

Original Research

# In Silico Optimization of SARS-CoV-2 Spike Specific Nanobodies

Xiaohong Zhu<sup>1,2,†</sup>, Ke An<sup>1,2,†</sup>, Junfang Yan<sup>1</sup>, Peiyi Xu<sup>1</sup>, Chen Bai<sup>1,3,\*</sup>

<sup>1</sup>Warshel Institute for Computational Biology, School of Life and Health Sciences, School of Medicine, The Chinese University of Hong Kong, 518172 Shenzhen, Guangdong, China

<sup>2</sup>School of Chemistry and Materials Science, University of Science and Technology of China, 230026 Hefei, Anhui, China

<sup>3</sup>Chenzhu Biotechnology Co., Ltd., 310005 Hangzhou, Zhejiang, China

\*Correspondence: [baichen@cuhk.edu.cn](mailto:baichen@cuhk.edu.cn) (Chen Bai)

†These authors contributed equally.

Academic Editor: Amedeo Amedei

Submitted: 24 September 2022 Revised: 1 December 2022 Accepted: 12 December 2022 Published: 6 April 2023

## Abstract

**Background:** The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has spread worldwide, caused a global pandemic, and killed millions of people. The spike protein embedded in the viral membrane is essential for recognizing human receptors and invading host cells. Many nanobodies have been designed to block the interaction between spike and other proteins. However, the constantly emerging viral variants limit the effectiveness of these therapeutic nanobodies. Therefore, it is necessary to find a prospective antibody designing and optimization approach to deal with existing or future viral variants. **Methods:** We attempted to optimize nanobody sequences based on the understanding of molecular details by using computational approaches. First, we employed a coarse-grained (CG) model to learn the energetic mechanism of the spike protein activation. Next, we analyzed the binding modes of several representative nanobodies with the spike protein and identified the key residues on their interfaces. Then, we performed saturated mutagenesis of these key residue sites and employed the CG model to calculate the binding energies. **Results:** Based on analysis of the folding energy of the angiotensin-converting enzyme 2 (ACE2)-spike complex, we constructed a detailed free energy profile of the activation process of the spike protein which provided a clear mechanistic explanation. In addition, by analyzing the results of binding free energy changes following mutations, we determined how the mutations can improve the complementarity with the nanobodies on spike protein. Then we chose 7KSG nanobody as a template for further optimization and designed four potent nanobodies. Finally, based on the results of the single-site saturated mutagenesis in complementarity determining regions (CDRs), combinations of mutations were performed. We designed four novel, potent nanobodies, all exhibiting higher binding affinity to the spike protein than the original ones. **Conclusions:** These results provide a molecular basis for the interactions between spike protein and antibodies and promote the development of new specific neutralizing nanobodies.

**Keywords:** SARS-CoV-2 spike protein; nanobody; coarse-grained (CG) model; binding free energy

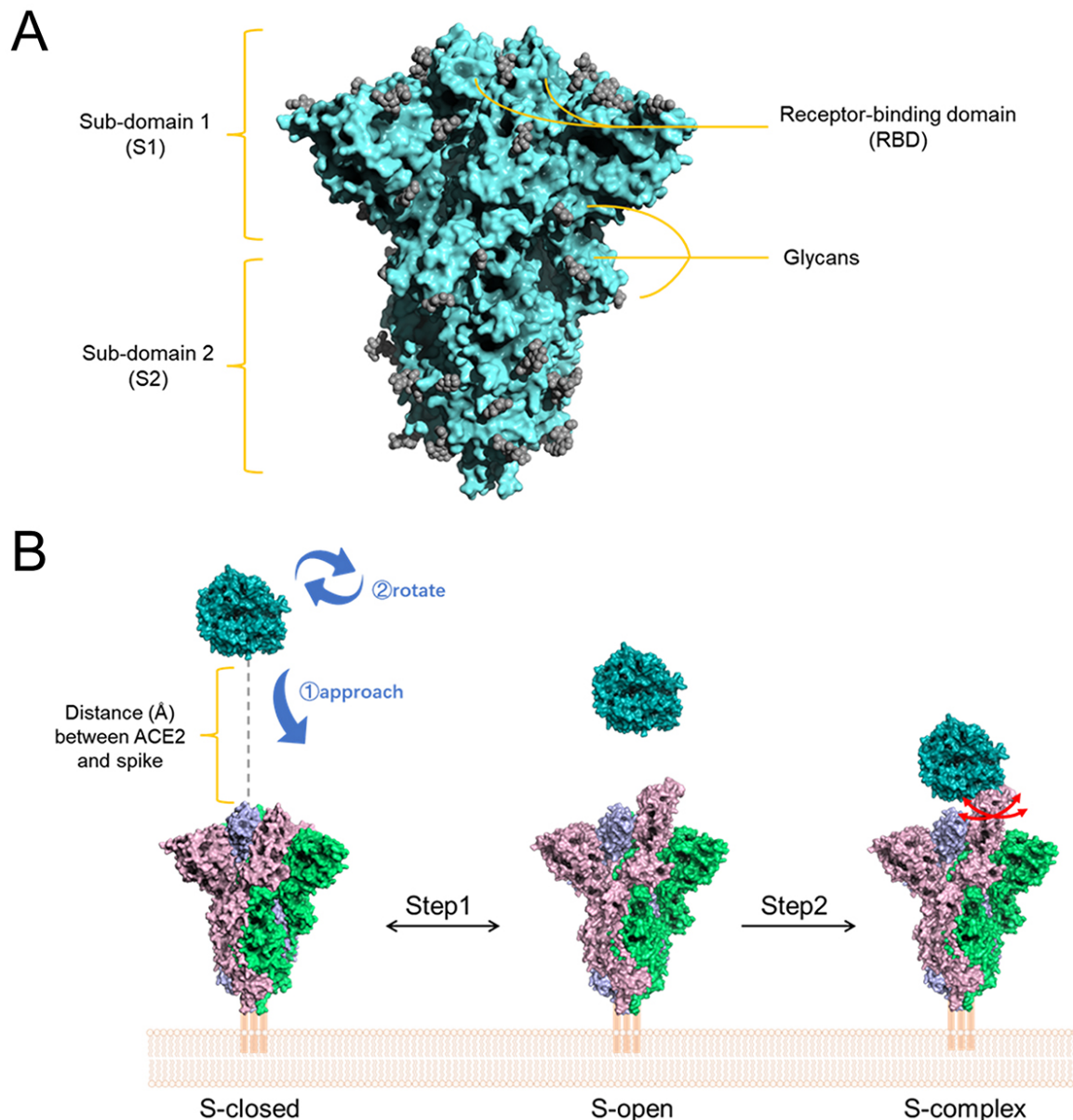
## 1. Introduction

The spread of the coronavirus disease 2019 (COVID-19) pandemic caused by severe acute respiratory syndrome, has resulted in over 608 million infections and more than 6.50 million deaths worldwide as of September 16th, 2022 (<https://www.who.int/emergencies/diseases/novel-coronavirus-2019>). The COVID-19 pandemic has emerged as a global international health crisis world with far-reaching implications for the global economy, science, peace, and security [1]. Therefore, to meet this challenge, tremendous efforts have been devoted to developing therapeutic approaches against SARS-CoV-2.

The SARS-CoV-2 spike protein is the focus of therapeutic and vaccine developmental efforts. The spike protein of SARS-CoV-2 is a large class I trimeric fusion protein, which consists of two subunits, S1 and S2 (Fig. 1A, Ref. [2]) [3–5]. The S1 subunit contains a receptor-binding domain (RBD) which is responsible for the interaction with angiotensin-converting enzyme 2 (ACE2) to gain entry into the host [6], while the S2 subunit mediates membrane fu-

sion and viral entry [7]. Another key structural feature of the spike protein is its extensive glycosylation which plays a crucial role in viral pathogenesis (gray part in Fig. 1A) [8]. The activation of SARS-CoV-2 spike protein is closely related to the approach and binding of the ACE2 receptor. Xu *et al.* [9] described the process of spike activation and elucidated the Cryo-EM structures of three key conformational states of the spike trimer (Fig. 1B). Their results suggested that ACE2 facilitates the capture of the pre-existing open conformation of S trimers rather than triggering a trimer opening event. Under the ACE2-free condition, the majority of the spike trimers (94%) are in the tightly closed ground prefusion state (S-closed), and only a minority (6%) are in the intrinsically transient open state with one RBD up representing the fusion-prone state (S-open), forming a dynamic balance between the two states under equilibrium conditions. The ACE2 would trap the RBD and then overcome the energy barrier, break the balance, and shift the conformational landscape toward the open state. Once the ACE2 traps the up RBD, the associated ACE2-RBD





**Fig. 1. The structure and the activation process of SARS-CoV-2 spike protein.** (A) The structure of SARS-CoV-2 spike protein. The dashed line indicates the rotation symmetry axis of the spike trimers. (B) The activation process of SARS-CoV-2 spike protein. Three conformations of the spike protein of SARS-CoV-2 (trimers in forest green, glass green, and lavender pink), and the position of the ACE2 receptor (yellow). The shallow yellow one represents the initial position of ACE2 when binding to the spike protein. The binding poses of ACE2 to the spike protein are obtained by using a protein-protein docking method, HDOCK [2]. The red dot indicates the center of the ACE2. The distance between ACE2 and the spike is defined as the distance between the center of the distant ACE2 (yellow) and the center of ACE2 in the initial position (shallow yellow).

exhibits combined continuous swing motions on the top-most surface of the S trimer. After ACE2 binding, there are 26.2% in S-open state, 73.8% in S-ACE2 state (an open state when the spike protein binds to the ACE2 receptor, S-complex state), and 0% in S-closed state. After the spike protein binds to the receptor, TM protease serine 2 (TM-PRSS2) [10], a type 2 TM serine protease located on the host cell membrane, promotes virus entry into the cell by activating the spike protein. In order to fulfill the function

of the spike protein, the spike protein of SARS-CoV-2 first binds to the receptor ACE2 through the RBD and then is proteolytically activated by human proteases. Therefore, blocking the interaction between RBD and ACE2 plays a vital role in inhibiting the infection of pathogenic SARS-CoV-2 in the host cells.

Nanobodies are single-domain antibodies derived from camelids and sharks. They show a large sequence identity with the human VH gene family III [11]. Nanobod-

ies have favorable biomedical properties, including high thermostability, high solubility, and deep tissue penetration because of their small size (~15 kDa). Thus, nanobodies are popular for many biotechnology and medical applications [12–17]. Recently, the potency of nanobodies against SARS-CoV-2 infection has been demonstrated in cell-based assays [18–27] and most recently in animal studies [28,29]. The high preclinical efficacy of an ultrapotent nanobody construct (PiN-21) has been demonstrated to prevent viral pneumonia at a very low dose (0.2 mg/kg) [28]. Zhou *et al.* [27] employed nanobody maturation technology to develop several nanobodies targeting SARSCoV-2 spike protein. Their crystal structures showed that the nanobodies successfully block the interaction between RBD and ACE2 [27]. Thus, stable and potent nanobodies that target the RBD of SARS-CoV-2 are promising therapeutics to help mitigate the evolving pandemic.

In recent years, numerous efforts have been made in the development of vaccines, but no completely efficient treatment has yet been found. Therefore, in this study, we report a computational approach to optimize and design potent nanobodies that broadly target all SARS-CoV-2 variants. First, we employed the coarse-grained (CG) model to simulate the activation process of SARS-CoV-2 spike protein. Our results give an adequate explanation of the activation mechanism of the spike protein based on energy. Next, fourteen different nanobodies were used to identify their binding modes to the viral spike protein, and we identified the critical residues for their binding. Then, we selected the 7KSG nanobody as a template for further optimization. We performed saturated mutagenesis of these key residue sites and employed a CG model to calculate the binding energies. We found that mutations V27D, L29E, Y32E, S49D, C50K, S53D, R57E, A97D, T102E, Y104E, S105E, N107K, H109K, Y110E, C112D, S113K, M116D, and Y118D in complementarity determining regions (CDRs) of the nanobody can strengthen the binding between the spike protein and the nanobody. Based on the results of single mutants, we employed multiple mutations on nanobodies and designed four potent nanobodies exhibiting a higher binding affinity than the original one.

## 2. Materials and Methods

### 2.1 Modelling the S Trimers

In this work, we used Modeller [30,31] to perform homology modeling in constructing the binding complexes of ACE2-SARS-CoV-2 and nanobody-SARS-CoV-2. First, the structures of the three key states of S trimers were extracted from the Cryo-EM structure (Protein Data Bank identification (PDB ID): 7DF3, 7DK3, and 7DF4) resolved by Xu *et al.* [9]. The experiment structures underwent a repair process that includes completing the missing residues, removing extra ligands, and trimming all structures to the same length. After repairing, a targeted molecular simulation (TMD) method was conducted to construct the confor-

mational pathways between different structures and sample a series of intermediate conformations representing the transition process. At each TMD step, the initial structure was aligned to the target structure using the backbone heavy atoms, then a force was applied to the initial structure to move them toward the corresponding atom of the target structure. The system was restrained throughout the simulation to prevent abnormal translation and rotation. For these structures, the solvent was treated implicitly. Extensive MD relaxation by Molaris-XG software (version 9.15, Los Angeles, CA, USA) [32,33] was carried out until reached the convergence. This software is developed by Dr. Warshel and his team at the University of Southern California.

There is no conformational transition of ACE2, only position and pose changes, so the intermediate conformations of ACE2 are obtained by an angle-distance interpolation method. The method first calculated 4 parameters between the initial structure and the target structure in the Cartesian coordinate system: the center-of-mass distance, and the differences of their Euler angles. Then, these four parameters are divided equally depending on the number of intermediates to be obtained. The intermediate structure is obtained by moving/rotating the initial structure according to the values of the four parameters after divination.

### 2.2 Coarse-Grained (CG) Model and the Total Energy Calculation

The coarse-grained model was employed to calculate the free energy of each structure and the relevant binding energies. The CG model we employed was developed by Arieh Warshel not only gives a reliable description for protein stability and functions, but also considers the importance of electrostatic effects of proteins [34,35]. In CG model, the side chain is reduced to a simplified united atom and the backbone atoms of each residue are treated explicitly. The total CG energy is defined as follows:

$$\begin{aligned}\Delta G_{\text{fold}}^{CG} &= \Delta G_{\text{side}}^{CG} + \Delta G_{\text{main}}^{CG} + \Delta G_{\text{main-side}}^{CG} \\ &= c_1 \Delta G_{\text{side}}^{vdw} + c_2 \Delta G_{\text{solv}}^{CG} + c_3 \Delta G_{HB}^{CG} + \Delta G_{\text{side}}^{\text{elec}} \\ &\quad + \Delta G_{\text{side}}^{\text{polar}} + \Delta G_{\text{side}}^{\text{hyd}} + \Delta G_{\text{main-side}}^{\text{elec}} + \Delta G_{\text{main-side}}^{vdw}\end{aligned}\quad (1)$$

Here the terms are the side chain van der Waals energy, main chain solvation energy, main chain hydrogen bond energy, side chain electrostatic energy, side chain polar energy, side chain hydrophobic energy, main chain/side chain electrostatic energy, and main chain/side chain van der Waals energy, respectively. The scaling coefficients  $c_1$ ,  $c_2$ , and  $c_3$  are 0.10, 0.25, and 0.15, respectively, in this work [34,36].

To evaluate the CG energy, we first calculated the reliable charges for the protein ionized groups using the Monte Carlo Proton Transfer algorithm (MCPT) [37,38]. This method allows a proton transfer between pair of ionizable residues or within an ionizable residue and bulk sol-

vent. The transferring is repeated until the electrostatic interaction of the folded protein converges, then the ionization states of the protein residues are obtained to evaluate the CG free energy.

### 2.3 The Binding Free Energy Change Calculation

The binding free energies for the nanobody-spike protein are defined as follows.

$$\Delta G_{\text{binding}} = \Delta G_{\text{nanobody-spike}} - \Delta G_{\text{nanobody}} - \Delta G_{\text{spike}} \quad (2)$$

For the mutated binding free energy change of the nanobody-spike protein:

$$\begin{aligned} \Delta\Delta G_{\text{binding}} &= \Delta G_{\text{binding mutant}} - \Delta G_{\text{binding WT}} \\ &= (\Delta G_{\text{nanobody-spike mutant}} - \Delta G_{\text{nanobody mutant}} - \Delta G_{\text{ACE}}) \\ &\quad - (\Delta G_{\text{nanobody-spike WT}} - \Delta G_{\text{nanobody WT}} - \Delta G_{\text{ACE}}) \\ &= (\Delta G_{\text{nanobody-spike mutant}} - \Delta G_{\text{nanobody-spike WT}}) \\ &\quad + (\Delta G_{\text{nanobody WT}} - \Delta G_{\text{nanobody mutant}}) \\ &= \Delta G_1 + \Delta G_2 \end{aligned} \quad (3)$$

WT: Wild type.

## 3. Results

### 3.1 The Activation Process of the SARS-CoV-2 Spike Protein

Based on the structures of spike trimer in the different conformational states, Xu *et al.* [9] revealed the mechanism of ACE2-induced conformational transitions of S trimer from the ground prefusion state toward the post-fusion state. During the S trimer's activation, the S-protein's conformation transitioned from a "closed" to an "open" state, with one of its receptor binding domains up, obtaining the ability to infect host cells. The presence of ACE2 alters the conformational distribution of the S-protein and promotes its activation, however, these findings lack a structural/energetic explanation.

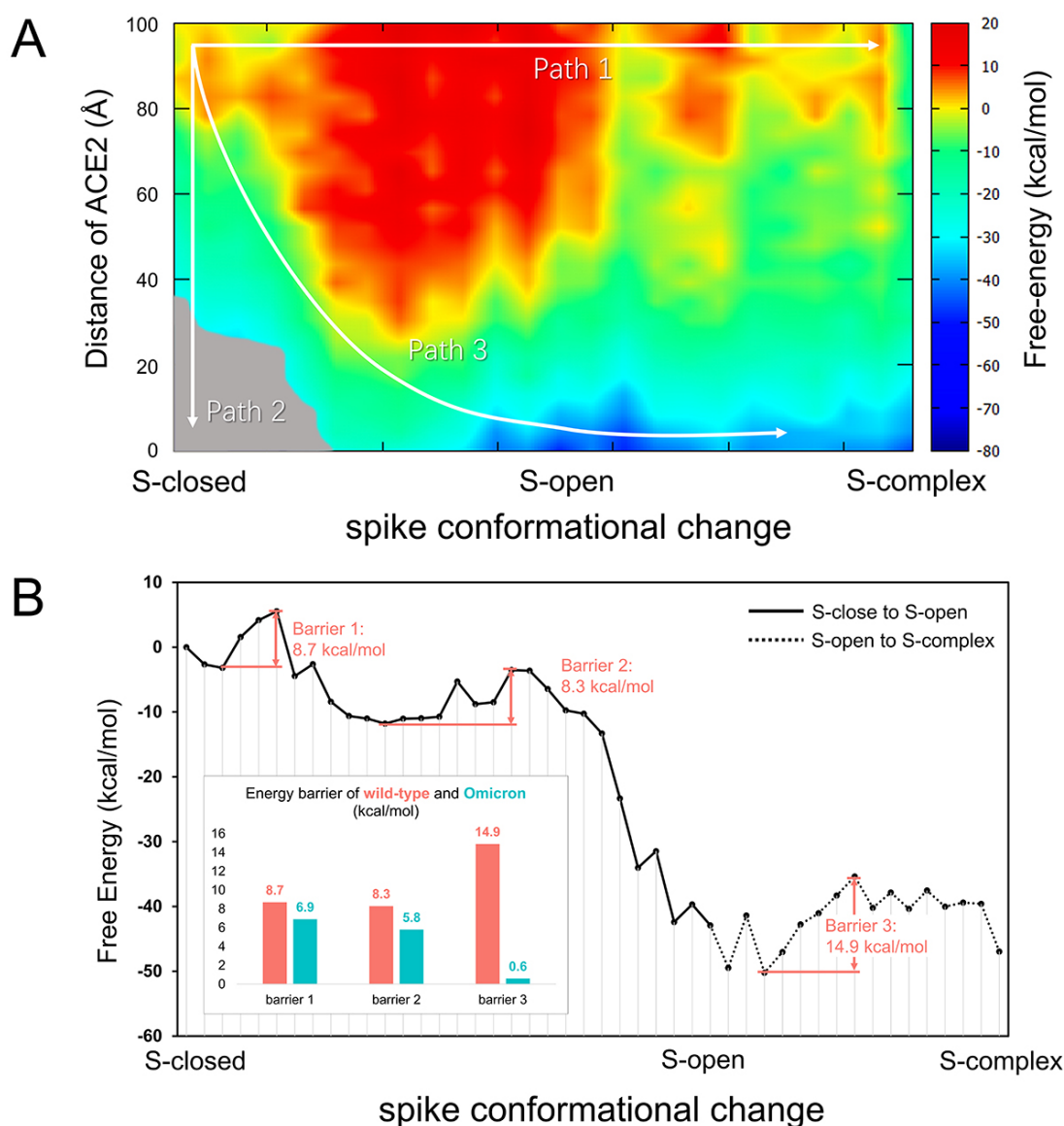
In order to understand the energetic mechanism of spike protein activation, we constructed a series of structural models of the coupling process of ACE2 approaching and conformational changes of S trimers. First, we extracted the structures of the three key states of S trimers from the Cryo-EM structure (PDB ID: 7DF3, 7DK3, and 7DF4) resolved by Xu *et al.* [9]. After repairing the experimental structures (See methods), we used a targeted molecular simulation (TMD) [39] method to obtain the intermediate structures connecting the three states (Fig. 1B) to form the conformational change trajectory. For the three key states, we found the optimal binding poses of ACE2 to their RBD by using a protein-protein docking method, HDOCK [2]. This method is developed by Huang's group at the Huazhong University of Science and Technology. Then, we identified the center of the ACE2 in the optimal binding mode as the initial position and pulled ACE2 away along the S trimers' rotation symmetry axis (the dashed line in

Fig. 1A). The Y-axis in Fig. 2B represents the distance between the center of the pulled ACE2 and the center of ACE2 in the initial position (Fig. 1B and Fig. 2). The X-axis represents the conformational change trajectory of the S trimer (Fig. 2A). Since the binding position of ACE2 is different between the three key states, the intermediate conformation of ACE2 was obtained by an angle-distance interpolation method (Fig. 1B, See methods). By this modeling, we simulated the complete process that ACE2 gradually approaches to the S trimer, during the conformation changes of the S trimer from S-closed to S-open and the S-complex.

We calculated the folding energy of each ACE2-spike complex and obtained the energy landscape (Fig. 2). The color indicates the relative folding free energy, with the point in the upper left corner as the zero point (Fig. 2A). There are three paths depicted as white lines on the energy landscape (Fig. 2A). Path 1 indicates that, when ACE2 is far enough away (100 Å) from the S trimer, the conversion of the spike protein to the activate conformation (S-complex) requires crossing a larger energy barrier. The result is consistent with the conclusions we obtained from our previous calculation that the energy barrier is 25.44 kcal/mol of spike conformational change in the absence of ACE2 [40]. Path 2 demonstrates the energy change resulting from the approach of ACE2 when the S trimer conformation does not change. The lower left corner of the energy landscape indicates that ACE2 has a stable interaction with the spike protein in the S-closed state. In the S-closed state, the surface of the spike is densely packed with glycan ligands [3,41]. The glycan ligands block the binding pathway of ACE2 near the region (the gray region in Fig. 2A). According to the energy profile, path 3 is the most reasonable pathway which the spike protein will take. As ACE2 approaches, the conformation of spike protein bypasses the high-energy barrier region to reach the S-open state, where ACE2 can bind to the spike protein. The spike protein continues to convert to the activated S-complex state (**Supplementary Movie 1**). We extracted the energy profile of the path 3 and identified three possible energy barriers (Fig. 2B). These three energy barriers are lower than the barrier in the absence of ACE2 (25.44 kcal/mol), which confirms the induction role of ACE2 in the activation of the spike protein. We also calculated the change of energy barriers when introducing the mutations of the Omicron variant to the spike protein. The result shows that the mutations led to significant decreases in all three energy barriers (Fig. 2B). In addition, we also calculated the change of the energy barrier for other SARS-CoV-2 variants (**Supplementary Fig. 1**). Compared with the wild-type, almost all the energy barriers are decreased from other SARS-CoV-2 variants. This indicates that the spike protein is more readily activated in the SARS-CoV-2 variants and may explain the high transmission of this variant.

Overall, we calculated a 2D energy landscape, identified a least energy pathway from the S-closed to the S-





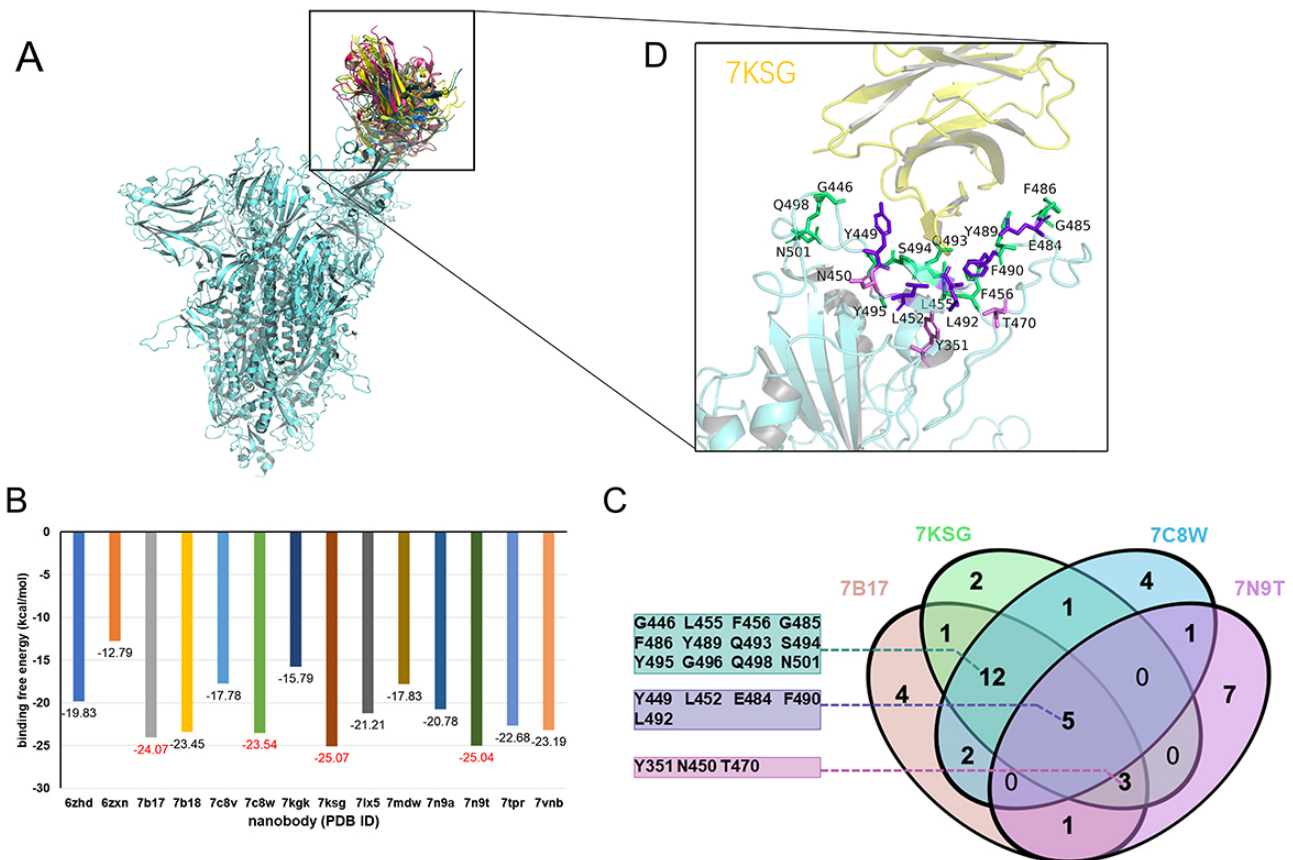
**Fig. 2. The energetics of the activation of SARS-CoV-2.** (A) The energy landscape of the activation of spike protein coupling with the distance of ACE2. The gray region indicates the area that is covered by glycan ligands. The Y-axis represents the distance between the center of the pulled ACE2 and the center of ACE2 in the initial position (see Fig. 1B). The X-axis represents the conformational change trajectory of the S trimer, from S-closed to S-open and then to S-complex. (B) The energy profile of Path 3 and the comparison of the energy barrier during the activation process between the wild-type and omicron variant (inset figure).

complex and calculated the changes of the energy barrier of all the SARS-CoV-2 variants. Our results suggest that the distance between the ACE2 and the spike protein is vital for spike protein activation. Compared to the wild type, the new coronavirus variant has a lower energy barrier, which can explain why the new variants show higher transmissibility. Moreover, these new variants have higher viral infectivity, and higher potential for immune evasion. Therefore, it is necessary to design a vaccine against these new variants. Stable and potent nanobodies that target the RBD of SARS-CoV-2 are promising therapeutics to help neutralize the new variants. The RBD of the spike protein is a prime target for therapeutic nanobodies. By blocking

the interaction between the ACE2 and the spike protein, nanobodies can inhibit the entry of SARS-CoV-2. Thus, it is important for us to identify key residues of the RBD in the spike protein for nanobody binding.

### 3.2 Identify Key Residues of the RBD in the Spike Protein for Nanobody Binding

The region of the RBD surface in contact with nanobodies differs between the structures. Therefore, it is necessary to explore the detailed structural information of their epitopes and binding modes to the viral spike protein. To compare the nanobodies-interacting residues on the SARS-CoV-2 RBD, we obtained the spike protein of the

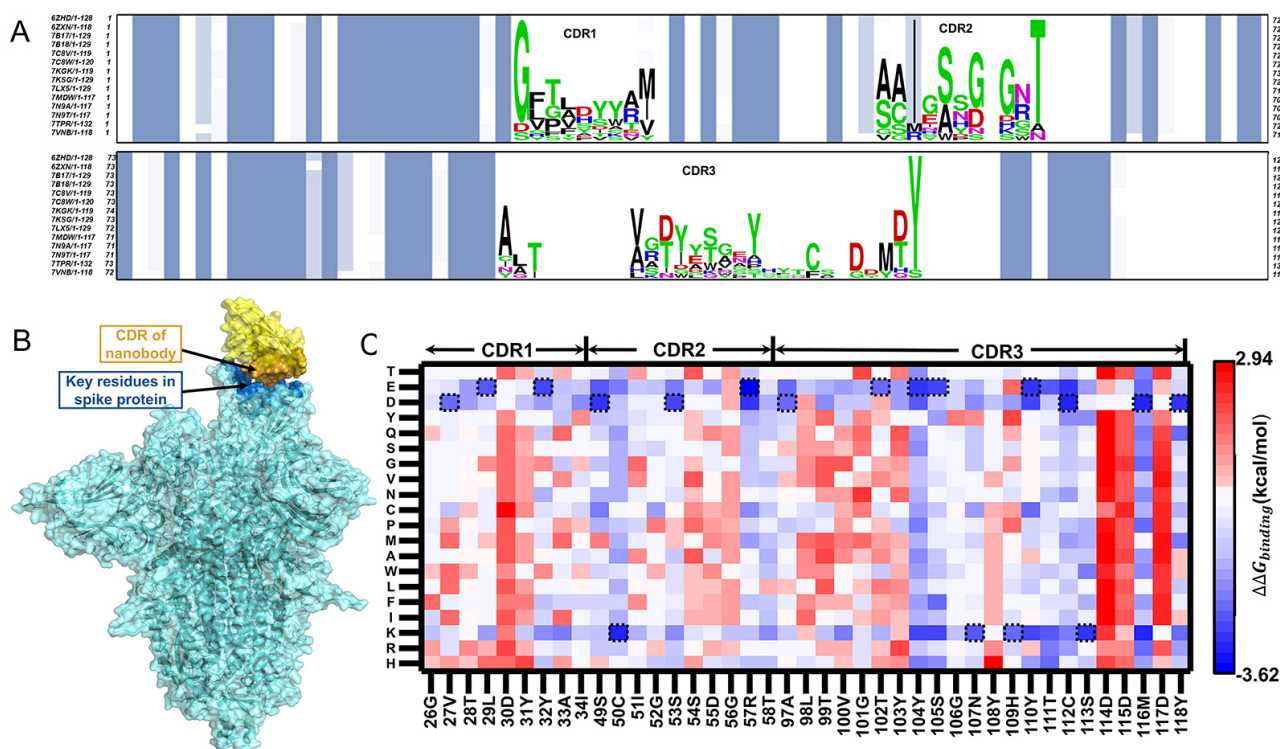


**Fig. 3. The interaction mode and the binding free energy between the spike protein and nanobodies.** (A) The binding sites on the spike protein of different nanobodies. (B) The binding free energy of nanobodies. (C) The key residues of spike protein binding to four high-affinity nanobodies. (D) The binding interface of the highest affinity nanobody (PDB ID: 7KSG) and spike protein. The same residues in the interacting interface of the nanobody-spike complex are highlighted by different colors. Green indicates the same residues appear in the 7B17, 7KSG, and 7C8W; purple indicates the same residues present in the 7B17, 7KSG, 7C8W, and 7N9T; pink indicates the identical residues appear in the 7KSG, and 7C8W and 7N9T. Key residues are shown in stick representations.

SARS-CoV-2 at the “S-complex” state (PDB ID: 7DF4), which contains an “up” conformation of the RBD. Steric effects play an important role in blocking the ACE2 binding to spike by nanobodies. When nanobodies bind to the side sites, the nanobodies cannot generate sufficient steric hindrance to block the interaction between ACE2 and the spike protein [42]. Furthermore, the resolved crystal structures of those nanobodies binding to the side sites only have the structure “up” conformation of the RBD, and not the whole structure of the spike protein. When we superimposed them to the wild type, their binding sites spatially clash with the rest of the structures of the spike protein. Thus, in this study, we only considered the binding sites in the top positions of the RBD. In this study, 14 different nanobodies were used to identify the nanobody binding sites of the spike protein. All of them targeted the RBD by aligning the “up” RBD structure in the spike protein (Fig. 3A). The PDB ID of all 14 nanobodies are 6ZHD, 6ZXN, 7B17, 7B18, 7C8V, 7C8W, 7KGK, 7KSG, 7LX5, 7MDW, 7N9A, 7N9T, 7TPR, and 7VBN respectively [19,21,29,43–47].

The results for the binding free energy among these 14

different nanobodies are shown in Fig. 3B. They all show high binding affinity to the wild SARS-CoV-2. We then selected four nanobodies with the highest affinity (PDB ID: 7B17, 7C8W, 7KSG, and 7N9T) to further analyze their binding modes. As shown in Fig. 3C, we counted all the residues of the spike protein at the RBD-nanobodies interfaces to find the key residues that drive the binding process. As expected, many interacting residues are identical between different nanobodies. Five identical residues (Y449, L452, E484, F490, and L492) appeared in all the interfaces of these four high-binding nanobodies. In the intersection of the three high-binding nanobodies, twelve identical residues including G446, L455, F456, G485, F486, Y489, Q493, S494, Y495, G496, Q498, and N501, appear in the interfaces of 7B17, 7KSG, and 7C8W, while only three identical residues (Y351, N450, and T470) are present in the interfaces of 7KSG, 7C8W, and 7N9T. There is no identical residue in the interfaces of 7B17, 7KSG, and 7C8W, or the interfaces of 7KSG, 7C8W, and 7N9T, or the interfaces 7B17, 7C8W, and 7N9T. In total, there are 20 same residues present in at least three different interfaces of



**Fig. 4. Analysis of the mutation effects.** (A) The multiple sequence alignment of nanobody. The color code designates the conserved region in nanobodies while the residues in CDRs are shown in the logo. (B) The mode of nanobody binding to spike protein for nanobody optimization. (C) The heatmap of the binding free energy change ( $\Delta\Delta G_{\text{binding}}$ ) due to the amino acid substitution of CDRs. Mutations V27D, L29E, and Y32E in CDR1, mutations S49D, C50K, S53D, and R57E in CDR2, and mutations A97D, T102E, Y104E, S105E, N107K, H109K, Y110E, C112D, S113K, M116D, and Y118D in CDR3, 17 mutations were selected for further optimization (circled by black dashed box).

high-binding nanobodies. These residues are key sites that contribute to the high nanobody-binding affinity of SARS-CoV-2. Among them, residues G446, Y449, L455, F456, F486, Y489, Q493, G496, Q498, and N501 also are the key sites within the RBD involved in ACE2 binding [48,49]. Among these 14 different nanobodies, 7KSG has the highest binding affinity ( $\Delta G_{\text{binding}} = -25.07$  kcal/mol). The interface of the 7KSG-spike complex also contains all 20 key residues, which are critical sites for nanobody binding (Fig. 3D). Therefore, we chose 7KSG as an initial template for further nanobody optimization. The details of the structure of 7KSG are shown in **Supplementary Fig. 2**.

### 3.3 Mutation Impacts on SARS-CoV-2 Nanobodies

To better understand whether there is any connection between these 14 nanobodies, we carried out multiple sequence alignments (MSA) to further understand the similarity and differences among these nanobodies. The MSA was performed using the Kalign web server of EMBL-EBI services [50]. As shown in Fig. 4A, all nanobodies are very similar to each other. The differences among these 14 nanobodies are mainly concentrated in the CDR1, CDR2, and CDR3.

The interface between the spike protein (orange re-

gions in Fig. 4B) and the nanobody (blue region in Fig. 4B) is a major factor in determining whether the nanobody can bind well to the RBD of the spike protein. We introduced mutations into the nanobody to find positions that can improve the complementarity with higher binding affinity. In saturated mutagenesis of residues in CDRs (G26-I34, S49-T58, and A97-Y118), a total of 779 mutations were considered in nanobodies (PDB ID: 7KSG). The binding free energy changes of the spike protein and mutated nanobodies are shown in Fig. 4C. Overall, most mutations on CDRs lead to mild negative binding free energy changes. Compared with the mutation in CDR1 and CDR2, most mutations in CDR3 lead to the strengthening of the spike protein and nanobody binding (blue squares in Fig. 4C). Mutations A97D, T102E, Y104E, S105E, N107K, H109K, Y110E, C112D, S113K, M116D, and Y118D in CDR3 all give rise to negative binding free energy changes for nanobodies, which strengthen the bindings. Some disruptive mutations, such as D30C, G101N, Y108H, D114A, D115C, and D117A in CDRs, lead to positive binding free energy changes (red squares in Fig. 4C), indicating weakening bindings between the spike protein and nanobodies.

Some residues that are directly in contact with or close to the spike protein, have a larger impact on increas-



ing the binding energy than those residues that do not directly contact the spike protein. For example, residues T102, Y104, and S105 are closer to the spike protein than residues G101, D114, and D115 (**Supplementary Fig. 3**). After single-site mutation, mutations T102E, Y104K, and S105E, all lead to the strengthening of the spike protein and nanobody binding, while mutations on G101, D114, D115 have little or negative effect on the binding affinity between the spike protein and nanobodies. We found that most residues in CDRs were mutated to negatively charged residues in favor of strengthening the binding affinity. Previous studies have demonstrated that ACE2 has many negatively charged residues, which resulted in a large increase in Coulomb's force between the spike protein and ACE2 [51,52]. Therefore, the more negative charges on the epitope of the nanobodies, the higher the attraction with the spike protein. This is consistent with our results. Our calculations confirmed the concept that the binding energy change is a practical approach for predicting mutational effects.

The variants of SARS-CoV-2 with multiple mutations in RBD are the major factor in the development of resistance against vaccines. In order to design potent nanobodies with broad-spectrum activity neutralizing SARS-CoV-2 variants, not only did we consider the single mutation, but also the multiple mutations in nanobodies. We selected those residues that make the  $\Delta\Delta G_{\text{binding}}$  is less than  $-2$  kcal/mol after a single mutation. If a residue is mutated to other residues, and the  $\Delta\Delta G_{\text{binding}}$  of many mutated nanobodies is less than 2 kcal/mol, then we chose the one with the highest binding affinity. Therefore, mutations V27D, L29E, Y32E in CDR1, mutations S49D, C50K, S53D, and R57E in CDR2, and mutations A97D, T102E, Y104E, S105E, N107K, H109K, Y110E, C112D, S113K, M116D, and Y118D in CDR3 were selected for multiple mutations. Next, we split these mutations into three categories based on their locations in the nanobodies and built four new mutated nanobodies, 7KSG\_CDR1 (mutations V27D, L29E, and Y32E in CDR1), 7KSG\_CDR2 (mutations S49D, C50K, S53D, and R57E in CDR2), 7KSG\_CDR3 (mutations A97D, T102E, Y104E, S105E, N107K, H109K, Y110E, C112D, S113K, M116D, and Y118D in CDR3), and 7KSG\_ALL (all 17 mutations in CDRs). As expected, these four nanobodies exhibit higher binding affinity than the original one (**Supplementary Table 1**). 7KSG\_ALL had the highest binding affinity ( $\Delta\Delta G_{\text{binding}} = -40.04$  kcal/mol). The specific mutation combination can facilitate the optimization of a potent nanobody with a higher binding affinity.

## 4. Discussion

Nanobodies are composed of the target-binding fragment of monoclonal antibodies. Compared with traditional antibodies, nanobodies have more advantages. For example, they are significantly smaller in size so they are able to

access and lodge onto conventionally inaccessible regions on therapeutic targets [53]. Also they exhibit favorable biophysical properties. In addition, nanobodies can be efficiently produced in prokaryotic expression systems at a low cost. Thus, the search for potent nanobody therapies on an industrial-scale is becoming one of the most feasible strategies for combating SARS-CoV-2.

However, in this early-stage trial, the nanobodies optimized in this study are mainly against the wild type of coronavirus. Considering the future variants and escape mutants, we will systematically analyze all the variants of coronavirus through computational approaches in our future studies in an attempt to work: find the key sites of their binding interface, computationally design a single-site saturated mutagenesis library in epitopes of the nanobodies, calculate their binding free energy, and further perform the combinations of mutations to design novel, potent nanobodies with broad-spectrum activity. In this study, we only calculated the conformational free energy of the spike protein in the S-complex state for the nanobody design. The lower the free energy, the more stable the structure. This suggests that the spike protein and antibody are not easy to dissociate. In the future, we will expand from only considering the binding energy of nanobody and spike protein in the S-complex state to considering its energy barrier changes according to the energy landscape. If the energy barrier of the designed nanobody is lower than that in the activation path of the spike protein and ACE2, the designed nanobody may be a potential candidate against SARS-CoV-2. This may be a promising method to design stable and potent nanobodies in the future. The SARS-CoV-2 spike protein is composed of the S1 and S2 subunits. Compared to S1 subunit, the S2 subunit contains more conserved residues [54]. Recent studies [55,56] have found that some antibodies are designed to bind the conserved fusion peptide region adjacent to the S2 subunit and they can broadly target the SARS-CoV-2 variants. However, because of the steric constraints of the spike density, it is hard for antibodies to access these regions [57]. Therefore, it is important to design nanobodies that can target the conserved and functionally essential sites on coronaviruses. We plan to study these concepts in the future.

Artificial intelligence technologies have been widely applied in the development of antibodies, such as recognizing antigen epitope [58–61], exploring the sequence space of CDRs [62–64], optimizing CDR sequences, predicting antibody structure [65], predicting binding modes [66], and predicting the binding affinity of antibodies to antigens [67]. Many attempts in the development of SARS-CoV-2 antibodies have also been reported [68–71]. Using expanded data sets and new deep learning technologies resulting in the potential for the development of better antibodies [72]. Our work provides a data set for the relationship of residue substitutions of nanobody CDRs with the binding affinity to spike protein. The dataset can be used to de-



velop models predicting the neutralization ability of artificially designed antibodies. The sequence space of CDRs is too large to be fully explored. Therefore, based on the assumption that the effect of multiple point mutations is approximately equal to the cumulative effect of single point mutations, we designed a combination of mutations to obtain antibodies with stronger affinity. By using artificial intelligence methods such as reinforcement learning and/or genetic algorithm, we can try as many mutation combinations as possible at an affordable computational cost. Overall, integrating artificial intelligence technologies expands our abilities to conduct these research studies.

## 5. Conclusions

In this work, we constructed a series of intermediate structures of the coupling process of ACE2 approaching and S trimers to explore the energy basis of the activation of the spike protein. By utilizing these structures, we have generated the free energy profile for conformational changes and found one possible lower energy pathway. To investigate the key residues in the nanobody-spike interface, we compared 14 nanobody-bound structures and analyzed the binding modes of 4 nanobodies with the highest binding affinity. We found that there are 20 conserved residues (Y449, L452, E484, F490, L492, G446, L455, F456, G485, F486, Y489, Q493, S494, Y495, G496, Q498, N501, Y351, N450, and T470) appearing at the interface of three or four nanobodies. Some of these residues (G446, Y449, L455, F456, F486, Y489, Q493, G496, Q498, and N501) are also the key sites within the RBD involved in the interface of the ACE2-spike complex. Next, we selected the one with the best binding affinity among the 14 nanobodies as a preliminary structure to optimize and design novel nanobodies. We introduced a single-site saturated mutagenesis library of CDR position to explore the effect of various mutations on binding affinity. After calculating the binding free energy changes followed by mutations, we found that most residues in CDRs were mutated to negatively charged residues in favor of strengthening the binding affinity. Based on the results of the single-site mutation, we employed a combination of mutations on CDRs and designed four novel nanobodies. The optimized nanobodies all exhibit higher binding affinity than the original ones.

In conclusion, studying the mechanism of the activation process gives us a more comprehensive understanding of the coronavirus infection and immune evasion. Identifying the key residues in the interface between the nanobody and the spike protein can provide useful information for understanding the binding mechanism of the nanobody-spike complex. Our results suggest that this approach can be a promising method to develop nanobodies with high binding affinity and broad-spectrum activity to neutralize SARS-CoV-2 variants.

## Availability of Data and Materials

The datasets used during the current study are available from the corresponding author on reasonable request.

## Author Contributions

XZ and KA designed the research study. XZ, KA, JY, and PX performed the research. XZ and KA analyzed the data and drafted the manuscript. CB provided help and advice on conception, acquisition of data and supervision. All authors contributed to editorial changes in the manuscript. All authors read and approved the final manuscript. All authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

## Ethics Approval and Consent to Participate

Not applicable.

## Acknowledgment

Not applicable.

## Funding

This research was funded by the National Natural Science Foundation of Youth Fund Project (grant no. 22103066), the 2021 Basic Research General Project of Shenzhen, China (grant no. 20210316202830001) and Warshel Institute for Computational Biology at the Chinese University of Hong Kong, Shenzhen (grant no. C10120180043).

## Conflict of Interest

CB is the founder of Chenzhu Biotechnology Co., Ltd.. CB took participated in this research. All authors declare that they have no conflict of interest.

## Supplementary Material

Supplementary material associated with this article can be found, in the online version, at <https://doi.org/10.31083/j.fbl2804067>.

## References

- [1] Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. *Lancet*. 2020; 395: 470–473.
- [2] Yan Y, Tao H, He J, Huang S. The HDock server for integrated protein-protein docking. *Nature Protocols*. 2020; 15: 1829–1852.
- [3] Casalino L, Gaieb Z, Goldsmith JA, Hjorth CK, Dommer AC, Harbison AM, *et al.* Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Central Science*. 2020; 6: 1722–1734.
- [4] Dong Y, Dai T, Wei Y, Zhang L, Zheng M, Zhou F. A systematic review of SARS-CoV-2 vaccine candidates. *Signal Transduction and Targeted Therapy*. 2020; 5: 237.
- [5] Ke Z, Oton J, Qu K, Cortese M, Zila V, McKeane L, *et al.* Structures and distributions of SARS-CoV-2 spike proteins on intact virions. *Nature*. 2020; 588: 498–502.

- [6] Wang Q, Zhang Y, Wu L, Niu S, Song C, Zhang Z, *et al.* Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. *Cell*. 2020; 181: 894–904.e9.
- [7] Tortorici MA, Walls AC, Lang Y, Wang C, Li Z, Koerhuis D, *et al.* Structural basis for human coronavirus attachment to sialic acid receptors. *Nature Structural & Molecular Biology*. 2019; 26: 481–489.
- [8] Walls AC, Park Y, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*. 2020; 181: 281–292.e6.
- [9] Xu C, Wang Y, Liu C, Zhang C, Han W, Hong X, *et al.* Conformational dynamics of SARS-CoV-2 trimeric spike glycoprotein in complex with receptor ACE2 revealed by cryo-EM. *Science Advances*. 2021; 7: eabe5575.
- [10] V'kovski P, Kratzel A, Steiner S, Stalder H, Thiel V. Coronavirus biology and replication: implications for SARS-CoV-2. *Nature Reviews Microbiology*. 2021; 19: 155–170.
- [11] Hosseindokht M, Bakherad H, Zare H. Nanobodies: a tool to open new horizons in diagnosis and treatment of prostate cancer. *Cancer Cell International*. 2021; 21: 580.
- [12] Ardekani LS, Gargari SLM, Rasooli I, Bazl MR, Mohammadi M, Ebrahimizadeh W, *et al.* A novel nanobody against urease activity of *Helicobacter pylori*. *International Journal of Infectious Diseases*. 2013; 17: e723–e728.
- [13] Bakherad H, Gargari SLM, Sepehrizadeh Z, Aghamollaei H, Taheri RA, Torshabi M, *et al.* Identification and in vitro characterization of novel nanobodies against human granulocyte colony-stimulating factor receptor to provide inhibition of G-CSF function. *Biomedicine & Pharmacotherapy*. 2017; 93: 245–254.
- [14] Bakherad H, Mousavi Gargari SL, Rasooli I, Rajabibazl M, Mohammadi M, Ebrahimizadeh W, *et al.* In vivo neutralization of botulinum neurotoxins serotype E with heavy-chain camelid antibodies (VHH). *Molecular Biotechnology*. 2013; 55: 159–167.
- [15] Hu Y, Liu C, Muyldermans S. Nanobody-Based Delivery Systems for Diagnosis and Targeted Tumor Therapy. *Frontiers in Immunology*. 2017; 8: 1442.
- [16] Muyldermans S. Nanobodies: natural single-domain antibodies. *Annual Review of Biochemistry*. 2013; 82: 775–797.
- [17] Zare H, Rajabibazl M, Rasooli I, Ebrahimizadeh W, Bakherad H, Ardakani LS, *et al.* Production of nanobodies against prostate-specific membrane antigen (PSMA) recognizing LnCaP cells. *The International Journal of Biological Markers*. 2014; 29: e169–e179.
- [18] Custódio TF, Das H, Sheward DJ, Hanke L, Pazicky S, Pieprzyk J, *et al.* Selection, biophysical and structural analysis of synthetic nanobodies that effectively neutralize SARS-CoV-2. *Nature Communications*. 2020; 11: 5588.
- [19] Hanke L, Vidakovic Perez L, Sheward DJ, Das H, Schulte T, Moliner-Morro A, *et al.* An alpaca nanobody neutralizes SARS-CoV-2 by blocking receptor interaction. *Nature Communications*. 2020; 11: 4420.
- [20] Huo J, Le Bas A, Ruza RR, Duyvesteyn HME, Mikolajek H, Malinauskas T, *et al.* Neutralizing nanobodies bind SARS-CoV-2 spike RBD and block interaction with ACE2. *Nature Structural & Molecular Biology*. 2020; 27: 846–854.
- [21] Koenig P, Das H, Liu H, Kümmerer BM, Gohr FN, Jenster L, *et al.* Structure-guided multivalent nanobodies block SARS-CoV-2 infection and suppress mutational escape. *Science*. 2021; 371: eabe6230.
- [22] Schoof M, Faust B, Saunders RA, Sangwan S, Rezelj V, Hoppe N, *et al.* An ultrapotent synthetic nanobody neutralizes SARS-CoV-2 by stabilizing inactive Spike. *Science*. 2020; 370: 1473–1479.
- [23] Valenzuela Nieto G, Jara R, Watterson D, Modhiran N, Amarilla AA, Himelreichs J, *et al.* Potent neutralization of clinical isolates of SARS-CoV-2 D614 and G614 variants by a monomeric, sub-nanomolar affinity nanobody. *Scientific Reports*. 2021; 11: 3318.
- [24] Wrapp D, De Vlieger D, Corbett KS, Torres GM, Wang N, Van Breedam W, *et al.* Structural Basis for Potent Neutralization of Betacoronaviruses by Single-Domain Camelid Antibodies. *Cell*. 2020; 181: 1004–1015.e15.
- [25] Xiang Y, Nambulli S, Xiao Z, Liu H, Sang Z, Duprex WP, *et al.* Versatile and multivalent nanobodies efficiently neutralize SARS-CoV-2. *Science*. 2020; 370: 1479–1484.
- [26] Xu J, Xu K, Jung S, Conte A, Lieberman J, Muecksch F, *et al.* Nanobodies from camelid mice and llamas neutralize SARS-CoV-2 variants. *Nature*. 2021; 595: 278–282.
- [27] Zhou D, Duyvesteyn HME, Chen C, Huang C, Chen T, Shih S, *et al.* Structural basis for the neutralization of SARS-CoV-2 by an antibody from a convalescent patient. *Nature Structural & Molecular Biology*. 2020; 27: 950–958.
- [28] Nambulli S, Xiang Y, Tilston-Lunel NL, Rennick LJ, Sang Z, Klimstra WB, *et al.* Inhalable Nanobody (PiN-21) prevents and treats SARS-CoV-2 infections in Syrian hamsters at ultra-low doses. *Science Advances*. 2021; 7: eabh0319.
- [29] Pymm P, Adair A, Chan L, Cooney JP, Mordant FL, Allison CC, *et al.* Nanobody cocktails potentially neutralize SARS-CoV-2 D614G N501Y variant and protect mice. *Proceedings of the National Academy of Sciences of the United States of America*. 2021; 118: e2101918118.
- [30] Marti-Renom MA, Stuart AC, Fiser A, Sánchez R, Melo F, Sali A. Comparative protein structure modeling of genes and genomes. *Annual Review of Biophysics and Biomolecular Structure*. 2000; 29: 291–325.
- [31] Webb B, Sali A. Comparative Protein Structure Modeling Using MODELLER. *Current Protocols in Bioinformatics*. 2016; 54: 5.6.1–5.6.37.
- [32] Lee FS, Chu ZT, Warshel A. Microscopic and Semimicroscopic Calculations of Electrostatic Energies in Proteins by the Polaris and Enzymix Programs. *Journal of Computational Chemistry*. 1993; 14: 161–185.
- [33] Kamerlin SCL, Vicatos S, Dryga A, Warshel A. Coarse-grained (multiscale) simulations in studies of biophysical and chemical systems. *Annual Review of Physical Chemistry*. 2011; 62: 41–64.
- [34] Vicatos S, Rychkova A, Mukherjee S, Warshel A. An effective coarse-grained model for biological simulations: recent refinements and validations. *Proteins*. 2014; 82: 1168–1185.
- [35] Vorobyov I, Kim I, Chu ZT, Warshel A. Refining the treatment of membrane proteins by coarse-grained models. *Proteins*. 2016; 84: 92–117.
- [36] Bai C, Wang J, Mondal D, Du Y, Ye RD, Warshel A. Exploring the Activation Process of the  $\beta$ 2AR-G<sub>s</sub> Complex. *Journal of the American Chemical Society*. 2021; 143: 11044–11051.
- [37] Messer BM, Roca M, Chu ZT, Vicatos S, Kilshtain AV, Warshel A. Multiscale simulations of protein landscapes: using coarse-grained models as reference potentials to full explicit models. *Proteins*. 2010; 78: 1212–1227.
- [38] Beroza P, Fredkin DR, Okamura MY, Feher G. Protonation of interacting residues in a protein by a Monte Carlo method: application to lysozyme and the photosynthetic reaction center of *Rhodobacter sphaeroides*. *Proceedings of the National Academy of Sciences of the United States of America*. 1991; 88: 5804–5808.
- [39] Schlitter J, Engels M, Krüger P. Targeted molecular dynamics: a new approach for searching pathways of conformational transitions. *Journal of Molecular Graphics*. 1994; 12: 84–89.
- [40] Shi D, An K, Zhang H, Xu P, Bai C. Application of Coarse-Grained (CG) Models to Explore Conformational Pathway of Large-Scale Protein Machines. *Entropy*. 2022; 24: 620.

- [41] Grant OC, Montgomery D, Ito K, Woods RJ. Analysis of the SARS-CoV-2 spike protein glycan shield reveals implications for immune recognition. *Scientific Reports*. 2020; 10: 14991.
- [42] Guo H, Gao Y, Li T, Li T, Lu Y, Zheng L, *et al.* Structures of Omicron spike complexes and implications for neutralizing antibody development. *Cell Reports*. 2022; 39: 110770.
- [43] Ahmad J, Jiang J, Boyd LF, Zeher A, Huang R, Xia D, *et al.* Structures of synthetic nanobody-SARS-CoV-2 receptor-binding domain complexes reveal distinct sites of interaction. *The Journal of Biological Chemistry*. 2021; 297: 101202.
- [44] Gu J, Liu T, Guo R, Zhang L, Yang M. The coupling mechanism of mammalian mitochondrial complex I. *Nature Structural & Molecular Biology*. 2022; 29: 172–182.
- [45] Hong J, Kwon HJ, Cachau R, Chen CZ, Butay KJ, Duan Z, *et al.* Dromedary camel nanobodies broadly neutralize SARS-CoV-2 variants. *Proceedings of the National Academy of Sciences of the United States of America*. 2022; 119: e2201433119.
- [46] Li T, Cai H, Yao H, Zhou B, Zhang N, van Vliissingen MF, *et al.* A synthetic nanobody targeting RBD protects hamsters from SARS-CoV-2 infection. *Nature Communications*. 2021; 12: 4635.
- [47] Sun D, Sang Z, Kim YJ, Xiang Y, Cohen T, Belford AK, *et al.* Potent neutralizing nanobodies resist convergent circulating variants of SARS-CoV-2 by targeting diverse and conserved epitopes. *Nature Communications*. 2021; 12: 4676.
- [48] Yi C, Sun X, Ye J, Ding L, Liu M, Yang Z, *et al.* 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*. 2020; 17: 621–630.
- [49] Jackson CB, Farzan M, Chen B, Choe H. Mechanisms of SARS-CoV-2 entry into cells. *Nature Reviews. Molecular Cell Biology*. 2022; 23: 3–20.
- [50] Lassmann T. *Kalign 3: Multiple Sequence Alignment of Large Datasets*. Oxford University Press: Oxford. 2020.
- [51] Lu Y, Zhao T, Lu M, Zhang Y, Yao X, Wu G, *et al.* The Analyses of High Infectivity Mechanism of Sars-Cov-2 and Its Variants. *COVID*. 2021; 1: 666–673.
- [52] Xie Y, Karki CB, Du D, Li H, Wang J, Sobitan A, *et al.* Spike Proteins of SARS-CoV and SARS-CoV-2 Utilize Different Mechanisms to Bind With Human ACE2. *Frontiers in Molecular Biosciences*. 2020; 7: 591873.
- [53] Dumoulin M, Conrath K, Van Meirhaeghe A, Meersman F, Heremans K, Frenken LGJ, *et al.* Single-domain antibody fragments with high conformational stability. *Protein Science*. 2002; 11: 500–515.
- [54] Premkumar L, Segovia-Chumbez B, Jadi R, Martinez DR, Raut R, Markmann A, *et al.* The receptor binding domain of the viral spike protein is an immunodominant and highly specific target of antibodies in SARS-CoV-2 patients. *Science Immunology*. 2020; 5: eabc8413.
- [55] Dacon C, Tucker C, Peng L, Lee CD, Lin T, Yuan M, *et al.* Broadly neutralizing antibodies target the coronavirus fusion peptide. *Science*. 2022; 377: 728–735.
- [56] Low JS, Jerak J, Tortorici MA, McCallum M, Pinto D, Cassotta A, *et al.* ACE2-binding exposes the SARS-CoV-2 fusion peptide to broadly neutralizing coronavirus antibodies. *Science*. 2022; 377: 735–742.
- [57] Amitai A. Viral surface geometry shapes influenza and coronavirus spike evolution through antibody pressure. *PLoS Computational Biology*. 2021; 17: e1009664.
- [58] Chen B, Khodadoust MS, Olsson N, Wagar LE, Fast E, Liu CL, *et al.* Predicting HLA class II antigen presentation through integrated deep learning. *Nature Biotechnology*. 2019; 37: 1332–1343.
- [59] Mason DM, Friedensohn S, Weber CR, Jordi C, Wagner B, Meng SM, *et al.* Optimization of therapeutic antibodies by predicting antigen specificity from antibody sequence via deep learning. *Nature Biomedical Engineering*. 2021; 5: 600–612.
- [60] Tubiana J, Schneidman-Duhovny D, Wolfson HJ. ScanNet: an interpretable geometric deep learning model for structure-based protein binding site prediction. *Nature Methods*. 2022; 19: 730–739.
- [61] Xu Z, Davila A, Wilamowski J, Teraguchi S, Standley DM. Improved Antibody-Specific Epitope Prediction Using AlphaFold and AbAdapt. *Chembiochem*. 2022; 23: e202200303.
- [62] Lim YW, Adler AS, Johnson DS. Predicting antibody binders and generating synthetic antibodies using deep learning. *MAbs*. 2022; 14: 2069075.
- [63] Mahajan SP, Ruffolo J, Frick R, Gray JJ. Towards Deep Learning Models for Target-Specific Antibody Design. *Biophysical Journal*. 2022; 121: 528a.
- [64] Prihoda D, Maamary J, Waight A, Juan V, Fayadat-Dilman L, Svozil D, *et al.* BioPhi: A platform for antibody design, humanization, and humanness evaluation based on natural antibody repertoires and deep learning. *MAbs*. 2022; 14: 2020203.
- [65] Ruffolo JA, Gray JJ. Fast, Accurate Antibody Structure Prediction from Deep Learning on Massive Set of Natural Antibodies. *Biophysical Journal*. 2022; 121: 155a–156a.
- [66] Schneider C, Buchanan A, Taddese B, Deane CM. DLAB-Deep learning methods for structure-based virtual screening of antibodies. *Bioinformatics*. 2022; 38: 377–383.
- [67] Myung Y, Pires DEV, Ascher DB. CSM-AB: graph-based antibody-antigen binding affinity prediction and docking scoring function. *Bioinformatics*. 2021; 38: 1141–1143.
- [68] Taft JM, Weber CR, Gao B, Ehling RA, Han J, Frei L, *et al.* Predictive Profiling of Sars-Cov-2 Variants by Deep Mutational Learning. *bioRxiv*. 2021. (preprint)
- [69] Shan S, Luo S, Yang Z, Hong J, Su Y, Ding F, *et al.* Deep learning guided optimization of human antibody against SARS-CoV-2 variants with broad neutralization. *Proceedings of the National Academy of Sciences of the United States of America*. 2022; 119: e2122954119.
- [70] Taft JM, Weber CR, Gao B, Ehling RA, Han J, Frei L, *et al.* Deep mutational learning predicts ACE2 binding and antibody escape to combinatorial mutations in the SARS-CoV-2 receptor-binding domain. *Cell*. 2022; 185: 4008–4022.e14.
- [71] Zhang R, Ghosh S, Pal R. Predicting binding affinities of emerging variants of SARS-CoV-2 using spike protein sequencing data: observations, caveats and recommendations. *Briefings in Bioinformatics*. 2022; 23: bbac128.
- [72] Shaver JM, Smith J, Amimeur T. Deep Learning in Therapeutic Antibody Development. *Methods in Molecular Biology*. 2022; 2390: 433–445.