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

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 was measured through the Neighbor Search 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.

This three-steps 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**

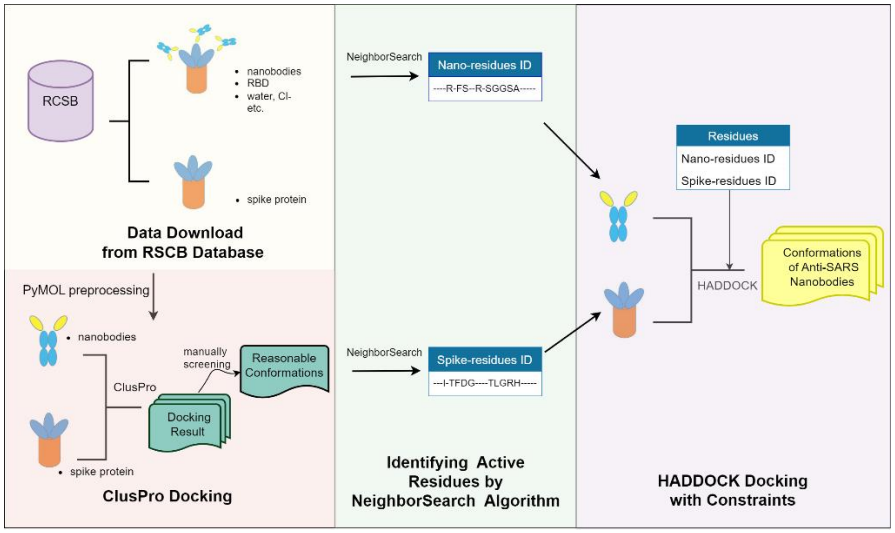

Fig. 1 The flowchart of three-steps in Silico screening strategy. Left panel: the first screening step by ClusPro. Middle panel: Identifying spike active residues by NeighborSearch Algorithm. Right panel: Further screening by HADDOCK with active residues information.

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. Effective nanobodies have been successfully generated in biological experiments [1,2], which could be used for targeting the SARS-CoV-2 virus for the

* corresponding author.
E-mail: wangxin@sztu.edu.cn

diagnosis and treatment of COVID-19. However, there is a lack of in silico screening techniques to promptly identify a large number of nanobodies.

The surface of the SARS-CoV-2 is charactered by the spike protein (S protein), which is a large membrane protein comprising approximately 1,273 amino acid residues [3]. The S protein contains the Receptor Binding Domain (RBD), which binds to the ACE2 receptor on the surface of host cells and plays a crucial role in the virus's ability to infect human cells [4, 5]. Some nanobodies can target the RBD of the spike protein, then potentially impede virus-cell fusion, which shows promise in reducing viral loads and pathogenicity [6].

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. Notably, we identified the active residues on both the antigen and nanobody, subsequently integrated them into HADDOCK to enhance the performance.

METHOD

A. Data Mining

In this study, we searched for PDB files containing nanobody complexes related to SARS viruses in the RCSB [7] 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 [8] was used to determine protein-protein interactions, which is a useful and validated technique for antibody screening.

We firstly employed ClusPro [9–12] (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.

However, when only employed ClusPro, we noticed that some antibody binding sites were not within the virus's RBD in protein pairing configurations predicted. Although they displayed good stability, these protein pairing configurations are not correct for screening nanobodies against the novel coronavirus.

B2. Further docking with HADDOCK

To solve this problem, we added the manually selected step that selected the reasonable conformation to the next step. Then, we also added HADDOCK [13, 14] for the next step iterative screening (https://wenmr.science.uu.nl/haddock2.4/), which can perform rigid, semi-flexible, and fully flexible docking. 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 select reasonable conformations to calculate their active residue. Then, with the active residue information aqquare by Method C, HADDOCK could be used to further evaluate the performance of the docking.

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 [15] is a Python module used to find distances between atoms in PDB structures. For antigen-antibody, NeighborSearch algorithm traverses the atoms of both chains. For each pair of atoms, the program uses the algorithm to find the distance between chains. If the distance found is not greater than the set threshold, it predicate that the atom in the antibody chain is "close atoms", saves its ID for the HADDOCK step.

RESULTS

We preprocessed the nanobodies and spike proteins using PyMOL [16], 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 [17].

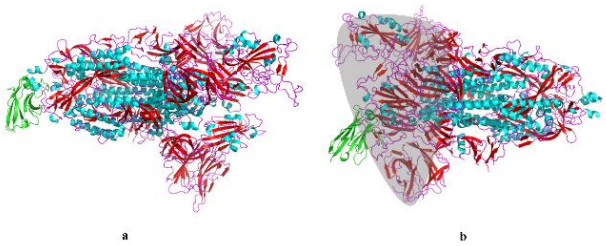

Fig. 2 A comparison of two ClusPro-predicted models for nanobody binding sites(a) Nanobodies bind outside the SARS-CoV-2 RBD domain. (b) Nanobodies bind within the SARS-CoV-2 RBD domain (shaded area).

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 [18].

The fig.3 shows the predicted active residues in the docking of the S protein and nanobody, indicating the positions and corresponding residue IDs of active residues.

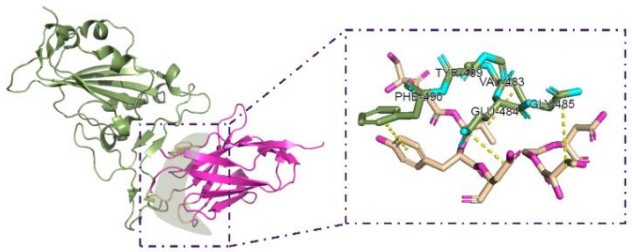

Fig. 3 Visualization of Program-Computed Active Residues by NeighborSearch algorithm

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 [14]. 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.

Table 1. Top 30 results of nanobodies screening by Haddock Score

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
7b14	Nanobody	-137.7	+/- 4.3
8cyc	Nanobody 2-34	-137.7	+/- 9.2
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
7mej	Nb21	-129	+/- 5.0
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

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

CONCLUSION

Compared to using only ClusPro docking or only default Haddock docking settings, this three-steps approach provided a more comprehensive and efficient way of analyzing the molecular interactions between the S protein and the nanobodies. This strategy not only provides 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.

DISCUSSION

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 file 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 [19] 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. 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] 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 therapy, vol. 7, no. 1, p. 44, 2022.

[2] 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.

[3] S. K. Samrat, A. M. Tharappel, Z. Li, and H. Li, "Pr ospect of sars-cov-2 spike protein: Potential role in vaccine and therapeutic development," Virus research, vol. 288, p. 198141, 2020.

[4] 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.

[5] C. Peng, Z. Zhu, Y. Shi, X. Wang, K. Mu, Y. Yang, X. Zhang, Z. Xu *et al.*, "Exploring the binding mechanism and accessible angle of sars-cov-2 spike and ace2 by molecular dynamics simulation and free energy calculation," 2021.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] S.Vajda, C. Yueh, D. Beglov, T. Bohnuud, S. E. Mottarella, B. Xia, D. R. Hall, and D. Kozakov, "New additions to the cluspro server motivated by capri," *Pro- teins: Structure, Function, and Bioinformatics*, vol. 85, no. 3, pp. 435–444, 2017.

[11] 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 protocols*, vol. 12, no. 2, pp. 255–278, 2017.

[12] 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.

[13] R. V. Honorato, P.I. Koukos, B. Jiménez-García,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.

[14] 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.

[15] 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.

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

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

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

[19] 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