The Novel Coronavirus SARS-CoV-2

The Novel Coronavirus SARS-CoV-2:

An Overview

By

Thangam Menon

Cambridge
Scholars
Publishing

The Novel Coronavirus SARS-CoV-2: An Overview

By Thangam Menon

This book first published 2022

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN (10): 1-5275-8427-5
ISBN (13): 978-1-5275-8427-3

TABLE OF CONTENTS

CHAPTER 1

HISTORY AND TAXONOMY OF CORONAVIRUSES

Introduction

Coronaviruses (CoVs) are RNA viruses that infect human beings and animals. They were first described in the 1960s in embryonic tracheal organ cultures obtained from a patient with a common cold. There are four genera of human and animal coronaviruses (α-, β-, γ- and δ-CoV). The viruses HCoV-229E, HCoV-OC43, HCoV-NL63 and HCoV-HKU-1 circulate in the human population. SARSCoV-2 is classified under the genus Betacoronavirus.

History

The first coronavirus was described in 1965 by Tyrrell and Bynoe1 who cultured it from the respiratory tract of an adult with a common cold. It was proved that the cultures contained an infectious agent, because symptoms of the common cold were seen in a significant proportion of subjects who were exposed to the culture. It was cultivated in human embryonic tracheal organ cultures but could not be grown in tissue culture. The virus was named B814.

Discovery of OC43 and 229E

During the same period, in 1966, Hamre and Procknow[2] isolated a virus from medical students with upper respiratory tract infections and called it 229E. It was later adapted to grow in WI-38 lung cell lines. This strain of the virus produced common cold symptoms, including headache, sneezing,

malaise, sore throat, and less frequently fever and cough (seen only in 10~20% cases). While most patients exhibited only mild symptoms, some patients who were immunocompromised presented with severe lower respiratory tract infection. Both B814 and 229E were ether-sensitive and therefore assumed to have a lipid envelope. They were found to be unrelated to any known myxovirus or paramyxoviruses.

McIntosh et al. in 1967 grew similar viruses from nasopharyngeal washings collected from adults with upper respiratory tract infections by inoculation into human embryonic tracheal organ cultures.[3] They were named 'OC' because they were grown in organ cultures. The clinical features of the HCoV-OC43 infection were similar to those caused by HCoV-229E and were generally indistinguishable from infections caused by other respiratory viruses, such as influenza viruses and rhinoviruses. Organ culture harvests of OC43 when examined by electron microscopy showed that these viruses exhibited an unusual morphology closely resembling that of avian infectious bronchitis virus (IBV) and the two other ether-labile agents: strain 229E, described by Hamre and Procknow, and strain B814, described by Tyrrell. It appeared that a new group of viruses was emerging, and members of the group could infect the respiratory tract of birds and humans. Although one of them, strain 229E, could grow and produce cytopathic effects in tissue culture, strains B814 and OC43 could only be grown in the laboratory in human ciliated respiratory tract organ cultures. Though they resembled the myxoviruses in size, ether lability, and nucleic acid type, they were morphologically different in that, the spikes were club-shaped and more widely spaced, in contrast to the closely spaced narrow surface projections of the myxo- and paramyxoviruses. The OC43 and 229E strains were the only human coronaviruses studied in the next 30 years. They showed periodicity, with large epidemics occurring at 2- to 3-year intervals.[4] Strain 229E caused epidemics throughout the United States, whereas strain OC43 often caused localized outbreaks. Infections were more common in children and reinfections were frequent. Serological tests showed that there were other related strains of coronaviruses causing infections which were not identical to OC43 and 229E. In fact, serologically, strain B814, was found to be different from OC43 and 229E.

Human coronavirus OC43 (HCoV-OC43) is the coronavirus commonly associated with human infections. There are four genotypes of HCoV-OC43, A to D. Genotype D is said to have evolved from recombination. The earliest ancestor of all the genotypes was traced back to the 1950's using molecular clock analysis. The evolution of different genotypes occurred over several years, leading to the emergence of the new recombinant genotype D in 2004. This genotype is associated with severe respiratory disease, particularly in the elderly.[5]

SARS, NL63, HKU1, and MERS

SARS-CoV, human coronavirus NL63, and HKU1 were first described in 2003, 2004, and 2005, respectively.[6] The virus that caused severe acute respiratory syndrome (SARS), SARS-CoV, caused outbreaks in 2002-2003 and affected more than 8000 people.[7] The outbreak was first identified in Foshan, Guangdong, China, in November 2002. The disease had a mortality rate of about 10%. The infection spread to 29 countries. Though the epidemic was contained by July 2003, cases continued to be reported until May 2004.

The human coronavirus NL63 (HCoV-NL63), was reported independently by two groups in 2004.[8,9] It was the fourth human coronavirus to be discovered and was isolated from a 7-month-old child who presented with bronchiolitis, conjunctivitis, and fever in January 2003 in an Amsterdam hospital.[10] A nasopharyngeal aspirate specimen collected from the patient (sample NL63) which was inoculated onto tertiary monkey kidney cells showed diffuse cytopathic effects and cell detachment. Genomic studies showed that the virus (HCoV-NL63) was a new member of the Alphacoronaviridae. When it was cultivated in the LLC-MK2 monkey kidney cell line, it caused enlargement and rounding of cells. It was identified by a new method for virus discovery based on the cDNA-amplified restriction fragment–length polymorphism technique (cDNA-AFLP). It is distributed throughout the world and has been frequently isolated from patients with respiratory tract infections, particularly children, the elderly, and the immunosuppressed, and it accounts for approximately 4.7% of common respiratory diseases.[11]

In 2005, HCoV-HKU1 was isolated in Hong Kong from a 71-year-old man hospitalized with pneumonia and bronchiolitis who had recently returned from Shenzhen, China.[12]

The virus which caused the Middle East Respiratory Syndrome (MERS) caused an epidemic in the Arabian Peninsula with sporadic cases in other parts of the world.[13] The outbreak first occurred in Saudi Arabia in 2012, after which it spread to more than 26 countries. A secondary outbreak occurred in South Korea in 2015. MERS was more severe and had a higher mortality rate (over 30%) when compared to SARS. Clinically MERS was similar to SARS, and was characterized by progressive acute pneumonia however, many patients with MERS also developed acute renal failure.

SARS-CoV-2

SARS-CoV-2 is the causative agent of the current pandemic that began in Wuhan, Hubei Province, China, in December 2019, where groups of patients presented with pneumonia.[14] It was declared an international public health emergency by the World Health Organization and the disease was named COVID-19. As of 1 March 2022, 437,476,521 cases have been confirmed worldwide. The case fatality differs in different parts of the world. SARS-CoV-2 was similar to SARS-CoV and MERS-CoV, in that it caused severe respiratory infection, with common presentations being fever, cough, and dyspnea. Some patients may progress to pneumonia and acute respiratory distress syndrome. Patients may also have gastrointestinal symptoms. SARS-CoV-2 is more transmissible than SARS-CoV and MERS-CoV, but less pathogenic. Asymptomatic infections are common with SARS-CoV-2 and such patients, who do not exhibit symptoms, are responsible for the rapid spread of the disease in the community.

The symptoms of COVID-19 may vary between the severity of SARS-CoV and the less severe community-acquired HCoVs (i.e., HCoV-229E, HCoV-OC43, HCoV-HKU1 and HCoV-NL63). On the one hand, the presentation may be nonspecific, mild, or even without symptoms. On the other hand, it may present as a severe disease, similar to SARS-CoV infection. SARS-CoV-2 appeared to have a transmission rate similar to that of community-acquired HCoVs, but whether the transmissibility of SARS-CoV-2

decreases after passages in humans as in the case of SARS-CoV and MERS-CoV remains to be seen.

NL63, 229E, OC43, and HKU1 are all found throughout the world and cause mild respiratory diseases. All four are community-acquired HCoVs that are well adapted to humans and rarely mutate into pathogenic strains. The betacoronaviruses are known to cause severe disease and fatalities. HKU1 and OC43 are betacoronaviruses, but are associated with milder symptoms similar to the disease caused by the alphacoronaviruses NL63 and 229E.

Taxonomy

Coronaviruses are members of the subfamily Coronavirinae under the family Coronaviridae and the order Nidovirales. They are enveloped viruses with the largest known RNA genomes, about 30 kb in length (varying between 26 kb and 32 kb). The virus is spherical in shape with a core shell and surface projections resembling a crown, hence termed coronavirus (corona is the Latin word for crown).

The family Coronaviridae was represented by a single genus Coronavirus for over two decades but later it was observed that the genus Torovirus shared similar characteristics, and hence in 1993**,** the International Committee for the Taxonomy of Viruses (ICTV) formally expanded the Coronaviridae to include Torovirus**.**[15]

Based on phylogenetic relationships the subfamily Coronavirinae, which contains both animal and human viruses, is divided into the four genera – Alphacoronavirus, Betacoronavirus, Gammacoronavirus and Deltacoronavirus (α-, β-, γ- and δ-CoV).

The four coronaviruses HCoV-229E, HCoV-OC43, HCoV-NL63 and HCoV-HKU-1 circulate in the human population. HCoV-229E and HCoV-NL63 are alphacoronaviruses; HCoV-OC43 and HCoV-HKU-1 are betacoronaviruses. SARSCoV-2 is classified under the genus Betacoronavirus, the subgenus Sarbecovirus, and is the seventh CoV that has been found to cause infections in humans. Other important beta-coronaviruses which originate in animals and cause serious disease in human beings are MERS-CoV and SARS-CoV.

Origin of coronaviruses

The alpha- and beta-coronaviruses originate from mammals, such as bats, while the gamma- and delta-viruses originate from pigs and birds.[16] The alphacoronaviruses generally cause asymptomatic or mild infections.

Surveillance of coronaviruses in wild animals has shown that there is a great diversity of coronaviruses in bat and avian species, suggesting that these animals are natural reservoirs of viruses.[17] Molecular clock dating analyses of coronaviruses based on the RNA-dependent RNA polymerase (RdRp) gene suggest that the most recent common ancestor of these viruses existed around 10,000 years ago.[18] Population dynamics studies have revealed that various coronaviruses are endemic in different bat species, with repeated introductions to other animals and occasionally infecting other species. There are studies that suggest that bats may be the natural hosts of all currently known coronavirus lineages and that all coronaviruses recognized in other species were derived from viruses residing in bats. Bat coronaviruses are said to be older than other animal coronaviruses.[19]

Coronaviruses in animals

Animal coronaviruses were known before human coronaviruses. The first animal virus to be discovered was the Avian Infectious Bronchitis Virus (IBV) in 1937, which was isolated from a flock of diseased chicken. It causes respiratory infections in commercial poultry and mortality rates are high in unvaccinated flocks. Outbreaks of infectious bronchitis have declined due to the extensive use of vaccines; however, new variants of the virus can emerge resulting in vaccine failures.[20] After the discovery of IBV, other related viruses were discovered in rodents and other domestic animals.

There are several animal coronaviruses known today which were initially classified into three groups (Table 1). Human coronaviruses HCoV-229E and HCoV-OC43 were included in the first and second groups, respectively. These groups are now classified into four genera, Alphacoronavirus, Betacoronavirus, Gammacoronavirus, and Deltacoronavirus. Mammalian CoVs are included in the first two genera; HCoV-229E and HCoV-OC43

are included in the first and second genera respectively, while avian CoVs are included in the third and fourth genera.

Table 1: Animal coronaviruses

Group	Virus
1	Porcine epidemic diarrhoea virus (PEDV)
1	Porcine transmissible gastroenteritis virus (TGEV)
1	Feline coronavirus and feline infectious 1 peritonitis virus (FCV)
1	Canine coronavirus (CCov)
2	Bovine coronavirus (BCoV)
2	Porcine hemagglutinating encephalomyelitis virus (HEV)
2	Rat coronavirus (RtCoV)
2	Equine coronavirus (ECoV)
3	Avian infectious bronchitis virus (IBV)
3	Turkey coronavirus (TCV)
3	Pheasant coronavirus

Coronaviruses have restricted host ranges. They infect only their natural hosts and others who are closely related. Occasionally, cross-species infection may occur showing some flexibility of host range. Animal coronaviruses have restricted tissue tropism. The ability to cause infection in a particular tissue depends on the distribution of cellular receptors and other related molecules that regulate virus entry. The replication mechanisms of the virus allow for the generation of variants, which may be able to cause infection in different species. Studies of such interspecies infections are important in the control of emerging zoonotic infections and transdisciplinary research of this nature is now an important part of the One Health concept.[21]

CHAPTER 2

ORIGIN OF SARS-COV-2

Introduction

Coronaviruses have been in existence for a very long time, but were not given much importance until recently. Human coronaviruses have animal origins. Zoonotic infections are a major public health problem. Host switching is an important phenomenon that helps a pathogen spread from animals to humans.

Host switching

Host switching is the mechanism by which many of the well-known microorganisms such as influenza, cholera, smallpox, malaria, dengue etc have existed for centuries. A typical example is the influenza virus. The 1918 pandemic was caused by enzootic viruses in birds that were transmitted to humans from their natural reservoir. This was proven by genetic studies that compared the avian viruses with the 1918 virus and its descendants, which have caused subsequent epidemics and pandemics. Avian influenza viruses are also known to have host-switched into other animals such as dogs, pigs, horses, etc. When RNA viruses are transmitted to different hosts, they exhibit genetic differences and are often not identical virions. A complex set of related but genetically distinct viruses may be present in the host at any point of time.

A betacoronavirus named SARS-CoV emerged almost two decades ago in China, spread to 29 countries, and nearly caused a pandemic. It was finally controlled with the use of aggressive public health measures and has not reappeared since then. In 2012, a related betacoronavirus named Middle East respiratory syndrome coronavirus (MERS-CoV), emerged to cause severe human infections with high fatality rates. However, transmission rates were low and limited to the Middle East. The intermediary host was

the dromedary camel, and the infections were limited to geographic areas where these hosts were found in large numbers. In 2016, an alphacoronavirus emerged in China that caused a new disease in pigs, and it was termed the swine acute diarrhoea syndrome coronavirus (SADS-CoV). In November 2019, the new coronavirus SARS-CoV-2 emerged. It had never been found in humans or animals prior to 2019. It became the third fatal bat virus–associated human disease appearing in the last two decades.

Some of the coronaviruses such as HCoV-OC43 and HKU1 appear to have originated from rodents,[1] whereas others such as SARS-CoV, MERS-CoV, HCoV-NL63 and HCoV-229E are considered to have originated in bats.

The betacoronaviruses SARS-CoV, MERS-CoV, and SARS-CoV-2 are all closely related; they are SARS-like or MERS-like and belong to two adjacent phylogenetic groups: sarbecovirus and merbecovirus. The two SARS viruses, SADS-CoV, are known to have originated from horseshoe bats (genus Rhinolophus).

The genetic similarity between SARS-CoV-2 and a bat coronavirus of the subgenus Sarbecovirus has been confirmed by several researchers. The human strain has 96.2% similarity to a coronavirus (RaTG13) isolated from the bat *Rhinolophus affinis*. Phylogenetic analysis of the full-length genome and the gene sequences of the RNA-dependent RNA polymerase and spike protein has shown that RaTG13 is closely related to SARS-CoV-2, suggesting that SARS-CoV-2 may have originated in bats.[2]

Animal reservoirs

Bats are important reservoirs of animal coronaviruses. They are of different genera and species and are distributed throughout the world. Bats of one species may co-roost with bats of another species, giving rise to gene exchanges, interspecies transmission, and evolution of new recombinants. However, genetic studies on coronaviruses in bats appear to be limited, and less than 10% of the CoV sequences in GenBank are from bats. A study on the diversity of CoV in different hosts (animals and humans) as well as "high-risk" interfaces from 20 countries, proved that bats were the major reservoirs and 91% of phylogenetic sequences identified were from bats.[3]

Hence the geographic distribution of bats was related to CoV diversity globally and the possibility of host-switching and zoonotic emergence was known even before the occurrence of this pandemic. Human-animal interactions are potential risk factors. Bat-watching tourism, wet markets, changes in land management practices with increased natural resource exploitation have all increased the threat of zoonotic infections.[4] Scientists had warned that dangerous viruses may re-emerge, and risk mapping studies have indicated the possibility of more coronavirus infection outbreaks in the future. It was therefore important to identify risk factors and enhance preventive measures.[5]

The first patient who had COVID-19 reported no exposure to the seafood market. Like SARS-CoV and MERS-CoV, which have intermediate hosts, such as civets or camels, it was assumed that the SARS-CoV-2 virus would also have an intermediate host; however, the specific route of transmission of SARS-CoV-2 from natural reservoirs to humans remains unclear. The Malayan pangolins (*Manis javanica*) contain coronaviruses similar to SARS-CoV-2. These coronaviruses belong to multiple sublineages, and the receptor binding domain in one of the lineages is very similar to SARS-CoV-2.[6]

A study that reported autopsy findings in dead pangolins showed pulmonary fibrosis of the lungs and evidence of a SARS-CoV-2-like CoV. This virus, named Pangolin-CoV, was 91.02% identical to SARS-CoV-2 and 90.55% identical to BatCoV RaTG13. The amino acid composition of the receptor binding domain of the spike protein, in the strain isolated from a pangolin (GD410721) and SARS CoV-2, was similar, with five key amino acid residues involved in the interaction with human angiotensin-converting enzyme receptors being identical. The Pangolin-CoV, thus was found to be the most closely related coronavirus to SARS-CoV- 2.[7] Thus, pangolins as possible hosts may play an important role in the zoonotic transmission of SARS-CoV- 2.

Genome sequence data

Genetic studies carried out in the early period of the pandemic indicated that viruses from countries other than China appeared to be more diverse,

indicating increased mutation rates in these strains.[8] Data from full genomes of strains available in the online database (https://www.gisaid.org/), when analyzed, identified three variants based on differences in amino acid composition. They were named A, B, and C types. Type A, the ancestral type was very similar to the bat outgroup coronavirus. Type B, the 'Wuhan type', was the most common type in East Asia, and seemed to be adapted to the population in that region. Types A and C were found in Europe and America. Only mutated strains of the B type, occurred in regions beyond East Asia.[9]

Initial dating studies have used the sequence divergence of SARS-CoV-2 and closely related coronaviruses isolated from bats and pangolins. These studies dated the most recent common ancestor (MRCA) to approximately 50 years ago. The MRCA represents the point from where all the sampled cases had descended. A study conducted in March 2020 that analyzed 176 full-length genomes of SARS-CoV-2 showed limited diversity indicating a recent common ancestor for all these viruses. The phylogenetic tree of the recently isolated strains showed greater divergence, as is characteristic of RNA viruses. Estimating the date of MRCA in this study indicated a period between early and mid-December 2019 and this appeared to match the first reported cases in Wuhan.[10] A recent study using a phylogenetic tree with more than 80,000 full-length SARS-CoV-2 genomes has dated MRCA as August 2019 with an evolutionary rate of 0.05526 mutations/genome/day.[11]

The current pandemic has been driven entirely by human-to-human transmission. The virus may continue to exist in one or more animal species, but is unlikely to affect the current pandemic.[12]

Mutations

As the current pandemic evolved, surveillance improved because genomic sequences of the virus were being shared at an unprecedented rate. Sequencing studies also enabled the detection of variants and the presence of specific mutations which changed the properties of the virus. Most mutations were neutral or mildly deleterious. Others which were highly deleterious were rapidly purged. Mutations that improved the fitness of the virus were detected early in the course of the pandemic, within a few months

of the detection of the virus in the human population. When the evolution of the virus was studied in the first year of the pandemic, approximately two mutations per month were reported throughout the world.[13]

SARS-CoV-2 viruses were initially classified into two major types (L and S) based on single nucleotide polymorphisms (SNPs) at location 8,782 and 28,144. The S type causes milder infections and is the ancestral type.[14] The L type was more prevalent than the S type in the early period of the pandemic but decreased subsequently. The two lineages appeared to have faced different selection procedures and the L lineage had a higher number of mutations.

The mutation N439K, in the receptor binding motif, was observed in March 2020. This was designated as the B.1.141 lineage but quickly became extinct, and soon another lineage (B.1.258) acquired the same mutation. This mutation resulted in an enhanced receptor binding efficacy of the virus and a reduction in the neutralizing activity of antibodies. The mutation in the spike protein that caused the amino acid change D614G was observed in April 2020. It was advantageous for the virus as it increased transmissibility. Similarly, another mutation in the receptor binding motif, Y453F, also increased the efficacy of receptor binding and was associated with the B.1.1.298 lineage. This lineage also had a deletion in the amino terminal domain (Δ69–70), changing its conformation and increasing the infectivity of the virus.[15]

Mutations in residue E484 (E484A, E484D, E484G and E484K) in the receptor binding domain led to drastic reductions in the effect of neutralizing antibodies, including combination monoclonal antibodies (mAb) used in therapy. E484K was identified as an escape mutation which occurred due to exposure to mAb; others being K444E, G446V, L452R, and F490S. S477G, S477N, and S477R were escape mutations at position 477 of the spike protein.[16]

Tracing the origin based on phylogenetic analysis

It is likely that a progenitor of SARS-CoV-2 may have been circulating in the population causing asymptomatic infections; the mutations that it

acquired over this period of time would have helped it to maintain transmission in humans. Several short chains of human-to-human transmission occurring over an extended period can give rise to widespread disease. SARS-CoV-2 is different from other coronaviruses such as SARS-CoV and MERS-CoV because it has a much more efficient human-to-human transmission, which is the reason for the rapid spread of the disease across countries.

There is still no clarity on how COVID-19 emerged. It appears that many bat viruses are functionally preadapted because of similarities between the angiotensin-converting enzyme 2 (ACE2) receptors on the cells of numerous mammals such as human beings, bats, minks, etc. This phenomenon may have occurred in nature and may continue to occur to cause further pandemics. After the first report of the disease from China in December 2019, the whole genome sequence of SARS-CoV-2 was published in January 2020 (http://virological.org/t/novel-2019-coronavirus-genome/319; Wuhan-Hu-1, GenBank accession No. MN908947). By mid 2021, over 1.2 million coronavirus genome sequences from 172 countries were shared on an online data platform.[17] Genomic data has helped in understanding the origin of the virus and its transmissibility in the human population. Sharing of sequence data is crucial for understanding the occurrence and spread of SARS-CoV-2 variants across the globe, and to understand the re-emergence of other pathogens, including those causing zoonotic infections.

Human coronaviruses have been linked to animal trading in the past. All four human endemic human coronaviruses, HCoV-OC43, HCoV-HKU1, HCoV-229E, and HCoV-NL63, are known to have zoonotic origins. Bat viruses are the closest known relatives of SARS-CoV-2; however, there is a significant distance between the place where bat viruses were isolated (Yunnan) and the location of the first human case, indicating that the path of travel of the virus and intermediate hosts in the chain is still not known. The reason why so far, no bat reservoir or intermediate animal host has been identified is because a sufficient number of animal species and populations have not been sampled. Moreover, most of the spill overs from animals to human beings do not result in major outbreaks and index case infections do not result in sustained transmission.[18] In fact, the animal origins of several

other viruses, including the coronaviruses HCoV-HKU1 and HCoV-NL63, have not yet been identified.

The bat virus RaTG13 was found to have a genetic distance of ~4% (~1,150 mutations) to the reference sequence of SARS-CoV-2, which indicates several decades of evolutionary divergence.[19] Because of widespread genomic recombination it is often difficult to assign the closest matching virus; other bat viruses (e.g., RmYN02, RpYN06, and PrC31) appear to be closer in genome relatedness (particularly ORF1ab) and may share a more recent common ancestor with SARS-CoV-2. Therefore, it is unlikely that RaTG13 is the progenitor of SARS-CoV-2. The furin cleavage site (FCS) in the SARS-CoV-2 spike protein consists of multiple basic amino acids that promote fusion. The furin cleavage site is absent from the closest known relatives of SARS-CoV-2. This may be due to recombination occurring in these viruses or because of insufficient studies on the lineage. Furin cleavage sites are seen in the spike proteins of other coronaviruses including the endemic human coronaviruses HCoV-OC43 and HCoV-HKU1.

There has been much debate as to whether this virus could have escaped from the laboratory. However, its low pathogenicity in laboratory animals and the absence of genomic markers associated with rodent adaptation make this unlikely. Further, there is insufficient evidence to show that the virus could have infected laboratory workers while they were involved in gain-of-function experiments.[20]

CHAPTER 3

STRUCTURE OF SARS-COV-2, GENOME ORGANIZATION AND REPLICATION

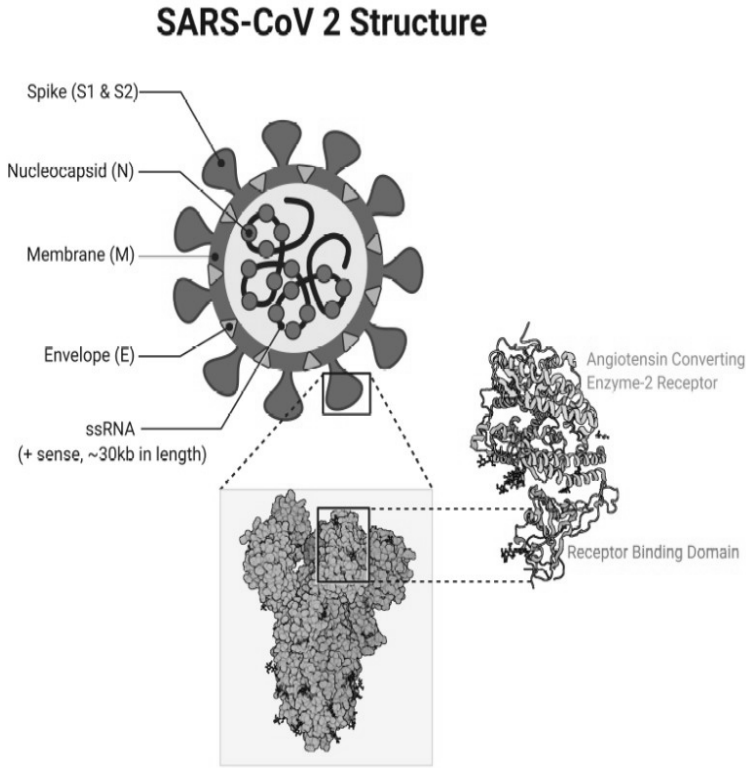
Introduction

SARS-CoV-2 belongs to the Coronavirus family. They are large, spherical, pleomorphic, enveloped particles, 80-120 nm in size, with a prominent fringe of petal-shaped surface projections, which are 20 nm long. These spikes are composed of a trimeric glycoprotein known as the spike protein (S), which projects over the envelope, giving it the appearance of a crown. They are positive stranded RNA viruses and replicate their large genomes in the cell cytoplasm, transforming the endoplasmic reticulum of the host cell into organelles suitable for viral replication. A special viral transmembrane protein plays an important role in replication. Mutations of the virus were rare in the early phase of the pandemic but were observed later giving rise to variants with different antigenic properties.

Structure of the virus

The virus has a helical nucleocapsid within the envelope, which consists of the nucleocapsid protein (N), bound to the single-stranded positive sense RNA, which is about 30 kb in size. The gene encoding the surface glycoprotein is 4.4 kb in size. The viral envelope has a lipid bilayer seen as a double shell under the electron microscope where the membrane (M), envelope (E) and spike (S) proteins are anchored. N, M, E, and S are the main structural proteins (Fig. 1). Apart from this, the virus also has 16 nonstructural proteins.

Fig. 1 -SARS- CoV 2 structure.



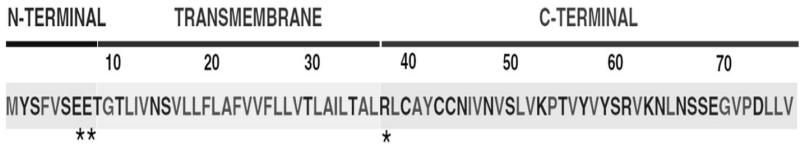
Contributed by Rohan Bir Singh, MD; Made with Biorender.com
From: Features, Evaluation and Treatment Coronavirus (COVID-19)
StatPearls Publishing; 2020 Jan.

See centrefold for this image in colour.

The E protein is the envelope protein. It is a small, membrane protein that plays a role in assembly, budding, and envelope formation. It can also function as an ion-channelling viroporin.[1] It is 76–109 amino acids in length and 8.4 to 12 kDa in size. It has a short hydrophilic amino terminus consisting of 7–12 amino acids, a large hydrophobic transmembrane domain (TMD) of 25 amino acids, and a long hydrophilic carboxyl

terminus. A large portion of the TMD consists of the two nonpolar, neutral amino acids, valine, and leucine, making it strongly hydrophobic (Fig. 2).

Fig. 2 – Envelope protein of Coronavirus



Amino Acid Sequence and Domains of the SARS-CoV E Protein. The SARS-CoV E protein consists of three domains, i.e., the amino (N)-terminal domain, the transmembrane domain (TMD), and the carboxy (C)-terminal domain. The amino acid properties are indicated: hydrophobic (red), hydrophilic (blue), polar, charged (asterisks).

See centrefold for this image in colour.

The M protein is the membrane protein. It is the most abundant structural protein and is responsible for the shape of the envelope.[2] It is 25 kDa in size and is shaped like a triple helix. It interacts with all major structural proteins and organizes the assembly of the virus. M and E are two structural proteins that play a role in the production and release of virus-like particles (VLP). The sequence of the M protein reveals that the M polypeptides comprise 225-230 amino acids. The M protein binds to the N protein and stabilises the nucleocapsid and subsequently completes the viral assembly.[3]

The N protein is the nucleocapsid protein. It is a 50 to 60-kDa phosphoprotein and functions primarily to bind to the CoV RNA genome, making up the nucleocapsid.[4] The N proteins are 377 to 455 amino acids in length, and are highly basic. Although N is largely involved in processes related to the viral genome, it is considered multifunctional because it is also involved in other aspects of the CoV replication cycle and the host's cellular response to viral infection.[5] Its location in the endoplasmic reticulum suggests that it may be necessary for complete virion formation.

The S protein is the spike protein. It mediates the attachment of the virus to the host cell and is a large protein with a length of 1,225 amino acids. Two important biological activities of the S protein are induction of membrane fusion and receptor binding. After the virus binds to the surface receptors, fusion occurs between the viral and host cell membranes, which facilitates viral entry into the host cell. Expression of the S protein at the cell membrane can mediate cell-to-cell fusion between infected cells and adjacent uninfected cells. Such giant, multinucleated cells, or syncytia, allows the virus to spread between cells, without coming into contact with virus-neutralising antibodies. The spike protein is antigenic, and the host's immune system responds to the virus by producing neutralizing antibodies to the S protein, which are protective.

The virus is sensitive to ultraviolet rays and heat. It can be effectively inactivated by lipid solvents including ether, ethanol, chlorine-containing disinfectant, peroxyacetic acid, and chloroform.

The spike protein

ACE2 acts as a receptor for SARS-CoV-2 and mediates the entry of the virus into the host cell. The spike protein of the virus plays an important role in recognising this receptor.

The coronavirus spike contains three segments: a large ectodomain, a transmembrane anchor, and an intracellular tail. There are two main functional domains, S1 and S2. S1 has two domains, the amino N-terminal domain and the carboxy C-terminal domain. The N-terminal S1 subunit contains the receptor binding domain (RBD) that interacts with the ACE2 receptor, and the C-terminal S2 subunit, which contains the fusion peptide responsible for the fusion of virus and host cell membrane. For effective function of the spike protein, the two subunits must be separated. A polybasic cleavage site is present at the S1/S2 boundary (PRRAR, amino acids 681 to 685) and is responsible for increased transmissibility. Multiple proteases including serine proteases such as the transmembrane protease, serine 2 (TMPRSS2), and the cathepsin protease are responsible for the cleavage.

Electron microscopy studies have revealed that the spike is clove-shaped; it is glycosylated with three receptor binding heads (S1) and a trimeric stalk (S2). S2 is a six-helix bundle with exposed fusion peptides. S2 polypeptide is more conserved than S1 and there may be regions showing up to 30% amino acid identity between the S2 polypeptides of different antigenic groups of coronaviruses. The S1 subunit is on the surface of the S glycoprotein and is a polypeptide composed of 672 amino acids (residues 14–685). It shows considerable variation, and when S1 sequences of different strains of a species are compared, there are many hypervariable regions. S1 has four domains: an N-terminal domain (NTD), a C-terminal domain (CTD), and two subdomains (SD1 and SD2). The CTD is also known as the receptor-binding domain. This is the region of the spike protein which recognizes the ACE2 receptor and is a short immunogenic fragment. This region has a core entity and a receptor binding motif that binds to ACE2 on the cell surface. ACE2 has a claw-like structure and has two virus-binding hot spots. The amino acids in this region determine the host range of the virus. There are specific amino acids at positions 442,472,479,480 and 487 that increase the ability of the virus to bind human ACE2. These amino acids are different for the virus that binds to receptors in animals such as civets. The receptor binding domain is a large molecule and has 144 amino acid residues distributed in varying ways. All amino acids are present except methionine. Tyrosine and valine have the highest distribution frequencies (13 each) followed by serine (12). The amino acids with the lowest frequencies are glutamic acid, tryptophan, and histidine (frequencies of 3, 2, and 1 respectively). There are cysteine residues at five positions (336, 361, 379, 391, 432); phenylalanine at position 486 on the flexible loop is important because it fits into a deep pocket in the ACE2.[6]

The transmembrane S2 subunit is composed of 588 amino acids (residues 686-1273). It consists of five regions, a hydrophobic fusion peptide (FP), two heptad repeats (HR1 and HR2), a transmembrane domain (TM) and a cytoplasmic tail (CT). These regions are arranged as FP-HR1-HR2-TM-CT.[7]

The genome of the virus

The genome of the virus is a positive-sense, single-stranded RNA genome, with a 5′-cap structure and 3′-poly-A tail. It is the largest viral RNA genome known. The G + C content varies from 32% to 43%. The viral RNA functions as an mRNA. It contains approximately 7-10 functional genes, 4 or 5 of which encode structural proteins.[8] Both the 5 ' and 3′ ends of the coronavirus genomes contain short untranslated regions. The genome size of the SARS-CoV-2 varies from 29.8 kb to 29.9 kb.

At least two thirds of the genome capacity is occupied by the two large open reading frames, ORF1ab and ORF1a, which together constitute the replicase gene located at the 5′ terminus of the genome and encode the proteins pp1ab and pp1a. ORF1a encoded proteases initiate the proteolytic cleavage of pp1a and pp1ab to release 16 functional nonstructural proteins (nsp1-16) that have specific functions such as RNA polymerase (nsp12) and helicase (nsp13), some unique domains involved in mRNA capping (nsp14, nsp16) and fidelity control (nsp14). Several smaller subunits (nsp7–nsp10) act as crucial cofactors of these enzymes contributing to the “nsp interactome”.[9] While two-thirds of the 5′ end of the genome comprises ORF1ab, the remaining one third of the genome at the 3′ end consists of genes encoding structural proteins: spike (S), envelope (E), membrane (M) and nucleocapsid (N) proteins. Furthermore, SARS-CoV-2 contains eight accessory proteins (3a, 3b, p6, 7a, 7b, 8b, 9b, and orf14).

Viral replication

The virus replicates in the cytoplasm. The compartment between the endoplasmic reticulum and the Golgi apparatus is the space in which viral assembly and budding of virus particles takes place. Coronaviruses are known to modify cellular membranes to create sites for viral replication. They form double membrane vesicles (DMV) from the cisternae of the endoplasmic reticulum using non-structural proteins. Studies using high resolution microscopy show that DMVs have two membranes separated by an 18-nm space and the inner lumen is about 340 nm.[10] The space between the inner and outer membranes is the lumen of the endoplasmic reticulum, while the cytoplasmic space enclosed by the DMV contains viral RNA.

Viral RNA synthesis occurs in DMVs that are connected to the endoplasmic reticulum. This transcription process helps the virus to evade the innate immune system. Several nonstructural proteins are involved in the creation of DMVs such as nsp3, nsp4, nsp6, and nsp10. The transcription process synthesizes the full-length viral RNA as well as small sub genomic RNA strands. Replication of the genome is mediated by an RNA-dependent RNA polymerase that associates with other nonstructural proteins and forms a replication–transcription complex (RTC). The synthesis of viral RNA, capping, and proofreading are all performed by the RTC. The nsp12 seen in the centre of the RTC, is the active site of the enzyme RdRp. It requires nsp7 and nsp8 to function. These three nonstructural proteins (nsp7, nsp8, and nsp12) thus form the core RdRp complex. Other proteins such as the nsp13 (helicase), the nsp14 (exonuclease), the nsp16 (methyltransferase) as well as nsp9 and nsp10, bind to the core to form the RTC.[11]

The SARS-CoV-2 nucleocapsid protein N is in the cytosol. It is 46 kDa in size with 419 amino acids and has different domains, each having different roles in virus packaging. The viral RNA comes out of the DMVs through membrane pores and condenses with the nucleocapsid protein. The N-RNA complex is G-shaped, and together with other proteins forms the ribonucleoprotein. The interaction with the M protein occurs later and is important for viral assembly. The subunits of the enzyme replicase are associated with the molecular pore complex to guide the viral RNA to the cytosolic side. The dsRNA, which is an intermediary, remains inside the DMV. The RNP complexes containing the genome go to the cell membrane, and progeny viruses are made with the help of the viral envelope protein, which bud into membrane-bound compartments derived from the endoplasmic reticulum and are released out of the cell. Multiple copies of the molecular pore complex, spanning both membranes, connect the space enclosed by the DMV to the cytosol, facilitating the export of newly synthesized viral RNA from the DMVs to the cytosol. The nonstructural protein nsp3 appears to be the central component of the molecular pore and therefore may serve as a possible drug target.[12]

Mutations

In the early period of the pandemic, mutations in the viral genome were rare and known as a period of "relative evolutionary stasis." A year after the emergence of the virus, mutations were observed, giving rise to variants of concern. These mutations changed the characteristics of the virus, including its antigenicity and its ability to spread in the community.

Genome annotations on the first three sequenced genomes of SARS-CoV-2, when compared to related coronaviruses, have identified 380 amino acid substitutions between these coronaviruses.

No amino acid substitutions occurred in nsp7, nsp13, envelope, matrix, or accessory proteins p6 and 8b. There were 102 amino acid substitutions in nsp3 and 61 amino acid substitutions in nsp2. Furthermore, 27 amino acid substitutions were found in the spike protein, including six substitutions in the RBD in amino acid region 357-528 and six substitutions in amino acid region 569-655. No amino acid substitutions were present in receptor binding motifs that interact directly with the human receptor ACE2 protein, but six mutations occurred in other regions of the RBD.[13]

Mutations can change the phenotype of the virus, giving them a fitness advantage. A mutation in the spike protein that resulted in the amino acid change D614G was discovered in April 2020 and became the dominant genotype in a few months. G614 increased globally, replacing the original D614 form, and patients infected with this variant had lower CT values in RT-PCR. This was associated with higher viral loads, but not always with more severe disease.[14] D614G was found to augment viral replication and enhance cell entry. It is also known to increase transmissibility. Other amino acid substitutions were noticed in the receptor binding domain such as N439K and Y453F which increased the ability of the virus to bind to host receptors. The SARS-CoV-2 variants have mutations in the RBD. The five common mutations are K417N, K417T, N501Y, E484K, and S477N. All of these, except K417T, were found to increase the affinity of the RBD/ACE2 interaction. K417T was found to facilitate immune escape.[15] In addition to this, mutations can also be present in the ACE2 receptor of the host cell.