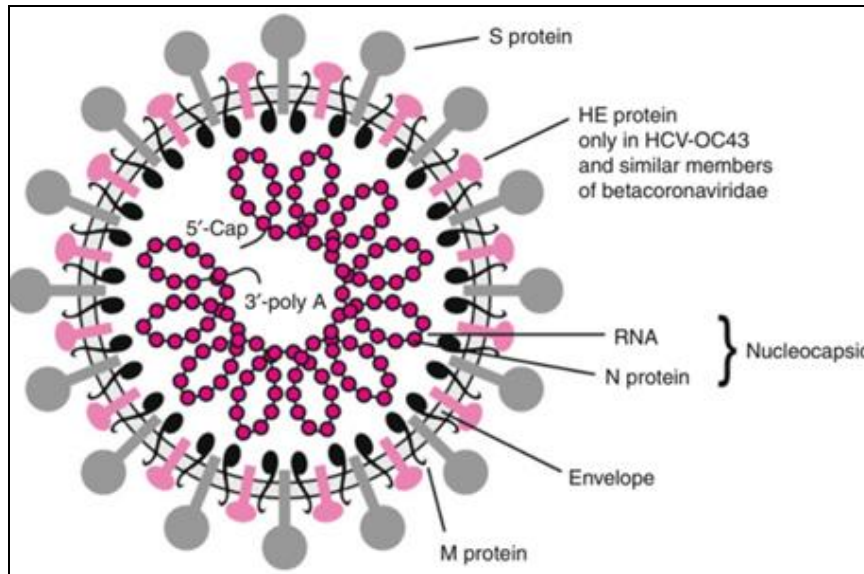


Fig 1: CoV Particle (Source *Molecular Virology* Ed. Susanne Modrow *et al* Springer)

Molecular Genetics

These viruses have the largest genome of all known RNA viruses: it has a length of 27,000-32,000 nucleotides. A general, overall and simplified structure of genomic RNA is given in following figure.

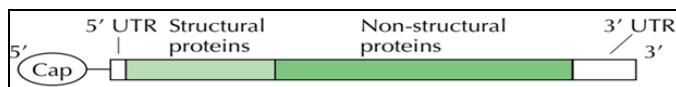


Fig 2: Overall Genome Organization of SARS- CoV

These ssRNAs genomes of are capped and polyadenylated like other eukaryotic mRNAs. The ssRNA is of positive strand genome so virions can replicate without making intermediate DNA in the viral replication cycle. The RNA

genome contains multiple coding regions. There are two relatively large open reading frames (ORFs), *1a* and *1b*, which overlap at their ends by 40-60 nucleotides (*1a* starts immediately adjacent to the 5' terminus). A total of 14 ORFs have been identified. 2 large 5'-terminal ORFs, *1a* and *1b*, they forms about two-thirds of the entire genome and constitute the replicase gene which encodes proteins required for viral RNA synthesis. 4 ORFs which encode structural proteins in this gene order, i.e., spike (S), envelope (E), membrane (M) and nucleocapsid (N) and 8 ORFs encoding accessory proteins. Following illustration represent viral genome and their protein products. Here readers are required to notice a ribosomal frame shift between ORF1a and ORF1b, these two ORFs are entire source of all nonstructural proteins (NSP) [3].

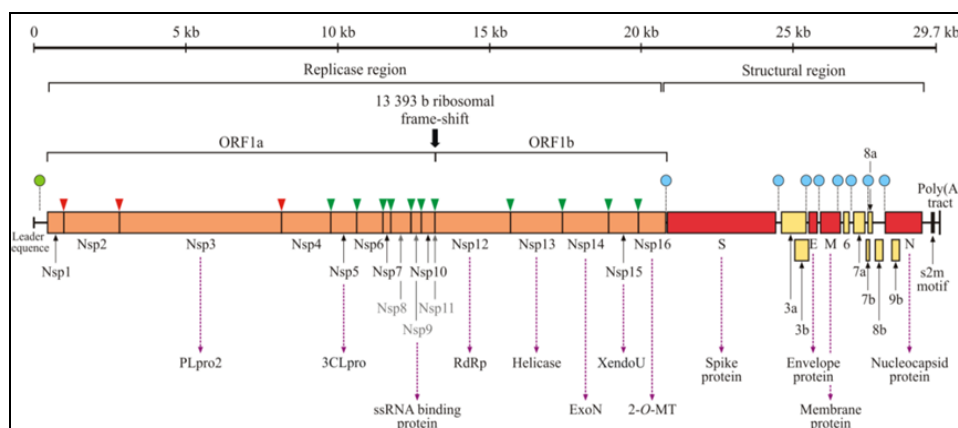


Fig 3: Genome organization and replication of SARS CoV (Source: Vincent T. K. Chow and W. F. Leong, Chapter 13 in book *RNA Viruses, Host Gene Response to Infections* Editor: Decheng Yang)

There is a 72 nucleotide leader RNA region which is at upstream the 5' untranslated region (UTR), this is also found in a nested set of eight subgenomic (sg) mRNAs. Due to ribosomal frame-shifting into the -1 reading frame occurring just upstream of the ORF1, a stop codon, polyprotein (pp) 1a can be extended with the ORF1b-encoded sequences to form *pp1ab*. The *pp1a* and *pp1ab* are cleaved by viral proteinases, papain-like cysteine protease (PLpro) and 3C-like cysteine protease (3CLpro), into individual polypeptides necessary for viral RNA replication and transcription. 3CLpro processes the central and C-terminal

regions and produces the key viral replicative enzymes which are RNA-dependent RNA polymerase (RdRp) and helicase with ATPase and DNA/RNA duplex unwinding activity [4].

Non-Structural Proteins (NSPs)

The large, not isolable precursor protein *pp1ab* is autocatalytically cleaved by the activity of the proteases into 13-16 cleavage products. Some important NSPs are summarized in following table.

Table 2: Shows NSP/Accessory proteins, Functions and Remarks

NSP/Accessory proteins	Functions and Remarks
NSP12	RNA-dependent RNA polymerase (RdRp) is NSP12 that is required for both replication of the viral RNA genome and synthesis of subgenomic mRNAs
NSP13, NSP14 and NSP16	NSP 12 forms with the RNA helicase (NSP13) that probably binds Zn ²⁺ ions, an exoribonuclease (NSP14), an endoribonuclease (NSP15) and a 2'-O-ribose methyltransferase (NSP16)
NSP5/3CL ^{Pro}	Serine protease
NSP3/PL2 ^{Pro}	Papain like cysteine protease, causes protein deubiquitylation; ADP phosphatase activity
NSP1	Virulence factor, causes degradation of cellular RNAs, blocks the synthesis of IFN-α and IFN-β
NSP4	Viral morphogenesis

Structural Proteins

Three types of envelope proteins are found in all coronaviruses: the M protein (formerly also known as E1 protein), the S protein (also called E2 protein) and the E

protein (also called M protein). The following table summarizes basic function of structural proteins in viral life cycle:

Table 3: Shows Structural Protein, Functions and remarks

Structural Protein	Functions and remarks
S protein (Surface or spike protein)	<ul style="list-style-type: none"> Molecular mass of 180-200 kDa. Glycosylated and anchored in the envelope of the virus, and in the cytoplasmic membrane by a transmembrane domain at the carboxy terminus, which is modified by an aliphatic acid. Exists as a dimer or trimer and forms club-like protuberances on the surface of the virus Neutralizing antibodies are produced against the S protein in the course of an infection. Helps the virus attaches to host cell surface molecules by specific domains, Fusogenic activity of virus
HE protein	<ul style="list-style-type: none"> Present only in betacoronaviruses. Glycosylated, has a molecular mass of about 65 kDa and forms dimers by disulphide bonds. Provide capability of haemagglutinating and binding to erythrocytes. (In this process, the HE protein interacts with the 9-O-acetylated neuraminic acid (sialic acid), which is a modification of lipid and protein components on cell surfaces) its Esterase activity, which enables the virus to remove acetyl groups from sialic acid molecules
E protein	<ul style="list-style-type: none"> (9-12 kDa is found within infectious virus particles in different concentrations; it is Necessary for particle assembly and morphogenesis. Proapoptotic function and acts as viroporin, forming ion channels and altering the membrane permeability
M protein (M for Matrix)	<ul style="list-style-type: none"> Surface protein with a molecular mass of 20–30 kDa and is glycosylated at the amino-terminal domain. The C terminus resides inside the virus particle and interacts with the N protein of the nucleocapsid Not transported to the plasma membrane via the Golgi apparatus, but it remains in the ER membrane By the interaction of the M protein with the nucleocapsid, the first steps of viral assembly occur at those sites, which initiate the budding process into the ER lumen. The E protein is also involved in viral morphogenesis
N protein (N for nucleic acid binding)	<ul style="list-style-type: none"> Interacts with the viral genome; it is rich in basic amino acids and phosphorylated. It can also interact specifically with the carboxy-terminal regions of the M protein. Interaction with smad3, interferes with cellular transcription and cell cycle regulation
3a protein	<ul style="list-style-type: none"> Up-regulates the expression of fibrinogen in lung epithelial cells, which may contribute to pathogenesis via excessive formation of fibrin Have Nuclear localization signal (NLS) so induce apoptosis and cell cycle arrest at G2-M phase
7a protein	<ul style="list-style-type: none"> 7a protein inhibits cellular protein synthesis and cell cycle progression With 3a and a non-structural protein 1 contribute to chemokine dysregulation via NF-κB activation

Host entry of CoV via ACE2

sSARS-related coronavirus and human coronavirus NL63 bind to target cell by a domain of the S1 protein to a

metalloprotease, angiotensin-converting enzyme 2 (ACE-2). This protein is present on the surface of pneumocytes, enterocytes and cells of heart, kidney and endothelium. CoV

is also known to bind some lectins such as DC-SIGN, L-SIGN and LSECtin, and facilitates entry of CoV into the cell. In addition to the S protein the viral HE protein in the membrane, can also interact with 9-O-acetylated neuraminic acid residues on the host cell surface. Penetration of the particle seems to occur by receptor-mediated endocytosis, and subsequent fusion of the endosomal membranes with the viral membrane. Binding between S1 and ACE-2 proteins induces conformational changes in the S proteins of CoV, and the fusogenic domain of S2 moiety promotes fusion between the viral envelope and the cell membrane. A protease, cathepsin L, which is associated with host endosomal membranes further promotes the infectivity of the virus by processing uncleaved S polypeptides into S1 and S2 proteins. Entire replication cycle of virus takes place in the host cell cytoplasm. Following flow chart depicts sequential events of host entry of CoV into host cells.

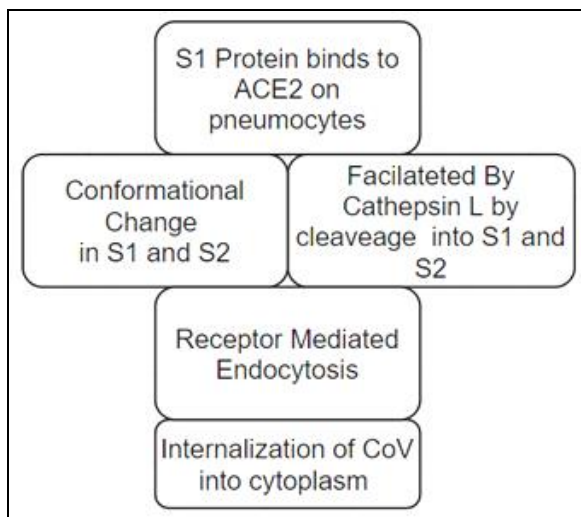


Fig 4: Host entry of CoV into host cells via ACE2

Molecular Pathology of CoV

1. Growth inhibition and genetic dysregulation

Table 4: Shows Upregulated Genes and Down regulated Genes

Upregulated Genes	Down regulated Genes
<ul style="list-style-type: none"> DUSP1 and FANCC: Dual specificity phosphatase 1 (DUSP1) plays an important role as a protein phosphatase act as a negative regulator of cellular proliferation by inactivating mitogen-activated protein kinases (MAPK). KLF5 is a member of the Kruppel like factor subfamily of zinc finger proteins. Some reports indicate that KLF5 inhibits cell proliferation and lung cells 	<p>Cathepsin L (CTSL) a lysosomal cysteine proteinase that plays a major role in intracellular protein recycling.</p>

2. ACE 2 Modulated Effects: Susceptibility to lung Infections

CoV infects and proliferates in the epithelial cells of the respiratory tract. The ACE-2 proteins, to which acts as receptor for the viral S proteins, are found on the surface of cells in several tissues; they are an integral part of the renin-angiotensin system and regulate it. The interaction of S

proteins with ACE-2 decreases its concentration on the infected cell surface. This induces a susceptibility to inflammations and lung pathogenesis.

3. Immune Hyper Response

In early phase of infection, increased production of CC chemokines and chemokine receptors, proinflammatory interleukins and Toll-like receptor 9 causes mobilization of monocytes and macrophages, which migrate into the lungs, and initiate the severe inflammatory process. A rapid decrease in CD4+ and CD8+ T lymphocytes is also found in acute phase of infection [6].

Conclusions

Molecular genetics of CoV reveals most critical proteins which are potential for drug and vaccine targets. CoV affects host cell biology by disturbing and dysregulation genes which affect apoptosis and cell cycle, further infection leads to severe inflammatory response and lung fibrosis which clinically manifested as difficult breathing and pneumonia and multiple organ failure including myocardial infarction. Being highly infective and devastating as we are witnessing the Corona calamity must be knocked out with earlier development of drug designing and vaccination.

References

1. Yang, Decheng (Ed). RNA viruses, Host Gene response to Infections. World Scientific Publishing, Singapore, 2009.
2. Douglas D. Richman. Clinical Virology. ASM Press, Washington DC, 2017.
3. Zeibuhr John (Ed) Coronaviruses. Elsevier, Cambridge United States, 2016.
4. Lal Sunil K (Ed). Molecular Biology of SARS Coronaviruses. Springer
5. Grandoff A, Webster RG (Ed). Encyclopedia of Virology. Academic Press. USA
6. Medrow S (Ed) Molecular Virology. Springer, 2010.