MINISTRY OF EDUCATION AND SCIENCE OF UKRAINE
KYIV NATIONAL UNIVERSITY OF TECHNOLOGIES AND DESIGN
Faculty of Chemical and Biopharmaceutical Technologies
Department of Biotechnology, Leather and Fur

QUALIFICATION THESIS

on the topic **Effectiveness of biotechnological methods in the development of human antibodies to CD47**
First (Bachelor's) level of higher education
Specialty 162 "Biotechnology and Bioengineering"
Educational and professional program "Biotechnology"

Completed: student of group BEBT-20
Xinyue XU

Scientific supervisor
Tetiana SHCHERBATIUK, Dr. Sc., Prof.

Reviewer
Iryna VOLOSHYNA, Ph.D., As. prof.

Kyiv 2024

KYIV NATIONAL UNIVERSITY OF TECHNOLOGIES AND DESIGN

Faculty: Chemical and Biopharmaceutical Technologies
Department: Biotechnology, Leather and Fur
First (Bachelor's) level of higher education
Specialty: 162 Biotechnology and Bioengineering
Educational and professional program Biotechnology

APPROVE
Head of Department of Biotechnology,
Leather and Fur, Professor,
Doctor of Technical Science
Olena MOKROUSOVA

_____
«___»_____2024

ASSIGNMENTS
FOR THE QUALIFICATION THESIS
Xu Xinyue

1. Thesis topic **Effectiveness of biotechnological methods in the development of human antibodies to CD47**
scientific supervisor Tetiana Shcherbatiuk, Dr. Sc., Prof.
approved by the order of KNUTD "__"_____2024, №____

2. Initial data for work: assignments for qualification thesis, scientific literature on the topic of qualification thesis, materials of Pre-graduation practice

3. Content of the thesis (list of questions to be developed): literature review; object, purpose, and methods of the study; experimental part; conclusions

4. Date of issuance of the assignments_____

EXECUTION SCHEDULE

№	The name of the stages of the qualification thesis	Terms of performance of stage	Note on performance
1	Introduction	From 01 April 2024 to 11 April 2024	
2	Chapter 1. Literature review	From 06 April 2024 to 20 April 2024	
3	Chapter 2. Object, purpose, and methods of the study	From 21 April 2024 to 30 April 2024	
4	Chapter 3. Experimental part	From 01 May 2024 to 10 May 2024	
5	Conclusions	From 07 May 2024 to 12 May 2024	
6	Draw up a bachelor's thesis (final version)	From 20 May 2024 to 25 May 2024	
7	Submission of qualification work to the supervisor for feedback	From 25 May 2024 to 27 May 2024	
8	Submission of bachelor's thesis to the department for review (14 days before the defense)	27 May 2024	
9	Checking the bachelor's thesis for signs of plagiarism (10 days before the defense)	31 May 2024	
10	Submission of bachelor's thesis for approval by the head of the department (from 7 days before the defense)	03 June 2024	

I am familiar with the task:

Student _____Xinyue XU

Scientific supervisor _____Tetiana SHCHERBATIUK

SUMMARY

Effectiveness of biotechnological methods in the development of human antibodies to CD47 – Manuscript.

Qualification thesis on the specialty 162 «Biotechnology and Bioengineering». – Kyiv National University of Technologies and Design, Kyiv, 2024.

The traditional approach to antibody design requires a large number of experiments and time, which is inefficient. This may delay the therapeutic window, consume a lot of human resources and increase the economic cost of research and development. With the rapid development of artificial intelligence technology in recent years, it has been widely used in the biomedical field, which greatly improves the efficiency of drug development, especially in antibody design and function optimization shows great potential. Therefore, many researchers now choose to use artificial intelligence methods to design humanized CD47 antibodies and compare them with those obtained from experimental screening to measure the feasibility of AI methods for antibody design.

In this experiment, the target antibody was designed by remotely logging into the server using Finalshell, opening the DiffAb and AlphaPanda software environments using conda, and designing the target antibody to contain three heavy chains and three light chains of the antibody. The t-test was performed to compare the Root Mean Square Deviation, Sequence Identity and ddG of the DiffAb and AlphaPanda designs, and Pymol graphs were used to show a more intuitive design result. In this way, the feasibility of the AI method for antibody design was evaluated, the performance of the design software was tested, and its strengths and weaknesses were analyzed.

By analyzing the data, it can be seen that the performance of DiffAb in designing human CD47 antibody is better than AlphaPanda; in the design of L_CDR2, the antibody designed by AlphaPanda is better than DiffAb in thermodynamic stability.

It was found that the AI method is able to rapidly screen a large amount of data in terms of efficiency, which greatly improves the speed of antibody discovery. At the same time, AI can successfully design human CD47 antibodies that can achieve the atomic precision and high sequence consistency of natural antibodies in terms of structure and sequence. Although AI can generate a large number of candidate antibodies in the design phase, the shortcoming of AI-designed antibodies is the lack of experimental validation, as the current experimental validation process is relatively difficult. Experimental validation of the biological properties and functions of these candidate antibodies is needed to ensure their feasibility and safety in clinical applications. However, the experimental validation process may be constrained by a variety of factors, such as the complexity of experimental conditions, the difficulty in obtaining experimental materials, and the high cost of experiments. Therefore, continuous technological innovation and interdisciplinary collaboration will promote the wide application of AI in antibody design in the future.

Key words：*human CD47 antibody; artificial intelligence; antibody design*

TABLE OF CONTENTS

INTRODUCTION 7

CHAPTER 1 8

LITERATURE REVIEW 8

1.1 Background and significance of the study 8

1.2 Research content 9

1.3 Biological significance of human CD47 antibodies 9

1.4 Current status of research 10

1.5 The significance of design through AI 11

Conclusions to chapter 1 12

CHAPTER 2 13

OBJECT, PURPOSE, AND METHODS OF THE STUDY 13

2.1 Models used in antibody design 13

2.1.1 Transformer 13

2.1.2 3DCNN 13

2.1.3 Diffusion 14

2.2 Research object 14

2.3 Research purpose 15

2.4 Operation steps 15

Conclusions to chapter 2 17

CHAPTER 3 18

EXPERIMENTAL PART 18

3.1 Available data 18

3.2 H_CDR1 19
3.3 H_CDR2 21
3.4 H_CDR3 22
3.5 L_CDR1 24
3.6 L_CDR2 26
3.7 L_CDR3 27
3.8 Data Analysis 29
3.9 Performance Evaluation of Artificial Intelligence Methods 30
3.9.1 Efficiencies 30
3.9.2 Accuracy 31
3.9.3 Ingenuity 31
3.10 Application prospects and challenges 31
Conclusions to chapter 3 32
CONCLUSIONS 34
LIST OF REFERENCES 36

INTRODUCTION

The primary aim of this research is to evaluate the efficacy of artificial intelligence (AI) methods in designing human-origin CD47 antibodies. CD47 is a highly glycosylated membrane protein implicated in various cellular processes, including immune response and angiogenesis. It is a significant therapeutic target for cancers and other diseases due to its overexpression in malignancies and its role in preventing macrophage-mediated phagocytosis. The goal of this study is to benchmark AI-designed CD47 antibodies against those derived from traditional experimental methods to assess their specificity, affinity, and clinical potential.

The key theme of this research revolves around comparing AI's efficiency and accuracy in antibody design with conventional techniques. By leveraging advanced machine learning algorithms and bioinformatics tools, this study aims to simulate and optimize the immune process to rapidly identify antibodies with desirable properties.

Methodologically, the research will employ a combination of computational analysis using machine learning models, data mining of known antibody structures, and comparative assessments of antibody sequences, structures, and functions. Software like Diffab and AlphaPanda will be used for structure prediction and validation.

The novelty of this study lies in its comprehensive assessment of how AI algorithms can outperform traditional methods by identifying high-affinity CD47 antibodies with unprecedented speed and specificity. Moreover, it seeks to address challenges such as data quality, interpretability, and clinical applicability, offering valuable insights into the current and future landscape of AI-driven antibody design.

This research is significant as it can lead to improved therapeutic strategies, reducing the time and cost required to discover effective antibodies. By exploring AI's transformative potential, it contributes to personalized medicine and paves the

way for developing next-generation therapeutics that target CD47 and other immune checkpoint inhibitors.

CHAPTER 1
LITERATURE REVIEW

1.1 Background and significance of the study

Malignant tumors are one of the highly lethal diseases in human beings, and their safe and effective treatment is the focus of extensive attention. The treatment of malignant tumors is either alone or in combination, including surgery, radiotherapy, chemotherapy and immunotherapy [1]. Tumor immunotherapy can effectively activate immune cells in the body to produce an immune response to tumors and exert anti-tumor effects[2]. Many immunotherapies are currently under clinical investigation, including targeted antibody therapy, tumor vaccines and immunomodulators [3].

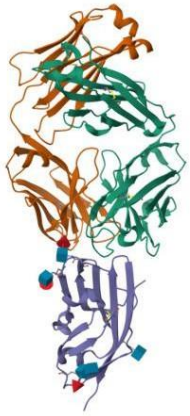

Figure 1.1 – CD47 Structure Diagram

However, the traditional method of designing humanized Cluster of differentiation 47 (hereinafter referred to as CD47) antibodies requires a lot of experiments and time, and is inefficient, so many researchers now choose to use artificial intelligence (AI) methods for designing humanized CD47 antibodies, and compare them with those obtained by experimental screening to measure the AI method's The feasibility of the AI method for designing antibodies is tested by comparing it with the antibodies obtained from experimental screening. AI algorithms are used to mimic and surpass the natural immune process of the human

body, thus accelerating the design of antibody molecules with precise targeting and high exploitability.

1.2 Research content

To check the experimental research and technical methods on the preparation of human CD47 antibody at home and abroad, and to compare the antibody prepared by traditional experiments with the antibody designed by artificial intelligence method, and to analyze the feasibility of designing antibodies by artificial intelligence method by comparing the sequences, structures, and performances of the antibodies with the two software, Diffab and AlphaPanda. It is also necessary to test the performance of the AI software AlphaPanda, analyze its strengths and weaknesses and discuss the current challenges faced in this field, such as data quality, algorithmic interpretability, the difficulty of clinical translation, etc., and look forward to the future direction of AI in the field of biomedicine[1].

1.3 Biological significance of human CD47 antibodies

CD47, also known as integrin-associated protein, is a highly glycosylated membrane-penetrating protein that is widely distributed on the cell surface, including an amino-terminal extracellular variable region, a membrane-penetrating region consisting of three to five highly hydrophobic membrane-penetrating fragments, and a hydrophilic carboxy-terminal cytoplasmic tail region[4]. And it belongs to the immunoglobulin superfamily, which can interact with integrins, platelet reactivators, and signal-regulatory protein alpha (SIRP α)[5].CD47 has the potential to be a therapeutic target for a number of cancers as a signaling molecule that can prevent phagocytosis by macrophages, and recent studies have shown that it may also be associated with pulmonary fibrosis[6].

CD47 is involved in a range of cellular activities and plays an important role in immunity and angiogenesis, including apoptosis, proliferation, adhesion and migration.CD47 is expressed in a large number of cells and is overexpressed in a wide range of cancer cells. Therefore, the development of anti-CD47 antibody has

significant clinical significance and vast application space for tumor-targeted therapy[7].

1.4 Current status of research

Artificial Intelligence can generate text content, images and videos, its ability to generate new proteins is another important direction of the development of AI technology. the development of AI protein generation technology will become a key common technology to lead the development of biomedical innovation and may bring about a change in synthetic biology, change in medicine, change in bioengineering, change in the R & D paradigm. AI subversion of the design of proteins to solve the problems that can not be solved by traditional methods, activate the potential for industrial applications of protein technology. AI protein design in the era of biotechnology will promote the development of biomedicine, materials, agriculture, food, environmental protection and other areas of change.

In recent years, the development of advanced algorithms, the accumulation of big data and the computational power of computer hardware have been improving, and AI technology has been actively developed and applied to the field of protein design[8]. In recent years, artificial intelligence has made rapid progress in protein structure design, and Zhihang Chen, Menglin Ji, and Yifei Qi reviewed the latest protein structure design algorithms in three directions, namely, fixed framework design, variable backbone design, and sequence structure generation, and elucidated the novelty and innovation compared with the traditional computational methods[9]. With the availability of artificial intelligence technology, the success rate and rationality of protein design have been significantly improved, and the era of functional protein design has come to an end.

Internationally, artificial intelligence techniques, particularly machine learning methods, have made significant progress in antibody design and discovery[10]. These methods are used to computationally predict the interactions between antibody structures and antigen interfaces, as well as to assess the exploitability of antibodies. For example, RFdiffusion[11], a comprehensive improvement over current protein

design methods, enables the design of proteins with a total length of up to 600 amino acid residues from scratch and achieves unprecedented complexity and accuracy. More importantly, RFdiffusion can design proteins that bind to target proteins. in July 2023 protein design pioneer David Baker and his team published a research paper in the journal Nature entitled: De novo design of protein structure and function with RFdiffusion in a research paper[12]. They developed and described RFdiffusion, a deep learning method for designing entirely new proteins from scratch, which generates a wide variety of functional proteins, including topologies never before seen in natural proteins[13].

1.5 The significance of design through AI

The use of AI is significant in that it can significantly improve the efficiency and accuracy of antibody design. Using advanced algorithms to simulate and analyze complex biological data, these software programs are able to rapidly identify antibody candidates with high affinity and specificity. This process dramatically reduces the time required for traditional antibody development, thereby accelerating the development of new drugs. In addition, these tools support fine-tuning of antibody structures to optimize their biological activity and stability, which is critical to improving the success of antibody drugs in clinical applications[14].

The use of software such as DiffAb and AlphaPanda also helps to reduce development costs. By reducing the number of experiments and identifying potential problems in advance in the computer simulation stage, the waste of resources can be effectively reduced and the economic efficiency of R&D can be improved. In the current rapidly developing biomedical field, this efficient and low-cost R&D model can accelerate the speed of new drugs to market and better respond to rapidly changing medical needs.

Conclusions to chapter 1

1. The literature demonstrates that AI has made significant strides in the accurate prediction and design of protein structures, with advanced machine learning models like RFdiffusion achieving unprecedented levels of complexity and accuracy. By comparing the sequences, structures, and performances of antibodies designed using both traditional methods and AI-based tools, researchers can assess the feasibility of these cutting-edge technologies in biomedical applications[1]. The improvements in computational power and the availability of large datasets have enabled AI algorithms to mimic natural immune processes, significantly improving the efficiency of antibody design.
2. Artificial intelligence has emerged as a pivotal tool in the design of human-origin CD47 antibodies due to its ability to accelerate and enhance the discovery process[3]. Traditional methods of antibody development are labor-intensive and time-consuming, necessitating numerous experiments to identify suitable candidates. In contrast, AI algorithms can swiftly analyze vast datasets to identify antibodies with high specificity and therapeutic potential.
3. The significance of using AI is that it can significantly improve the efficiency and accuracy of antibody design, while also reducing design costs.

CHAPTER 2
OBJECT, PURPOSE, AND METHODS OF THE STUDY

2.1 Models used in antibody design

2.1.1 Transformer

The Transformer model is a deep learning model widely used in the field of natural language processing, originally proposed by Vaswani et al. in 2017. The core of their model is Self-Attention Mechanism, which automatically learns to assign different attention weights based on different parts of the input. This means that the model is able to synchronize the integration of information from all locations when responding to sequential data, unlike the step-by-step processing of Recurrent Neural Networks and Convolutional Neural Networks. This parallel processing feature empowers Transformer to show higher processing efficiency when facing long sequence data. When constructing antibodies, the Transformer model can be applied to explore the connections between protein sequences, especially when identifying and predicting antibody-antigen binding sites. With the input of a sequence of amino acids, the Transformer model learns the similarities and correlations between sequences, and then deduces the antibody sequences that are most likely to bind to the antigen.

2.1.2 3DCNN

3DCNN[15] Protein Design Software is a computational tool developed by Anand Namrata that utilizes 3D convolutional neural networks to analyze protein structures and perform protein design. The tool creates new ways of analyzing protein structures, using deep learning to gain insight into the complex world of proteins[16]. In designing antibodies, the 3DCNN model can be used to process protein structural data, with a particular focus on antigen-antibody binding site prediction and the resolution of complex spatial structures of antibodies. Protein structures are usually

represented in the form of three-dimensional coordinates, in which each amino acid residue has its coordinates in three-dimensional space. 3DCNN can effectively utilize these three-dimensional coordinate data to learn the three-dimensional structural features of proteins and perform antibody design tasks based on them. Moreover, by recognizing and predicting the sites in the protein structure where the antibody binds to the antigen, 3DCNN can help design antibodies with higher affinity and specificity, and accordingly predict the amino acid residues that are most likely to be involved in antigen binding.

2.1.3 Diffusion

Diffusion is a diffusion-based protein structure generation model that can generate new protein structures by learning the distribution characteristics of known protein structures. These newly generated structures show potential for a wide range of applications in antibody design and protein structure improvement. In addition, the Diffusion model can also be used to design amino acid sequences for antibodies, and by learning a large amount of antibody sequence data, the model can generate new antibody sequences and optimize existing antibody sequences.The Diffusion model is often used in conjunction with other models to improve the accuracy and efficiency of antibody design. For example, Diffusion can be combined with deep learning models (such as recurrent neural networks or convolutional neural networks) to process different types of protein structural data to extract a more comprehensive characterization, thus deepening the scientific and practicality of the design.

2.2 Research object

The object of this experiment are natural human CD47 antibodies and antibodies designed by AlphaPanda and DiffAb.

AlphaPanda[17] is an algorithm that integrates a transformer model, a 3DCNN model, and a diffusion generation model. It uses the transformer model to capture global information, the 3DCNN model to capture local structural features of

antibody-antigen complexes, and then the diffusion generation model to generate the sequence and structure of the antibody.The 3DCNN model has the ability to capture both pairwise and non-pairwise interactions and requires fewer data for training samples while avoiding some of the pitfalls of the autoregressive and autocorrelative iterative model generation processes.

DiffAb is a cutting-edge tool for antigen-specific antibody design and optimization using diffusion-based protein structure generation models. It was presented at NeurIPS 2022 and is one of the first deep learning-based methods to generate antibodies against specific antigen structures.The DiffAb model operates by co-modeling the sequence and structure of antibody complementary decision regions (CDRs). These CDRs are critical because they determine the binding affinity of antibodies to antigens such as viruses and bacteria. By utilizing diffusion probability models and isovariant neural networks, DiffAb can perform several tasks: sequence-structure co-design, sequence design, structure prediction and antibody optimization. The model has now been extensively evaluated and shows promising results in biophysical energy functions and other protein design metrics.

2.3 Research purpose

Design human source CD47 using AlphaPanda and DiffAb software respectively, and perform T-test on the obtained data. Analyze the data and evaluate the performance of the software to verify whether the antibodies designed by the software achieve atomic accuracy. Which software designs antibodies with higher sequence similarity and more stable energy. And compare the antibodies designed by artificial intelligence with natural antibodies to analyze the feasibility of designing antibodies using artificial intelligence.

2.4 Operation steps

1. First identify the antibody under study and search the protein database (RCSB PDB) to filter out the most compatible one. Here the antibody I studied has the ID 7xjf in the database, the three-dimensional structure is shown in Figure 2-1.

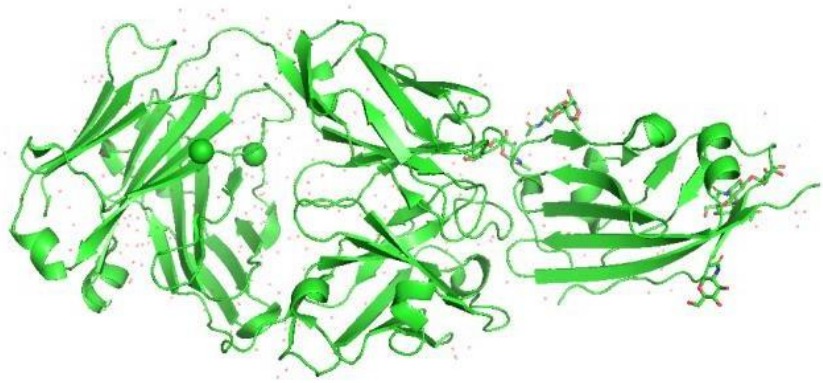

Figure 2.1 – Human Source CD47 3D Map

2. Use Finalshell to log in to the server remotely.

3. Use conda to open the DiffAb software environment, enter the command: conda activate diffab

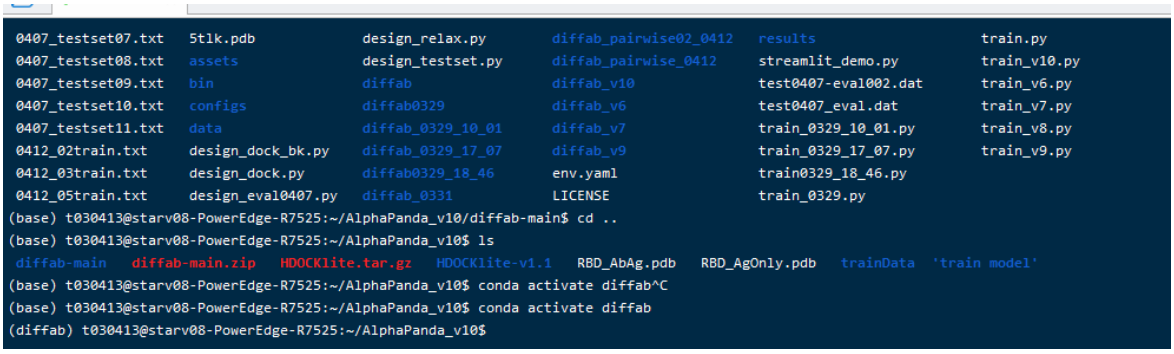

Figure 2.2 – Finalshell operation interface

4. Commands for entering design antibodies:python design_pdb.py /home/data/t030413/diffab-origin/diffab-main/7xjf/7xjf.pdb --heavy A --light B --config/home/data/t030413/AlphaPanda_v3_Vcnn/diffab-main/configs/YueTest/codesign_single_yueTest119.yml -d cpu

5. Enter the command to analyze the antibody:python design_eval_single.py --root=/home/data/t030413/diffab-origin/diffab-main/results/codesign_single06/7xjf.pdb_2023_10_08___10_37_06 --pfx="

6. The antibodies designed by DiffAb and AlphaPanda were compared in terms of Root Mean Square Deviation, Sequence Identity, and Energy Difference, and the data obtained were t-tested and analyzed in detail and the performance of the design procedure was analyzed using SPSS.

Conclusions to chapter 2

1. The models involved in antibody design and the specific software used in this experiment, AlphaPanda[17] and DiffAb, are described.
2. Specific steps of the experimental manipulation are described in detail, six CDRs for the CD47 antibody are designed, and the data are analyzed.

CHAPTER 3
EXPERIMENTAL PART

3.1 Available data

The analysis leads to the important data of RMSD (Root Mean Square Deviation), seqid (sequence identity) and ddG (energy difference).

Among them, RMSD is a metric commonly used to quantify the difference between the structures of two molecules, especially in biomolecular modeling and structural biology, which measures the average deviation of the positions between two sets of corresponding atoms.The smaller the value of RMSD, the higher the similarity between the two structures. It is commonly used to compare the three-dimensional structures of proteins or other macromolecules, for example, in simulated folding or structure prediction studies.RMSD is a very intuitive metric for assessing the accuracy of a model's predictions or the structural differences before and after a change. In drug design and protein engineering, by comparing the RMSD of different models or mutants, scientists can assess the effect of alterations on the structure and thus understand functional changes.

The seqid value usually represents Sequence Identity. Sequence similarity is the degree of similarity between two protein sequences. It measures the proportion of amino acid residues that are identical in two sequences. Higher sequence similarity means that two sequences are likely to be more similar in structure and function because they share more amino acid residues. In antibody design, the seqid value may be used to assess the degree of similarity of a designed antibody sequence to a target sequence, such as an antigen associated with a specific disease. Also a higher seqid value may indicate that the designed antibody sequence has a higher similarity to the target sequence, which may help to improve the affinity and efficacy of the antibody. Therefore, considering sequence similarity is an important factor when designing antibodies in AI.

The ddG represents the difference in free energy of antibody and antigen binding, i.e., the amount by which the free energy of the system changes after the formation of an antibody-antigen complex relative to the free energy of the antibody and antigen alone. Typically, a lower value of ddG indicates a more stable and strong binding between the antibody and antigen, as this means that the free energy of the system is reduced to a greater extent after the formation of the complex. Therefore, in antibody design, it is one of the goals to design antibodies with lower ddG values because these antibodies are more likely to have higher affinity and specificity and thus recognize and bind the target antigen more efficiently.

3.2 H_CDR1

In terms of rmsd value, the mean value of the antibody designed by DiffAb software (0.536) is lower than that of AlphaPanda (0.933). t-test results show that the t-value between both DiffAb and AlphaPanda is -8.221 with a p-value of 0. This suggests that there is a significant difference in terms of rmsd between the two software-designed antibodies and the The mean value of DiffAb is significantly lower than that of AlphaPanda, and a p-value of 0 implies that this difference is extremely significant.The structural similarity of the antibodies designed by DiffAb is higher than that of AlphaPanda.

In terms of seqid metrics, the mean value of DiffAb software (70.714) is higher than the mean value of AlphaPanda (55.714). Meanwhile, the results of t-test showed that the t-value was 2.748 and the p-value was 0.022, which was lower than the 0.05 level, indicating that there was a significant difference between DiffAb and AlphaPanda in terms of seqid, and that the antibody designed by DiffAb was closer to the reference standard in terms of sequence similarity.

In terms of ddG index, the mean value of DiffAb (128.337) was significantly lower than that of AlphaPanda (1123.686). Also the t-test showed a t-value of -4.706 and a p-value of 0.001, which is much lower than the 0.05 level. This indicates that there is a statistically significant difference between DiffAb and AlphaPanda in terms

of ddG, and the antibody designed by DiffAb is superior to AlphaPanda in terms of thermodynamic stability.

Table 3.1 – **H_CDR1**

	software	Number of cases	average value	standard deviation	t	p
rmsd	DiffAb	100	0.536	0.139641505	-8.221	0
	AlphaPanda	10	0.933	0.200611801		
seqid	DiffAb	100	70.714	7.425851504	2.748	0.022
	AlphaPanda	10	55.714	17.10312842		
ddG	DiffAb	100	128.337	263.8601799	-4.706	0.001
	AlphaPanda	10	1123.686	663.5917497		

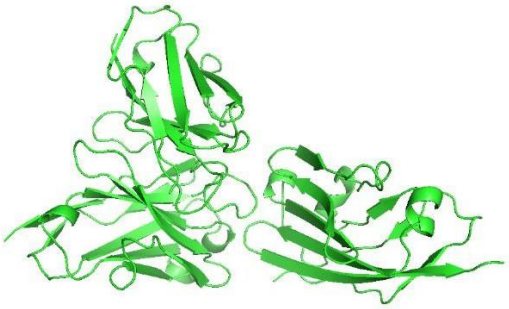

Figure 3.1 – Human CD47 antibody reference

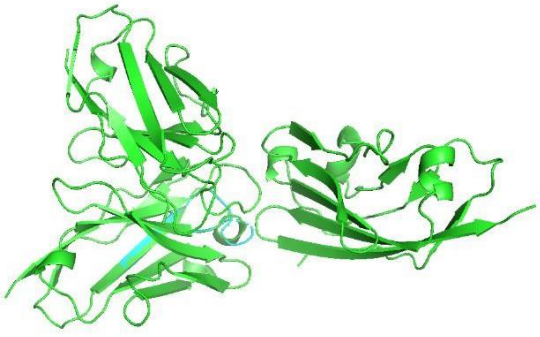

Figure 3.2 – H_CDR1 designed by AlphaPanda

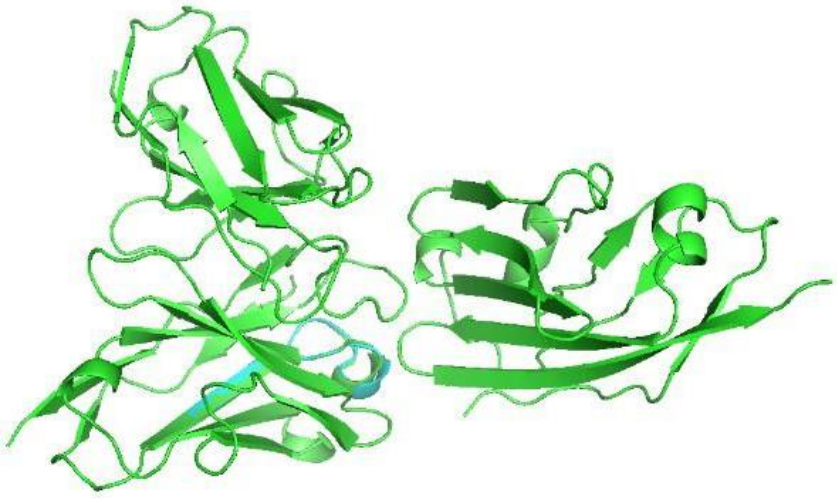

Figure 3.3 – H_CDR1 designed by DiffAb

3.3 H_CDR2

In terms of rmsd values, the mean value of the DiffAb software-designed antibody (0.648) was lower than the mean value of AlphaPanda (0.768). t-test results showed a t-value of -1.756 and a p-value of 0.082 between DiffAb and AlphaPanda, indicating that there is no significant difference in terms of rmsd between the two software-designed antibodies. significant differences exist between the two software-designed antibodies.

In terms of seqid index, the mean value of DiffAb software (26.740) is higher than that of AlphaPanda (14.651). Meanwhile, the results of t-test showed that the t-value was 4.054 and the p-value was 0.001, which was lower than the 0.05 level, indicating that there was a significant difference between DiffAb and AlphaPanda in terms of seqid, and that the antibody designed by DiffAb was closer to the reference standard in terms of sequence similarity.

In terms of ddG index, the mean value of DiffAb (211.201) was lower than that of AlphaPanda (383.745). Also the t-test showed that the t-value was -0.769 and the p-value was 0.461, which is above the 0.05 level. This indicates that there is no significant difference between DiffAb and AlphaPanda in terms of ddG.

Table 3.2 – **H_CDR2**

	software	Number of cases	average value	standard deviation	t	p
rmsd	DiffAb	100	0.648	0.209954816	-1.756	0.082
	AlphaPanda	10	0.768	0.149720565		
seqid	DiffAb	100	26.740	13.89353959	4.054	0.001
	AlphaPanda	10	14.651	8.345057645		
ddG	DiffAb	100	211.201	306.2642778	-0.769	0.461
	AlphaPanda	10	383.745	703.1522346		

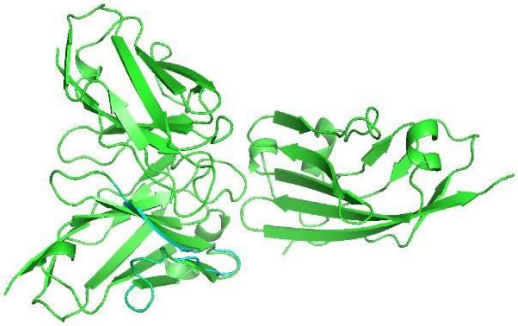

Figure 3.4 – H_CDR2 designed by AlphaPanda

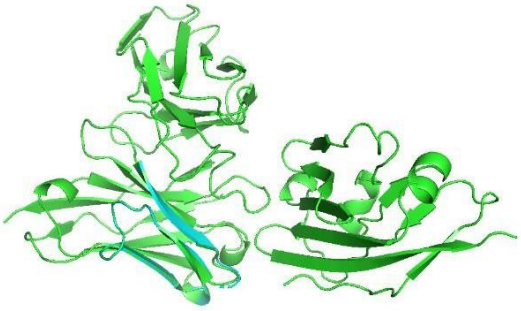

Figure 3.5 – H_CDR2 designed by DiffAb

3.4 H_CDR3

In terms of rmsd value, the mean value of DiffAb software-designed antibody (4.868) is lower than the mean value of AlphaPanda (6.060). t-test results show that the t-value between the two, DiffAb and AlphaPanda, is -3.192, with a p-value of 0.002, which is lower than 0.05. This indicates that in terms of rmsd, there is a

significant difference between the two software-designed antibodies are significantly different from each other; moreover, the mean value of DiffAb is lower than that of AlphaPanda, indicating that the structural similarity of DiffAb-designed antibodies is higher than that of AlphaPanda.

In terms of seqid index, the mean value of DiffAb software (19.195) was higher than that of AlphaPanda (16.930). Meanwhile the result of t-test shows that the t-value is 0.699 and p-value is 0.486, which is higher than 0.05 level, indicating that there is no significant difference between DiffAb and AlphaPanda in terms of seqid.

In terms of ddG index, the mean value of DiffAb (2926.268) is lower than the mean value of AlphaPanda (8059.134). Also the t-test showed that the t-value was -2.979 and the p-value was 0.015, which is below the 0.05 level. This indicates that there is a significant difference between DiffAb and AlphaPanda in terms of ddG; and the mean value of DiffAb is much lower than that of AlphaPanda, suggesting that the antibody designed by DiffAb is superior to AlphaPanda in terms of thermodynamic stability.

Table 3.3 – **H_CDR3**

	software	Number of cases	average value	standard deviation	t	p
rmsd	DiffAb	100	4.868	1.124740754	-3.192	0.002
	AlphaPanda	10	6.060	1.134088043		
seqid	DiffAb	100	19.195	9.712495336	0.699	0.486
	AlphaPanda	10	16.930	10.34727597		
ddG	DiffAb	100	2926.268	2771.653295	-2.979	0.015
	AlphaPanda	10	8059.134	5377.890469		

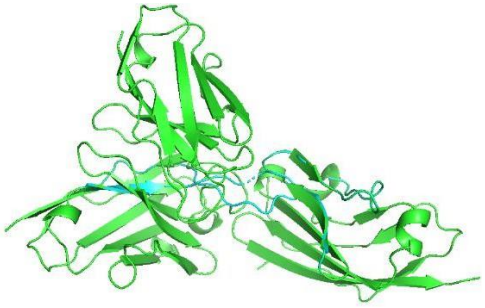

Figure 3.6 – H_CDR3 designed by AlphaPanda

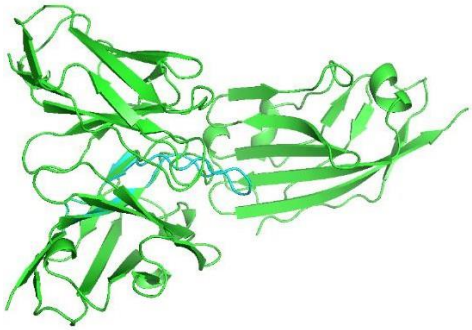

Figure 3.7 – H_CDR3 designed by DiffAb

3.5 L_CDR1

In terms of rmsd value, the mean value of the antibody designed by DiffAb software (0.596) is lower than the mean value of AlphaPanda (1.011). t-test results show that the t-value between DiffAb and AlphaPanda is -10.127 and the p-value is 0, which is lower than 0.05. This suggests that there is a significant difference in rmsd between the antibodies designed by both software; moreover, the mean value of DiffAb is lower than AlphaPanda, meaning that this difference is extremely significant. This indicates that in terms of rmsd, there is a significant difference between the two software-designed antibodies; moreover, the mean value of DiffAb is lower than that of AlphaPanda, with a p-value of 0 implying that this difference is extremely significant.The structural similarity of the DiffAb-designed antibodies is higher than that of AlphaPanda.

In terms of seqid metrics, the mean value of DiffAb software (69.455) is higher than the mean value of AlphaPanda (62.727). Meanwhile, the results of t-test showed that the t-value was 2.05, and the p-value was 0.043, which was lower than the 0.05 level, indicating that there were significant differences between DiffAb and

AlphaPanda in terms of seqid; and the antibody designed by DiffAb was closer to the reference standard in terms of sequence similarity.

In terms of ddG index, the mean value of DiffAb (1.185) was lower than that of AlphaPanda (307.400). Also the t-test showed that the t-value was -6.56 and the p-value was 0, which is below the 0.05 level. This indicates that there is a significant difference between DiffAb and AlphaPanda in terms of ddG; moreover, the mean value of DiffAb is much lower than that of AlphaPanda, and the p-value of 0 implies that this difference is extremely significant, which suggests that DiffAb-designed antibodies are superior to AlphaPanda in terms of thermodynamic stability.

Table 3.4 – **L_CDR1**

	software	Number of cases	average value	standard deviation	t	p
rmsd	DiffAb	100	0.596	0.120069383	-10.127	0
	AlphaPanda	10	1.011	0.15554073		
seqid	DiffAb	100	69.455	9.79973094	2.05	0.043
	AlphaPanda	10	62.727	10.88380916		
ddG	DiffAb	100	1.185	5.684579535	-6.56	0
	AlphaPanda	10	307.400	147.600621		

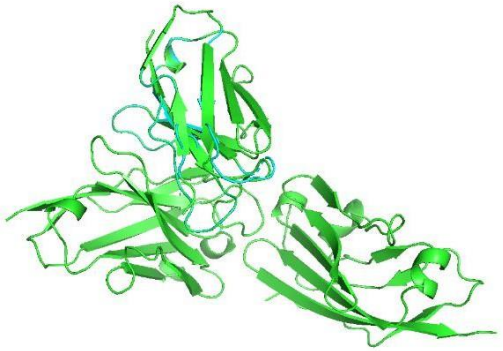

Figure 3.8 – L_CDR1 designed by AlphaPanda

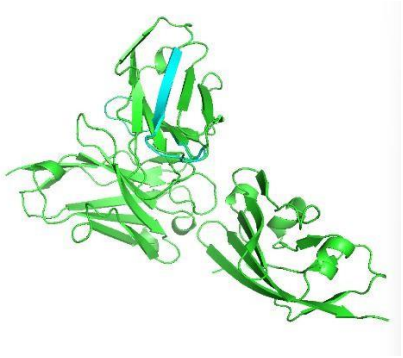

Figure 3.9 – L_CDR1 designed by DiffAb

3.6 L_CDR2

In terms of rmsd value, the mean value of the antibody designed by DiffAb software (0.540) is lower than the mean value of AlphaPanda (0.759). t-test results show that the t-value between DiffAb and AlphaPanda is -5.668 and the p-value is 0, which is lower than 0.05. This suggests that there is a significant difference between the two software-designed antibodies in terms of rmsd; moreover, the mean value of DiffAb is lower than AlphaPanda, and a p-value of 0 means that this difference is extremely significant. This indicates that in terms of rmsd, there is a significant difference between the two software-designed antibodies; moreover, the mean value of DiffAb is lower than that of AlphaPanda, and a p-value of 0 implies that this difference is extremely significant.The structural similarity of DiffAb-designed antibodies is higher than that of AlphaPanda.

In terms of seqid metrics, the mean value of DiffAb software (43.051) is higher than the mean value of AlphaPanda (26.402). Meanwhile, the results of t-test showed that the t-value was 3.049 and the p-value was 0.013, which was lower than the 0.05 level, indicating that there was a significant difference between DiffAb and AlphaPanda in terms of seqid; and the antibody designed by DiffAb was closer to the reference standard in terms of sequence similarity.

In terms of ddG index, the mean value of DiffAb (48.759) was higher than the mean value of AlphaPanda (19.403). Also the t-test showed a t-value of 0.552 and a p-value of 0.582, which is above the 0.05 level. This indicates that there is no significant difference between DiffAb and AlphaPanda in terms of ddG; the mean

value of DiffAb is higher than that of AlphaPanda, indicating that the antibody designed by AlphaPanda is better than DiffAb in terms of thermodynamic stability.

Table 3.5 – **L_CDR2**

	software	Number of cases	average value	standard deviation	t	p
rmsd	DiffAb	100	0.540	0.108129857	-5.668	0
	AlphaPanda	10	0.759	0.183954406		
seqid	DiffAb	100	43.051	8.617295698	3.049	0.013
	AlphaPanda	10	26.402	17.0522474		
ddG	DiffAb	100	48.759	165.4564049	0.552	0.582
	AlphaPanda	10	19.403	87.48451068		

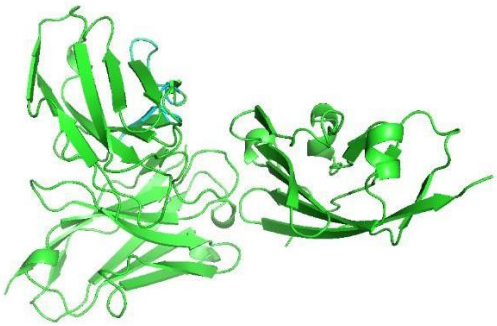

Figure 3.10 – L_CDR2 designed by AlphaPanda

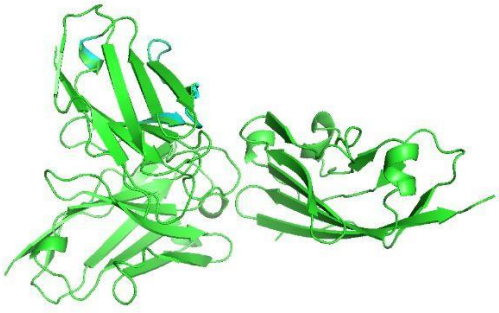

Figure 3.11 – L_CDR2 designed by DiffAb

3.7 L_CDR3

In terms of rmsd value, the mean value of the antibody designed by DiffAb software (0.670) is lower than the mean value of AlphaPanda (1.156). t-test results show that the t-value between DiffAb and AlphaPanda is -9.88, and the p-value is 0, which is lower than 0.05. This suggests that there is a significant difference between

the two software-designed antibodies in terms of rmsd; moreover, the mean value of DiffAb is lower than AlphaPanda, and a p-value of 0 means that this difference is extremely significant. This indicates that in terms of rmsd, there is a significant difference between the two software-designed antibodies; moreover, the mean value of DiffAb is lower than that of AlphaPanda, and a p-value of 0 implies that this difference is extremely significant.The structural similarity of DiffAb-designed antibodies is higher than that of AlphaPanda.

In terms of seqid metrics, the mean value of DiffAb software (49.778) is higher than the mean value of AlphaPanda (49.444). Meanwhile the result of t-test shows that the t-value is 0.134 and the p-value is 0.894, which is higher than the 0.05 level, indicating that there is no significant difference between DiffAb and AlphaPanda in terms of seqid.

In terms of ddG metrics, the mean value of DiffAb (4.950) was lower than the mean value of AlphaPanda (196.321). Also the t-test showed that the t-value was -2.421 and the p-value was 0.039, which is below the 0.05 level. This indicates that there is a significant difference between DiffAb and AlphaPanda in terms of ddG; the mean value of DiffAb is much lower than that of AlphaPanda, suggesting that the antibody designed by DiffAb is superior to AlphaPanda in terms of thermodynamic stability.

Table 3.6 – **L_CDR3**

	software	Number of cases	average value	standard deviation	t	p
rmsd	DiffAb	100	0.670	0.143953784	-9.88	0
	AlphaPanda	10	1.156	0.190097181		
seqid	DiffAb	100	49.778	7.319342189	0.134	0.894
	AlphaPanda	10	49.444	9.240722524		
ddG	DiffAb	100	4.950	4.004352125	-2.421	0.039
	AlphaPanda	10	196.321	249.9383563		

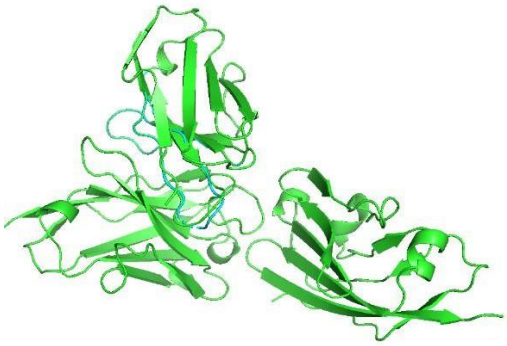

Figure 3.12 – L_CDR3 designed by AlphaPanda

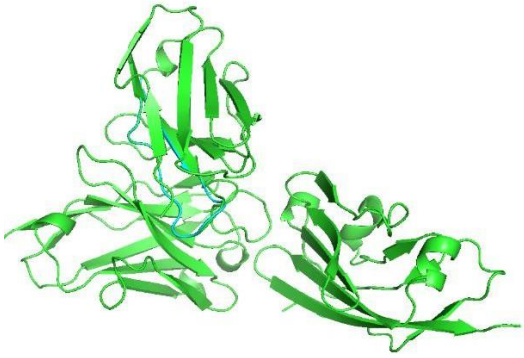

Figure 3.13 – L_CDR3 designed by DiffAb

3.8 Data Analysis

Generally speaking, the rmsd is about 1.5 Å to achieve atomic precision, and in the six CDRs designed by AlphaPanda and DiffAb in this experiment, H_CDR1, H_CDR2, L_CDR1, L_CDR2, L_CDR3 all achieve atomic precision (all less than 1.5 Å).

A seqid of about 30% indicates that the antibody obtained from the surface design has some homology with the natural antibody. Among the six CDRs designed by AlphaPanda in this experiment, the seqid values of H_CDR1, L_CDR1, L_CDR3 are all greater than 30%, which have some sequence homology with the natural antibody; H_CDR2 has the lowest homology of only 14%. Among the six CDRs designed by DiffAb, the seqid values of H_CDR1, L_CDR1, L_CDR2, and L_CDR3 were all greater than 30%, which had some sequence homology with the natural antibody; among them, the homology obtained from the design of H_CDR1 was the highest, which reached 70%, and the homology of H_CDR3 was the lowest, which was only 19%.

The lower the ddG, the more stable the resulting antibody is. By analyzing the data, the percentage of the energy of the designed antibody lower than the energy of the natural antibody is shown in Table 3-7. It can be concluded that DiffAb is more stable and better than AlphaPanda in the design of H_CDR2 and L_CDR1; AlphaPanda is more stable and better than DiffAb in the design of L_CDR2.

Table 3.7 – **Percentage of antibody energy obtained by design that is lower than natural antibody energy**

	H_CDR1	H_CDR2	H_CDR3	L_CDR1	L_CDR2	L_CDR3
DiffAb	5%	53%	1%	49%	39%	4%
AlphaPanda	0%	30%	0%	0%	90%	0%

3.9 Performance Evaluation of Artificial Intelligence Methods

3.9.1 Efficiencies

Traditional antibody discovery and preparation takes a significant amount of time, such as hybridoma technology or B-cell cloning, and typically takes months or even longer to identify and optimize effective antibodies. In contrast, AI approaches can quickly sift through large amounts of sequence and structural data to identify the most likely successful antibody candidates. By using deep learning and other machine learning techniques, AI is able to accomplish design tasks in a short period of time, dramatically speeding up the process of antibody discovery and initial evaluation[18].

AI is able to significantly shorten the development cycle by automating the screening of a large number of antibodies and rapidly identifying those that bind to specific targets. By combining AI technology with phage display technology, diagnostic and therapeutic antibodies against specific pathogens or toxins are developed[19]. This demonstrates that AI technology allows researchers to optimize the structure and function of antibodies, providing new avenues for the diagnosis and treatment of infectious diseases and toxins, while narrowing the scope of antibody screening and reducing experimental time.

3.9.2 Accuracy

Accuracy is the most basic and important indicator of AI performance. As derived from experiments using DiffAb and AlphaPanda to design human CD47 antibodies, AI methods can take into account the structural stability of an antibody at the design stage, for example, by predicting the behavior of the antibody during expression and purification to avoid aggregation tendencies or instability. Natural antibodies, on the other hand, may have limitations in terms of stability and manufacturability and require subsequent modification and optimization to be suitable for clinical applications.

3.9.3 Ingenuity

While natural antibodies excel in recognizing and neutralizing a wide range of pathogens thanks to their diversity and the optimization of evolutionary processes, these antibodies may lack the innovative structures required in certain highly specific applications. In contrast to the possible limitations of natural antibodies, AI approaches demonstrate unique advantages. Using AI techniques, researchers can explore the use of unnatural amino acids or non-standard antibody frameworks for antibody design, which allows for the creation of entirely new antibody entities that are difficult to find under natural conditions. This approach not only extends the structural and functional range of antibodies, but may also lead to more effective therapeutic options for complex diseases[20].

3.10 Application prospects and challenges

The application of artificial intelligence in the field of antibody design has shown great potential, foreshadowing a more important role for this technology in biopharmaceutical research and development in the future.AI can improve the speed and accuracy of antibody discovery, as well as process a large amount of biological data to quickly identify protein targets associated with specific diseases and design antibodies with high specificity and affinity. With increased computing power and improved algorithms, the speed and accuracy of AI prediction is expected to increase further, significantly shortening the cycle from antibody discovery to optimization. It

can also open up new therapeutic areas by identifying new immune targets that are difficult to discover by traditional methods and designing antibodies against them. For example, in under-explored areas such as neurodegenerative diseases and rare diseases, AI can provide new therapeutic strategies by simulating and predicting the effects of mutations in antibodies to optimize their expression, stability and efficacy[21].

Although AI approaches have brought many innovations in the field of antibody design, there are several challenges to its development and application. In the antibody design stage, AI can efficiently generate a large number of candidate antibodies, providing researchers with a rich selection space. However, the shortcomings of antibody design by AI are the lack of experimental validation and the relative difficulty of the validation process. This is because the experimental validation process may be constrained by a variety of factors, such as complex experimental conditions, difficulties in obtaining experimental materials, and high experimental costs.

The main disadvantage of AI-designed antibodies is that it greatly depends on the quality and diversity of the input data. If the data used to train the AI model is biased or incomplete, the antibody produced by the model may not be sufficiently generalizable or achieve the desired functional effect[22]. In addition, while AI can speed up the antibody design process, this speed may come at the expense of biological validation, as the laboratory validation step still takes time, and over-reliance on AI may lead to misclassification. Moreover, data in the biomedical field often involves sensitive personal information, so data sharing is limited by strict privacy laws and ethical considerations[23].

Conclusions to chapter 3

1. The use of artificial intelligence in antibody design offers significant advantages in efficiency, accuracy, and ingenuity. AI accelerates the traditionally long process of antibody discovery by quickly analyzing vast

amounts of data to identify promising candidates. It also enhances accuracy through deep learning models, which can predict antibody structures with improved stability and manufacturability, reducing the need for extensive post-optimization. Additionally, AI approaches enable innovative designs by exploring non-standard frameworks and unconventional amino acids.AI is able to significantly shorten the development cycle by automating the screening of a large number of antibodies and quickly identifying those that bind to specific targets. The development of diagnostic and therapeutic antibodies against specific pathogens or toxins has been achieved by combining AI technology with phage display technology. This demonstrates that AI technology allows researchers to optimize the structure and function of antibodies, providing new avenues for the diagnosis and treatment of infectious diseases and toxins, while narrowing the scope of antibody screening and reducing experimental time.

2. Although AI methods have brought many innovations in the field of antibody design, there are some challenges to its development and application[22]. The shortcomings of AI-designed antibodies are the lack of experimental validation and the relative difficulty of the validation process. And the main disadvantage of AI-designed antibodies is that it greatly depends on the quality and diversity of the input data. If the data used to train the AI model is biased or incomplete, the antibody produced by the model may not be sufficiently generalizable or may not achieve the desired functional effect[23].
3. By analyzing the data, it can be seen that DiffAb has an overall better performance than AlphaPanda in the design of human CD47 antibody; in the design of L_CDR2, the antibody designed by AlphaPanda is better than DiffAb in thermodynamic stability.

CONCLUSIONS

1. Most of the six CDRs designed by AlphaPanda and DiffAb have RMSDs of less than 1.5 Å in most of the parts, which is at the atomic precision level. In terms of sequence homology, most of the antibodies designed by AlphaPanda and DiffAb showed some similarity to natural antibodies, especially H_CDR1, L_CDR1, and L_CDR3. However, in the design of H_CDR2, AlphaPanda's homology was lower, only 14%; in the design of H_CDR3, DiffAb's homology was lower, only 19%. In addition, the analysis of ddG data revealed that the energy of the antibody obtained from the design was lower than that of the natural antibody, indicating that the designed antibody was more stable. Overall, DiffAb performs more stable and better than AlphaPanda in the design of H_CDR2 and L_CDR1; while in the design of L_CDR2, AlphaPanda performs more stable and better than DiffAb.

2. Artificial intelligence methods have overturned traditional methods and greatly reduced the time to design and optimize antibodies, but their application still faces multiple challenges. Artificial intelligence can generate a large number of candidate antibodies in the design phase, but the shortcoming of AI-designed antibodies is that these designs lack the necessary experimental validation support, which is mainly due to the difficulty of implementing the current experimental validation. The experimental validation process may be constrained by a variety of factors, such as the complexity of experimental conditions, the difficulty of obtaining experimental materials, and the high cost of experiments. Therefore, the continuous innovation of technology and the strengthening of interdisciplinary collaboration in the future will be the core driving force to promote the wide and deep application of AI in the field of antibody design.

3. This study demonstrates that the AI method can rapidly screen a large amount of data in terms of efficiency, which greatly improves the speed of antibody discovery. Meanwhile, AI can successfully design human CD47 antibodies, which

can achieve the atomic precision and high sequence consistency of natural antibodies in terms of structure and sequence.

LIST OF REFERENCES

[1] Qi L. (2020). Structure prediction and humanization modification of anti CD47 antibodies based on molecular simulation technolog Master (Dissertation, Beijing Union Medical College).Master https://link.cnki.net/doi/10.27648/d.cnki.gzxhu.2020.000537 doi:10.27648/d.cnki.gzxhu.2020.000537.

[2] Jie X L, Kong Y Y & Zhou G M. (2023). Research progress on the role of CD47 in tumor immune escape and targeted treatment strategies. Journal of Sichuan University (Medical Edition) (03), 455-461.

[3] Zheng W J, Zhang H, Huo Y, Zhang X Q & Wang T. Application of macrophage phagocytosis checkpoint blockade in anti-tumor immunotherapy. Journal of Air Force Medical University 1-12.

[4] Huang Q H, Zhang W T & Jin H. (2019). Research progress in targeted CD47 anti-tumor therapy. Chinese Journal of Tumor Biotherapy (10), 1161-1166.

[5] Zheng Hu,Wanqiong Li,Shaomeng Chen,Danhong Chen,Ran Xu,Danlu Zheng... & Yanfeng Gao. (2023). Design of a novel chimeric peptide via dual blockade of CD47/SIRPα and PD-1/PD-L1 for cancer immunotherapy. Science China(Life Sciences) (10), 2310-2328.

[6] Zheng K F, Xiong Y L, Zhang J, Lu Q, Tang X Y, Liu Y J... & Zhao J B. The expression characteristics and significance of immune checkpoint genes in non-small cell lung cancer. Journal of Air Force Medical University 1-19.

[7] Liu Q Q, Tang C C, Li H & Cheng Y. (2021). Research progress in targeted CD47 anti-tumor therapy. Tumour (09), 640-649.

[8] Zhang H, Wang H J, Lu R J, Lan J J, Chen L J & He X B. (2024). Progress in the Application of Protein Structure Prediction Model AlphaFold2. Journal of Biotechnology (05), 1406-1420. doi:10.13345/j.cjb.230677.

[9] Chen Z H, Ji M L & Qi Y F. (2023). Research progress in artificial intelligence protein structure design algorithms. Synthetic Biology (03), 464-487.

[10] Ben Shor & Dina Schneidman Duhovny. (2024). CombFold: predicting structures of large protein assemblies using a combinatorial assembly algorithm and AlphaFold2. Nature methods (3), 477-487.

[11] Nathaniel R Bennett,Joseph L Watson,Robert J Ragotte,Andrew J Borst,Déjenaé L See,Connor Weidle... & David Baker. (2024). Atomically accurate de novo design of single-domain antibodies.. bioRxiv : the preprint server for biology

[12] Watson Joseph L,Juergens David,Bennett Nathaniel R,Trippe Brian L,Yim Jason,Eisenach Helen E... & Baker David. (2023). De novo design of protein structure and function with RFdiffusion.. Nature (7976), 1089-1100.

[13] Hao Z W, Wang H P, Liu C, Ding C C, Liu L,He T... & Dong K. (2020). Preparation of anti-CD47 monoclonal antibody and identification of its pro macrophage phagocytic effect. Modern Oncology Medicine (18), 3114-3118.

[14] Wang F. (2024-03-21). Artificial intelligence is producing new antibodies from scratch for the first time. China Science Daily, 002.

[15] Anand Namrata, Eguchi Raphael, Mathews Irimpan I, Perez Carla P, Derry Alexander, Altman Russ B. & Huang Po Ssu. (2022). Protein sequence design with a learned potential. Nature Communications (1), 746-746.

[16] Isert Clemens, Atz Kenneth & Schneider Gisbert. (2023). Structure-based drug design with geometric deep learning. Current opinion in structural biology 102548-102548.

[17] Hu, Y, F. Tao, W. Lan. (2024). Combining transformer and 3DCNN models to achieve co-design of structures and sequences of antibodies in a diffusional manner. 2024.04.25.587828.

[18] (2019). Artificial intelligence AlphaFold successfully predicts protein 3D structure. Journal of Medical Informatics (01), 93.

[19] Roth Kristian Daniel Ralph,Wenzel Esther Veronika,Ruschig Maximilian,Steinke Stephan,Langreder Nora,Heine Philip Alexander... & Hust Michael. (2021). Developing Recombinant Antibodies by Phage Display Against Infectious Diseases and Toxins for Diagnostics and Therapy
. Frontiers in Cellular and Infection Microbiology 697876-697876.

[20] Li J J, Chen D X, Chai R J & Yu W W. (2023). The potential and application of protein attribute prediction based on artificial intelligence. Pharmaceutical progress (10), 733-740. doi:10.20053/j.issn1001-5094.2023.10.003.

[21] Kaczmarski, J., H. Ashley, J. Dalton. (2023). Applying ProteinMPNN, RFDiffusion & ColabFold tools for the generation & optimisation of useful protein-tools. Protein Science. 32(12).

[22] Miao H J, Dong K Z, Xiang Q R & Xue G R. (2024). Research progress of artificial intelligence protein design technology and its application and challenges in innovative development of biomedicine. Shanghai Pharmaceutical (07), 1-9+55.

[23] Wen L L. (2024-03-12). Protein design safety guidelines have been released. China Science Daily, 002.