Iterative In Silico Screening for Optimizing Stable Conformation of Anti-SARS-CoV-2 Nanobodies

Wenyuan Shang[1*], Xiujun Hu[1*], Xiaoman Lin[1], Shangru Li[1], Shuchang Xiong[1], Bingding Huang[1#], Xin Wang[1#]

[1]College of Big Data and Internet, Shenzhen Technology University, Shenzhen, 518118, China

* Co-first authors.

Corresponding authors. E-mail: huangbingding@sztu.edu.cn and wangxin@sztu.edu.cn

ABSTRACT

***Abstract*—Nanobodies (Nbs or VHHs) are single-domain antibodies (sdAbs) derived from camelid heavy-chain antibodies. The variable region of these nanobodies has special and unique characteristics, such as small size, good tissue penetration, and cost-effective production, making nanobodies a good candidate for the diagnosis and treatment of viruses. Identifying effective nanobodies against the COVID-19 would help us defeat this dangerous virus or other unknown variants in future. Herein, we introduce an in silico screening strategy for optimizing stable conformation of anti-SARS-CoV-2 nanobodies.**

Firstly, various complexes containing nanobodies were downloaded from the RCSB database, which were identified from immunized llamas. The primary docking between nanobodies and the SARS-CoV-2 spike protein receptor-binding domain was performed through ClusPro program, with the manually screening that leaving the reasonable conformation to the next step. Then, the binding distances of atoms between the antigen-antibody interfaces were measured through the NeighborSearch algorithm. Finally, filtered nanobodies were acquired according to HADDOCK scores through HADDOCK docking the COVID spike protein with nanobodies under restrictions of calculated molecular distance between active residues and antigenic epitopes less than 4.5 Å. In this way, those nanobodies which with more reasonable conformation and with stronger neutralizing efficacy were acquired.

To validate the efficacy ranking of the nanobodies we obtained, we calculated the binding affinities (ΔG) and dissociation constants (Kd) of all screened nanobodies using the PRODIGY web tool, and predicted the stability changes induced by all possible point mutations in nanobodies using the MAESTROWeb server. Furthermore, we examined the performance of the relationship between nanobodies' ranking and their number of mutation-sensitive sites (Spearman correlation > 0.68), the results revealed a robust correlation, indicating that the superior nanobodies identified through our screening process exhibited fewer mutation hotspots and higher stability. This correlation analysis demonstrates the validity of our screening criteria, underscoring the suitability of these nanobodies for future development and practical implementation.

In conclusion, this three-step screening strategy iteratively in silico greatly improved the accuracy of screening desired nanobodies compared to using only ClusPro docking or default HADDOCK docking settings. It provides new ideas for the screening of novel antibodies and computer-aided screening methods.

***Index Terms*—nanobody, in silico screening, protein docking**

INTRODUCTION

Broke out in late 2019, COVID-19 was a major health problem worldwide, owing to the lack of effective treatment methods. With the rapid global spread of Omicron and other notable variants, the development of more potent antibodies has been a global concern.

There are three main proteins on the surface of the SARS-CoV-2 virus, namely spike protein (S), envelope protein (E), and membrane protein (M) [1,2]. Among them S protein has a key role in virus entry into host cells and is an important target for SARS-CoV-2 infection. The S protein is a class of trimeric transmembrane glycoproteins, which has a full size of 1,273 amino acid residues and consists of two functional subunits (S1 and S2) [3]. S1 subunit is responsible for binding to the host cell receptor and S2 subunit can fuse with the viral and cellular membranes. The S1 subunit can be further divided into two functional domains, an N-terminal domain and a C-terminal domain (CTD) involving the receptor-binding domain (RBD) , which is the main receptor used by SARS-Cov-2 for cell entry and directly binds to the receptor angiotensin-converting enzyme 2 (ACE2) on host cells [4,5]. The C-terminal domain is involved in the assembly and release of viral particles. Thus, S

protein is a key molecule for viral infection of host cells by decorating the virion surface as a major antigen and inducing neutralizing antibody responses.

Tremendous progress in the structure and function of SARS-CoV-2 spike protein has been made since the initial outbreak of the COVID-19. Some treatments have shown some benefits in certain patients to date [6,7,8], however, there are no generally proven effective therapies for antivirals against SARS-CoV-2. Studies reported that specific monoclonal antibody therapy is an effective immunotherapy for SARS-CoV-2 infection [9] and vaccination is the most effective method for a long-term strategy for prevention and control of COVID-19 [10,11]. However, the wide application of therapeutic monoclonal antibodies may be restricted by the high cost and limited capacity of manufacturing, as well as the problem of bioavailability.

Nanobodies (Nbs) are a special type of antibodies with no light Chains (L) in their structure and composed of only heavy chains (H). It is derived from the peripheral blood of Camelidae family animals, including alpacas. The peculiar properties of Nbs include nanoscale size, stability, high affinity and specificity, water solubility, high tissue penetration and a sustainable source, et al. These Nbs have attracted considerable interest and applications of Nbs depend on reliable, cost-effective and high-volume production. These unique properties of Nbs over conventional antibodies make them powerful therapies against SARS-CoV-2 [12,13]. Effective nanobodies have been successfully generated in biological experiments [14,15], which could be used for targeting the SARS-CoV-2 virus for the diagnosis and treatment of COVID-19. Computational studies on Nbs targeting SARS-CoV-2 S protein have attracted enormous attention [16]. However, there is a lack of in silico screening techniques to promptly identify a large number of nanobodies which can effectively neutralizing SARS-CoV-2 virus.

In this paper, we highlight an iterative screening strategy: the flowchart in fig.1 depicts the methodology used to predict the molecular docking and binding free energy between nanobodies and the SARS-CoV-2 Spike protein RBD using ClusPro and HADDOCK molecular docking servers with the algorithm of NeighborSearch, which predicted the active binding sites for HADDOCK to calculate the final results. Notably, we identified the active residues on both the antigen and nanobody, subsequently integrated them into HADDOCK to enhance the performance of the in silico screening.

METHOD

A. Data Mining

In this study, we searched for PDB files containing nanobody complexes related to SARS viruses in the RCSB [17] database (https://www.rcsb.org/). We selected nanobodies obtained through biological experiments and especially cryo-electron microscopy since these nanobodies are more reliable. Additionally, we also retrieved a PDB file containing the natural spike protein of the SARS-CoV-2 from the RCSB database.

B. Molecular Docking and Determination of Biophysical Interactions

B1. Pre-docking with ClusPro

Molecular docking [18] was used to determine protein-protein interactions, which is a useful and validated technique for antibody screening.

We firstly employed ClusPro [19–22] (https://cluspro.org/) for docking. CluPro is a computational tool for molecular docking of diverse molecular interactions, including protein-protein, protein-nucleic acid, and protein-small molecule complexes. Massive information including interaction force, binding free energies and combine modes between molecules can be predicted by molecular mechanics and Monte Carlo simulations techniques. For each antigen-antibody complex, cluspro retained 1,000 lowest energy models and evaluated the top 30 models (centers of the 30 largest clusters) using the DockQ program [23]. When some complexes have fewer than 30 clusters, a DockQ result of 0.23 (acceptable) and above (better) would be considered a good solution (Cluspro considers a DockQ result of 0.23 and above as a good solution [19]). Top 15 predictions of each docking conformation were considered to the next manual selection step.

B2. Manual screening step

However, when only employed ClusPro, we noticed that some predicted antibody binding sites were not within the virus's RBD, although they displayed good stability, these protein pairing confirmations are not useful and should be filtered out manually. So, we added a manual screening step here after the ClusPro docking to filter out the docking results which docked outside the virus's RBD region.

C. Identifying active residues

The stability of antigen-antibody binding can be evaluated by factors such as the interface area of interaction, hydrogen bonds, van der Waals interactions etc. at the interface. In addition, the distance between atoms at the antigen-antibody interface is also an important indicator for evaluating binding stability. Bio.PDB.NeighborSearch is a powerful Python module that is part of the Biopython [24] library. It seamlessly integrates with other protein structure analysis tools and data structures. It can interact with objects such as Structure, Model, Chain, and Residue from Bio.PDB, providing a fast and efficient method to search for neighboring atoms or residues in protein structures.

The Bio.PDB.NeighborSearch accomplishes rapid nearest neighbor search by constructing k-d trees [25]. A k-d tree (k-dimensional tree) is a binary tree data structure used for partitioning and organizing points in k-dimensional space. In the context of protein structures, each atom can be represented as a k-dimensional vector, where k is the dimensionality of the space (typically 3, representing three-dimensional spatial coordinates). The k-d tree divides the space into multiple hyper-rectangular regions, where each node represents a hyper-rectangular region and is stored in the tree nodes.

The algorithm steps as follows: Firstly, a partitioning dimension is selected (typically the dimension exhibiting the highest variance), and a partitioning value is chosen to segregate the data into left and right subtrees. This partitioning process is recursively applied to the left and right subtrees until each leaf node exclusively encompasses a singular atom or residue.

The search commences from the root node. By comparing the query point with the partitioning value of the current node, the direction of the search, whether within the left subtree or the right subtree, is determined. The same search process is repeatedly executed within the selected subtree until the nearest neighbor atom is identified, and finally, the program returns the found neighboring atoms.

For antigen-antibody, we used NeighborSearch algorithm traverse the atoms of both chains. For each pair of atoms, the program uses the NeighborSearch algorithm to find the distance between chains. If the distance found is not greater than the set threshold, it predicates that the atom in the antibody chain is a "close atoms". The residue, composed of multiple atoms, is the fundamental structural unit of a protein. We identified key residues as active residues at the binding interfaces of the antigen-nanobody from the close atoms, and saved its ID for the HADDOCK step.

(The code of active residues identification could be downloaded at: https://github.com/WangLabforComputationalBiology/SARS-COV-nanobody-screening)

D. Further docking with HADDOCK

We employed HADDOCK [26, 27] for the next step iterative screening, which can perform rigid, semi-flexible, and fully flexible docking. (https://wenmr.science.uu.nl/haddock2.4/) The screening steps of HADDOCK provide several advantages for our further docking and screening.

Firstly, HADDOCK incorporates information from identified or predicted protein interfaces, utilizing ambiguous interaction restraints (AIRs) to guide the docking process. This allows for a more informed and targeted exploration of the protein-protein interaction space. Additionally, HADDOCK enables the inclusion of specific unambiguous distance restraints, such as those derived from Mass Spectrometry cross-links, as well as various other experimental data sources, including NMR residual dipolar couplings and cryo-EM maps. This integration of diverse experimental data enhances the accuracy and reliability of the docking predictions [28].

Moreover, by incorporating active residues to constrain the prediction range, the accuracy of docking prediction results can be improved and the computational complexity can be reduced.

By first using ClusPro for a primary screening, we were able to quickly select reasonable conformations to calculate their active residue. Then, with the active residue information acquired by Method C, HADDOCK could be used to further evaluate the performance of the docking more precisely.

E. Determination of Binding Affinity and Stability Validation of Screening

The energy of protein-protein interaction between the spike protein fragment and nanobodies is an important factor to evaluate the stability of different nanobodies upon binding. Binding affinity represents the strength of protein-protein interaction and is associated with the cellular functions of these proteins. This binding energy can be quantified as a physicochemical parameter known as the dissociation constant (Kd). The prediction of binding affinity between nanobodies and the spike protein was performed using the PRODIGY[29] server, which is based on an intermolecular contacts prediction method for protein-protein complex binding affinity.

In addition to evaluating the binding affinity in diverse conformations, assessing the robustness of nanobodies is a crucial factor in determining their suitability for industrial production. While protein mutations can have beneficial effects on physicochemical or biological properties, they can also result in structural variations and loss of function. Therefore, estimating the stability changes caused by all possible point mutations can provide an additional validation of the exceptional quality of the screened nanobodies. For prediction, we used the 'Calculate a mutation sensitivity profile' module on the MAESTROweb [30]server, which shares similarities with the PoPMuSiC [31] web server.

In the MAESTROweb server, the stability changes resulting from point mutations are denoted as $\Delta\Delta G$ (s_w, s_m), which are used to estimate the folding free energy change when residue s_w is mutated to s_m. It is calculated based on the wild-type protein structure and a set of energy functions:

$$\Delta\Delta G_p = \sum_{i=1}^{13} \alpha_i(A)\Delta\Delta W_i + \alpha_{14}(A)\Delta V_+ + \alpha_{15}(A)\Delta V_- + \alpha_{16}(A)$$

The coefficients α_i depend on the solvent accessibility A of the wild-type amino acid s_w. $\Delta\Delta W_i$ is a linear combination of 13 statistical potentials, which describe the torsional angles of amino acid type s, defined backbone conformation t, solvent accessibility a, and the correlation between the spatial distance d of the average geometric centers of each residue pair and their side chains. $\Delta V\pm$ is related to the difference in volume between the mutant amino acid and the wild-type amino acid: $\Delta V = V_m - V_w$. They are defined as $\Delta V\pm = \Delta V H(\pm\Delta V)$, where H represents the Heaviside function. They offer a depiction of the effects associated with the formation of a cavity (when $\Delta V<0$) or the accommodation of a larger side-chain within the protein structure (when $\Delta V>0$).

For each position *i* in the protein, $\Delta\Delta G$ can be calculated for all 19 possible mutations. ($\Delta\Delta G_{pred}<0.0$ indicates a confirmed stabilizing mutation.) Mutations occurring at positions with lower $\Delta\Delta G_{pred}$ values have a lesser impact, indicating a reduced sensitivity to mutation. We conducted a statistical analysis of the MAESTROweb results for each nanobody, counting the number of sensitive positions where $\Delta\Delta G_{pred}$ exceeded 1.8 (kcal/mol). Additionally, the MAESTROweb server provides a visualization service, which can show us the impact of each mutation at all potential positions (Fig. 2).

RESULTS

We preprocessed the nanobodies and spike proteins using PyMOL [32], removed unwanted molecules such as water and chloride ions, as well as impurities such as small-molecule ligands or other proteins. We also checked structural issues and then used ClusPro to dock them. Multiple sets of predictive conformations were obtained by docking each pair of spike proteins and nanobodies. However, some predicted conformations did not exhibit nanobody binding within the RBD region of the spike protein, which is a critical area to screen for nanobodies capable of effectively competing with ACE2 for binding to the RBD domain [33] (Fig. 3).

Further screening was then performed using HADDOCK, and active residues were added for precise localization, and the conformation of the binding site within the RBD region was obtained. The active residues are situated on the interface of interaction between the nanobody and S protein. We set the distance threshold (4.5Å) between the molecules at the antigen-antibody interface to clarify if these regions are active residues. Within this 4.5Å distance range, the residue interactions between protein chains is regarded as strong interactions. Active residues play a crucial role in maintaining the stability of the protein complex and conferring binding specificity [34]. Fig. 4 shows the predicted active residues in the docking of the S protein and nanobody, indicating the positions and corresponding residue IDs of active residues.

Finally, Haddock docking was completed, and each predicted conformation was assigned an evaluation score. The HADDOCK score is calculated using a linear function that combines various energies and buried surface area [27]. The lower the score, the better the conformation. Moreover, with the ClusPro pre-screening step, the binding sites of the antibodies are reasonably located within the RBD region. Based on the best score ranking from each predicted conformation, we selected the top 30 excellent nanobodies. In Table 1, we listed the source PDBID, the best HADDOCK Score, and their HADDOCK Score ERROR Range for each nanobody.

To validate our findings, we used Spearman correlation to assess the correlation between the rankings of nanobodies and the number of mutation-sensitive positions they exhibited (Fig.5). And in the PRODIGY, our findings indicate that the top-ranking nanobodies also displayed concurrent high binding affinities.

CONCLUSION

Compared to using only ClusPro docking or only default Haddock docking settings, this three-step approach provided a more comprehensive and efficient way of analyzing the molecular interactions between the S protein and the nanobodies. This strategy not only providing a valuable tool for the selection of nanobodies but also presents new ideas for the screening of novel antibodies and the development of computer-aided screening methods. The iterative process and the incorporation of multiple algorithms with customized settings can be applied to other antibody screening projects and may lead to the identification of more effective therapeutic antibodies.

Moreover, we conducted a comprehensive investigation on the stability of nanobodies in relation to their rankings. To accomplish this, we used MAESTROWeb server to predict the stability changes resulting from point mutations in nanobodies. We performed a statistical analysis to quantify the number of potential mutation sites. Our analysis demonstrated a significant correlation between the number of mutation-sensitive sites that exhibited significant energy fluctuations and the rankings of the nanobodies. In conclusion, our three-step screening strategy effectively identifies nanobodies with superior mutation resistance, rendering them highly suitable for practical production applications.

DISCUSSION

COVID-19 is an infectious disease caused by novel coronavirus known as SARS-CoV-2. As the COVID-19 pandemic erupted globally from late 2019, it caused tremendous impacts on human health, economies and other various aspects to many countries. As of May 2023, there have been over 760 million confirmed cases and more than 69,000 deaths [35].

The rapid spread of the COVID-19 pandemic can be attributed to the primary modes of transmission of the SARS-CoV-2 virus, which are airborne and contact transmission. The virus can spread through respiratory secretions, droplets, and aerosols. Simply being in contact with an individual infected with SARS-CoV-2 poses a risk of transmission. People may experience various symptoms, including fever, cough, shortness of breath, fatigue, muscle pain, sore throat, loss of taste or smell, and others. Severe cases may require hospitalization and even respiratory support with ventilators.

The COVID-19 pandemic has had a profound impact, and we are working to develop effective strategy. We hope to extend this approach to screen out a bunch of effective antibodies, especially nanobodies. This study aimed to improve the accuracy of docking predictions for the spike protein and nanobody complex by introducing HADDOCK for iterative filtering of ClusPro pre-screened docking results. The calculated active residues were used as a constraint output to refine the ClusPro docking predictions and reduce computational complexity.

This approach not only increased the accuracy of ClusPro/HADDOCK docking but also addressed the issue of how to filter out unrealistic binding sites that might result from pure rigid docking. Then, HADDOCK score [36] was used as a comprehensive evaluation index to compare the quality and accuracy of different docking conformations, with a lower score indicated a more stable and reliable docking conformation.

In future studies, we aim to employ this iterative filtering approach to dock a wider range of viral variants and identify high-quality nanobodies to specific pathogens.

In the next step, we would also integrate this iterative filtering module into our NanoLAS online platform [37], providing an integrated and efficient service for screening nanobodies. In NanoLAS 2.0, we will incorporate additional rigid docking programs, including ZDOCK [38], GRAMM-X [39], pyDOCK [40], and FTDock [41]. These programs exhibit comparable performance to ClusPro and can serve as alternative options for the initial screening step. After the initial rigid docking, if further precise screening of higher quality nanobodies is required, the user can directly input the output of the first screening step into the next module. The module will calculate the active residues of the antigen-antibody binding site and use it for more precise and targeted screening via algorithms such as HADDOCK or RosettaDOCK [42].

Furthermore, the residue IDs of binding sites on the spike protein chain in the models may provide important information on potentially binding epitopes. In our experience, it is crucial to find nanobody structures that match the epitopes with the active residues' information. We expect our method to be extended to nanobodies screening of other diseases, especially in related neoplastic diseases such as lung cancers and brain tumors.

Currently, when screening for reasonable conformations from ClusPro, we still rely on manual selection based on our experience. Manually screening could also be replaced by the programs with appropriate algorithms in future. This technological progress will further standardize the screening flow and improve efficiency.

REFERENCES

[1] P. Zhou, X.-L. Yang, X.-G. Wang, B. Hu, L. Zhang, W. Zhang, H.-R. Si, Y. Zhu, B. Li, C.-L. Huang *et al.*, "A pneumonia outbreak associated with a new coronavirus of probable bat origin," *nature*, vol. 579, no. 7798, pp. 270–273, 2020.

[2] S. P. Singh, M. Pritam, B. Pandey, and T. P. Yadav, "Microstructure, pathophysiology, and potential therapeutics of covid-19: A comprehensive review," *Journal of medical virology*, vol. 93, no. 1, pp. 275–299, 2021.

[3] A. C. Walls, Y.-J. Park, M. A. Tortorici, A. Wall, A. T. McGuire, and D. Veesler, "Structure, function, and antigenicity of the sars-cov-2 spike glycoprotein," *Cell*, vol. 181, no. 2, pp. 281–292, 2020.

[4] J. Lan, J. Ge, J. Yu, S. Shan, H. Zhou, S. Fan, Q. Zhang, X. Shi, Q. Wang, L. Zhang *et al.*, "Structure of the sars-cov-2 spike receptor-binding domain bound to the ace2 receptor," *nature*, vol. 581, no. 7807, pp. 215–220, 2020.

[5] J. Shang, G. Ye, K. Shi, Y. Wan, C. Luo, H. Aihara, Q. Geng, A. Auerbach, and F. Li, "Structural basis of receptor recognition by sars-cov-2," *Nature*, vol. 581, no. 7807, pp. 221–224, 2020.

[6] Z. Zhu, Z. Lu, T. Xu, C. Chen, G. Yang, T. Zha, J. Lu, and Y. Xue, "Arbidol monotherapy is superior to lopinavir/ritonavir in treating covid-19," *Journal of Infection*, vol. 81, no. 1, pp. e21–e23, 2020.

[7] M. Hoffmann, H. Kleine-Weber, S. Schroeder, N. Kru¨ger, T. Herrler, S. Erichsen, T. S. Schiergens, G. Herrler, N.-H. Wu, A. Nitsche *et al.*, "Sars-cov-2 cell entry depends on ace2 and tmprss2 and is blocked by a clinically proven protease inhibitor," *cell*, vol. 181, no. 2, pp. 271–280, 2020.

[8] X. Xu, M. Han, T. Li, W. Sun, D. Wang, B. Fu, Y. Zhou, X. Zheng, Y. Yang, X. Li *et al.*, "Effective treatment of severe covid-19 patients with tocilizumab," *Proceedings of the National Academy of Sciences*, vol. 117, no. 20, pp. 10 970–10 975, 2020.

[9] S. J. Zost, P. Gilchuk, J. B. Case, E. Binshtein, R. E. Chen, J. P. Nkolola, A. Scha¨fer, J. X. Reidy, A. Trivette, R. S. Nargi *et al.*, "Potently neutralizing and protective human antibodies against sars-cov-2," *Nature*, vol. 584, no. 7821, pp. 443–449, 2020.

[10] T. R. Smith, A. Patel, S. Ramos, D. Elwood, X. Zhu, J. Yan, E. N. Gary, S. N. Walker, K. Schultheis, M. Pur- war *et al.*, "Immunogenicity of a dna vaccine candidate for covid-19," *Nature communications*, vol. 11, no. 1, p. 2601, 2020.

[11] F.-C. Zhu, Y.-H. Li, X.-H. Guan, L.-H. Hou, W.-J. Wang, J.-X. Li, S.-P. Wu, B.-S. Wang, Z. Wang, L. Wang *et al.*, "Safety, tolerability, and immunogenicity of a recombinant adenovirus type-5 vectored covid-19 vaccine: a dose-escalation, open-label, non-randomised, first-in- human trial," *The Lancet*, vol. 395, no. 10240, pp. 1845– 1854, 2020.

[12] L. Hanke, L. Vidakovics Perez, D. J. Sheward, H. Das, T. Schulte, A. Moliner-Morro, M. Corcoran, A. Achour, G. B. Karlsson Hedestam, B. M. Ha¨llberg *et al.*, "An alpaca nanobody neutralizes sars-cov-2 by blocking recep- tor interaction," *Nature communications*, vol. 11, no. 1, p. 4420, 2020.

[13] S. Ezzikouri, J. Nourlil, K. Tsukiyama-Kohara, M. Ko- hara, H. El Ossmani, M. P. Windisch, and S. Benjelloun, "Nanobodies: an unexplored opportunity to combat covid-19," *Journal of Biomolecular Structure and Dynamics*, vol. 40, no. 7, pp. 3129–3131, 2022.

[14] X. Chi, X. Zhang, S. Pan, Y. Yu, Y. Shi, T. Lin, H. Duan, X. Liu, W. Chen, X. Yang *et al.*, "An ultrapotent rbd- targeted biparatopic nanobody neutralizes broad sars-cov-2 variants," *Signal transduction and targeted ther- apy*, vol. 7, no. 1, p. 44, 2022.

[15] M. Li, Y. Ren, Z. Q. Aw, B. Chen, Z. Yang, Y. Lei, L. Cheng, Q. Liang, J. Hong, Y. Yang *et al.*, "Broadly neutralizing and protective nanobodies against sars-cov-2 omicron subvariants ba. 1, ba. 2, and ba. 4/5 and diverse sarbecoviruses," *Nature Communications*, vol. 13, no. 1, p. 7957, 2022.

[16] J.-P. Salvador, L. Vilaplana, and M.-P. Marco, "Nanobody: outstanding features for diagnostic and therapeutic applications," *Analytical and bioanalytical chemistry*, vol. 411, pp. 1703–1713, 2019.

[17] S. K. Burley, C. Bhikadiya, C. Bi, S. Bittrich, L. Chen, G. V. Crichlow, C. H. Christie, K. Dalen- berg, L. Di Costanzo, J. M. Duarte *et al.*, "Rcsb protein data bank: powerful new tools for exploring 3d structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences," *Nucleic acids research*, vol. 49, no. D1, pp. D437–D451, 2021.

[18] M. R. Elmorsy, E. Abdel-Latif, H. E. Gaffer, S. E. Mahmoud, and A. A. Fadda, "Anticancer evaluation and molecular docking of new pyridopyrazolo-triazine and pyridopyrazolo-triazole derivatives," *Scientific Reports*, vol. 13, no. 1, p. 2782, 2023.

[19] I. T. Desta, K. A. Porter, B. Xia, D. Kozakov, and S. Vajda, "Performance and its limits in rigid body protein-protein docking," *Structure*, vol. 28, no. 9, pp. 1071–1081, 2020.

[20] S. Vajda, C. Yueh, D. Beglov, T. Bohnuud, S. E. Mottarella, B. Xia, D. R. Hall, and D. Kozakov, "New additions to the c lus p ro server motivated by capri," *Proteins: Structure, Function, and Bioinformatics*, vol. 85, no. 3, pp. 435–444, 2017.

[21] D. Kozakov, D. R. Hall, B. Xia, K. A. Porter, D. Pad- horny, C. Yueh, D. Beglov, and S. Vajda, "The cluspro web server for protein–protein docking," *Nature proto- cols*, vol. 12, no. 2, pp. 255–278, 2017.

[22] D. Kozakov, D. Beglov, T. Bohnuud, S. E. Mottarella, B. Xia, D. R. Hall, and S. Vajda, "How good is automated protein docking?" *Proteins: Structure, Function, and Bioinformatics*, vol. 81, no. 12, pp. 2159–2166, 2013.

[23] S. Basu and B. Wallner, "Dockq: a quality measure for protein-protein docking models," *PloS one*, vol. 11, no. 8, p. e0161879, 2016.

[24] P. J. Cock, T. Antao, J. T. Chang, B. A. Chapman, C. J. Cox, A. Dalke, I. Friedberg, T. Hamelryck, F. Kauff, B. Wilczynski *et al.*, "Biopython: freely available python tools for computational molecular biology and bioinfor- matics," *Bioinformatics*, vol. 25, no. 11, p. 1422, 2009.

[25] J. L. Bentley, "Multidimensional binary search trees used for associative searching," *Communications of the ACM*, vol. 18, no. 9, pp. 509–517, 1975.

[26] R. V. Honorato, P. I. Koukos, B. Jime´nez-Garc´ıa, A. Tsaregorodtsev, M. Verlato, A. Giachetti, A. Rosato, and A. M. Bonvin, "Structural biology in the clouds: the wenmr-eosc ecosystem," *Frontiers in Molecular Bio- sciences*, vol. 8, p. 729513, 2021.

[27] G. Van Zundert, J. Rodrigues, M. Trellet, C. Schmitz, P. Kastritis, E. Karaca, A. Melquiond, M. van Dijk, S. De Vries, and A. Bonvin, "The haddock2. 2 web server: user-friendly integrative modeling of biomolecular complexes," *Journal of molecular biology*, vol. 428, no. 4, pp. 720–725, 2016.

[28] C. Dominguez, R. Boelens, and A. M. Bonvin, "Haddock: a protein-protein docking approach based on biochemical or biophysical information," *Journal of the American Chemical Society*, vol. 125, no. 7, pp. 1731– 1737, 2003.

[29] L. C. Xue, J. P. Rodrigues, P. L. Kastritis, A. M. Bonvin, and A. Vangone, "Prodigy: a web server for predicting the binding affinity of protein–protein complexes," *Bioinformatics*, vol. 32, no. 23, pp. 3676–3678, 2016.

[30] J. Laimer, J. Hiebl-Flach, D. Lengauer, and P. Lackner, "Maestroweb: a web server for structure-based protein stability prediction," Bioinformatics, vol. 32, no. 9, pp. 1414–1416, 2016.

[31] Y. Dehouck, J. M. Kwasigroch, D. Gilis, and M. Rooman, "Popmusic 2.1: a web server for the estimation of protein stability changes upon mutation and sequence optimality," *BMC bioinformatics*, vol. 12, no. 1, pp. 1–12, 2011.

[32] W. L. DeLano *et al.*, "Pymol: An open-source molecular graphics tool," *CCP4 Newsl. Protein Crystallogr*, vol. 40, no. 1, pp. 82–92, 2002.

[33] H. Aria, F. Mahmoodi, H. S. Ghaheh, H. Zare, M. Heiat, H. Bakherad *et al.*, "Outlook of therapeutic and diag- nostic competency of nanobodies against sars-cov-2: A systematic review," *Analytical Biochemistry*, vol. 640, p. 114546, 2022.

[34] S. Vishwanath, A. Sukhwal, R. Sowdhamini, and N. Srinivasan, "Specificity and stability of transient protein–protein interactions," *Current Opinion in Struc- tural Biology*, vol. 44, pp. 77–86, 2017.

[35] P. Relan, N. V. Motaze, K. Kothari, L. Askie, O. Le Po- lain, M. D. Van Kerkhove, J. Diaz, and B. K. T. Vija- yaraghavan, "Severity and outcomes of omicron variant of sars-cov-2 compared to delta variant and severity of omicron sublineages: a systematic review and metanaly- sis," *BMJ Global Health*, vol. 8, no. 7, p. e012328, 2023.

[36] N. C. Das, P. Chakraborty, J. Bayry, and S. Mukherjee, "Comparative binding ability of human monoclonal antibodies against omicron variants of sars-cov-2: An in silico investigation," *Antibodies*, vol. 12, no. 1, p. 17, 2023.

[37] S. Xiong, Z. Liu, X. Yi, K. Liu, B. Huang, and X. Wang, "Nanolas: a comprehensive nanobody database with data integration, consolidation and application," *Database*, vol. 2024, p. baae003, 2024.

[38] R. Chen, L. Li, and Z. Weng, "Zdock: an initial-stage protein-docking algorithm," *Proteins: Structure, Func- tion, and Bioinformatics*, vol. 52, no. 1, pp. 80–87, 2003.

[39] A. Tovchigrechko and I. A. Vakser, "Gramm-x public web server for protein–protein docking," *Nucleic acids research*, vol. 34, no. suppl 2, pp. W310–W314, 2006.

[40] T. M.-K. Cheng, T. L. Blundell, and J. Fernandez-Recio, "pydock: electrostatics and desolvation for effective scor- ing of rigid-body protein–protein docking," *Proteins: Structure, Function, and Bioinformatics*, vol. 68, no. 2, pp. 503–515, 2007.

[41] G. Moont, H. A. Gabb, and M. J. Sternberg, "Use of pair potentials across protein interfaces in screening predicted docked complexes," *Proteins: Structure, Function, and Bioinformatics*, vol. 35, no. 3, pp. 364–373, 1999.

[42] S. Lyskov and J. J. Gray, "The rosettadock server for local protein–protein docking," *Nucleic acids research*, vol. 36, no. suppl 2, pp. W233–W238, 2008.

Fig. 1: The flowchart of three-step in Silico screening strategy. Left panel: the first screening step by ClusPro.Middle panel: Identifying active residues of spike by NeighborSearch Algorithm. Right panel: Further screening by HADDOCK with active residues information.

Fig. 2: Mutation sensitivity profile of nanobody (PDBID: 7tpr). The vertical axis represents ΔΔG, and the horizontal axis depicts amino acid residues.

Fig. 3: A comparison of two ClusPro-predicted models for nanobody binding sites(a) Nanobodies bind outside the SARS-CoV-2RBD domain. (b) Nanobodies bind within the SARS- CoV-2 RBD domain (shaded area).

Fig. 4: Visualization of Program-Computed Active Residues by NeighborSearch algorithm.

Fig. 5: Correlation plot depicting the predicted count of mutation-sensitive sites (hot spot sites) and the ranking order of the screened nanobodies.

Table I: Top 30 results of nanobodies screened by HADDOCKScore

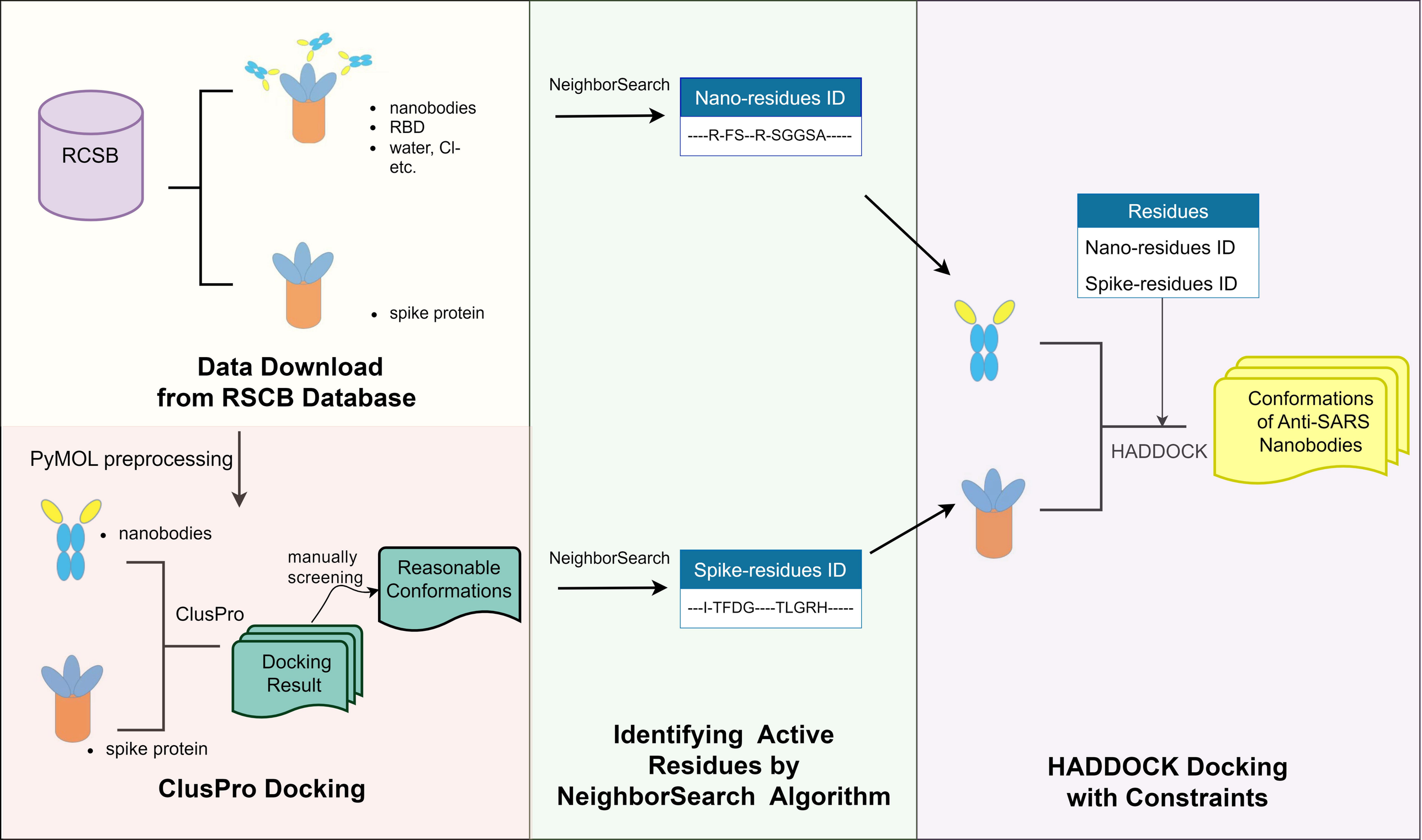
(Full list could be downloaded at https://www.nanolas.cloud/?#/download)

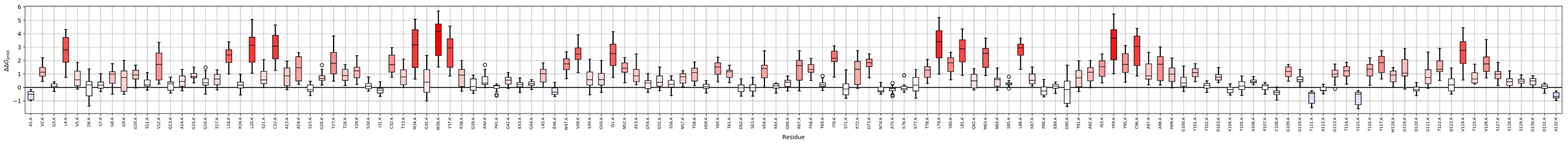
Source (PDB ID)	Nanobody Names	Best Score	Error Range
7tpr	Nanobody 8A2	-175.8	±5.9
7tpr	Nanobody 7A3	-162.2	±5.1
7rby	Nanobody nb-112	-154	±2.9
7rxd	Nb_RBD	-153.9	±3.3
7vq0	Nanobody P86	-149.4	±13.2
7vnb	n3113	-148.8	±1.9
8cyb	Nanobody 1-8	-148.7	± 1.2
7a25	Sybody 23	-148	± 5.2
8cy7	Nanobody 2-38	-143.9	± 4.8
7r4r	Nanobody 1.10	-141.5	±0.6
7r4q	Nanobody 1.29	-141.3	±3.0
7x4i	Nanobody aSA3	-140.1	± 2.6
8cyc	Nanobody 2-34	-137.7	± 9.2
7b14	Nanobody	-137.7	± 4.3
8bev	Nanobody W25	-137.5	± 4.6
7whi	Bn03_nano2	-136.1	± 3.0

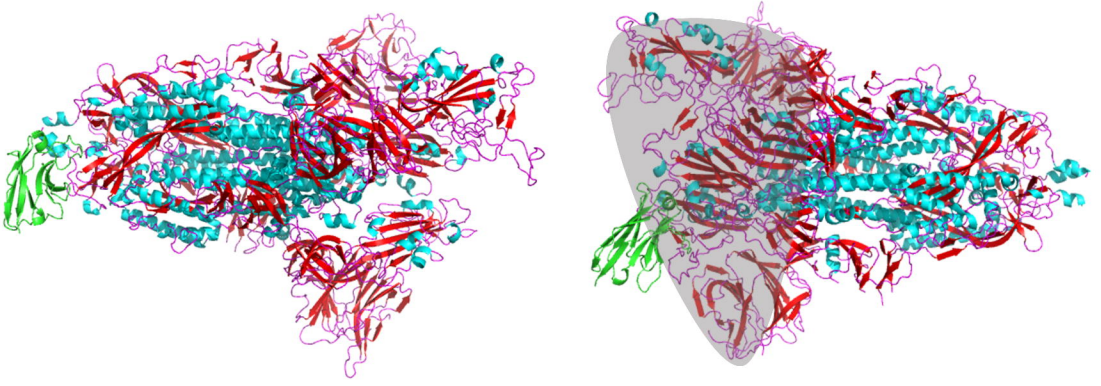
8cy9	Nanobody 1-23	-133.7	± 4.3
7x2j	Nb70	-133.5	± 2.2
7whi	Bn03_nano1	-133.1	± 5.9
7zf4	Nanobody F2	-132.5	± 3.4
7wpf	Nanobody	-132.2	± 4.6
8dqu	Nanobody	-130.9	± 2.2
8cya	Nanobody 2-67	-130.8	± 3.2
7xrp	C5G2 nanobody	-129.4	± 7.7
8cwu	VHH 1-21	-128.3	± 4.8
7xod	Nanobody	-128.1	±7.9
7voa	Nanobody	-127.9	±2.3
7wd2	Nanobody	-125	± 3.7
7r98	Nanobody B6	-124.1	±4.7
7z85	Nanobody H11-B5	-123.6	± 3.2

Table Ⅱ: Affinity of top-ranking nanobodies against spike protein

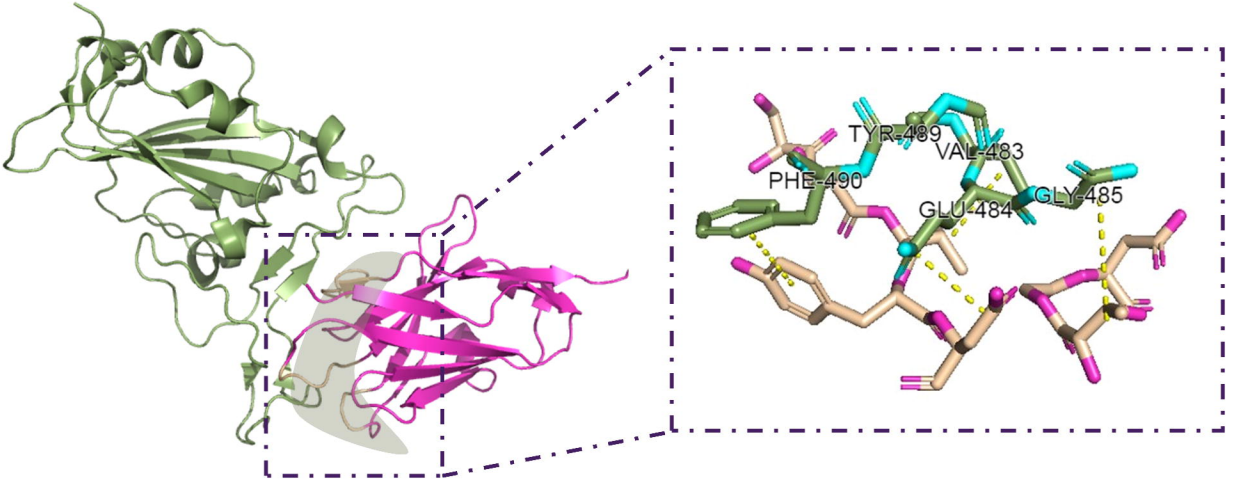
Source PDB	**ΔG (kcal mol^{-1})**
8cy7	-15.8
7tpr	-15.4
8cyb	-13.6
7tpr	-13.4
7vnb	-13.1

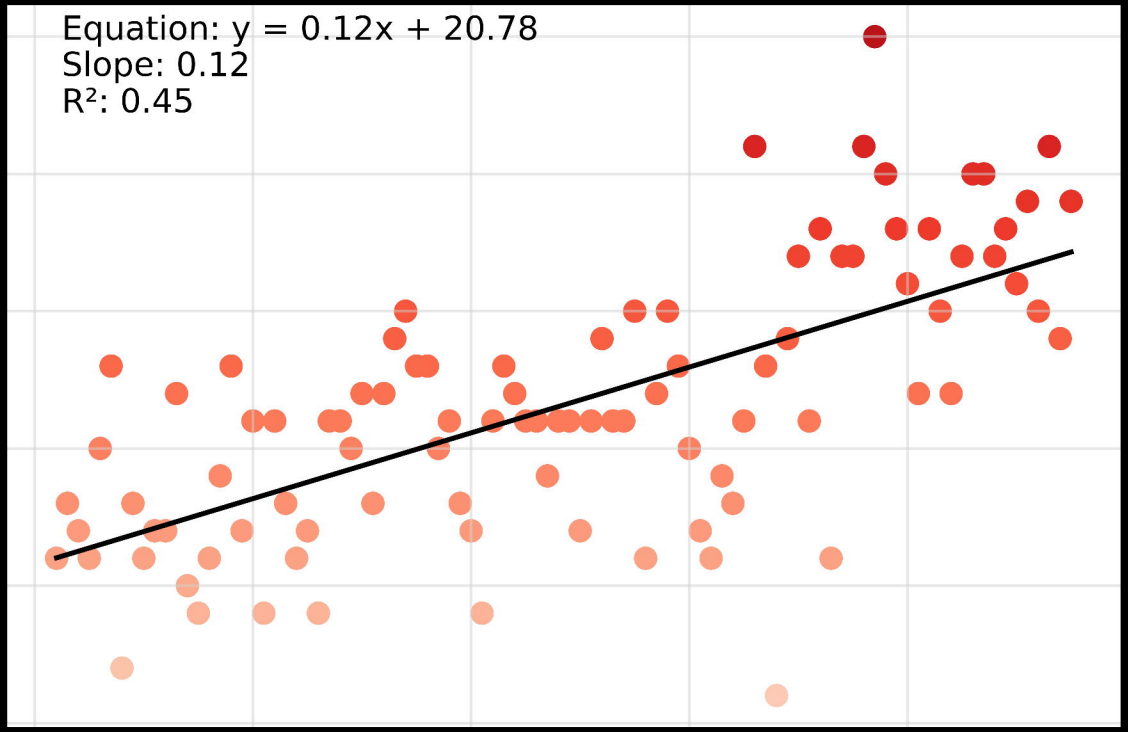

RCSB
nanobodies
RBD
water, Cl-
etc.
spike protein
Data Download
from RSCB Database
PyMOL preprocessing
nanobodies
spike protein
ClusPro
manually
screening
Reasonable
Conformations
Docking
Result
ClusPro Docking
NeighborSearch
Nano-residues ID
----R-FS--R-SGGSA-----
NeighborSearch
Spike-residues ID
---I-TFDG----TLGRH-----
Identifying Active
Residues by
NeighborSearch Algorithm
Residues
Nano-residues ID
Spike-residues ID
HADDOCK
Conformations
of Anti-SARS
Nanobodies
HADDOCK Docking
with Constraints


ΔΔG_pred
Residue
A1.A
V2.A
Q3.A
L4.A
V5.A
D6.A
S7.A
G8.A
G9.A
G10.A
S11.A
V12.A
Q13.A
A14.A
G15.A
G16.A
S17.A
L18.A
R19.A
L20.A
S21.A
C22.A
A23.A
A24.A
S25.A
G26.A
Y27.A
T28.A
Y29.A
S30.A
I31.A
C32.A
T33.A
M34.A
G35.A
V36.A
Y37.A
R38.A
Q39.A
A40.A
F41.A
G42.A
E43.A
G44.A
L45.A
E46.A
W47.A
V48.A
S49.A
G50.A
I51.A
N52.A
A53.A
D54.A
G55.A
S56.A
N57.A
T58.A
H59.A
Y60.A
T61.A
D62.A
S63.A
V64.A
K65.A
G66.A
R67.A
F68.A
T69.A
I70.A
S71.A
R72.A
D73.A
N74.A
A75.A
K76.A
K77.A
T78.A
L79.A
Y80.A
L81.A
Q82.A
M83.A
N84.A
S85.A
L86.A
K87.A
P88.A
E89.A
D90.A
T91.A
A92.A
I93.A
Y94.A
Y95.A
C96.A
A97.A
A98.A
H99.A
G100.A
T101.A
Y102.A
D103.A
K104.A
Y105.A
A106.A
P107.A
C108.A
G109.A
G110.A
F111.A
A112.A
G113.A
T114.A
Y115.A
T116.A
Y117.A
W118.A
G119.A
Q120.A
G121.A
T122.A
Q123.A
V124.A
T125.A
V126.A
S127.A
S128.A
S129.A
G130.A
Q131.A
A132.A



a b


TYR-489
VAL-483
PHE-490
GLU-484
GLY-485


Equation: y = 0.12x + 20.78
Slope: 0.12
R²: 0.45