

SARS-COV-2 Spike Protein: Characteristics and Treatments

Ming Chit Calvin Chick *

Department of Life Sciences, Imperial College London, London, United Kingdom

* Corresponding Author Email: calvin.chick20@imperial.ac.uk

Abstract. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) belongs to the family of beta coronavirus. The virus led to the outbreak of COVID-19 in early 2020. Characterized by its Spike (S) proteins, SARS-CoV-2 can recognize the human angiotensin-converting enzyme 2 (ACE2) receptor. Viral membrane fusion is initiated by the interaction between the S protein and the human ACE2 receptor. Due to its crucial role in receptor recognition, the S protein attracted great interest in being a drug target. The structure and functions of different domains of the S protein are presented in this article. Current progress in therapeutic compounds which targets the Receptor Binding Domain (RBD), Heptapeptide repeat sequence (HR) and the Fusion Loop (FL) domain are highlighted. Acquiring information on the S protein from a structural perspective and understanding treatment opportunities against SARS-CoV-2 will assist in combatting the disease.

Keywords: Protein Structure; SARS-CoV-2 Spike Protein; Spike Protein Treatments.

1. Introduction

COVID-19 is caused by a beta coronavirus later known as SARS-CoV-2. The virus caused a pandemic, where an estimated 607,745,726 confirmed cases were reported, while 6,496,721 deaths were caused worldwide as of 15 September 2022 [1]. The pandemic took a heavy toll on the global economy, despite efforts to combat it.

SARS-CoV-2 is an enveloped, single-stranded, positive-sense RNA virus. Accessory, non-structural, and structural proteins are encoded within the genomes of the virus [2]. Airborne respiratory droplets are the main median for SARS-CoV-2 transmission. Viral infection begins with viral entry, where the surface S glycoprotein binds to the human ACE2 receptor. The host transmembrane, protease serine 2 (TMPRSS2), then cleaves the S protein, facilitating viral membrane fusion. After membrane fusion, the viral genome is released and translated [2].

Despite being effective at preventing individuals from contracting the disease, vaccines do not protect the ones who already have the disease [3]. Moreover, vaccine rollout was far from sufficient and satisfactory in less developed countries. Effective drugs against SARS-CoV-2, therefore, are needed to contain the rampant infection.

Currently, drugs that directly target the virus are classified into two types: ones that target the structural portion of the virus and those that target the non-structural part of the virus [4]. Remdesivir is an example of a drug that targets the non-structural protein of the virus. Remdesivir limits viral replication by targeting the RNA Dependent RNA Polymerase [4]. Despite being one of the earlier drugs approved by FDA to combat SARS-CoV-2, previous trials indicated that it had no statistically significant clinical improvement for patients burdened with the disease [5].

Targeting the S protein to prevent viral binding is an approach to combatting SARS-CoV-2, given its importance in initiating viral infection. The structure of the S protein is presented in this article. A summary of current findings on therapeutic compounds targeting different domains of the S protein and their clinical implications is also presented.

2. Structural overview of the S protein

The S protein has a length of 1,273 amino acid residues and comprises three distinct domains: the N-terminal signal peptide (residue 1-13), the S1 subunit (residue 14-685) and the S2 subunit (residue 686-1273) [2]. The S protein is found on the virus's surface and exists as a trimer when not bound to its target receptor. S protein is in its native state (adopting the "pre-fusion" conformation) prior to

receptor binding. Upon receptor binding and proteolytic cleavage, the S protein will experience conformational changes, adopting the “post-fusion” conformation. Viral membrane fusion is initiated only when the “post-fusion” conformation is adopted [6].

The S1 subunit has an N-terminal domain (residue 14-305) and an RBD (residue 319-541). The S2 subunit comprises 6 domains: FL (residue 788-806), heptapeptide repeat sequence 1 (HR1) (residue 912-984), heptapeptide repeat sequence 2 (HR2) (residue 1163-1212), transmembrane domain (TM) (residue 1213-1236) and cytoplasmic domain (CD) (residue 1237-1273) (Figure 1) [2].

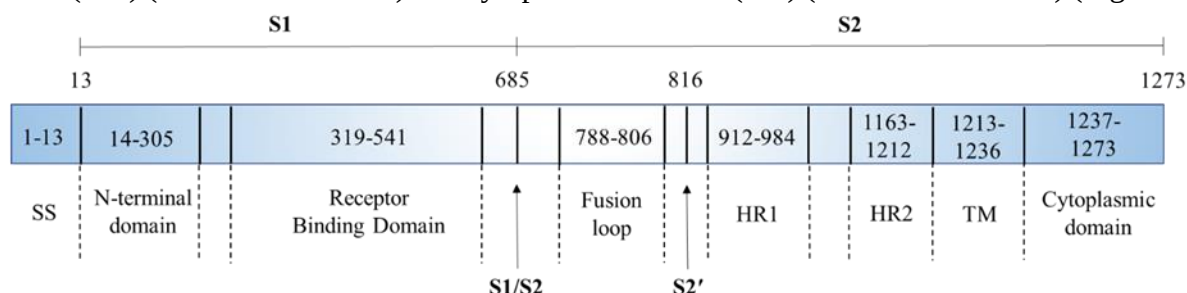


Figure 1. Representation of the S protein. SS: Signal sequence. HR1: Heptapeptide repeat sequence 1. HR2: Heptapeptide repeat sequence 2. TM: Transmembrane domain. S1/S2: S1/S2 cleavage site. S2': S2' cleavage site.

The crystal structure of the S protein trimer adopting the “prefusion state” is depicted in Figure 2. The RBD of each S protein can adopt two unique conformations: “up” and “down”. In the “up” conformation, the S protein is active and readily available to accept ACE2 receptor binding, while in the “down” state, the S protein is inactive [6].

Upon receptor binding, the S1/S2 and S2' sites will be cleaved by the host proteases (Figure 1). Proteolytic cleavage will lead to the dissociation of the S1 subunit, and act as a signal for the S protein to undergo a conformational change. The S protein will now exist in the “post-fusion” state (Figure 3) [6]. In the “post-fusion” state, the HR domains will interact and form a helical bundle. The helical bundle then functions by bringing the virus near the host membrane, enabling FP to insert [6].

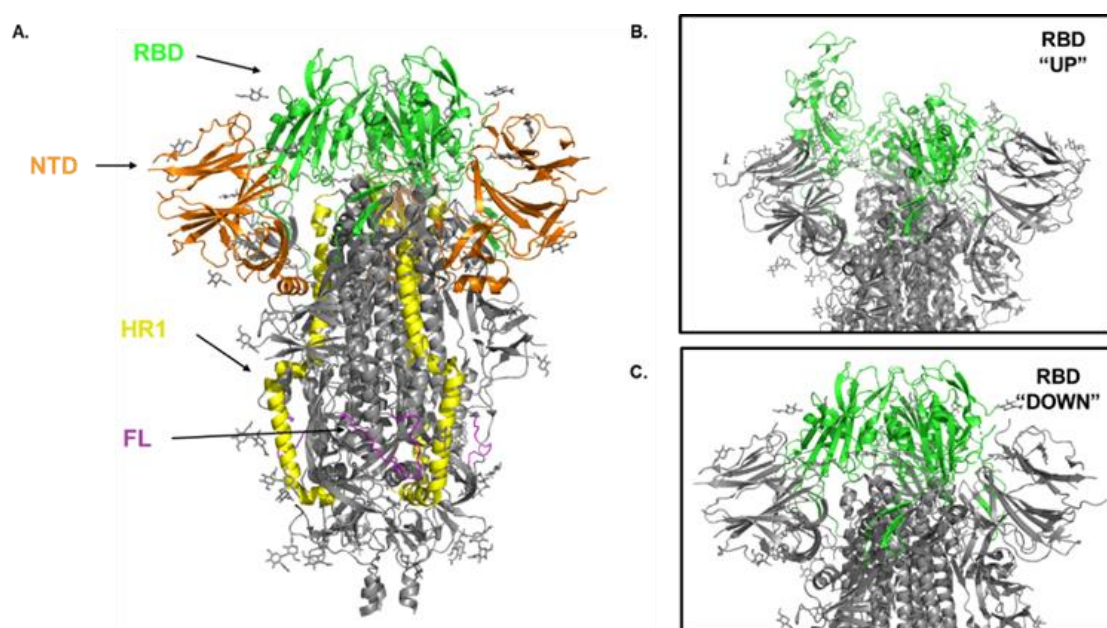


Figure 2. S protein in the “pre-fusion” state. Figures were illustrated using PyMol [7]. (A) S protein adopting the “pre-fusion” conformation (PDB ID: 6VXX). (B) S protein with one RBD adopting the “up” conformation and two RBD adopting the “down” conformation (PDB ID: 7KRR). (C) S protein with all 3 RBD adopting the “down” conformation (PDB ID: 6VXX).

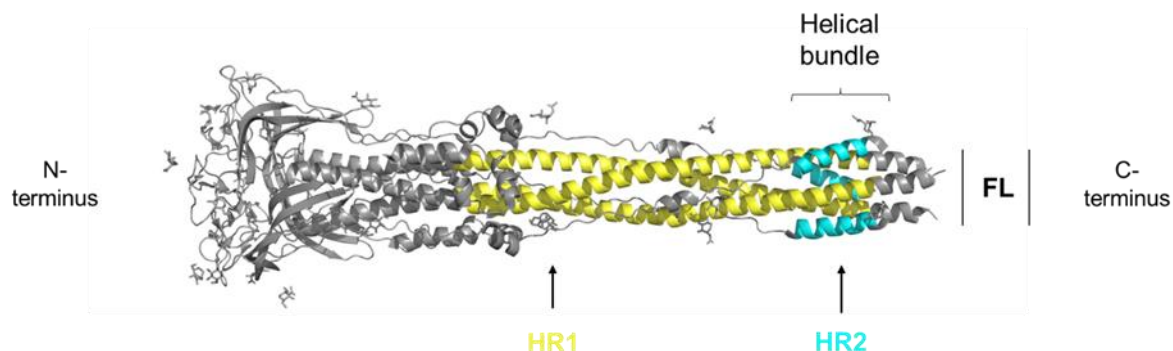


Figure 3. S2 subunit adopting the “post-fusion” conformation. Figure illustrated using PyMol [7] (PDB ID: 6M3W).

3. Functions of S protein domains and treatments against them

3.1. RBD

The RBD is situated at the S1 subunit and comprises two distinct domains: a 5-stranded antiparallel β -pleated sheet and a receptor binding motif (RBM) [6]. SARS-CoV-2 utilises the RBM domain to make direct contact with the human ACE2 receptor.

Crystal structures indicate that the S protein trimer could bind to a maximum of 3 human ACE2 receptors simultaneously (Figure 2). Three binding and mutational hotspots for SARS-CoV-2 RBD were previously defined: hotspot 31, hotspot 353 and hotspot ridge [8]. These three areas are where most interactions take place (Figure 4). The most common interactions between the amino acids are hydrogen bonds and salt bridges, which function by stabilising protein binding. Instances would be the hydrogen bonds formed between TYR449, GLN493, and GLN493 of the RBM with ASP38, GLU35 and GLU35 of the human ACE2 receptor (Figure 4). Hydrophobic interactions between PHE486 of the RBM and TYR83 of the human ACE2 receptor are less common but also enhance binding stability [6].

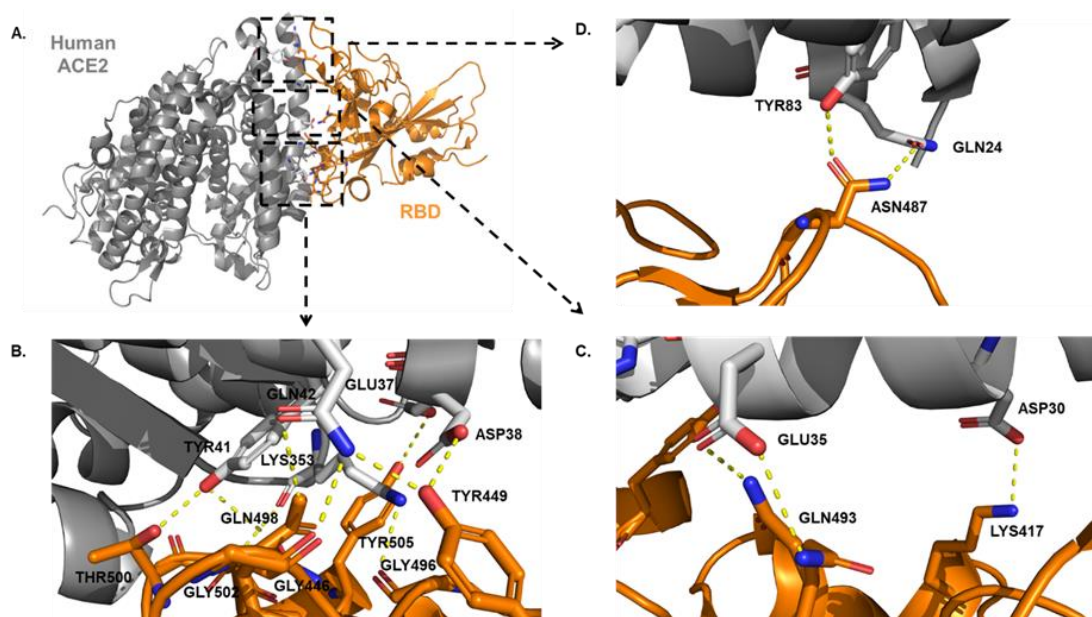


Figure 4. Binding between the human ACE2 receptor and the RBD. The RBD is coloured orange, while the human ACE2 receptor is grey. All Hydrogen bonds are limited to a maximum length of 3.5Å. Figures were illustrated using PyMoL [7] (PDB ID: 6m0j). (A) RBD in complex with the human ACE2 receptor. (B) Hydrogen bonds formed at Hotspot 353. (C) Hydrogen bonds formed at Hotspot 31. (D) Hydrogen bonds formed at Hotspot ridge.

Due to its importance in viral infection initiation, the RBD received significant interest as a potential drug target. The goal of the medication is to inhibit the interaction of the virus with the host cell by competing with human ACE2 receptors. Monoclonal antibodies (mAb) were designed to target the RBD. There are four distinct classes of mAb (Class I, II, III and IV) targeting the RBD. They are classified based on their ability to bind to the different conformations adopted by the RBD [9].

Bamlanivimab (LY-CoV555) is an example of a Class II mAb which binds to the RBD at both “up” and “down” conformation. Bamlanivimab was authorized by the FDA to be used against patients with the potential of developing severe illnesses caused by SARS-CoV-2 [9]. The effects of Bamlanivimab are indicated through crystal structures generated from X-ray crystallography, showing the interaction between Bamlanivimab with the RBD-ACE2 complex (Figure 5, withdrawn from [10]). Bryan et al. showed that Bamlanivimab could bind to the RBD at the hotspot ridge, overlapping its binding site with the human ACE2 receptor [11]. Moreover, it was shown that the binding site for Bamlanivimab is accessible on both the “up” and “down” conformation of the RBD, rendering the drug even more efficacious [11].

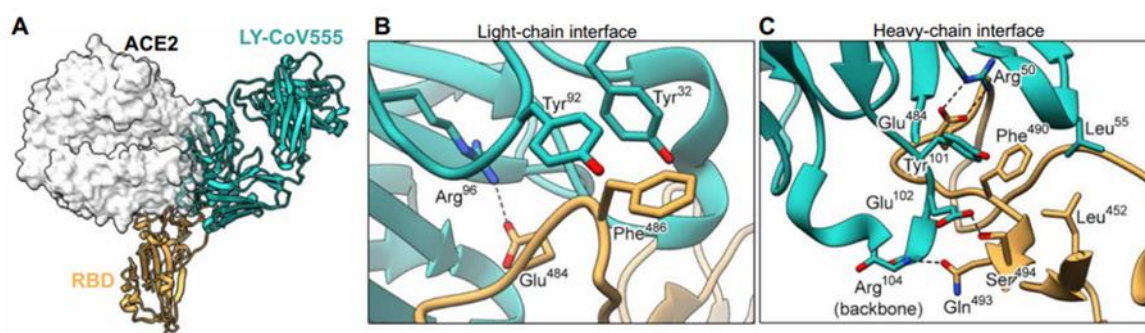


Figure 5. Bamlanivimab competes with the human ACE2 receptor at the hotspot ridge. (A) Crystal structures of Bamlanivimab overlapping with RBD-human ACE2 receptor complex. (B) Interactions made by the light chain of Bamlanivimab with the RBD. (C) Interactions made by the heavy chain of Bamlanivimab with the RBD. Figure withdrawn from [10].

The efficacy of Bamlanivimab was further illustrated through animal trials. Bryan et al. used rhesus macaque as the model and compared the number of sub-genomic RNA (sgRNA) and viral genomes (gRNA), an indication of viral replication, between control and treated samples. Researchers discovered that samples treated with Bamlanivimab had significantly reduced sgRNA and gRNA counts in both lung and bronchoalveolar lavage fluid tissues, indicating that viral replication was suppressed by Bamlanivimab [11]. Results from phase II trials of Bamlanivimab were in line with this. Hospitalisation rates, on day 29 of the test, for patients who received a placebo and patients who received Bamlanivimab were 6.3% and 1.6%, respectively. In addition, patients who received 2800mg of Bamlanivimab experienced a higher degree of viral load reduction, measured through reverse-transcriptase PCR (RT-PCR), compared to patients who received a placebo [10].

3.2. HR1 Domain

HR1 and HR2 domain is situated on the S2 subunit, between the fusion loop and the TM domain. HR1 and HR2 are two essential domains for viral membrane fusion as they function by bringing the virus into proximity with the human ACE2 receptor. HR1 and HR2 form a six-membrane helical bundle upon receptor binding and protease activation (Figure 3).

Both domains are composed of repetitive heptapeptides [2]. (HPPHCPC)_n is a representation of the repeat sequence. The notation for the sequence goes as follows: H denotes hydrophobic residues, P denotes polar residues, and C denotes charged residues [2]. Previous studies also indicated that HR1 and HR2 are poor in glycan, suggesting that the structure is relatively rigid [12].

Interestingly, the amino acid sequence for HR1 and HR2 are highly conserved among coronaviruses [12]. This and its crucial role in viral membrane fusion made HR1 and HR2 promising drug targets.

Fusion inhibitors were previously created to bind to the HR domains, abolishing their ability to enhance membrane fusion. An example would be the lipoprotein-linked drug EK1C4, developed based on EK1 (a fusion inhibitor designed initially to combat SARS-CoV) [13]. Inhibitory activities were measured using a cell-cell fusion assay. The results showed that against SARS-CoV-2 infection, EK1 had an IC_{50} of 2,375 nM, while EK1C4 had an IC_{50} of 15.8 nM. This indicates that EK1C4 is 149-fold more effective at inhibiting viral membrane fusion than EK1 [13]. Furthermore, the efficacy of EK1C4 was determined using transgenic mouse models. Xia and colleagues measured the viral titer within the lungs and intestines of mouse control samples and mouse samples treated with EK1C4. Their results indicated that nasal administration of EK1C4 grants mice protection from SARS-CoV-2 [14].

3.3. FL Domain

The FL domain is situated within the S2 subunit of the S protein, between the RBD and the HR1 domain. Conformational changes within the S2 subunit are initiated through ACE2 receptor binding, followed by proteolytic cleavage at the S1/S2 site [12]. The FL will then act as an anchor, penetrating the host membrane. The FL domain, therefore, is a critical first step to viral membrane fusion.

Crystal structures of the FL domain showed that the structure is abundant in non-polar amino acid residues [2]. A high abundance of nonpolar residues, such as tryptophan, made the domain hydrophobic, which is expected to assist with membrane penetration and fusion. Figure 6 shows the potential binding pocket for the FL domain (Figure withdrawn from [15]).

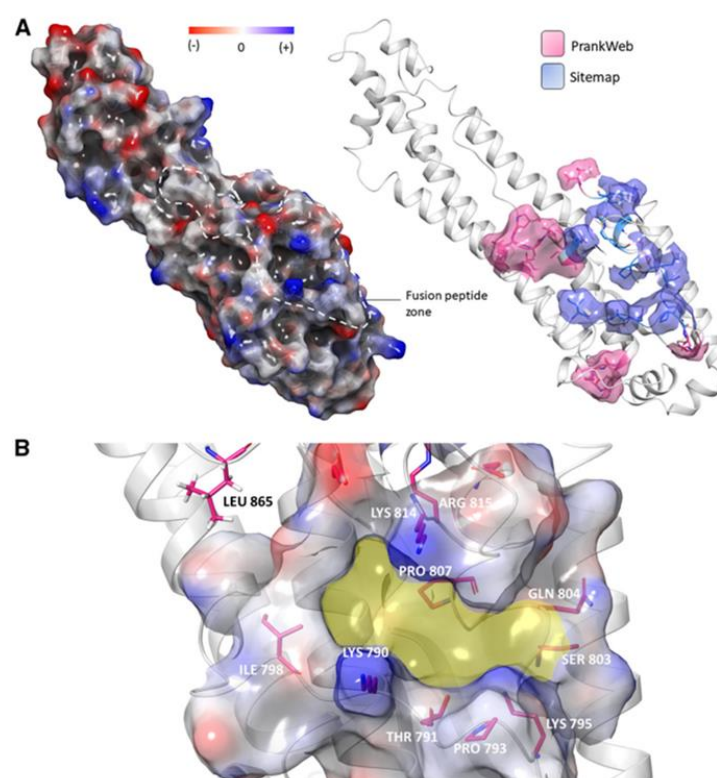


Figure 6. Crystal structure of the FL domain and its binding pocket. (A) Left: Structure illustrating the polarity of residues within the FL domain. Right: Residues predicted to be a part of the binding pocket, predicted with PrankWeb and Sitemap. (B) The estimated position of the binding pocket (labelled in yellow). Figure withdrawn from [15].

Like the HR1 domain, the FL domain is crucial in membrane fusion and is seen as a potential drug target. COV44-62 and COV44-79 are mAbs previously designed to bind to the FL. In-vitro binding

assay indicated that both COV44-62 and COV44-79 could prevent SARS-CoV-2 S protein from fusing with cells expressing human ACE2 receptor [16]. Moreover, test results derived from Syrian hamster models indicated that recovery speeds were higher for hamsters treated with COV44-62 and COV44-79 than untreated hamsters [16].

4. Conclusions

Extensive research and efforts have been made to pursue an effective treatment against SARS-CoV-2 ever since its first outbreak in early 2020. Despite the development of vaccines that could protect individuals against the virus, there is still a demand for specific drugs which would render the virus harmless.

S1 and S2 subunits are the 2 domains that comprise the S protein. Unique domains within the subunits function differently to assist with membrane fusion. SARS-CoV-2 infection initiates through S protein-directed membrane fusion. The S protein will adopt the “pre-fusion” conformation prior to receptor binding and undergoes conformation changes triggered by receptor binding and proteolytic cleavage to adopt the “post-fusion” conformation.

S protein and its crucial role in viral infection made it a promising target for anti-COVID-19 drugs. The S protein RBD makes direct contact with the human ACE2 receptor. Drugs such as Bamlanivimab, previously approved by the FDA, function by competing with the human ACE2 receptor for RBD. Alternatively, drugs designed for the HR1 domain act as a fusion inhibitor, abolishing the virus’s ability to fuse with the host membrane. An instance would be EK1C4, where experimental results from mouse models indicated its efficiency. Other fusion inhibitors include COV44-62 and COV44-79, which are mAbs targeting the FL domain. These two drugs were found to bolster the recovery speed of SARS-CoV-2-infected hamsters.

There is still a persistent and urgent demand for an effective drug against SARS-CoV-2. Further efforts placed into targeting the S protein are of the utmost importance.

References

- [1] WHO Coronavirus (COVID-19) Dashboard | WHO Coronavirus (COVID-19) Dashboard With Vaccination Data [EB/OL]. [2022-09-18]. <https://covid19.who.int/>.
- [2] HUANG Y, YANG C, XU X feng, et al. Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19 [J]. *Acta Pharmacologica Sinica*, 2020, 41(9): 1141-1149.
- [3] Liu H, Lin H, Shen C, et al. A network representation approach for COVID-19 drug recommendation [J]. *Methods*, 2022, 198: 3-10.
- [4] Gil C, Ginex T, Maestro I, et al. COVID-19: Drug Targets and Potential Treatments [J]. *Journal of Medicinal Chemistry*, 2020, 63(21): 12359-12386.
- [5] Wang Y, Zhang D, Du G, et al. Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial [J]. *Lancet*, 2020, 395(10236): 1569-1578.
- [6] Zhang J, Xiao T, Cai Y, et al. Structure of SARS-CoV-2 spike protein [J]. *Current opinion in virology*, 2021, 50: 173-182.
- [7] PyMOL | pymol.org [EB/OL]. [2022-09-22]. <https://pymol.org/2/>.
- [8] Shang J, Ye G, Shi K, et al. Structural basis of receptor recognition by SARS-CoV-2 [J]. *Nature*, 2020, 581(7807): 221-224.
- [9] Kumar S, Chandele A, Sharma A. Current status of therapeutic monoclonal antibodies against SARS-CoV-2 [J]. *PLoS Pathogens*, 2021, 17(9): e1009885.
- [10] Chen P, Nirula A, Heller B, et al. SARS-CoV-2 Neutralizing Antibody LY-CoV555 in Outpatients with Covid-19 [J]. *The New England journal of medicine*, 2021, 384(3): 229-237.

- [11] Jones B E, Brown-Augsburger P L, Corbett K S, et al. The neutralizing antibody, LY-CoV555, protects against SARS-CoV-2 infection in nonhuman primates [J]. *Science translational medicine*, 2021, 13(593): eabf1906.
- [12] Xia X. Domains and Functions of Spike Protein in Sars-Cov-2 in the Context of Vaccine Design [J]. *Viruses*, 2021, 13(1): 109.
- [13] Xia S, Liu M, Wang C, et al. Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion [J]. *Cell Research*, 2020, 30(4): 343-355.
- [14] Xia S, Lan Q, Zhu Y, et al. Structural and functional basis for pan-CoV fusion inhibitors against SARS-CoV-2 and its variants with preclinical evaluation [J]. *Signal Transduction and Targeted Therapy*, 2021, 6(1): 288.
- [15] Shekhar N, Sarma P, Prajapat M, et al. In Silico Structure-Based Repositioning of Approved Drugs for Spike Glycoprotein S2 Domain Fusion Peptide of SARS-CoV-2: Rationale from Molecular Dynamics and Binding Free Energy Calculations [J]. *mSystems*, 2020, 5(5): e00382-20.
- [16] Dacon C, Tucker C, Peng L, et al. Broadly neutralizing antibodies target the coronavirus fusion peptide [J]. *Science*, 2022, 377(6607): 728-735.