An Overview of the SARS-CoV-2 Spike Protein in Current Research

Alexander M. Washington and Timothy M. Reichart

Department of Chemistry, Hampden-Sydney College, Hampden-Sydney, VA 23943

Background

Sars-CoV-2 is the virus responsible for the Covid-19 pandemic that has altered daily life in 2020. It is in the family of Coronaviruses which are known for their unique crown shape on their spike protein (1). Much like the original SARS-CoV, it attacks its host via the respiratory pathway and is namely characterized by various symptoms like a cold (1). In addition, the virus is also enveloped and thus has a membrane containing a phospholipid by-layer with various embedded proteins (2). Due to the unique features of having an envelope, the genome of these kinds of viruses will encode for proteins that can be embedded inside of the membrane that play distinct roles in pathogenicity (3). Viruses that contain a similar structure to this include ones such as the Human Immunodeficiency Virus (HIV) and Influenza, which have been important in the prediction of functions in SARS-CoV-2 (4). This means that the novel coronavirus is an RNA based virus that utilizes proteases (enzymes that digest proteins) to invade the host cell and introduce its genetic material into a foreign genome (4). Given the chance, similar techniques used to treat these specific viruses can be utilized to develop therapies for COVID-19.

SARS-CoV-2 contains a spike protein that is embedded in the membrane that is responsible for the virus to recognize the host (5). It is unique properties like these that are of interest to scientists because it will allow for the development of drugs that can actively inhibit the infection of cells and prevent further damage to the cell. Currently, there are already therapies designed to interact with the angiotensin-converting enzyme (ACE2) protein on SARS-CoV-2 that is responsible for protease activation to enter the host (6). Drug design for proteins can function on various aspects to stop function of the virus. One aspect that holds significance is the ability for protein subunits to oligomerize into bigger units that can be used for active processing (7). Through the use of various biochemical techniques, the active oligomerization state of SARS-CoV-2's Spike protein function can be inferred from an analysis of the protein's unique domains, one being the region that is embedded in the membrane of the envelope (8).

Traditionally, prediction software can look at the functions proteins and predict the use of domains based off of the environment they find themselves in. Unfortunately, there are massive limitations in looking at TM projected domains as this region typically is missed in crystallography experiments (9). To analyze a region embedded in the membrane more closely, one must have to synthetically synthesize the domain and analyze its unique properties separate from the entire protein. This presents an interesting challenge because one must synthesize a peptide (chain of amino acids) that correspond to the proper formation seen in a natural protein encoded by the virus's central dogma. Accuracy in each step of the synthesis would be required as one could only learn useful information from a peptide that would be identical to the natural TM domain (10). To do this, Solid Phase Peptide Synthesis (SPPS) must be done to synthetically make a transmembrane domain for analysis.

The main portion of our research done here at Hampden-Sydney college is dedicated to using SPPS so that way may answer vital questions about protein structure that could lead to the development of therapeutics. In just a short period of time, the worldwide dedication to fighting the pandemic has led to an acceleration of our understanding of the virus and the various proteins that make it up. In this short review article, we will take a synopsis of the various findings that have come out about SARS CoV-2 to appreciate the reach of modern science. Despite the arduous struggles of this past year, many advances have been made to expand our understanding of the virus.

Structure and Variants

Coronaviruses are composed of a single stranded positive-RNA genome that encodes directly the mRNA needed to translate the proteins of the virus from host ribosomes. This process of replication will occur directly in the endoplasmic reticulum of the cell and will form an envelope that embeds the key proteins needed to produce an infectious virus (3). These proteins are the spike, envelope, membrane, and nucleoprotein that make up the essential structure and are encoded by the genome (3). The function of these proteins carries unique features and must replicate appropriately to carry the function properly.

For the functioning of the SARS-CoV 2 spike protein, it contains a set of two subunits that make up the entire glycoprotein. Subunit one holds the receptor binding domain (RBD) where subunit 2 contains the fusion peptide needed to fuse the host cell and virus membranes together (11). This feature is important for the infectivity of a virus and requires the recognition of the spike to those proteases to allow for proper recognition, showing another level of reliance on host machinery (11). One protease is the furin protease that

cleaves the arginine residues linking both subunits together (12). In direct interaction with the host ACE2 receptor, the residue similarity in binding affinity to the original SARS-CoV-2 is conserved showing that evolution has decided that this a vital structure for infection (12). Furthermore, some studies have shown that there is a structural desire for sparse placement of the virus on the envelope to provide stability (13). These facts about the virus become important when looking at changes in structure on the spike between different variants.

Where SARS-CoV 2 differs from the original SARS in terms of animal host transmission and varying symptoms, the variants of coronavirus have unique modifications to their spike protein that give them an advantage. Like animals, viruses evolve over time and can be correlated to common ancestors based on its similarities. Since the virus genome is composed of RNA, it lacks the proofreading mechanisms to prevent mutations from occurring. This feature will provide the virus the capability of conferring an advantage from mutation and allow it to better infect its host (14). Thus, a large component of the pandemic's response is in sequencing the genomes of viruses collected from patients to see what mutations are circulating the population (14). When these studies were started, scientists were worried about the possibility of there being multiple introductions of the virus into populations that would result in the mixing of genomes known as recombination. The ACE2 receptor of human beings also exists as many variants as possible throughout the population which also in turn will result in a change of binding strength to the spike protein (15). An analysis of the ACE2 gene has yielded several variants that exist in wild type cells that contribute to the efficiency of SARS binding. With the mixing of different viral genomes, there is constant need for there to be mutations that will confer the best possible binding to its host (11).

Unfortunately, the variations in the spike gene pose the most significant threat to public health as this gene is the most responsible for invading the immune system. When antibodies are developed by the immune system to attack the pathogen, it will bind to the spike to inhibit the virus's activity based on which peptide was presented to the immune system. These "epitopes" can escape this response if there is a strong enough change to prevent the antibody bonding (16). One observed region is the cysteine residue 135 in the CoV-47 variants receptor binding domain as monoclonal antibodies were no longer able to recognize the epitope (16). Fortunately, the immune system can recognize a plethora of epitopes from viruses so a significant number of mutations will be required before total escape can occur (16). Such mutations will become a concern when epitopes of the virus's spike protein develop to lose their immunogenic response and evade current therapeutics (17).

Vaccine Potential

Due to the importance of the SARS-CoV-2 spike protein for the virus infection cycle, antibodies typically have the strongest immune response to it. Currently, vaccines on the market are designed to respond to this protein to inhibit the virus. In drug design, a particular issue that needs to be overcome is the potential for antibodies to enhance infection, which is a process known as antibody-dependent-enhancement (ADE) (3). This problem can be overcome by the presentation of epitopes that serve to neutralize the peptide either at where it fuses or at the receptor binding domain. To avoid potential issues with vaccination, the safest forms of viral immunization involve RNA and DNA vaccines as they will allow for the replication of only specific proteins from the virus (18). The Pfizer RNA liposome vaccine utilizes this technology to deliver an RNA sequence of the spike protein to the host that will be taken into cells to produce the sequence and be phagocytized by the immune system. Allowing for the immune system to respond to the whole protein will allow the body to develop natural antibody responses that goes through its own selection for safe antibodies (19). As a result, vaccines that incorporate the Spike protein are the safest possible vaccines people can be given (20).

Conclusion

Despite the extensive knowledge the scientific community has gained about the working of SARS-CoV-2 proteins, there is still much to be understood about the properties of the virus. The spike protein holds the most importance in terms of promising vaccine treatments. With this, the logistics of viral mutation must be understood to recognize conditions for the spike to no longer be recognized by antibodies. While no such mutation has yet to occur, having enough data on spike epitopes will allow for swift action to be taken on new variants. It is for this reason that the spike protein has been highlighted the most in research on this pandemic.

REFERENCES

1. Hu, B et al. "Characteristics Of SARS-CoV-2 and COVID-19". Nature Reviews Microbiology. 2020

2. Schoeman, D; Fielding, B. “Coronavirus envelope protein: Current Knowledge”. Virology Journal. 16:69. 2019
3. Nieto-Torre, J et al. “Subcellular location and topology of severe acute respiratory syndrome coronavirus envelope protein”. Virology. 415. 2011
4. Huang, Y et al. “Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19” Acta Pharmacologica Sinica. 2020
5. Cai, Y et al. “Distinct conformation states of SARS-CoV-2 Spike protein”. Research. 2020
6. Ortega, J et al. “ROLE OF CHANGES IN SARS-COV-2 SPIKE PROTEIN IN THE INTERACTION WITH THE HUMAN ACE2 RECEPTOR: AN IN-SILICO ANALYSIS”. Excli Journal. 2020
7. Reichart, T et al. “Trimerization of the HIV Transmembrane Domain in Lipid Bilayers Modulates Broadly Neutralizing Antibody Binding” Angewandte Chemie. 2016
8. Reichart, T et al. “The Conserved Positive charge in the transmembrane domain of HIV gp41 contributes to its trimerization.”
9. Gratkowski, H et al. “Polar Side chains drive the association of model transmembrane peptides”. PNAS. 2001
10. Palomo, J. “Solid-Phase peptide synthesis: an overview focused on the preparation of biologically relevant peptides”. Royal Society of Chemistry. 2014
11. Devaux, C. “ACE2 receptor polymorphism: Susceptibility to SARS-CoV-2, hypertension, multi-organ failure, and COVID-19 disease outcome”. Journal of Microbiology. 2020.
12. Yang, Y. “SARS-CoV-2: characteristics and current advances in research”. Viology Journal. 2020.
13. Ke, Z. “Structures and distributions of SARS-CoV-2 spike proteins on intact virions”. Nature. 2020.
14. Kirtipal, N. “From SARS to SARS-CoV-2, insights on structure, pathogenicity and immunity aspects of pandemic human coronaviruses”. Infection, Genetics and Evolution. 2020.
15. Hussain, M. “Structural variations in human ACE2 may influence its binding with SARS-CoV-2 spike protein”. Journal of Medical Virology. 2020.
16. Weisblum, Y. “Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants”. eLife. 2020
17. Yadav, P. “Full-genome sequences of the first two SARS-CoV-2 viruses from India”. Indian Journal of Medical Research. 2020
18. Li, Y. “Coronavirus vaccine development from SARS and MERS to COVID-19”. Journal of Biomedical Science. 2020.
19. CDC EUA of Pfeizer- Biontech COVID-19 Vaccine Fact sheet.
20. Grigoryan, L. “The immunology of SARS-CoV-2 infections and vaccines. Seminars in Immunology. 2020