Abstract. – OBJECTIVE: The pandemic of the coronavirus disease 2019 (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a highly transmissible and pathogenic coronavirus, has been representing enormous threats to the world in almost all aspects. S protein recognizing and binding to angiotensin-converting enzyme-2 (ACE2) receptor is the key step of viral infection. This review summarized the structure of S protein, and the difference among members of the coronaviridae, especially the different sites between SARS-CoV-2 and SARS-CoV S protein sequences. We reconstructed the phylogenetic tree of 18 coronaviruses based on S protein, and detected the conserved motif. We had a further discussion on various promising antiviral compounds, drugs or approaches for treatment. Considering that various virus mutants are rampant around the world, we introduced some SARS-CoV-2 mutants, which are more contagious and spread fast. It indicates the limitations of wide spectrum therapeutic research. We wish the information provided by this review can be helpful to the global battle against SARS-CoV-2 infection.

Key Words: SARS-CoV-2, COVID-19, Variants, Transmission, Virulence, Neutralization, Antibodies, Vaccines.

Introduction

Since the outbreak of the coronavirus disease 2019 (COVID-19) pandemic, the number of deaths worldwide has continued to rise. By the end of May 2021, the cumulative number of confirmed cases has exceeded 170 million, making emerging coronaviruses a new public health concern in the twenty-first century. At the end of April, the virus was rampant in India, and the number of newly diagnosed cases in a single day even exceeded 360,000. The viruses spread rapidly all over the world by mutation. Indian mutants have spread more than 40 countries, and new mutants have been continuously found in other countries.

The viral pathogen responsible SARS-CoV-2, is an emergent, zoonotic and highly transmissible pathogen first identified in China in late 2019. According to sequence analysis, SARS-CoV-2 belongs to Betacoronavirus genera and is thought to inherit from a bat coronavirus (BatCoV RaTG13), with 96.2% identity in the whole genome sequence. Its S protein specifically recognizes and binds to ACE2 receptor, leading to mild respiratory failure. Phylogenetic analysis based on S protein sequence divided 18 coronaviruses into three Clade (Clade I, II and III). SARS-CoV, SARS-CoV-2 and RaTG13 make Clade II; RaTG13 is still the most identical to SARS-CoV-2 in phylogeny. The mutation rate of the virus is so fast that various kinds of virus mutants spread all over the world. How to develop effective treatment and broad-spectrum therapy is still an urgent problem. Despite a flood of SARS-CoV-2 researches involving in pathology, treatment and prevention improve every day, current knowledge of this novel coronavirus is just the tip of the iceberg. Tackling this epidemic is a long-term job and requires efforts of international cooperation by scientists, authorities and the public.

S Protein – A Key Structure of Coronavirus

Coronaviridae is composed of four genera (Alpha-, Beta-, Delta-, Gamma-coronaviruses), containing a diverse group of viruses infecting mammals and birds that individually cause a variety of diseases, including pneumonia, enteritis, hepatitis, encephalomyelitis, Tracheobronchitis, nephritis and various other diseases. Alphacoronavirus and Betacoronavirus can infect mammals, including humans, pigs, cats, dogs and cattle, while Gammacoronavirus and Deltacoronavirus usually infect birds, although some can infect mammals.

SARS-CoV-2 is clustered into Betacoronavirus due to the phylogenetic proximity to SARS-CoV and Middle East Respiratory Syndrome Coronavirus (MERS-CoV) and it is the seventh coronavirus found to infect human, causing atypical
pneumonia that can lead to death; SARS-CoV, MERS-CoV are also severe disease-related virus and belong to Betacoronavirus, while Alphacoronavirus 229E (HCoV-229E), NL63 (HCoV-NL63), Betacoronavirus HKU1 (HCoV-HKU1) and OC43 (HCoV-OC43) are much milder. Phylogenetic analysis of complete genome sequences showed that SARS-CoV-2 shares 79% genome sequence identity with SARS-CoV, and only 50% with MERS-CoV, but interestingly, 96.2% with a bat coronavirus (BatCoV) named RaTG13.

Betacoronaviruses shares a similar genome organisation with six open reading frames (ORFs): replicase (ORF1a/ORF1b), spike(S), envelope(E), membrane(M) and nucleocapsid (N). Among these distinctive proteins, the S is responsible for recognizing and mediating the virus into the host cell, and usually the S sequences among coronaviruses belong to different genera diverse a lot, even SARS-CoV and SARS-CoV-2 diverse in this region. Coronavirus can also achieve the host receptor recognition diversity through genetic mutation and recombination events in their S gene. For example, although they all belong to betacoronavirus, SARS-CoV-2 and HCoV-NL63 both recognize human ACE2, while MERS-CoV recognizes dipeptidyl peptidase 4 (DPP4). Murine hepatitis virus (MHV), another betacoronavirus mediates cell entry by recognizing the carcinoembryonic antigen cell adhesion molecules 1 (CEACAM1).

S protein observed under the electron microscope is a clove-like trimer, composed of three S1 heads and a trimeric S2 stalk. The S1 subunit recognizes and binds to the receptor on the surface of host cells, then S2 protein fuses the membrane of viral and host cells, allowing virus entry. There are two domains in S1 subunit determine the host receptor recognition range, which are N-terminal domain of S1 (S1-NTD) and C-terminal domain of S1 (S1-CTD). These S1 domains can bind to receptors, usually sugars or protein receptors, and act as receptor binding domains (RBDs).

X-ray diffraction analyzed the crystal structure of human ACE2 protein complexed with SARS-CoV-2 S protein RBD. The structure confirmed that most of the residues that related to receptor binding in SARS-CoV-2 are highly conserved or similar in characteristics to those in SARS-CoV RBD. Among these amino acid residues, the six RBD amino acids in SARS-CoV-2(Y455, L586, N49, D494, T501) are critical for binding to ACE2 receptors and for determining the host range. Five of these six residues vary between SARS-CoV-2 and SARS-CoV(L455Y, F486L, Q493N, S494D, N501T), but the changed amino acids completely maintained the original conformation of the interaction between S protein and ACE2 receptor.

Additionally, SARS-CoV-2 S protein RBD not only binds with high affinity to ACE2 from humans, but also binds to homologous receptors from ferrets, cats and other species. The presence of such homology is to maintain the key sites on the interface (Figure 1) between ACE2 and SARS-CoV-2.

**Phylogenetic Analysis Based on S Protein**

So far, the phylogenetic analysis of coronaviruses is mainly based on the whole genome. Although S protein is a a key factor in coronaviral...
SARS-CoV-2: phylogenetic status, mutations and therapeutic research based on spike protein

recognition and invasion, the phylogeny of S protein in multiple coronavirus remains unclear.

We downloaded the S protein sequences of 18 coronaviruses (Table I) from NCBI (https://www.ncbi.nlm.nih.gov/), then, we used mafft\(^{18}\) for sequence alignment, and FastTree\(^{19}\) for reconstructing the phylogenetic tree based on S protein (Figure 2A). We analyzed the conserved motifs of the S protein sequences of these coronaviruses (Figure 2B). The phylogenetic tree can be roughly divided into three clades (Clade I, Clade II and Clade III). In general, the C-terminal of the S protein is more conservative than the N-terminal, which indicates that the C-terminal sequence determines the conservative characteristics of the S protein and makes coronavirinae distinctive. However, the N-terminal of S protein mutated a lot during the viral evolutionary history, which offered each virus peculiar characteristics. At the C-terminal, the motif 6, 7 and 9 of Clade I are not found in Clade II and III, and Clade III also lacks motifs 17 and 18. Clade I is the most conserved motif in the C-terminal sequence, with 5 or 6 conserved motifs. There are only 2 or 3 motifs in Clade II and 1 or 2 in Clade III.

It is worth noting that RaTG13 is still the most identical to SARS-CoV-2 based on the S protein sequence, and RaTG13 is, for now, considered to be the origin of the SARS-CoV-2. However, even if there are a few differences in the S protein,

<table>
<thead>
<tr>
<th>Genus</th>
<th>Short name</th>
<th>Full name</th>
<th>Disease/Symptoms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alphacoronavirus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PEDv</td>
<td>Porcine epidemic</td>
<td>Gastroenteritis, Watery diarrhea,</td>
<td></td>
</tr>
<tr>
<td>PEDv</td>
<td>diarrhea virus</td>
<td>vomiting, dehydration</td>
<td></td>
</tr>
<tr>
<td>FCoV</td>
<td>Feline coronavirus</td>
<td>Feline enteric coronavirus: mild gastroenteritis and diarrhea</td>
<td></td>
</tr>
<tr>
<td>MHV</td>
<td>Mouse hepatitis virus</td>
<td>Enterosis, hepatitis, demyelinating encephalomyelitis</td>
<td></td>
</tr>
<tr>
<td>RSDAV</td>
<td>Rat sialodacryoadenitis virus</td>
<td>Rhinitis, epiphora, pneumonia</td>
<td></td>
</tr>
<tr>
<td>PHev</td>
<td>Porcine hemagglutinating encephalomyelitis virus</td>
<td>Vomiting, wasting disease, encephalomyelitis.</td>
<td></td>
</tr>
<tr>
<td>BCV</td>
<td>Bovine coronavirus</td>
<td>Gastroenteritis with profuse or bloody diarrhea, dehydration, decreased milk, or respiratory disease</td>
<td></td>
</tr>
<tr>
<td>SARS-CoV</td>
<td>SARS-CoV</td>
<td>Respiratory disease; zoonotic with bats as natural reservoir</td>
<td></td>
</tr>
<tr>
<td>SARS-CoV-2</td>
<td>Severe acute respiratory syndrome coronavirus 2</td>
<td>Respiratory disease; fever, dry cough and fatigue; or upper respiratory and digestive tract symptoms such as nasal obstruction, runny nose and diarrhea</td>
<td></td>
</tr>
<tr>
<td>Betacoronavirus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MERS-CoV</td>
<td>Middle East respiratory syndrome coronavirus</td>
<td>Respiratory disease; zoonotic with camels and bats as a likey reservoir</td>
<td></td>
</tr>
<tr>
<td>HCoV-OC43</td>
<td>Human coronavirus OC43</td>
<td>Mild respiratory disease</td>
<td></td>
</tr>
<tr>
<td>HCoV-NL63</td>
<td>Human coronavirus NL63</td>
<td>Mild respiratory disease</td>
<td></td>
</tr>
<tr>
<td>HCoV-229E</td>
<td>Human coronavirus 229E</td>
<td>Mild respiratory disease</td>
<td></td>
</tr>
<tr>
<td>RaTG13</td>
<td>Bat coronavirus RaTG13</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>ECoV</td>
<td>Equine coronavirus</td>
<td>Gastroenteritis</td>
<td></td>
</tr>
<tr>
<td>CrCoV</td>
<td>Canine respiratory coronavirus</td>
<td>Respiratory disease</td>
<td></td>
</tr>
<tr>
<td>PRCV</td>
<td>Porcine respiratory coronavirus</td>
<td>Mild respiratory disease or subclinical</td>
<td></td>
</tr>
<tr>
<td>IBV</td>
<td>Avian infectious bronchitis virus</td>
<td>Tracheobronchitis, nephritis</td>
<td></td>
</tr>
</tbody>
</table>
RaTG13 seems not to effectively bind to ACE2 receptor\textsuperscript{16}. Deeper structural and bioinformatics studies have found two major genomic features of SARS-CoV-2: (i) the S protein of SARS-CoV-2 appears to be optimized for binding to the human receptor ACE2; and (ii) the S protein of SARS-CoV-2 has a functional polybasic (furin) cleavage site at the S1-S2 boundary through the insertion of 12 nucleotides\textsuperscript{20}, which forms a polybasic cleavage site (RRAR), that enables effective cleavage by furin and other proteases\textsuperscript{21}. This furin-cleavage site can reduce the stability of SARS-CoV-2 S protein and facilitate the conformational adaption that is required for binding to the RBD of its receptor\textsuperscript{22}. This explains why the S proteins of RaTG13 and SARS-CoV-2 are highly similar in sequence, but RaTG13 cannot infect human. The emergence of SARS-CoV-2 may have been selected and optimized for binding to human ACE2 receptor.

Although there are just several amino acid mutations, SARS-CoV-2 have evolved adaptively from bat to human and accelerated human-to-human transmission.

**Viral Mutants of SARS-CoV-2**

According to Centers for Disease Control and Prevention COVID-19 Response, since a G614 substitution in the S protein has emerged; virus containing this substitution has become the predominant circulating variant in the COVID-19 pandemic. G614 substitution had increased affinity to human ACE2, through regulating the stability of S protein trimer. It which may be another mechanism that underlies the increased replication and transmission of SARS-CoV-2. Some other mutants have continued to threat. The B.1.1.7 mutant in the UK has 10 mutational sites on the S protein, of which the key mutation Y501 is in the

---

**Figure 2.** Phylogenetic analysis based on S protein. (A) Phylogenetic tree of 18 coronaviruses; (B) Conserved motifs of the viral S protein. Detailed identification of variable motifs is highlighted with the colored square. Bar with different length represents the conserved position of each motif in different viruses.
RBD and the binding sequence. Mutation Y501 can increase the affinity between the viruses and ACE2 receptor of human cells, thus accelerating the spread of the virus in the population. In addition, the loss of two amino acids on S protein may help it escape immune response, and the H681 mutation may enhance its ability to fuse with cells. The 501.V2 mutant found in South Africa is more serious. There are nine mutations on S protein, three of which are located in the RBD region, namely Y501, N417 and K484. K484 is also located in the binding sequence of RBD. Previous detection showed that this mutation was related to the viral ability to escape the complex antibody, and it may affect the efficacy of existing vaccines targeting S protein. In addition, the loss of three amino acids on the S protein of 501.V2 mutant may affect the antigenicity, and then lead to “escape mutant” to escape the immune response of monoclonal antibody. The P.1 in Brazil contains mutations at T417, K484, Y501 and other sites. According to the SSI report, 5 mutants have been isolated from Danish minks, among which F453 is of great significance, because it may be more suitable for mink host. It indicates that F453 mutation may enhance its ability to fuse with cells again and evolve resistance to existing vaccines.

Inhibiting the Replication of SARS-CoV-2

Chemotherapy-Like Therapy

Utilizing quantitative proteomics and transcriptional analysis to study the host cells infected with SARS-CoV-2, the analysis showed that viral infection reshaped a variety of core regulatory pathways in host cells, such as translation, splicing, carbon metabolism and nucleic acid metabolism. Small molecule inhibitors targeting these pathways can inhibit viral replication in host cells. After testing, drugs targeting these essential signaling pathways, including cyclohexidine (translation elongation inhibitor), emetine (inhibition of 40S ribosomal protein S14), pladienolide, and NMS-873 (small molecule inhibitor of adenosine triphosphatase p97), can effectively inhibit the replication of SARS-CoV-2 at non-toxic concentration, which reveals that these drugs can be used as potential therapeutic strategies. However, these small molecule inhibitors not only inhibit virus replication, but also hinder the core regulatory pathways (translation, splicing, carbon metabolism and nucleic acid metabolism) of host cells to a certain extent. Self-damaging may need to be considered and avoided in further research.

Drugs Blocking Viral Replication

FDA (Food and Drug Administration) described a large-scale screening study to assess nearly 12000 drugs to block the replication of SARS-CoV-2, and eventually found 100 molecules that can inhibit the replication of the virus, of which 13 drugs showed significant characteris-
tives – especially effective at realistic dose levels. These drugs include: Anti-HIV drug R 82913; ds-6930, a member of the PPAR-γ agonist family for the treatment of diabetes; Ono 5334, a potential drug for the treatment of osteoporosis; apilimod, a potential drug for the treatment of autoimmune diseases such as Crohn’s disease. The three most effective drugs were Ono 5334, apilimod and MDL 28170 (it has been previously shown that MDL 28170 can attenuate Ebola virus infection). Using cultured lung tissue to test the ability of these three drugs to reduce the replication of SARS-CoV-2, the results showed that the number of infected cells reduced by 72% (Ono 5334), 65% (MDL 28170) and 85% (apilimod), respectively. On this basis, it has been shown that apilimod is well tolerated in human body and shows good safety in the dose range that may produce antiviral effect. Some of these drugs have been tested in clinical settings, so that can advance the preclinical evaluation and clinical assessment, and promote the understanding of their capacity as a therapeutic medicine.

Carmofur is valuable in the treatment of colorectal cancer, and inhibit main protease (Mpro) in the treatment of breast cancer, gastric cancer and bladder cancer. Since Mpro is a conserved sequence in all coronaviruses, SARS-CoV-2 Mpro are identified as a potential therapeutic target due to its involvement in viral replication. So Carmofur as well as drugs based on Carmofur may be effective against broad-spectrum coronavirus infection. The X-ray crystal structure of SARS-CoV-2 Mpro combined with Carmofur shows that Carmofur can directly modify the catalytic element Cys145 of SARS-CoV-2 Mpro. This and the further details of the interaction between Carmofur and Mpro revealed by this structure are expected to lay a foundation for the design of more powerful derivatives of Carmofur.

CVL218, an inhibitor of poly-ADP-ribose polymerase 1 (PARP1), exhibits effective inhibitory activity against SARS-CoV-2 replication, and able to suppress the CpG-induced IL-6 production in peripheral blood mononuclear cells, suggesting that it may also have anti-inflammatory effect that is highly relevant to the prevention immunopathology induced by SARS-CoV-2 infection. Sequence analysis method named Viral-Track was developed in May 2020, based on which, the host response induced by virus and the key host factors needed for virus replication can be revealed. For this convenience, more reliable assessment of the effectiveness of some inhibitors for viral replication, large-scale randomized controlled trials may be possible to conduct.

**Inhibition of Virus Entry**

One of the important ways to prevent SARS-CoV-2 from entering host cells is blocking the interaction between S protein and ACE2 receptor. Antibody neutralization is also an effective substitute. Antibodies P2C-1F11 and P2B-2F6 have nearly 100% binding efficiency to SARS-CoV-2, even much higher than ACE2, indicating that they can block the interaction between SARS-CoV-2 RBD and ACE2. Although some neutralizing antibodies were found to have strong binding with SARS-CoV-2 RBD, and may be promising candidates for prevention and treatment of SARS-CoV-2 interventions, they show no obvious cross reaction with RBD of SARS-CoV or MERS-CoV. It highlighted the immune difference of RBD among the three viruses. In order to elucidate the relationship between antibody response and disease progression, as well as its driving factors and effects, more clinical trials must be studied.

There are missense genetic variants, like p.His378Arg, p.Ser19Pro, p.Gly211Arg, p.Asp206Gly, p.Arg219His, p.Lys341Arg, p.Ile468Val, and p.Ser547Cys, might affect the structure and function of ACE2, and some might affect the binding capacity of SARS-CoV-2, so SARS-CoV-2 may have different susceptibilities, severity and mortality in those populations with these loci. It means that the variants of ACE2 affects the binding of SARS-CoV-2 to ACE2, thus slowing down or accelerating the spread of SARS-CoV-2. This reminds us that we can edit ACE2 specific sites to prevent it from interacting with the virus. The distribution of the linear epitopes of the S protein of the SARS-CoV-2 was analyzed, and three epitopes might induce neutralizing antibodies in the non-RBD region. The validation of a larger sample is in progress, and the ability of each epitope to induce neutralizing antibody is studying and expanding based on animal immunity. It will provide important support for neutralizing antibody and vaccine development.

Inhibitors like H014 can competitively bind to the SARS-CoV-2 and SARS-CoV, thus preventing the virus from attaching to the target cell surface. Additionally, it has been known that pH-dependent viruses, including enveloped coronaviruses (SARS-CoV-2, SARS-CoV and MERS-CoV), influenza A(H1N1), avian influenza A(H7N9) and uncoated rhinoviruses need endosome acidification to fuse with host membranes. A defensin a-
ologue, P9R, has strong antiviral activity against pH-dependent viruses, depending on the direct binding to viruses and the inhibition of virus-host endosomal acidification.

**IFN-I Response and Antiviral Drugs**

Induction of Type I Interferon (IFN-I) is a central event of the immune defense against viral infection. Upon exposure to RNA viruses, an intracellular antiviral response is initiated by activation of retinoic acid inducible gene I (RIG-I) like receptors. Particularly, when RIG-I/MDA5 (melanoma differentiation-associated gene 5) detects viral RNA, they trigger a signaling complex on the mitochondrial outer membrane, including the adapter proteins mitochondria antiviral signaling protein/TNF receptor associated factors 3/TNF receptor associated factors 6/translocase of the outer membrane70 (TOM70). This process ultimately leads to IFN-β production and induction of a host antiviral state. The most prominent feature of SARS-CoV-2, in terms of immune responses as compared to that of other viruses such as influenza A, is that it strongly inhibits the level of IFN-I in patients. In addition, it has also been found that the chemical, Liquiritin, can inhibit SARS-CoV-2 by mimicking IFN-I. SARS-CoV-2 ORF9b inhibits IFN-I response by binding to TOM70, thereby significantly inhibits the induction of IFN-I in cells, and the overexpression of TOM70 can largely offset the inhibition of IFN-I mediated by ORF9b. Based on this, a potential drug design strategy is developing inhibitors targeting the interaction interface between ORF9b and TOM70 to restore IFN-I response.

SARS-CoV, MERS-CoV and other respiratory viruses such as rhinovirus and influenza influenza A can up-regulate ACE2 in host cells, and cytokines such as IFN-γ can also stimulate the expression of ACE2. It is speculated that virus infection can activate the immune system and induce the expression of a variety of cytokines including IFN. These cytokines can promote the transcription and expression of ACE2 by activating downstream signaling pathways like JNK pathway. Viruses induce inflammatory response thus promoting the expression of ACE2 receptor and aggravating viral infection and transmission. All these are instances of an issue that “inflammatory factor storm” caused by SARS-CoV-2 infection can not only damage human organs, but also promote the further infection and transmission of virus. Therefore, antiviral drugs may also need to be combined into anti-infection therapy.

**Future Perspectives**

SARS-CoV-2 is the third highly pathogenic human coronavirus disease after SARS-CoV and MERS-CoV. Since the epidemic continued for more than one year, some countries and regions have carried out large-scale vaccination in order to achieve herd immunity. However, the precipitous epidemic situation in India this April posed challenges to the efficiency, the insufficient supply and uneven distribution of vaccines. Virus genetic drift occurs all the time. Although this is a random result, we has to realize that the mutants reserved by natural selection is often more virulent and spreads more rapidly. The mutation rate is far faster than that of vaccine research and development. Therefore, it is necessary to further discover the origin of animal SARS-CoV-2, the cross species infection pathway, the pathogenesis of SARS-CoV-2 infection and the interaction between virus and host. In view of these biological processes, we should design treatment strategies that can still play an effective role in the case of viral mutation.

**Acknowledgments**

This work was supported by College of Life Sciences, Northwest A&F University, and we thank the editor and the reviewers for their useful feedback that improved this paper.

**Authors’ Contribution Statement**

Designed the experiments: YZ  Analyzed the data and plot: YZ. Wrote the paper: YZ. All authors read and approved the final manuscript.

**Conflict of Interest**

The Authors declare that they have no conflict of interests.

**References**


