A Review of SARS-CoV-2, Responsible for COVID-19: History, Biology, Infection Mechanism, Antigenic Vaccines, their Risks and How an Alternate RBD Vaccine is Safer?

ZMG Sarwar Jahangir and Arleta Helena Marnik

Department of Biological Sciences, Kingsborough Community College, The City University of New York, 2001 Oriental Boulevard, Brooklyn, NY 11235.

Abstract: The SARS (severe acute respiratory syndrome)-CoV (Coronavirus)-2 S(spike)-protein mRNA/cDNA currently being used as vaccines are antigenic but not antigens against SARS-CoV-2, that causes COVID (Coronavirus Disease) -19. Furthermore, the mRNA and cDNA antigenic vaccines also have potentials for homologous as well as heterologous recombination, primarily into the somatic cell DNA of the vaccine recipients. On the contrary, a SARS-CoV-2 RBD-protein antigen, a part of the S-protein, will directly stimulate antibody production against SARS-CoV-2. Hence, a vaccine composed of SARS-CoV-2 RBD-protein as a safer, fast acting, and effective vaccine against SARS-CoV-2 and thus against COVID-19. This is also useful for some immune compromised individuals.

Introduction

1. History of coronavirus human infection, mortality and socioeconomic impacts:

Based on studies of RNA-dependent RNA polymerase molecular clock, it was concluded that a common ancestor of avian coronavirus (CoV) appeared about 10,000 years ago and a mammalian CoV appeared about 7,000–8,000 years ago [1]. As human travel increased over a long period of time, such as the extensive trade missions in early Mesopotamia around 5000 years ago, like many other diseases, CoV also spread from regions to regions more frequently [1]. The first human CoV, HCoV (human coronavirus)-229E, strain B814, was isolated from human upper respiratory tract infections (URTI) in 1965 [2-4]. Since then, many more strains of coronaviruses, such as HCoV-NL63, HCoV-229E, HCoV-OC43 and HCoV-HKU1, have been identified causing human URTI leading from mild to moderate common cold infections [5]. Some of them were known to be more severe and even fatal for humans [4].

In November 2002, a coronavirus infection started an epidemic in Foshan Municipality in Guangdong Province, China, causing SARS (severe acute respiratory syndrome) [6]. It ultimately spread...
over other countries in North America, South America, Europe, and Asia infecting at least 8,098 people, out of whom 774 died [4,6]. The patients were diagnosed with flu like symptoms which progressed into atypical pneumonia followed by increasing shortness of breath and oxygen desaturation [7]. This human disease was caused by SARS-CoV virus which has 88-95% homology to coronavirus originated in horseshoe bat (*Rhinolophus pusillus, R. macrotis, R. pearsoni, and R. sinicus*) and 99.6% homology to coronavirus found in masked palm civets, *Paguma larvata* [8-10].

The above incidence was followed by another outbreak of SARS coronavirus infection known as Middle East Respiratory Syndrome (MERS), diagnosed in September 2012 in Saudi Arabia, that extended over 27 countries including East, North and West Africa and South Korea [11,12]. It was determined that the MERS coronavirus (MERS-CoV) originated in the Egyptian tomb bat, *Taphozous perforates*, transmitted to the dromedary camel, *Camelus dromedarius*, and then to humans was responsible for the disease [13,14]. The MERS-CoV infected 2,400 persons including >850 (35%) deaths, in addition to causing diseases in heart, kidneys, lower respiratory tracts and thus multiorgan failures to those survived the infection [11,15,16].

Again, in November 2019, SARS emerged in Wuhan, in Hubei Province in Central China, and ultimately causing a global pandemic [17]. Zhang Jixian was the first doctor who noticed uncommon abnormalities in the thorax CT scan images of lungs of an elderly couple and reported it to the head of the hospital on December 27, 2019, which was recorded by the hospital officials to the Chinese Center for Disease Control and Prevention and China’s National Health Commission [18]. The causative coronavirus agents for the recent outbreaks in Guangdong, Arabian Peninsula and Wuhan are called SARS-CoV, MERS-CoV and SARS-CoV-2, respectively; and their outbreaks are known to cause human mortality [19-21]. The genome sequences of SARS-CoV-2, SARS-CoV, and MERS-CoV have revealed their close resemblance to each other [22]. They also reported that the cytokine profiling as well as inflammatory signaling in response to SARS-CoV-2 infection is different than SARS-CoV and MERS-CoV infections. This was also revealed by the ineffectiveness of several medications in controlling SARS-CoV-2, which were effective against SARS-CoV and MERS-CoV [22].

Due to the global spread of SARS-CoV-2 infections, that began in Wuhan, China, the World Health Organization (WHO) of the UN announced on February 11, 2020, a common name, COVID-19, to refer to the disease [23]. The SARS-CoV-2 has a wide range of host adaptability and is capable to cause diseases in humans, camels, pigs, mice, bats, pet dogs and cats, masked palm civets and chickens [24]. The novel coronavirus SARS-CoV-2 that emerged in the fourth quarter of 2019 in Wuhan, spread quickly around
the world and by March 19, 2021, the infection figures reached to 122,036,229 confirmed cases including 2,694,915 deaths [25]. The transmission of coronavirus outbreaks is shown in Figure 1, attached.

![Diagram of coronavirus transmission](image)

**Figure 1.** Transmission pattern of coronavirus outbreaks 2002 - 2019 [9,32]:
- SARS-CoV (2002) Bat → Palm civets → Human;
- MERS-CoV (2012) Bat → Camels → Human;

Statistically, this affected about 1.5% of the human population infected and also affected the global finance, as of May 2020, by $5.8 to $8.8 trillion [26]. Lockdowns, increase in unemployment, remote schooling, and quarantine led to global crisis with millions of families turned into the economical edge, social poverty and increased family stresses [27]. An excessive stress from stringent quarantine precautions caused elevation of plasma adrenal cortisol, resulting in sleep disorder, obesity, diabetes, and asthma [28]. The long-term exposure to hardship during maternity have the potential to cause delays in embryonic developmental due to an imbalance in the neuroendocrine immune system and telomere damages [29].
2. Origins of human coronavirus:

The genome of a coronavirus is composed of RNA molecules in the range of 26,000 to 32,000 nucleotide bases in length containing recombination of mammalian and avian coronavirus ancestors [30,31]. The SARS-CoV-2 is found to have 96.1% genetic homology to another coronavirus, SARSr-Ra-BatCoV-RaTG13, found in the horseshoe bat, *Rhinolophus affinis* [32,33].

The coronavirus spike (S) protein genome analysis supports a possible recombination of homologous coronavirus genomes found among bats and snakes [34]. The RBD domain of the SARS-CoV-2 S-protein has 97% symmetry in amino acid residues to a pangolin (*Manis javanica*) coronavirus, SARSr-CoV-MP789 [32,33,35]. The above literature suggests a natural recombination of RBD domains of the S-protein between horseshoe bat and pangolin coronaviruses. In general, diversity of viruses results from point mutations, genomic template-switch in replication, and genetic recombination [30,31]. In addition, several scientists experimentally confirmed that SARS-CoV-2 can bind to human, Chinese horseshoe bat, civet, and pig cell receptor ACE-2, supporting that the SARS-CoV-2 RBD can mediate infection in human and in other mammals too [36].

In a separate report, [37] described, taxonomically SARS-CoV-2 is closely related to SARS-CoV and SARS-related bat and pangolin coronaviruses that share a common receptor, ACE-2, binding. By amino acid analysis, they found that the SARS-CoV-2 S1 spike protein domain is closely related to the SARS-CoV and SARS-CoV-2 viruses, while the S2 domain is closer to MERS-COV and HCoV-OC43 [37,38]. Both S1 and S2 domains are the key components in the evolution of the structural loop affecting the virus’s stability and transmission [37].

3. Description of SARS-CoV-2:

SARS-CoV-2 is a single positive stranded RNA virus and belongs to the genus, *Betacoronavirus* [39]. Many SARS-CoV-2 RNA genome sequencing studies have found some variations in their lengths ranging >29,000 nucleotides. In a study of full-length SARS-CoV-2, the maximum length was determined to be 29,742 nucleotides, while the full lengths of the Wuhan strain SARS-CoV-2 were reported to vary between 29,881-29,903 nucleotides in length with 10-14 *ORF* encoding 9,860 amino acid residues [40-44].

The SARS-CoV-2 mRNA is housed in a spherical body, 60-140 nm in diameter, encapsulated by a bilayer phospholipid membrane containing envelop (E), membrane (M) and spike (S) proteins [40,45,
The E and M proteins are composed of 76 to 109 amino acids while the spike protein is 1273 amino acids in length [40,45,46]. The number of proteins per SARS-CoV-2 virion varies as follows: E proteins, 100-200; M proteins ~1100; nucleocapsid (N) proteins 730-2200, and ~90 S proteins [40,45,46]. The SARS-CoV-2’s rate of evolution is 0.27 to 2.38 substitutions X 10^{-3}/site/year; mutation is 3.5 X 10^{-6} per infection; infection time, 10-15 minutes; progeny in 7-8 hours and 6-7 X 10^2 pfu per burst [40]. In addition, there are several non-structural proteins, which were found to which play critical roles in SARS CoV replication and transcription [47].

4. Infection mechanism:

It is well known that SARS-CoV-2 infects humans by attaching its spike proteins to the ACE-2 receptor molecules present in human respiratory epithelial cells [48]. Once the virus reaches a person’s respiratory tract, an RBD epitope located in the S1 domain protruding out of a SARS-CoV-2 virion membrane binds to an ACE-2 receptor on a respiratory or nasal mucosal epithelial cell facilitating the entry of the SARS-CoV-2 virion into the cell [45, 49]. Specifically, it is the receptor binding motif (RBM), a specific part of the RBD S1 domain, attaches to the ACE-2 receptors on human cells. This occurs more frequently on the nasal mucosal cells, where the virus has easy and frequent access. More specifically, the RBM residue, located within the 227 amino acids long RBD region (p319 - p546 residue) binds to ACE-2 receptor protein present on the respiratory mucosal epithelial cells [50]. Soon, the entire S1 domain of the spike protein is cleaved by furin and TMPRSS2 enzymes hanging off from the mucosal cells making the S2 subunit free from S1 subunit. Soon after, with the assistance of mucosal cell proteases, the S2 domain undergoes proteolysis, losing some amino acids, Arg815-Ser833, and exposing some other amino acids, Ser861-Phe833 and becomes competent for fusion with the mucosal cell membrane of the host [51,52]. Following that, the S2 subunit attaches to the SARS-CoV-2 virion membrane with free C-terminal amino acids, Ser861-Phe833, and leads to insertion of SARS-CoV-2 virion into the cell [51-53].

5. SARS-CoV-2 replication mechanism inside the host:

Once inside the human cell, the SARS-CoV-2 viral genome is released into the cytoplasm of the mucosal cell. The SARS-CoV-2 positive mRNA genome is copied into a negative mRNA molecule which translates the viral proteins: N, M, E, and S proteins. Side by side, a positive SARS-CoV-2 mRNA is also transcribed from the negative SARS-CoV-2 mRNA. While the N proteins bind to the positive SARS-CoV-2 mRNA for providing and holding its structure, the E, M and S proteins attach to the ER membrane of the
host cell. This segment is soon transferred to the Golgi apparatus to pack the SARS-CoV-2 mRNA positive strand inside a phospholipid membrane carrying E, M and S proteins forming a vesicle. This in turn moves towards the mucosal cell membrane to be released outside as an active SARS-CoV-2 virion particle [54]. The infection time has been estimated to be 10-15 minutes, the replication time for generating progeny virions takes 7-8 hours and 6-7 X 10^2 pfu are released per burst [40].

6. **SARS-CoV-2 vaccine development:**

All types of viruses mutate, evolve as they replicate, and their prolonged presence in an uncontrolled environment favor development of new variants, some with potential deleterious impacts. Their ability to generate *de novo* diversity in a short period of time and the rate of spontaneous mutation vary among viruses: mutation rates in RNA viruses is higher than DNA viruses, single-stranded viruses are higher than double-strand viruses [55].

Hence, only vaccinating a smaller segment of a population against SARS-CoV-2 may favor generation of new variants with new infectivity. In that scenario, even the vaccinated individuals would face risks from arrivals of new variants. This would be enhanced further, since active SARS-CoV-2 may get transmitted back and forth between humans to pet animals, as already observed in Italy [56]. This may only be brought under control by administering a safe and effective vaccine as soon as possible to a significantly large part of the population. Successes of such efforts will reduce development of new variants and, thus, help the global human population attain herd immunity [57].

As cases of COVID-19 were growing globally, it brought together the efforts of worldwide biotechnologists, scientists, experts, pharmaceuticals, and investors to develop effective vaccines against SARS-CoV-2 as soon as possible. Hundreds of biotechnological and pharmaceutical institutions, individuals and companies across the globe sought to develop effective vaccines against SARS-CoV-2. More than 50 such vaccine candidates were put into human trials in 2020, and a total of 250 vaccine candidates were in the process [58]. Safety and effectiveness of a vaccine are measured by its long-term antigenicity and immunogenicity for its therapeutic application as a vaccine.

The SARS-CoV-2 mutates at a rate four times slower than influenza virus, and the new variants have a minimal effect on antigenicity of the virus [59,60]. Although the coronaviruses mutate at a slower rate, some of the new variants of SARS-CoV-2, D 614 G as an example, mutated outside the RBD residue, has a higher rate of infectivity by boosting viral replication in lung and respiratory tract tissues [61,62].
Since the SARS-CoV-2 S-protein molecules located on its outer surface is critical for its binding with ACE-2 receptors on human cells, the S-protein should be the most effective antigen in developing antibodies against the SARS-CoV-2 causing COVID-19. The currently administered COVID-19 vaccines are either antigenic SARS-CoV-2 S-protein mRNA or SARS-CoV-2 S-protein cDNA incorporated into an adenoviral vector DNA [63-65].

These vaccines express the entire viral S-protein once they enter the cells of the vaccine recipient which is later recognized by the immune system. The role of the vaccine protein is to trigger immune cells and produce immunity against future viral infections. Pfizer (BioNTech) vaccine contains a genetically modified mRNA molecule enclosed inside a lipid nanoparticle (LNP) as a vaccine that encodes SARS-CoV-2 spike S-protein. Recipient of this vaccine produces S-protein antigen, received by the T cells followed by antibody production by the B cells. The antibody has the potential to block SARS-CoV-2 infection by attaching to the RBD epitopes of S-protein [66].

7. Potential risks associated with mRNA and adenovirus mediated DNA vaccines:
7.1. Risks associated with mRNA vaccines:

Breakage and repair of double stranded DNA molecules are common inside human cells, and its repair with high fidelity is essential for genetic stability and good health [67]. In some cases, mRNA serves as a DNA damage code and promotes homologous recombination [68]. As a result, entrance of any foreign mRNA inside human cells poses a risk to the stability of the cellular genome of the recipient. Also, there are several evidences for live, attenuated viral RNA vaccines to recombine with similar or other viral genome in circulation in the recipient’s plasma, affecting the effectiveness of RNA vaccines [69,70].

DNA: RNA hybrids constitute a well-known source of recombinogenic DNA damage [71]. Spontaneous mitotic recombination is a well-known process, and it is also enhanced by transcription [72]. The frequency of transcription increases double strand breaks and/or single stranded DNA gaps which become susceptible to RNA-DNA recombination from bacteria to human. This takes place on the replication forks during transcription by provoking stalling and/or breakage. Furthermore, transcription – replication creates recombinogenic DNA breaks by forming co-transcriptional R-loops as a way where RNA influences genome dynamics. The above phenomenon poses a question on the potential interference and thus risks associated with SARS-CoV-2 S-protein mRNA vaccines on the cells and thus health of the vaccine recipients.
Zhang [73] published evidence of SARS-CoV-2 chimeric transcripts in primary cells of patients as well as SARS-CoV-2 infected cells in culture. There is also reported evidence of fusion and retro-integration of SARS-CoV-2 genome into human DNA and its subsequent expression. There is further evidence of RNAs of SARS-CoV-2 being reverse transcribed in cells of infected human patients either by their endogenous LINE-1 or HIV-1 reverse transcriptase, as LINE-1 expression is induced in patients following SARS-CoV-2 infection. It is also reported that cells in culture exposed to cytokine induces LINE-1 expression. Furthermore, human patients infected by SARS-CoV-2 were found to continue to produce the viral RNA even after recovery suggesting the existence of a different mechanism of viral RNA replication in human. These evidences support the existence of a molecular mechanism for SARS-CoV-2 retro integration into human genome and thus putting a challenge to the use of SARS-CoV-2 mRNA or cDNA as vaccines.

Another group of scientists reported their observation of increased retrotransposon expression and upregulation of 10-11 translocation enzymes in lung derived cells collected from SARS-CoV-2 infected human patients [74]. They hypothesized that coronavirus invades human cells, interacts with retrotransposons, and may harm through retrotransposon dysregulation. These authors support our view that SARS-CoV-2 S-protein mRNA as well as cDNA vaccines have the potential to enter the human cell genome that might harm the recipient.

7.2. Risks associated with adenoviral and adeno-associated viral vector mediated vaccines:

Although adenoviruses and adeno-associated viruses (AAV) are used as a vehicle for gene transfer and human gene therapy for some time, they also pose risks and danger to the recipients [75, 76]. Unfortunately, these vectors also integrate external genes into the recipient’s genome. Although rare, it was demonstrated that genes carried by AAV vectors integrate in a site-specific manner into the q arm of human Chromosome 19 which may also express when the virus enters the lytic cycle, induced by a helper virus, as recorded in AAV virus therapy [77].

In an in vivo experiment using mice, the median frequency of chromosomal integrations of adenoviral vector DNA into mice chromosomes were found to be $6.72 \times 10^{-5}$ as heterologous recombination and $3.88 \times 10^{-7}$ as homologous recombination [78]. Furthermore, recipients who were previously exposed to the adenoviral vector (very common) might have pre-existing immunity for the vector and hence, the efficacy of the COVID-19 cDNA vaccine using adenoviral vector may not be as
effective. This indicates the potential risks for using SARS-CoV-2 cDNA vaccines carried by the adenovirus vector.

8. Why did we choose RBD protein vaccine?

The RBD is a small segment of the S-protein of the SARS-CoV-2 virion, and it is critical for the binding of the virus to the ACE-2 receptors on human mucosal cells causing infection and the disease, COVID-19. Hence, using RBD as an antigen for producing a vaccine to protect human from SARS-CoV-2 infection and the disease will be highly effective.

According to Huang [42] the S-protein located on the SARS-CoV-2 virus envelop is composed of 1273 aa. The S-protein consists of a 13 aa long (1–13 residues) signal peptide, followed by a 672 aa long (14-685 residues) S1 domain, and a 588 aa long (686–1273 residues) S2 domain. The S1 domain contains a 292 aa long (14–305 residues) N—terminal domain and a 223 aa long (319–541 residues) RBD. The SARS-CoV-2 encapsulating membrane holds many trimeric S-proteins, each containing the RBD proteins as a trimer that binds to the ACE-2 receptor present in human cells [79]. Furthermore, Yang [80] also demonstrated that the S1 domain of the SARS-CoV-2 spike protein directly binds to ACE-2 receptors expressed by the undifferentiated human alveolar cells (A549) facilitating the SARS-CoV-2 entry into the cells. It was also reported that a SARS-CoV-2 receptor-blocking human antibody, HA001, attaches to amino acids A475 and F486 residues in the RBD of the SARS-CoV-2 spike protein [81]. In this process, a 71 amino acid long segment of the RBD, the RBM, tightly binds to the ACE-2 receptor [82]. Furthermore, the RBD contains nine cysteine residues making four pairs and one free. The three pairs are Cys336–Cys361, Cys379–Cys432 and Cys391–Cys525 residues that form the core RBD β sheet and the remaining pair, Cys480–Cys488 residue, binds to the N-terminal peptidase domain of ACE-2 [83]. All the above experimental evidence supports that the RBD protein will be a highly effective antigen for a SARS-CoV-2 vaccine.

After binding of RBD to ACE-2, two enzymes, furin and TMPRSS2 produced by human mucosal cells, cleave S1 domain from S2 domain leaving S2 domain free [49,84]. Then, the S2 domain protein (with SARS-CoV-2 virus attached) fuses with the human mucosal cell membrane allowing the entry of the virus into human cells followed by the release of SARS-CoV-2 mRNA into the cell cytoplasm.

The mRNA-1273 selected for an antigenic vaccine encodes for the entire spike protein [63,85]. Hence, it produces both the S1 and S2 domains of the S-protein as an antigen, although S2 of SARS-CoV has been reported to produce no immunity against SARS-CoV viral infection in mice.
Furthermore, after the removal of S1 domain, the S2 domain of the SARS-CoV-2 binds to human mucosal cells making the pathway for the SARS-CoV-2 virus to enter the cell.

Now the question remains to be answered, will the free S2 domain protein from the antigen produced by the mRNA vaccines allow entrance of any free SARS-CoV-2 virus into human mucosal cells and cause an infection and thus the disease? With minor differences another mRNA antigenic vaccine currently applied also contains the entire region of the mutant S-protein \[63,87\] gene. Hence, it will also produce both the S1 and S2 domain proteins. Once the S1 domain is separated by the furin and TMPRSS2 from the S2 domain, what is the potential for S2 domain to transfer free SARS-CoV-2 virus into the human mucosal cells causing infection and the disease? Furthermore, using S2 domain independently as an antigen in BALB/c mice failed to produce neutralizing antibodies against infection by SARS-CoV \[86\]. Hence, the free S2 domain released from the S1 domain produced by the mRNA vaccines plays no role to generate immunity against SARS-CoV-2 infections.

**Conclusion**

The RBD protein of SARS-CoV-2 is an excellent choice for developing vaccine to prevent COVID-19. The RBD itself is an antigen and hence, it will be direct and quick in stimulating the recipients’ systems to produce antibody against SARS-CoV-2 virions quickly. Furthermore, since it is not composed of any nucleotides or nucleic acids, it has no potential for homologous recombination in the recipient’s genome. The vaccines composed of antigenic mRNA have a short life in the recipients’ cells, hence, it is possible that some recipients of mRNA vaccines may develop limited immunity against SARS-CoV-2. Furthermore, both antigenic mRNA and DNA have the potential for homologous recombination through reverse transcriptase when present and active in the recipient’s system.

On the other hand, protein vaccines, like the SARS-CoV-2 RBD, can generate a strong immune response and can be used almost by everyone who needs them, including people with weakened immune systems and long-term health problems, according to \[87\]. Furthermore, the RBD protein vaccines cannot cause the COVID-19 disease. However, booster shots may be necessary to have a full protection against the ongoing SARS-CoV-2 infections \[88,89\].

**Acknowledgements**: We would like to thank Dr. Arthee E. Jahangir for her editorial support.
References


81. Yi, C., Sun, X., Ye, J., Ding, L., Liu, M., Yang, Z., Lu, X., Zhang, Y., Ma, L., Gu, W., Qu, A., Xu, J., Shi, Z., Ling, Z., & Sun, B. (2020). Key residues of the receptor binding motif in the spike protein of SARS-CoV-2 that interact with ACE2 and neutralizing antibodies. *Cellular & Molecular Immunology, 17*(6), 621-630. [https://doi.org/10.1038/s41423-020-0458-z](https://doi.org/10.1038/s41423-020-0458-z).


87. [https://www.hhs.gov/immunization/basics/types/index.html](https://www.hhs.gov/immunization/basics/types/index.html)
