The molecular biology of intracellular events during Coronavirus infection cycle

Juhi Jain[1] · Sharad Gaur[1] · Yash Chaudhary[1] · Rajeev Kaul[1]

Abstract CoV-2 which is the causative agent of COVID-19 belongs to genus *betacoronaviruses*. The sequence analysis of S protein of CoV-2 has shown that it has acquired a 'polybasic cleavage site' consisting of 12 aminoacids that has been predicted to enable its cleavage by other cellular proteases possibly increasing its transmissibility. The aminoacids present in receptor binding domain of S protein of SARS CoV which are critical for its binding to cellular receptor are different in CoV-2. The presence of heptanucleotide slippery sequence in ORF1 resulting in ribosomal frameshifting, and presence of transcription regulatory sequences between ORFs resulting in discontinuous transcription, are peculiar features of Coronavirus infection cycle. The exonuclease activity of nsp14 provides possible proofreading ability to RNA polymerase makes coronaviruses different from other RNA viruses allowing coronaviruses to maintain their relatively large genome size. This mini-review summarizes the peculiar features of Coronaviruses genome and the critical events during the infection cycle with focus on CoV-2.

Keywords CoV-2 · Coronavirus · COVID-19

Introduction

Coronaviruses are known to infect humans as well as animals. These viruses are classified in four genera named as *alpha-*, *beta-*, *gamma-* and *deltacornavirus*. Alpha- and *betacoronaviruses* infect only mammals, whereas *gamma-* and *deltacoronaviruses* mainly infect birds. Out of seven known human Coronaviruses, HCoV-NL63 and HCoV-229E belong to the genus *alphacoronavirus;* whereas remaining five belong to genus *betacoronavirus*. These include HCoV-OC43, HCoV-HKU1, SARS, MERS and CoV-2 which is the causative agent of COVID-19. The SARS and MERS Coronaviruses are no longer circulating in humans; however the other four human Coronaviruses had been circulating and are known to cause common cold [22]. Both SARS and MERS Coronaviruses originated from bats and were transmitted to humans via an intermediary host. For SARS, the intermediate host was Himalayan palm civet, whereas for MERS it was dromedary camels [19]. Subsequent studies have shown that bats are host to more than 500 species of Coronaviruses. However, nothing is known regarding the wild animal host or any intermediate host for CoV-2 which is the cause of current ongoing pandemic. SARS emerged in year 2002 and caused an epidemic that resulted in about 8000 infections with almost 10% case fatality rate, whereas MERS epidemic resulted in more than 2500 infections with approximately 34% case fatality rate thereby becoming the most pathogenic of all human Coronaviruses. The ongoing CoV-2 pandemic has so far infected more than 3 million people worldwide causing death of more than 200,000 people with case fatality rate of approximately 7% making it the deadliest of all Coronaviruses.

The virus entry into cells

The virus particle is pleomorphic and is packed with 30 Kb positive sense RNA genome which is coated with nucleocapsid (N) protein. Like all enveloped viruses, coronavirus
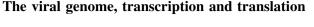
✉ Rajeev Kaul
rkaul@south.du.ac.in

1   Department of Microbiology, University of Delhi, South Campus, New Delhi, India

envelope is also derived from host cell into which several viral coded proteins are embedded. These include spike protein (S) which is a classic class I fusion protein important for entry of virus in susceptible cells, and envelope protein (E) which is important for virus assembly. Inside the envelope there is another membrane glycoprotein called matrix protein (M) which connects the virus envelope to nucleocapsid and is also important for virus assembly [2]. The crystal stricture of S protein of CoV-2 has been solved recently. The protein consists of two major domains- the 'receptor binding' domain and 'fusion machinery' domain. The aminoacid residues in receptor binding domain are highly variable, whereas the residues in fusion machinery domain are more conserved. Embedded deep inside the fusion machinery domain is the 'fusion peptide' which is important for fusion of viral membrane with host membrane [21].

During attachment and entry process, the S protein of CoV-2 interacts with cell surface receptor angiotensin-converting enzyme 2 (ACE2). Then cellular protease TMPRSS2 cleaves the S protein through a two stage cleavage process. The first cleavage event allows cellular protease TMPRSS2 to cleave the receptor binding domain of S protein thereby allowing its fusion subunit to form a hairpin which is embedded in membrane of target cell via fusion peptide of S protein. In second stage, the fusion peptide is activated. The fusion peptide consists of a sequence of hydrophobic aminoacids which allows it to be embedded in host cell membrane. Subsequent folding of this hairpin then allows the viral envelope to come in very close proximity to host cell membrane culminating in their fusion leading to release of viral nucleocapsid in to the cytoplasm of host cell [10]. This process occurs either at plasma membrane or through endosomal membrane. Previous studies have shown that there are six amino-acids which are present in receptor binding domain of S protein of SARS CoV which are critical for binding to cellular receptor ACE2. A recent study has now shown that five out of these six amino-acids are different in CoV-2 when compared to SARS-CoV. However the S protein of CoV-2 can still interact very efficiently with ACE2 receptor in host cell. In addition, the sequence analysis of S protein of CoV-2 has shown that it has also acquired a 'polybasic cleavage site' consisting of 12 aminoacids that has been predicted to enable its cleavage by other cellular proteases including furin [1]. In many other viruses including Influenza virus, the acquisition of similar polybasic site in viral envelope protein has been known to be associated with their increased transmissibility. It is not yet known if the presence of 'polybasic cleavage site' in S protein plays a similar role in CoV-2 pathogenesis.

The viral genome, transcription and translation

The analysis of CoV-2 genome has shown that it consists of 14 ORFs that encodes for at least 27 proteins. The largest ORF is called ORF1 which is present towards 5′ end of its genome and codes for a long polyprotein which is subsequently cleaved into multiple non-structural proteins (NSP). The remaining 13 ORFs are all present towards 3′ half of the genome and code for structural and accessory proteins [23]. The ORF1 is translated directly from the positive sense viral genomic RNA to make a polyprotein which also includes two proteases which cleave the polyprotein into the individual proteins. However, due to the presence of a stop codon in middle of ORF1, only the first half of polyprotein can be translated before translation terminates at this stop codon. This results in synthesis of partial polyprotein. Approximately half the time, the translation is able to proceed past the stop codon resulting in synthesis of full length polypprotein. This occurs due to the presence of a heptanucleotide slippery sequence around this stop codon, and a RNA pseudoknot structure downstream of the stop codon. These features of Coronavirus genome causes the ribosome to pause and allows it to occasionally slip out of frame resulting in frameshifting. This frameshifting event during translation allows the ribosome to read through the stop codon resulting in full length polyprotein to be synthesized approximately half of the time [8, 16].

The remaining 13 ORFs that code for structural and accessory proteins are translated from a nested set of sub-genomic RNAs of different sizes which have identical 3′ terminal end sequences. All these sub-genomic RNAa also have same 5′ end leader (L) sequence as in genomic RNA. A peculiar feature of Coronavirus genome is the presence of a series of conserved sequences called 'transcription regulatory sequences' (TRS) which are present at junction between each of these ORFs as well as at the 5′ end of the genomic RNA downstream of the leader (L) sequence. During the process of copying the genome when the polymerase reaches these TRSs, the structural properties of these sequences allows the polymerase to either continue copying the RNA or to jump from that TRS sequence to pair with TRS sequence down at 5′ end of the genomic RNA downstream of the leader (L) sequence. This occurs by a process called 'discontinuous transcription'. This results in generation of sub-genomic negative sense RNAs which are then used as template by same polymersase to generate positive sense sub-genomic mRNAs. Even though these sub-genomic mRNAs consist of multiple ORFs, all the ORFs do not get translated as polyproteins. Only the first ORF from the 5′ end of these sub-genomic RNAs get translated to synthesize individual viral protein, and the

remaining ORFs remains untranslated [16]. The polymerase jumping events associated with discontinuous transcription also results in high recombination rates in Coronaviruses.

The complex 'discontinuous transcription' mechanism of Coronaviruses is coordinated by a replicase complex consisting of different proteins sub-units responsible for polymerase, capping, and proof reading activities. The polymerase holoenzyme is made of nsp12, nsp7, and nsp8 subunits, and is capable of initiating de novo primer independent RNA synthesis. These subunits are also associated with nsp14 which plays key role in capping of viral subgenomic mRNAs. The nsp14 also has exonuclease activity that is critical for possible proofreading function in coronaviruses, a property which is missing in other RNA viruses. A variety of other viral processing proteins and activities are also associated with viral replicase complex [11]. The absence of proofreading activity in polymerases of RNA viruses has put a theoretical limit of approximately 30 Kb on genome size of RNA viruses due to constraints of evolution. However, the presence of exonuclease activity of nsp14 in Coronaviruses may be conferring the proof reading ability to RNA polymerase of these viruses thereby allowing them to cross the threshold and have genome sizes even above 30 Kb [15]. Studies performed on SARS-CoV have shown that mutation in its exonuclease gene results in 20-fold increase in mutation frequency as compared to the wild type virus [14]. The nsp14 protein is a bimodular protein which is composed of two domains: the ExoN domain for proofreading activity and N7-MTase domain for mRNA capping activity [13]. It has been shown that the exonuclease activity of nsp14 can efficiently excise the nucleoside analog chain terminator ribavirin thereby rendering this antiviral drug of no use against Coronaviruses [4].

Replication and transcription complex

Coronaviruses infection results in formation of multiple inter-connected double membrane vesicles in cells where viral replication and transcription complexes (RTC) are localized. These membranes are derived from endoplasmic reticulum (ER) and are continuous with rough ER [12]. The vesicles can protect viral genome from different anti-viral mechanism of cell, as well as from activity of other exonucleases present in the cytoplasm of infected cell. They also help the virus to concentrate the factors necessary for efficient transcription and replication of viral genome. The integral membrane proteins that are part of replicase complex play an important role in genesis of these vesicles. These include nsp3, nsp 4 and nsp6 which are directly involved in vesicle formation. Exogenous expression of nsp3 and nsp4 alone in a cell has been shown to be sufficient for formation of these vesicles indicating their crucial role in this process [7]. It has been suggested that this occurs due to the interaction between the luminal loops of these proteins which drives the membrane to fold and form vesicles.

Studies to investigate the viral and cellular proteome associated with Coronavirus replication and transcription complexes using proximity labeling approach have identified more than 500 host proteins that constitute the RTC microenvironment. These include the proteins involved in vesicular trafficking pathways, in ubiquitin-dependent and autophagy-related processes, and the translation initiation factors. The proteome also include most of non-structural proteins coded by virus as well. One of the viral coded proteins which are not detected in RTC microenvironment is nsp1 [20]. The nsp1 is a coronavirus pathogenicity factor that restricts host gene expression by directly interacting with 40S subunit of host ribosome and restricting translation of host mRNAs. It also mediates endonucleolytic cleavage of host mRNAs leading to extensive mRNA degradation in infected cells [24]. This results in hijacking of host translation machinery allowing it to be used exclusively for viral protein synthesis. The nsp1 does not act on viral transcripts as the leader sequence at 5′ end of viral mRNAs protects it from action of nsp1 [9]. In addition to nsp1, several other virus coded accessory proteins are not present in RTC microenvironment. These genes are usually dispensable for replication of virus in cell culture system, however they play an important role in virus host interaction in infected animals or humans [5]. CoV-2 and SARS CoV have a similar set of accessory genes with some differences among the interferon antagonists [23]. The function of most of the accessory genes of SARS CoV as well as CoV-2 has not been elucidated till date.

After the viral genome has replicated in vesicles, it gets assembled to generate infectious progeny virions. The viral assembly is initiated first by assembly of viral nucleocapsid protein with genomic RNA. This complex then associates with components of viral membrane including S, E and M proteins. The viral nucleocapid then buds in the membrane of ER-Golgi compartment. The viral particles are then released from cell by process similar to exocytosis.

Host immune response

The SARS and MERS coronaviruses induce very little interferon production in most cells when tested in cell culture system [17]. It is not known if CoV-2 also behaves similarly. There are a number of interferon antagonists that have been identified in SARS Coronavirus genome including non-structural protein nsp1, several accessory

factors, matrix protein and nucleocapsid protein. The virus is able to inhibit early interferon response of host against virus very efficiently. SARS pathogenesis has been shown to be linked to delayed IFN-1 signaling and subsequent immune toxicity. Experiments performed in IFN receptor knockout mice have shown that all of these mice can survive SARS-CoV infection whereas 80% of the wild type mice died indicating that IFN response is directly linked to pathogenesis of disease [3]. The delay in IFN response allows the virus to replicate to high titer after infecting the host. The delayed IFN response is subsequently unable to stop the virus infection; however it drives aberrant recruitment of pathogenic inflammatory monocyte-macrophages and activation of innate immune response leading to extensive cytotoxicity. A model proposed for SARS pathogenesis suggests that during acute virus infection, the cells in lung alveoli undergo rapid virus replication without any early IFN response. This result in inflammatory cells infiltration and release of pro-inflammatory cytokines from infected cells as well as from the infiltrating cells. These immune responses lead to acute lung injury and acute respiratory distress syndrome [6].

Retrospective studies on patients who recovered from SARS-CoV infection have shown that in these patients both the neutralizing antibody titers as well as the memory B cell responses against this virus were short lived. Though initially they mounted a robust antibody response against the virus, however the response was not sustained for long in most of these recovered people [18]. It is not clear how the immune response against CoV-2 is generated and for how long it will be sustained.

The ongoing pandemic

Travel restrictions, extensive contact tracing and quarantine of infected host allowed the SARS epidemic to be brought under control in approximately one year. The current CoV-2 pandemic is still ongoing and efforts are on to control it. Since the wild animal reservoirs for CoV-2 are yet to be identified, therefore it is possible that virus may again undergo animal to human transmission in future. Moreover, there has been widespread community transmission of CoV-2 in multiple countries all over the world. The problem has been further compounded by transmission of CoV-2 from asymptomatic carriers thereby making it extremely difficult to undertake contact tracing and breaking the chain of human to human transmission. A better understanding of molecular virology of COV-2 can help in formulating strategies for vaccine and anti-viral therapeutics development in immediate future.

References

1. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. Nat Med. 2020;26(4):450–2.
2. Barcena M, Oostergetel GT, Bartelink W, Faas FG, Verkleij A, Rottier PJ, Koster AJ, Bosch BJ. Cryo-electron tomography of mouse hepatitis virus: Insights into the structure of the coronavirion. Proc Natl Acad Sci U S A. 2009;106:582–7.
3. Channappanavar R, Fehr AR, Vijay R, Mack M, Zhao J, Meyerholz DK, Perlman S. Dysregulated Type I interferon and inflammatory monocyte-macrophage responses cause lethal pneumonia in SARS-CoV-infected mice. Cell Host Microbe. 2016;19:181–93.
4. Ferron F, Subissi L, Silveira De Morais AT, Le NTT, Sevajol M, Gluais L, Decroly E, Vonrhein C, Bricogne G, Canard B, Imbert I. Structural and molecular basis of mismatch correction and ribavirin excision from coronavirus RNA. Proc Natl Acad Sci U S A. 2018;115:E162–E171171.
5. Fung SY, Yuen KS, Ye ZW, Chan CP, Jin DY. A tug-of-war between severe acute respiratory syndrome coronavirus 2 and host antiviral defence: lessons from other pathogenic viruses. Emerg Microbes Infect. 2020;9:558–70.
6. Gralinski LE, Baric RS. Molecular pathology of emerging coronavirus infections. J Pathol. 2015;235:185–95.
7. Hagemeijer MC, Monastyrska I, Griffith J, van der Sluijs P, Voortman J, Henegouwen PM, Vonk AM, Rottier PJ, Reggiori F, de Haan CA. Membrane rearrangements mediated by coronavirus nonstructural proteins 3 and 4. Virology. 2014;458–459:125–35.
8. Irigoyen N, Firth AE, Jones JD, Chung BY, Siddell SG, Brierley I. High-resolution analysis of coronavirus gene expression by RNA sequencing and ribosome profiling. PLoS Pathog. 2016;12:e1005473.
9. Kamitani W, Narayanan K, Huang C, Lokugamage K, Ikegami T, Ito N, Kubo H, Makino S. Severe acute respiratory syndrome coronavirus nsp1 protein suppresses host gene expression by promoting host mRNA degradation. Proc Natl Acad Sci U S A. 2006;103:12885–90.
10. Kim IS, Jenni S, Stanifer ML, Roth E, Whelan SP, van Oijen AM, Harrison SC. Mechanism of membrane fusion induced by vesicular stomatitis virus G protein. Proc Natl Acad Sci U S A. 2017;114:E28–E36.
11. Kirchdoerfer RN, Ward AB. Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. Nat Commun. 2019;10:2342.
12. Knoops K, Kikkert M, Worm SH, Zevenhoven-Dobbe JC, van der Meer Y, Koster AJ, Mommaas AM, Snijder EJ. SARS-coronavirus replication is supported by a reticulovesicular network of modified endoplasmic reticulum. PLoS Biol. 2008;6:e226.
13. Ogando NS, Ferron F, Decroly E, Canard B, Posthuma CC, Snijder EJ. The curious case of the nidovirus exoribonuclease: its role in RNA synthesis and replication fidelity. Front Microbiol. 2019;10:1813.
14. Smith EC, Denison MR. Coronaviruses as DNA wannabes: a new model for the regulation of RNA virus replication fidelity. PLoS Pathog. 2013;9:e1003760.
15. Smith EC, Denison MR. Implications of altered replication fidelity on the evolution and pathogenesis of coronaviruses. Curr Opin Virol. 2012;2:519–24.
16. Sola I, Almazan F, Zuniga S, Enjuanes L. Continuous and discontinuous RNA synthesis in coronaviruses. Annu Rev Virol. 2015;2:265–88.
17. Spiegel M, Pichlmair A, Martinez-Sobrido L, Cros J, Garcia-Sastre A, Haller O, Weber F. Inhibition of Beta interferon

induction by severe acute respiratory syndrome coronavirus suggests a two-step model for activation of interferon regulatory factor 3. J Virol. 2005;79:2079–86.

18. Tang F, Quan Y, Xin ZT, Wrammert J, Ma MJ, Lv H, Wang TB, Yang H, Richardus JH, Liu W, Cao WC. Lack of peripheral memory B cell responses in recovered patients with severe acute respiratory syndrome: a six-year follow-up study. J Immunol. 2011;186:7264–8.

19. Totura AL, Bavari S. Broad-spectrum coronavirus antiviral drug discovery. Expert Opin Drug Discov. 2019;14:397–412.

20. V'Kovski P, Gerber M, Kelly J, Pfaender S, Ebert N, Braga Lagache S, Simillion C, Portmann J, Stalder H, Gaschen V, Bruggmann R, Stoffel MH, Heller M, Dijkman R, Thiel V. Determination of host proteins composing the microenvironment of coronavirus replicase complexes by proximity-labeling. Elife. 2019;8:e42037.

21. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. Cell. 2020;181:281.

22. Woo PC, Lau SK, Lam CS, Lau CC, Tsang AK, Lau JH, Bai R, Teng JL, Tsang CC, Wang M, Zheng BJ, Chan KH, Yuen KY. Discovery of seven novel Mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus. J Virol. 2012;86:3995–4008.

23. Wu A, Peng Y, Huang B, Ding X, Wang X, Niu P, Meng J, Zhu Z, Zhang Z, Wang J, Sheng J, Quan L, Xia Z, Tan W, Cheng G, Jiang T. Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. Cell Host Microbe. 2020;27:325–8.

24. Zust R, Cervantes-Barragan L, Kuri T, Blakqori G, Weber F, Ludewig B, Thiel V. Coronavirus non-structural protein 1 is a major pathogenicity factor: implications for the rational design of coronavirus vaccines. PLoS Pathog. 2007;3:e109.