Genomic and molecular analysis of SARS-CoV-2 and the possible strategies of Covid-19 treatment - A Review

Youssef LAHLOU¹, Belkassem El AMRAOUI¹,³, Toufiq BAMHAOUD¹
¹ControlQuality in Bio-control Industry & Bioactive Molecules Laboratory, Faculty of Sciences El Jadida, Morocco.
³Faculty Polydisciplinary of Taroudant, University IbnZohr, Agadir, Morocco.

Abstract:
The COVID-19 pandemic observed in Wuhan city and spread rapidly around the world is linked to a new coronavirus called SARS-CoV-2 (for Severe Acute Respiratory Coronavirus Syndrome 2). Doctors, pharmacologists, biochemists, chemists and virologists are now trying to find effective treatment for this virus. Better knowledge of the genetic and molecular characteristics of the SARS-CoV-2 can help researchers to find the effective treatment. The objective of this review is to present the current state of knowledge about the genomic and molecular characteristics of SARS-CoV-2 virus and to present the possible strategies for treating this pandemic. Until now, it can be concluded that SARS-CoV-2 virus, which is a retrovirus, has developed several strategies to escape the immune system during the infection, the most known are the mutations affecting its genomes and which still modify its antigenic determinants, this is why the production of a specific drug is a major challenge for scientists, which suggests other possible treatments such as the monoclonal antibodies, the CRISPR/Cas13d strategy or the extraction of bi-antiviral compounds from medicinal plants.

Key Word: Coronavirus, medicinal plants, immune system, antigenic determinants, genomic sequence, pandemic, Covid-19

I. Introduction

In Wuhan city in China, the COVID-19 epidemic has been detected, at the end of December 2019, quickly spread to the entire planet. More than 208 countries and territories around the world are affected by this pandemic with more than 2 million and 830 thousand cases of infection and more than 198 thousand deaths according to the WHO, up to the date of writing this article. Although this number of cases diagnosed, however, reflects only a fraction of the actual number of cases, as many countries are now testing only those cases who are treated in the hospital setting.

In early January, the pathogen was identified as a novel coronavirus called SARS-CoV-2 (for Severe Acute Respiratory Coronavirus Syndrome 2), the term COVID-19 (for Coronavirus Disease 2019) refers to the infection caused by this virus. Doctors, pharmacologists, biochemists, chemists and virologists are now trying to find effective treatment for this pandemic.

One of the most decisive steps in combating this virus, is the detailed knowledge of the genomic and molecular characteristics of the SARS-CoV-2, since it can allow researchers to better understand their infection strategies, their multiplication and destruction of host cells, in order to find the most effective treatment against this virus.

The purpose of this review article is to get an idea of the genomic and molecular characteristics of SARS-CoV-2 virus, which confers this high contagiosity on the pandemic, and to present an overview of possible treatment options other than conventional preventive vaccination options.

II. Description of SARS-CoV-2

The Severe Acute Respiratory Syndrome 2 coronavirus (SARS-CoV-2), is related to the SARS coronavirus, which is 125 nm in diameter (“slightly larger than influenza, SARS and MERS viruses”). It is a retrovirus (virus with a single strand positive RNA genome, not segmented and polyadenylated, the RNA is identical to the messenger RNA so that viral RNA can be immediately translated into viral proteins by the host cell. This virus consists of a helical-shaped nucleoside formed from its RNA genome interacting with N proteins. This nucleoside is surrounded by a capsid and shell, formed by the E structural proteins, M, S and lipids from virion budding. The origin of the virus is probably bats or pangolins.

According to the previous scientific data, there are two circulating strains of the virus SARS-CoV-2 (L and S).
• Strain S is believed to be older than strain L. At this stage, there is no evidence to confirm whether the mutation occurred in humans or in intermediate hosts (animals).

• Strain L would be the most severe and frequent circulating strain (70% of samples tested in the study), while strain S would be less aggressive and less frequent (30% of samples).

III. SARS-CoV-2 composition

Spike Glycoprotein (S)

Or protein (S) is a type I transmembrane protein strongly N-glycosylated, which assembles in homotrimers on the surface of the viral particle. It plays a double function in the viral input by allowing on the one hand the cellular receptor binding of the host cell at the advent of the infection process and on the other hand the fusion of the viral envelope with the membranes of the target cells. It has a determining role in cellular tropism and pathogenicity. It consists of two sub-units (S1 and S2) self-associates on the surface of the virus:

– Subunit S1: allows recognition and binding to its cellular receptor, the angiotensin-2 conversion enzyme (ACE-2), by its receptor-binding domain (RBD), but may recognize other receptors such as CD209L (L-SIGN) lectin type C, which could explain his tropism for negative ACE-2 cell types. It’s a major antigenic determinant.

– Subunit S2: intervenes in the fusion of the virus envelope with the cell membranes by HR1 and HR2 and leads to the virus entering the host cell. Interactions with different signalling pathways in the host cell, including induction of apoptosis, production of inflammatory cytokines, and cyclo-oxygenase-2, are also associated with S protein. It also appears to be responsible for negative regulation of ACE-2 after viral infection. It is also the main in vivo antibody-inducing protein.

Protein M

Membrane protein M is the most abundant protein in the skin virus, which explains the presence of antibodies against this protein in patients’ serum. It plays a major role in assembling the virion, transmembrane transport of nutrients, releasing the buds and forming the wrap. It is capable of interacting with itself and all other structural proteins: S, E and N. It is highly glycosylated.

Protein N

It is a phosphoprotein of 46 kDa, which self-associates to form a dimer. It contains a short part rich in serine and a putative bipartite nuclear localization signal, thus suggesting its involvement in other important functions besides assembling capsid during the viral life cycle. It is associated with multiple functions in SARS-CoV-2. First of all, it combines with viral RNA when assembling the virion and, in doing so, allows the packaging of the genome within it. These functions are made possible by a link domain of the RNA in the N-terminal part and a domain allowing the association of N in dimers and other forms of self-association. N also binds to different host proteins, notably cyclophilin A(Cyp A) and ribonucleopterine A1, suggesting a more important role than simple viral structure. In addition, the nucleocapsid protein would be involved in the modulation of different cellular signalling pathways by regulating the expression of certain actors of these pathways such as ERK, MAPK or JNK. These changes include apoptosis induction and cytoskeletal reorganisation. It is also involved in the activation of the AP1 signal transduction pathway (activating protein 1).

Protein E

The small E envelope protein has a role in the secretion of virions. It is an infrequent protein of 9 to 10 kDa. It has a short N-terminal end, followed by a transmembrane domain and then a C-terminal end that makes up the majority of the protein. The E protein does not have a clickable signal peptide, which suggests it is a transmembrane type II protein. It has a highly hydrophobic structure and this feature may allow it to modify the permeability of the infected cell by forming pores on the plasma membrane. Long 76 amino acids, E is mainly localized in the endoplasmic reticulum, Golgi apparatus and membrane in infected mammalian cells.

Hemagglutinin-esterase dimer (HE)

Hemagglutinin-esterase HE is a type I transmembrane protein of 70 kDa with a large ectodoma, a transmembrane domain and a short endodomain. As a haemagglutinin, this protein can bind to sialic acids. In addition, it has acetylerase activity in the acetyl groups of acetylated acids 9-Oou 4-O-neuramines.
SARS-CoV-2 genome

The genome of SARS-CoV-2 is a single strand positive polarity RNA, which ranks it among the largest known RNA virus genomes. The genome has a 5’ capped tip and a 3’ polyadenylated (polyA) tail. The 5’ extremity also contains a leading regulatory transcription sequence (transcriptional regulatory sequence TRS) and an untranslated region (untranslated region UTR) containing several secondary structures necessary for replication and transcription. The RNA structures essential for the replication and synthesis of viral RNA are also present in the 3’ extremity UTR. Several open reading frames are present. Two open reading frames Orf1a and Orf1b occupy the first two thirds of the genome. They encode cleaved polyprotein into 16 non-structural proteins (nsp) needed for viral replication. Mutations affecting NSP2 and NSP3 have an important role in the infectious capacity and differentiation mechanism of SARS-CoV-2. The last third of the genome codes for structural proteins S, M, N and E. The order of genes, unchanged among all members of the coronavirus family, is 5’-replicase–S–E–M–N–3’. Each gene is preceded by a sequence that regulates the TRS transcription necessary for the expression of each gene. TRS sequences consist of a “core” sequence and regulatory sequences in 5’ and 3’ of this core sequence. The core sequence of the 5’ leader SRT of the genome is identical to the core sequence of the SRT located at the 5’ extremity of the genes, which has an important role in the discontinuous synthesis of the sRNA.
A genetic analysis of 103 SARS-CoV-2 genomes showed two viral strains, one type L (~70%) and one type S (~30%). Type L strains are derived from type S, and are more aggressive and contagious evolutionarily. Indeed, virologists and epidemiologists need to closely monitor the new coronavirus in order to control virulence and the epidemic.

The genome of SARS-CoV-2 resembles 96.2% of the genome of the bat coronavirus BatCoV RaTG13. Other studies have shown that the receptor binding domain (RBD) in protein (S) is the most variable part of the genome of SARS-CoV-2. Six RBD amino acids were essential to bind to ACE2 receptors and to determine the host range of SARS-CoV viruses, they are Y442, L472, N479, D480, T487 and Y4911, which correspond to L455, F486, Q493, S494, N501 and Y505 in SARS-CoV-2. Five of these six residues differ between SARS-CoV-2 and SARS-CoV.

Fig. 3. SARS-CoV-2 genome isolated from a patient (Betacov/Wuhan/IVDC-HB-01/2020|EPI_ISL_402119) (29892 bp)

Based on structural studies and biochemical experiments appears to have a RBD that binds with a high affinity to ACE2 of humans and other species that have a structural homology in this receptor.

Fig. 4. Comparison of human SARS-CoV-2 RBD sequence with other coronaviruses

Although the above analyses suggest that SARS-CoV-2 may bind to human ACE2 with a high affinity, computational analyses predict that RBD and ACE2 interactions are not ideal, but are probably the result of a natural selection on an ACE2 of man in order to obtain a more optimal connection than possible.
IV. Phylogenetic analysis of SARS-CoV-2

According to Zhu et al., although SARS-CoV-2 is similar to some beta-coronaviruses detected in bats, it is distinct from SARS-CoV and MERS-CoV. Wuhan’s three 2019-nCoV coronaviruses, as well as two bat-derived SARS strains, ZC45 and ZXC21, form a distinct clade. SARS-CoV strains from humans and SARS-like coronaviruses from bats collected in southwestern China formed another clade in the Sarbecovirus subspecies. Since the identity of the sequence in the conserved replicate domains (ORF 1ab) is less than 90% between 2019-nCoV and the other members of the beta-coronavirus, 2019-nCoV, which probably causes viral pneumonia in Wuhan, is a new beta-coronavirus belonging to the sarbecovirus sub-genus of the Coronaviridae family) 27.

In general, coronaviruses (CoV) are divided into four genera, α/β/γ/δ-coronavirus 2. α- and β-CoV can infect mammals, while γ- and δ-CoV tend to infect birds. Previously, six CoV were identified as human-receptor viruses, including α-CoVs Hcov-229E and Hcov-NL63, and β-CoVs Hcov-HKU1 and Hcov-OC43 with low pathogenicity, causing mild respiratory symptoms similar to a common cold, respectively. The other two known β-CoVs, SARS-CoV and MERS-CoV lead to serious and potentially fatal respiratory tract infections 33. The genomic sequence of SARS CoV-2 was found to be identical to 96.2% of a CoV Ratg13 bat, while, it shares 79.5% identity with SARS CoV. Based on the results of virus genome sequencing and evolutionary analysis, bats are suspected to be naturally occurring viral hosts, and SARS-CoV-2 may be transmitted by bats via unknown intermediate hosts to infect humans. SARS-CoV-2 could use the angiotensin-converting (ACE2) enzyme 2 to infect humans, and it is the same receptor for SARS-CoV 34.

---

Fig. 5. Phylogenetic analysis of SARS-CoV-2 and genomes of other β-coronavirus 27

---

DOI: 10.9790/3008-1503013947
www.iosrjournals.org

V. Infection and pathology of SARS-CoV-2
Firstly, it is important to know that the viral proteins of SARS-CoV-2 are mostly N-glycosylated by the cellular machinery, which allows it to several advantages. N-glycans have a different function depending on their location: an N-glycan close to an epitope can allow immune escapement by disrupting recognition by antibodies. A N-glycan near a cleavage site may result in decreased virulence or change affinity to the receptor if it is close to a receptor-binding site.

The infection begins by binding the SARS-CoV-2 RBD domain to the ACE2 receptor (Angiotensin-converting enzyme 2), which has the normal function of lowering blood pressure. It exists extensively in the lungs, heart, kidneys and intestine.

After fusion of the virus-host membrane by the two domains (HR1) and HR2 of the S2 S protein subunit, the viral genome RNA is released into the cytoplasm, the uncoated RNA then translates two polyprotein, pp1a and pp1ab, which encode non-structural proteins (nsp) and form a replication-transcription complex (RTC). RTC continuously reproduces and synthesizes a nested set of sub-genomic RNA, which encode accessory and structural proteins. Next, neoformed viral RNA, nucleocapsid proteins and envelope glycoproteins assemble and form buds of viral particles. Finally, the vesicles containing the virion fuse with the plasma membrane to release the virus, releasing billions of copies of SARS-CoV-2 by budding, leading to the destruction of the host cell which explodes by osmotic shock.

Understanding the molecular mechanisms of each stage of the viral cycle is crucial for future antiviral treatment research.

Fig. 6. Attachment of SARS-CoV-2 RBD to the human ACE2 receptor

VI. Possible treatment strategies
The treatments used until now four COVID-19, are only symptomatic, but no specific antiviral drugs have yet been confirmed to be effective \(^{36}\).

**Antiviral drugs**

- **Chloroquine and its derivatives**
  
  Chloroquine is a common antimalarial drug. Clinical trials in France resulted in the cure of three-quarters of 24 infected patients within six days of receiving treatment. Other promoter results were also obtained following the administration of a dose of 600mg/day of Plaquenil (drug with a molecule derived from chloroquine), patients who did not receive Plaquenil are still 90% carriers of the virus after six days, while those who have received treatment carry only 25%. Chloroquine phosphate and hydroxychloroquine are already part of the therapeutic protocols used for symptomatic treatments. Chloroquine should not be considered a miracle drug for SARS-Cov-2, it may cause serious side effects if overdosed or misused (nausea, vomiting, immune system ailments, gastrointestinal ailments, liver or haematological disorders).

- **Other drugs**
  
  Thai doctors have used high doses of the anti-flu drug Oseltamivir combined with anti-HIV drugs, Lopinavir and Ritonavir, have observed an improvement in the health status of several patients. Baricitinib was also proposed to reduce inflammation and infection processes caused by virus \(^{36}\). Other US companies are collaborating to develop a vaccine in mRNA. Recently, Ivermectin showed a strong inhibition of the replication of SARS-Cov-2 clinical isolate in vitro \(^{37}\).

**Monoclonal antibodies**

Monoclonal antibodies that target the SARS-Cov-2 protein(s) and can potentially inhibit binding of the virus on the ACE2 cell receptor and then prevent viral infection \(^{38}\).

**Virus against virus**

SARS-Cov-2 acquires new mutations mainly at the level of membrane proteins which are antigenic determinants, allowing it to escape antibodies produced by the immune system and antiviral drugs. Researchers have developed a flexible and effective approach to targeting viral RNA and not virus membrane proteins using CRISPR/Cas13d technology to specifically chew viral RNA from SARS-Cov-2, limiting its reproductive capacity. The principle is to use an adeno-associated virus (AAV) that carries both a Cas13d protein and a matrix of RNA guides targeting different regions coding for viral genome peptides, the set binds specifically with viral RNA, Cas13d simultaneously climaxes viral RNA regions coded for Orf1ab (the replicase-transcriptase) and virus protein S \(^{39}\). In addition, the AAV has very specific stereotypes of the lungs which allow it to inject are RNA guide easily and it does not affect the transcriptome of the lungs \(^{40}\).

**Bio-antivirals from medicinal plants**

Instead of focusing on chemical synthesis of antiviral drugs that can have several side effects, medicinal plants are rich in photochemical substances with strong antiviral activity via many mechanisms such as inhibition of the assembly of viral particles, inhibition of the fixation of the virus on the cell receptor (infection) inhibition of polymerase RNA or polymerase DNA, viral neuraminidase, protease, reverse transcriptase and expression of viral proteins... \(^{41}\). The research has shown that the extracts of Cimicifuga rhizoma, Meliae cortex, Coptidis rhizoma, Phellodendron cortex and Sophora subprostrata for example decreased the replication MHV (mouse hepatitis virus), in vitro production of vesicular stomatitis virus (VSV) and PEDV virus (swine diarrhea epidemic virus), they can be a source of anti-coronavirus compounds \(^{32}\), other extracts of Rosa nutkana and Amelanchier alnifolia have a strong antiviral activity against coronavirus BCV (bovine coronavirus) \(^{43}\), phlorotannins isolated from Ecklonia cava also have an excellent anti-activitycoronavirus against PEDV by inhibiting virus entry and/or viral replication in host cell \(^{44}\). This may allow researchers to exploit other sources of bioactive molecules that can combat SARS-CoV-2.

**VII. Conclusion**

It can be concluded that the SARS-CoV-2 virus, as a retrovirus, has developed several strategies to escape the immune system during the most known infection are the mutations that affect its genomes and that still modify its antigenic determinants, This is why the production of a specific drug is a major challenge that pushes us to propose other possible treatments such as monoclonal antibodies, the CRISPR/Cas13d strategy or the extraction of bio-antiviral from medicinal plants.

**References**

DOI: 10.9790/3008-1503013947  www.iosrjournals.org  45 | Page


