Proceedings of the 2nd International Conference on Modern Medicine and Global Health
DOI: 10.54254/2753-8818/29/20240615

Sequence and structure statistics of the nanobody database and molecular dynamics simulation analysis of the nanobody VHH

Xusheng Zhang

Jilin University, Changchun, Jilin, 130012, China

zxs1227713525@gmail.com

Abstract. Antibodies are immunoglobulins produced in vivo by immune cells stimulated by exogenous molecules, and nano-antibodies are a class of molecules that are similar to conventional antibodies but smaller in size. Originally discovered as an antibody in camelidsand cartilaginous fishes, they are called heavy chain antibodies due to the lack of light and heavy chain constant regions in the CH1 region except for the retained heavy chain, and their binding region to the antigen consists of the heavy chain variable region only, making them the smallest antibodies available with complete antibody functional properties. In this experiment, we analyzed the commonalities and differences of the nanobodies by extracting all their amino acid sequences, performed protein modeling of the relevant nanobodies of coronaviruses among them, and then analyzed the data of RMSD and RMSF by using molecular dynamics simulation, and finally predicted and analyzed the protein conformation and the structures of VHH at all levels.

Keywords: nanobodies, molecular dynamics simulation, protein modeling

1. Introduction

1.1. Definition of Antibodies

An antibody (Abs), also known as Immunoglobulin, is a dimer consisting of two heavy chains and two light chains, as well as a protein molecule produced by the immune system that recognizes and binds to antigens.

1.2. Definition of Nanobodies

Nanobodies, a class of molecules similar to conventional antibodies but smaller in size. Originally discovered in camelids [1] and cartilaginous fishes [2, 3], they are called heavychain antibodies (HCAb)[4] due to the lack of light chain and heavy chain constantregion in the CH1 region except for the retained heavy chain, and their binding region to the antigen consists of the heavy chain variable region [5] (VH) only, which is currently a complete antibody with functional properties and the smallest antibody with complete antibody functional properties [6], and its molecular mass is about one-tenth of that of a general antibody [7], which is only about 15 kDa. In order to differentiate from the heavy chain variable region of the traditional antibody, the variable region of the heavy chain antibody was named as VHH or nanobody (Nb). Nanobodies contain four framework regions (FRs) and three

DOI: 10.54254/2753-8818/29/20240615

complementarity determining regions (CDRs), and the comparison of amino acid sequences shows that the structural domains of nanobodies differ from those of conventional antibodies mainly in FR2 and CDRs, with the former having more hydrophilic amino acids in FR2 [8], which enables it to have better water solubility and stability than conventional antibodies, while the CDR1 and CDR3 regions are longer than the latter, which not only have disulfide bonds to make the overall conformation stable, but also can form a more exposed convex ring structure, which is able to better recognize the size and position of the recognized target site [9]. Nanobodies can be produced in large quantities as they can be expressed in prokaryotic expression systems that lack post-translational modifications [10].

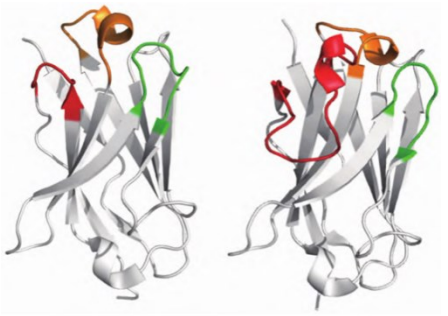

Figure 1. Structures of conventional antibody VH (left) and nanobody VHH (right). CDR1 is orange, CDR2 is green, and CDR3 is red

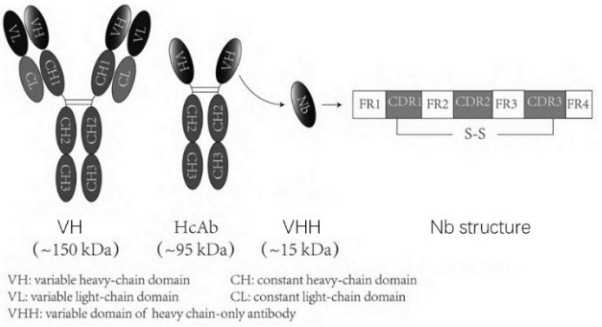

Figure 2. Structures of Monoclonal Antibodies and Nanobodies

Nanoantibodies usually have the following characteristics:
Small size, adequate database, abundant sample sources, high affinity, stability, low immunogenicity, high solubility, tissue permeability, easy editing and functional modification, high yield, and high specificity.

1.3. Classification of Nanobodies

Nanoantibodies can be categorised on the basis of different sources and can be divided into the following categories:

Naturally sourced nanoantibodies: these antibodies originate from the immune system of ruminants such as camels and ruminants. They have relatively small molecular size, high stability and are capable of producing binding against hidden sites of antigens.

Artificial generation of nano-antibodies: Specific nano-antibodies can be designed and obtained by in vitro screening techniques such as spot hybridisation and yeast surface display. This method allows targeted screening against specific antigens and helps to obtain more specific antibody molecules.
Engineering nanobodies: Scientists can engineer nanobodies to improve their affinity, stability and specificity. This includes means such as artificially introducing mutations and synthesising peptide tags to optimise the performance of nano-antibodies.

DOI: 10.54254/2753-8818/29/20240615

1.4. The Importance of Nanobodies

As an innovative biomolecular tool, nanobodies have important medical, biological and biomedical applications.

Nanoprobes have precise molecular recognition and binding capabilities. Due to their small size and high affinity, nanoprobes can bind highly to specific molecular targets, such as disease markers, tumour tips, etc.

As drug carriers, nanoparticles can precisely guide drugs to target cells or tissues, thus improving drug efficacy and reducing adverse reactions.

Nanobodies play a key role in immunotherapy. Nanobodies can activate or inhibit the function of the immune system for the treatment of cancer, autoimmune diseases and other diseases.

1.5. Prospects for the application of nanobodies

1.5.1. Coronavirus treatment. Daniel Wrapp [11] et al. performed two rounds of screening by phage display after immunisation of camels with MERS-CoV S proteins, with MERS VHH-55 having the strongest affinity. all three CDRs of MERS VHH-55 were in extensive contact with the receptor binding domain (RBD), resulting in RBD burial based on CDR2 up to CDR3. All three CDRs of MERS VHH - 55 were in extensive contacted with the receptor binding domain (RBD), leading to RBD burial based on CDR2 up to CDR3, which occupied the binding site of the MERS-CoV receptor dipeptidyl peptidase-4, thereby inhibiting its binding to the RBD.

Daniel Wrapp [11] et al. constructed a nanoantibody library by immunising camels with acute respiratory syndrome coronavirus 1 (SARS - CoV 1) and selected a nanoantibody called SARS VHH - 72 using phage display technology. This nanobody did not occupy the receptor site, but acted by generating hydrogen bonds between CDR 2 and CDR 3 and the SARS-CoV-1 RBD.

DongJ [12] et al. used the acute respiratory syndrome coronavirus 2 (SARS-CoV 2) S1 protein to screen a human VHH library from alpacas to obtain a fusion nanobody of the VHH gene to the structural domain of human IgG F to reduce immunogenicity. The generated VHH-FC antibody had good binding ability to the SARS-CoV 2 surface trimeric spiny (S) protein and blocked the binding of S protein to its receptor angiotensin-converting enzyme 2 (ACE2).

1.5.2. Cancer treatment. The promise of nanobodies in cancer therapy is particularly compelling. Targeted inhibition of oncogenic signals using Nb, by blocking the dysregulation of vascular endothelial growth factor receptor receptor tyrosine kinase (RTK) signal transduction or targeted delivery of killing substances using Nb or targeted radionuclide therapy using Nb, by labelling mAb, antibody fragments or other small molecule carriers with appropriate radioisotopes. thereby delivering cytotoxic radiation to cancer cells and targeting cancer cells for killing. or using Nb to direct immune cells to target and kill cancer cells [13].

1.5.3. Biosensor. Nanobodies can be used to develop highly sensitive and specific biosensors for the detection of disease markers, pathogens and other molecular targets. Benchao Su [14] developed an A1-ALP-mediated MnO2-TMB sensing system for the detection of A1-ALP.

1.5.4. Humanisation of nanobodies. If nano-antibodies are used in large doses for a long time in the clinic may produce different degrees of immune reactions, so to make nano-antibodies less immunogenic, it is necessary to humanise nano-antibodies, which can be done by transplanting the CDRs of specific nano-antibodies onto humanised scaffolds, and can also be done by modifying the 12 amino acid residues in the VHH according to the amino acid differences.

Due to the small molecular weight and short half-life of nanoantibodies, they are easily excreted from the body, resulting in loss of efficacy, and the stability can be improved by fusion with human lgG1 FC.

DOI: 10.54254/2753-8818/29/20240615

1.6. Definition of molecular dynamics

Molecular Dynamics (MD) is a computational physics method used to simulate the motion of atoms and molecules in the presence of classical force fields. By calculating the positions and velocities of particles at certain time intervals, molecular dynamics is able to simulate the time evolution of molecular systems, thus revealing the microstructure and macroscopic properties of matter.

1.7. The importance of molecular dynamics

Molecular dynamics is of great significance in the fields of materials science, biochemistry, drug design and so on.

1.7.1. Materials science. Molecular dynamics can be used to study the mechanical, thermal and transport properties of materials, which helps to design new materials, optimise material properties and explain experimental phenomena.

1.7.2. Biochemistry. At the biomolecular level, molecular dynamics can simulate the structure and function of biomolecules such as proteins, nucleic acids and membranes.

1.7.3. Drug design. In drug discovery, molecular dynamics can be used to assess the binding between candidate drug molecules and target molecules, and predict the affinity and selectivity of drugs.

In this experiment, Amber software was used to simulate the molecular dynamics of VHH, and the RMSD and RMSF curves were obtained by changing the time and temperature.

1.8. Innovativeness of the subject

Holistically investigated the PDB structures as well as commonalities and differences in amino acid sequences of more than 1,300 known nanoantibodies in the database.

Used deep learning techniques such as Alpha Fold2 to perform four-level structural modelling of proteins, and analysed the interactions between the various levels of protein structure and chemical bonds.

Structural analysis of VHH using molecular dynamics simulation.

1.9. Research purpose

Investigating commonalities and differences in nanobodies

Predicting the conformation and hierarchical structure of VHH proteins using molecular dynamics simulations and protein modelling

Provide a theoretical basis for protein prediction of nanobodies

1.10. Content of research

Extract the sequences of all nanobodies in the protein database for statistical analysis

Select the nanoantibodies of coronaviruses as the research objects for protein modelling.

Perform molecular dynamics simulation of the protein model using Amber software for structural analysis.

Analyse RMSD and RMSF.

Analyse the primary, secondary and tertiary structure of the modelled structure.

1.11. Significance of the experiment

This experiment provides an experimental basis for better analysis of the structure of nanoantibodies in the future, and proves that molecular dynamics simulation and protein modelling can be used to predict and analyse the protein structure of nanoantibodies.

2. Experimental steps

Search the antibody structure database SabDab-Nano to download the list of nanobodies and all antibody structures.

DOI: 10.54254/2753-8818/29/20240615

Use python to write a script to extract all the nanobody sequences from it.

Select the coronavirus nanobody VHH as the structure study object in the protein pdb library rcsb.org, and extract the pdb structure (A chain, PDB: 6waq).

Modelling using CharmmGuI website.

(1) Add hydrogen atoms to refine the structure 127 amino acids, disulfide bonds: C22-C92.

(2) Build simulated water box Nearest atom 15 Å from box boundary, size 85.2 Å*86.0 Å*86.2 Å.

(3) Add 0.15 mol neutralising ion (NaCl).

(4) Set up PBC conditions.

Total number of atoms: 57476

Molecular dynamics simulation using Amber.

(1) Optimise the energy regime.

(2) Optimise the heating regime.

(3) Run.

The protein structure obtained after the first step of this modelling is the original protein structure for this experiment, with the amino acid sequence of

QVQLQESGGGLVQAGGSLRLSCAASGRTFSEYAMGWFRQAPGKEREFVATISWSGGSTYY TDSVKGRFTISRDNAKNTVYLQMNSLKPDDTAVYYCAAAGLGTVVSEWDYDYDYWGQGTQ VTVSSGS

3. Results

3.1. Nanobody pdb structural commonalities and differences

After getting all the nanobody pdb sequences from SabDab-Nano by python, the frequency of amino acid sequences in them was counted, and it was found that alanine, leucine and serine appeared most frequently. Different nanobodies have different amino acid types and numbers due to different functions.

3.2. Original protein structure

The original protein structure was obtained from the first step of modelling the A-chain with pdb sequence number 6waq using the CharmmGuI website

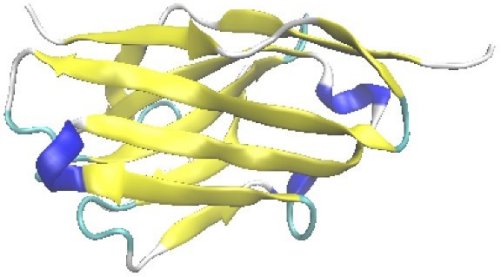

Figure 3. Protein conformation observed using VMD software

3.3. Protein conformational plausibility

The conformations of the proteins were analysed using the website saves.mbi.ucla.edu to analyse whether the protein conformation was reasonable or not using the score, pass or fail in Verify 3D and the pull-down plots therein.

The analysis can conclude that the protein structure constructed in this experiment scored 89 and passed Verify 3D.

Ramachandran is a tool for visualising the dihedral angle distribution of protein backbone residues, which is mainly used to assess the reasonableness of protein conformation. Ramachandran is divided into 3 main zones: permissive zone (dark red), maximum permissive zone (yellow zone) and impermissive zone (blank zone).

DOI: 10.54254/2753-8818/29/20240615

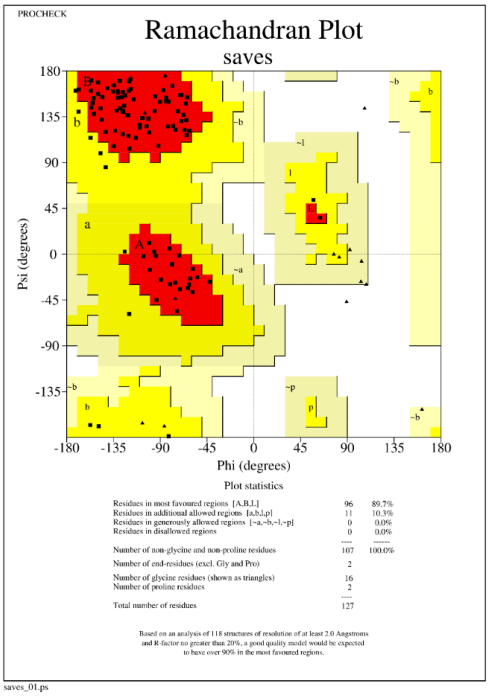

Figure 4. Ramachandran of protein conformations

It can be seen from Figure 4 that the vast majority of the black dots are distributed within the range of the permissive region, indicating that the constructed protein conformation is reasonable and can be used as the next step of the study.

3.4. Analysis of protein conformation after molecular dynamics simulations

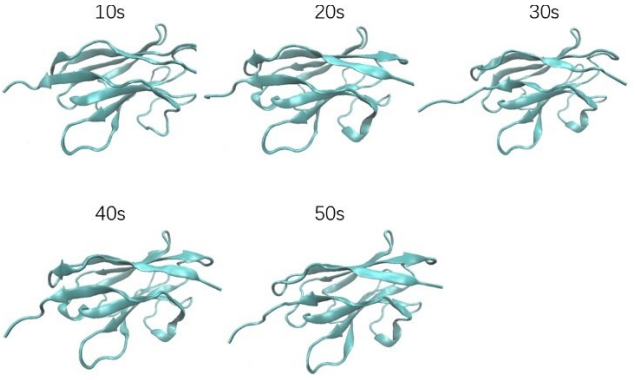

Figure 5. Protein conformation at 20 degrees Celsius for 10s, 20s, 30s, 40s, 50s starting from the top left to the bottom right using the VMD software

Molecular dynamics simulations of proteins were carried out using Amber, and the protein conformations at 10s, 20s, 30s, 40s, and 50s were listed separately using the VMD visualisation software, which shows that the molecular conformations are different at different times in a particular region.

3.4.1. RMSD analysis. The RMSD value is the Root Mean Square Deviation (RMSD). In statistics, this quantity is equivalent to the standard deviation, which reflects the extent to which the data deviates from

DOI: 10.54254/2753-8818/29/20240615

the mean. RMSD is a very common parameter used in protein structure analysis, modelling, structure alignment and molecular dynamics simulations to measure the extent to which atoms deviate from their relative positions.

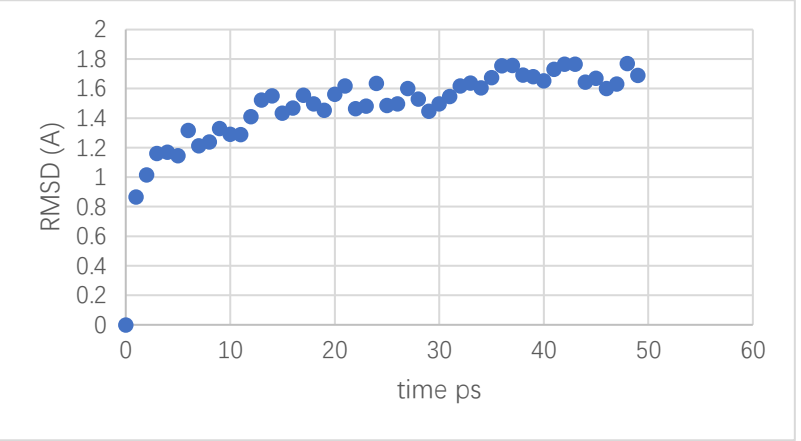

Figure 6. RMSD curves over time at 20°C

From Fig. 6, it can be seen that the RMSD curve first grows and then stabilises with increasing time, indicating that the protein structure first changes and then tends to stabilise.

3.4.2. RMSF analysis. RMSF (Root Mean Square Fluctuation) is a metric used to analyse molecular dynamics simulations or the structural flexibility of protein structures. It indicates the degree of fluctuation that keeps the positions of atoms in a molecule in a period of time. RMSF is used to study the stability and flexibility of different parts of the interior of a molecule, to understand and how the molecule responds to changes in the external environment or ligand binding.

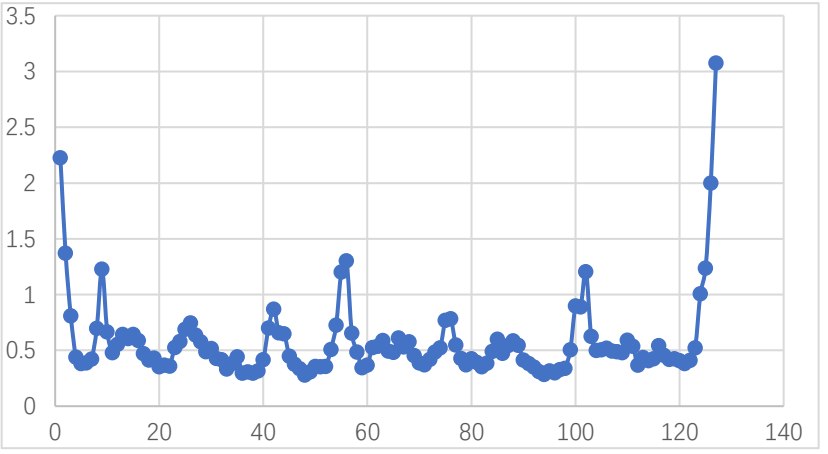

Figure 7. RMSF curves with temperature

From Fig. 7, it can be seen that the RMSF curve is constantly fluctuating, and there are drastic changes in the fluctuations in 0°C and 120°C, while there are large fluctuations in the curves in 10°C, 30°C, 40°C, 50°C, 75°C, and 100°C. The larger degree of fluctuation indicates that the effect on the proteins is larger, which means that the molecular composition is unstable and easy to change in these temperature regions.

DOI: 10.54254/2753-8818/29/20240615

3.5. Protein structure prediction and analysis

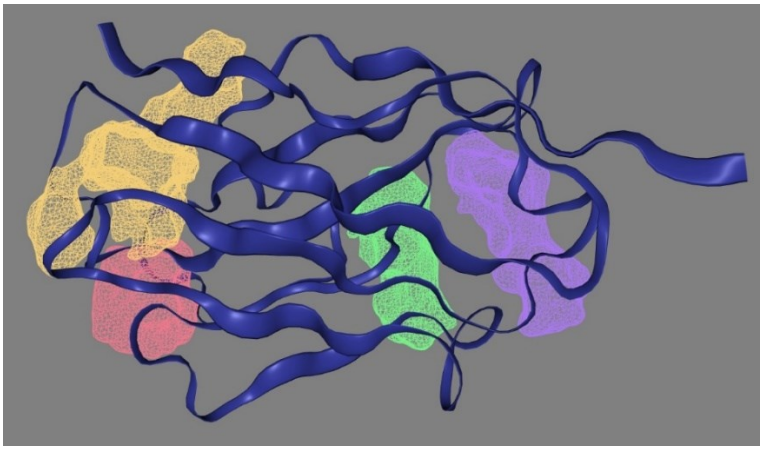

Figure 8. Protein activity pocket prediction

The activity pockets were predicted for this protein by the DoGSiteScorer method on the proteins. plus website, and it can be seen from Fig. 8 that the activity pockets within the protein are likely to be in one of the four positions.

Analysis of the original protein conformation at 20°C and the protein conformation obtained after molecular dynamics simulations.

Hydrophobic clusters, hydrogen bonding, and salt bridges were analysed in the structure of the protein via the website proteintools.uni-bayreuth.de.

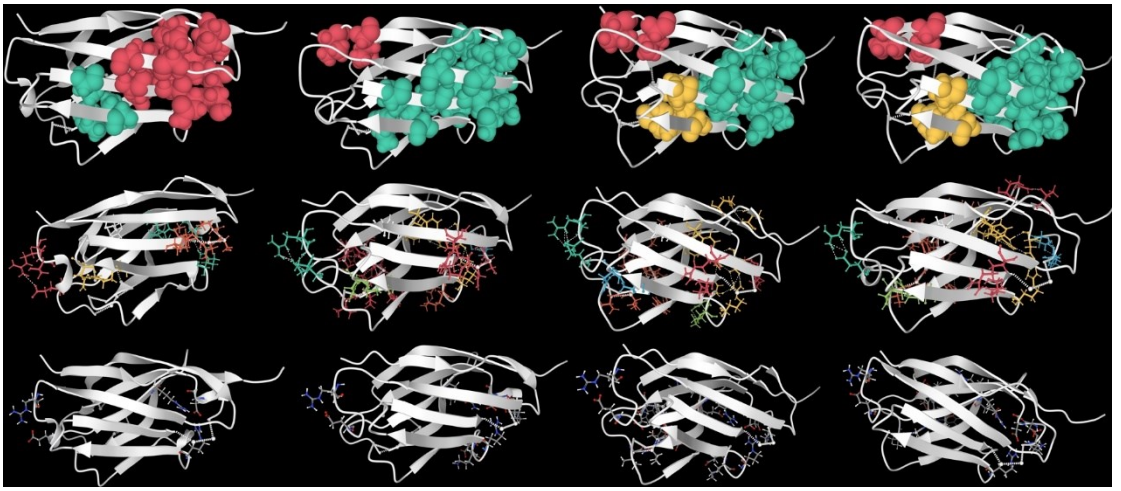

Figure 9. From left to right, the original, molecular dynamics simulated protein conformation at 20°C after 10s, 30s and 50s, respectively, with hydrophobic clustering, hydrogen bonding and salt bridging from top to bottom

It can be seen that after molecular dynamics simulation, the number of hydrophobic clusters, hydrogen bonds, and salt bridges in the active pocket region of the protein increases with time and then tends to be stable, indicating that the hydrophobic groups of the protein are exposed during the molecular dynamics simulation, and the protein binds to the ligand in the active pocket, forming more number of hydrogen bonds and salt bridges, and a more stable protein structure, which is roughly in agreement with the information contained in the RMSD curve. The locations where the changes in the number of hydrophobic clusters, hydrogen bonds, and salt bridges occur also expose the protein active pocket locations, which is consistent with the protein active pocket prediction.

DOI: 10.54254/2753-8818/29/20240615

Prediction of protein β-folding and α-helix using NetTurnP-1.0 and NetSurfP-3.0 methods respectively in services.healthtech.dtu.dk website reveals that sequential numbers 4 to 5, 7 to 17, 25 to 33, 39 to 45, 52 to 67, 72 to 78, 83 to 90, and 98 to 119 of the Amino acids present β-folding

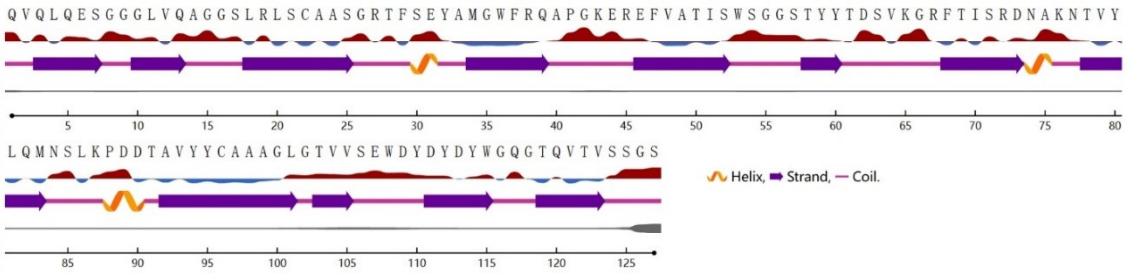

Figure 10. Alpha-helix prediction in protein conformation

Using the web.expasy.org website to analyse the components of the original protein structure, it was found that the molecular formula of the protein was C606H910N164O197S4, the total number of atoms was 1881, there were 12 positively charged residues and 10 negatively charged residues, and the instability index was 32.42, which indicated that the protein was a stable structure.

4. Conclusion

In this experiment, we analysed the pdb sequences of all the nanoantibodies to find out the commonalities and differences of the amino acid sequences, and then carried out protein modelling and molecular dynamics simulation of VHH to obtain the RMSD and RMSF curves, which led to the characteristics that VHH is very structurally unstable in 0°C and 120°C, and can easily change the molecular structure. The location of the active pocket of the original protein was also verified by comparing the number of hydrophobic clusters, hydrogen bonds and salt bridges between the original protein and the molecular dynamics simulated protein conformation. Finally, the secondary and primary structures of the original protein were analysed to have a detailed understanding of VHH.

Acknowledgments

I would like to thank Mr Ma Buyong for his teaching and Senior Ma for his help during the experiment.

References

[1] Hamers-Casterman C, Atarhouch T, Muyldermans S, et al. Naturally occurring antibodies devoid of light chains. Nature, 1993, 363: 446-8

[2] Muyldermans S, Atarhouch T, Saldanha J, et al. Sequence and structure of VH domain from naturally occurring camel heavy chain immunoglobulins lacking light chains. Protein Eng, 1994, 7: 1129-35

[3] Stanfield RL, Dooley H, Flajnik MF, et al. Crystal structure of a shark single-domain antibody V region in complex with lysozyme. Science, 2004, 305: 1770-3

[4] Muyldermans S, Atarhouch T, Saldanha J, et al. Sequence andstructure of VH domain from naturally occurring camel heavy chain immunoglobulins lacking light chains[J]. Protein Engineering, 1994, 7(9):1129-1135.

[5] Muyldermans S. Nanobodies: natural single - domain antibodies[J]. Annual Review of Biochemistry, 2013, 82(1) : 775 - 797

[6] Hassanzadeh — Ghassabeh G, Devoogdt N, De Pauw P, et al． Nano-bodies and their potential applications［J］. Nanomedicine, 2013, 8(6) : 1013 — 1026

[7] LI Jingying, LI Zhiwei, Xiao Shuqi. Nanobodies and disease control[J]. Advances in Animal Medicine, 2021, Vol.42 (8): 112-116

[8] Jovčevska, Ivana, Muyldermans, Serge. The Therapeutic Potential of Nanobodies. [J]. BioDrugs, 2020, Vol. 34(1): 11-26

DOI: 10.54254/2753-8818/29/20240615

[9] De Genst ESilence K, Decanniere K, et al. Molecular basis for the preferential cleft recognition by dromedary heavy-chain antibodies[J]. Proceedings of the National Academy of Sciences of the United States of America. 2006, 103(12): 4586-4591.

[10] Sun Baihe, Wu Yue, Zhao Rui, Lou Yuxin, Li Wanting, Li Yanfei, Ma Linlin. Progress of therapeutic nanobodies with different expression systems [J/OL]. Chinese Journal of Bioengineering.

[11] Wrapp D, De Vlieger D, Corbett K S, et al. Structural basis for po-tent neutralization of betacoronaviruses by single - domain camelid antibodies[J]. Cell, 2020, 181(5) : 1004 - 1015. e15.

[12] Jianbo Dong, Betty Huang, Zhejun Jia, Bo Wang, Sachith Gallolu Kankanamalage, Allison Titong, Yue Liu. Development of multi-specific humanized llama antibodies blocking SARS-CoV-2/ACE2 interaction with high affinity and avidity[J]. Emerging microbes & infections,2020,Vol.9(1): 1034-1036

[13] Zhou Huihui, CHEN Qu, Yang Xiaomei et al. Research progress of nanobodies in cancer therapy[J]. Life Science,2022,34(04):468-476.

[14] Su, Benchao. Research on the construction of an immunosensing system based on nanobodies and two-dimensional nanomaterials for the detection of human serum alpha-fetoprotein [D]. Hainan University,2022