

IsAb: a general in silico protocol for antibody design

by

Tianjian Liang

Bachelor of Engineering, Jiangnan University, 2019

Submitted to the Graduate Faculty of the
School of Pharmacy in partial fulfillment
of the requirements for the degree of
Master of Science

University of Pittsburgh

2021

UNIVERSITY OF PITTSBURGH

SCHOOL OF PHARMACY

This thesis or dissertation was presented

by

Tianjian Liang

It was defended on

March 24, 2021

and approved by

Xiang-Qun (Sean) Xie, PhD, EMBA, Department of Pharmaceutical Sciences

Levent Kirisci, PhD, Professor, Department of Pharmaceutical Sciences

Junmei Wang, PhD, Associate Professor, Department of Pharmaceutical Sciences

Zhiwei Feng, PhD, Assistant Professor, Department of Pharmaceutical Sciences

Thesis Advisor/Dissertation Director:

Zhiwei Feng, PhD, Assistant Professor, Department of Pharmaceutical Sciences

Copyright © by Tianjian Liang

2021

IsAb: a general in silico protocol for antibody design

Tianjian Liang, B.E

University of Pittsburgh, 2021

ABSTRACT

The design of therapeutic antibodies has attracted a large amount of attention over the past ten years. Antibodies are widely used to treat many diseases due to their high efficiency and low risk of adverse events. However, the experimental methods of antibody design are time-consuming and expensive. Although computational antibody design techniques have had significant advances in the past years, there are still some challenges that need to be solved, such as the flexibility of antigen structure, the lack of antibody structure data, and the absence of standard antibody design protocol. In the present work, we elaborated on an in-silico antibody design protocol for users to easily perform computer-aided antibody design. First, the Rosetta web server will be applied to generate the 3D structure of query antibodies if there is no structural information available. Then, two-step docking will be used to identify the binding pose of an antibody-antigen complex when the binding information is unknown. ClusPro is the first method to be used to conduct the global docking, and SnugDock is applied for the local docking. Sequentially, based on the predicted binding poses, in-silico alanine scanning will be used to predict the potential hotspots (or key residues). Finally, computational affinity maturation protocol will be used to modify the structure of antibodies to increase their affinity and stability. We also redesigned antibody D44.1 which is an anti-hen egg white lysozyme antibody and compared it with previously reported data to strengthen the persuasion of IsAb protocol. To further illustrate our proposed protocol, we used an

FDA-approved PD-1 checkpoint inhibitors antibody, cemiplimab as an example to showcase a step-by-step tutorial.

Key word: antibody design, protein engineering, protein-protein docking, computer-aided antibody protocol

Table of Contents

PREFACE.....	xi
1.0 INTRODUCTION.....	1
1.1 Antibody Structure.....	1
1.2 Antibody Development.....	3
1.3 Current Status of Antibody Design.....	3
1.4 The Usage of Computational Antibody Design	6
1.5 Common Techniques in Computational Antibody Design	8
1.5.1 Antibody Modeling	8
1.5.2 Affinity Maturation.....	10
1.6 Aim.....	13
2.0 MATERIAL AND METHODS	14
2.1 Generating and Preparing the Structures.....	14
2.2 Antibody-Antigen Global Docking	14
2.3 Antibody-Antigen Local Docking	15
2.4 Computational Alanine Scanning and Affinity Maturation.....	16
2.5 Molecular Dynamics (MD) Simulation and Molecular Mechanics/ Generalized Born Surface Area (MM/GBSA) Calculation.....	17
3.0 RESULTS	18
3.1 Workflow of Antibody Design Protocol	18
3.1.1 Generating and Preparing the Structures	20
3.1.2 Antibody-Antigen Global Docking	21

3.1.3 Antibody-Antigen Local Docking	22
3.1.4 Computational Alanine Scanning.....	23
3.1.5 <i>In silico</i> Affinity Maturation	24
3.2 Antibody D44.1 Redesign.....	25
3.2.1 D44.1 Structure and Binding Pose Generation	25
3.2.2 D44.1 Affinity Maturation.....	28
3.3 Case Study: Cemiplimab Design Procedure	31
3.3.1 Generating and Preparing the PD-1 and Cemiplimab Structures	34
3.3.2 Cemiplimab-PD-1 Global Docking.....	36
3.3.3 Cemiplimab-PD-1 Local Docking.....	38
3.3.4 Computational Alanine Scanning for Cemiplimab.....	43
3.3.5 <i>In silico</i> Affinity Maturation for Cemiplimab	44
4.0 CONCLUSION	48
APPENDIX.....	51
BIBLIOGRAPHY	60

List of Tables

Appendix Table 1 D44.1 local docking top10 I_sc decoys results	51
Appendix Table 2 Comparison of D44.1 redesign results	52
Appendix Table 3 Model1 local docking top10 I_sc decoys results.....	53
Appendix Table 4 Model2 local docking top10 I_sc decoys results.....	54
Appendix Table 5 PD-1 antibodies and their sequence alignment with cemiplimab	55

List of Figures

Figure 1. An illustration of the antibody structure.....	2
Figure 2. The number of approval antibodies and relative publications each year.	4
Figure 3. Computational antibody design process.....	6
Figure 4. Predicting possible paratopes from database.	15
Figure 5. The workflow of the IsAb.	19
Figure 6. The crystal structure of HyHEL-5-lysozyme complex.....	26
Figure 7. The I_{sc} vs Irmsd plot of D44.1-lysozyme local docking.	27
Figure 8. Superimposition of modeling and crystallization D44.1-lysozyme structures.	28
Figure 9. Boxplot of comparison of three metrics between antibody D44.1 control and design groups.....	30
Figure 10. The crystal structure of PD-1 (extracted from PDB:6jbt).....	32
Figure 11. Superimposition of PD-1 structures from three PD-1/antibody complexes.....	33
Figure 12. Three crystal structures of pembrolizumab, nivolumab, and toripalimab complex with PD-1.	34
Figure 13. Three most representative PD-1s structures.....	35
Figure 14. Three possible binding poses generated from global docking.....	37
Figure 15. The I_{sc} vs Irmsd plot of models1 and 2 local docking.....	39
Figure 16. Detailed hydrogen bonds in cemiplimab/PD-1 complex.	41
Figure 17. Hydrophobic interaction in cemiplimab/PD-1 complex.....	42
Figure 18. Score changing of mutated residues on antibody during affinity maturation....	44

Figure 19. Comparison of MD simulation results between four point mutations and control cemiplimab.....	46
Figure 20. Comparison of hydrogen bonds between the control and the design groups.	47
Appendix Figure 1. The script of Rosetta relax.	57
Appendix Figure 2. The script of Rosetta alanine scanning.	57
Appendix Figure 3. The script of defining antibody-antigen interface.....	58
Appendix Figure 4. The script of Rosetta affinity maturation.	58
Appendix Figure 5. The script of comparing controlled and designed antibody.....	58
Appendix Figure 6. The script of per residue energy analysis.	59
Appendix Figure 7. The script of residue energy breakdown analysis.....	59

PREFACE

I sincerely appreciate my advisor Dr. Zhiwei Feng who is an expert in computer-aided drug discovery. During the past two years, I studied in Pharmacometrics and Systems Pharmacology track, he taught me many things, from study to life. Dr. Feng hold the group meeting once a week. He let the group members discussed their projects and help with each other, which taught me the importance of teamwork. Dr. Feng also taught me the usage of the basic software and gave me the idea when I was stuck in the research. Most importantly, he always encouraged me and gave me lots of chance to show my research in front of the other people, which improved my presentation skill and made me more confident.

Also, I would like to express my thanks to Dr. Xiangqun Xie. As the director of Computational Chemical Genomics Screening Center (CCGS), he gave me the direction of my research. He always provided some latest findings in the scientific area to give us insight for our research.

Meanwhile, I would like to thank my committee member, Dr. Junmei Wang and Dr. Levent Kirisci. Dr. Junmei Wang is an expert in molecular dynamics simulation and force field study. The Computational and Systems Pharmacology course he taught gave me a general understanding for the software I could use in my research. Dr. Levent Kirisci is the professor in Statistical class. I have learnt the basic statistical knowledge and constructed the knowledge system from his class, which help me a lot in my later study.

Besides, I want to express my sincere acknowledgement to all the members in CCGS center. Especially Maozi Chen who is a programming expert in our center. He developed M CCS program, gave me step by step instruction for the basic software usage, and provided me lots of technical

supports. I also want to thank Dr. Lirong Wang, Dr. Terence McGuire, Dr. Jaden Jun, and Dr. Ying Xue, who are the professors in CCGS center. They taught me many new knowledge from the other areas and gave lots of useful advice for my research. Also, I would like to thank the support and assistance from my peers during my two-year studies.

Lastly, as an old saying goes “parents will always be your strong support”. My parents fully supported me during my master study, from life to study. I hope I can express my emotion to them at any moment. Now, I want to say “thank you” for them.

I offer my best regards for all of those who gave me a hand during my two-year master studies!

1.0 INTRODUCTION

1.1 Antibody Structure

Antibodies are Y-shaped proteins produced by B-cells in plasma and play a significant role in the immune system. Antibodies can recognize specific pathogens such as viruses and pathogenic bacteria, called antigens. Most of the antibodies have the similar structure (**Figure 1**) that consists of two symmetrical light and heavy chains, which are connected by two disulfide bonds. Classified by the heavy chains, antibodies can be divided into five different subtypes. The light chains have two types as well. Most of the antibodies used in therapy are IgG, especially IgG4 or IgG1. Antibodies use their fab region which is located on the two branches of the Y-shaped proteins to recognize and bind antigens. The variable regions are the sites inside the Fab region that contribute to the antigen-binding function and are located at the N-terminus of each chain. The hypervariable amino acid loops within the variable regions are called complementarity determining regions (CDR), and it is formed by six loops from heavy chains and light chains (H1, H2, H3, L1, L2, L3). Due to the variability of CDR, changes in this region can cause a significant change in specificity or affinity for the target, which makes the CDRs a region of interest for antibody engineering. Contrary to the variable regions are the constant regions where the amino acid sequences are relatively conserved. The base of the Y-shaped protein, called the Fc region, is composed of two heavy chains and plays an important role in antibody bioactivity. The Fc region can recognize various cell receptors and complement proteins, which is relative to the antibody's half-life and complement-dependent cytotoxicity (CDC).

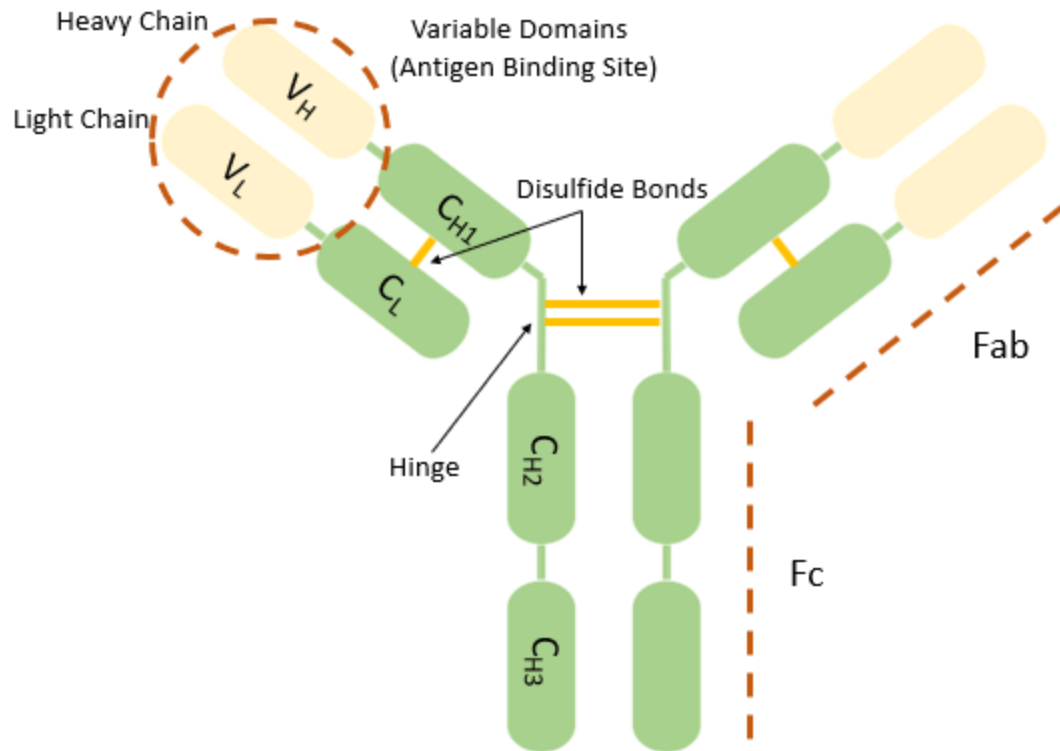


Figure 1. An illustration of the antibody structure.

The constant regions are in green, the variable regions are in light-yellow, and the disulfide bonds are highlighted in yellow.

1.2 Antibody Development

Due to the high affinity as well as specificity to a wide variety of macromolecules, antibodies are widely used in the immunotherapy treatments for cancers, autoimmune diseases, inflammatory diseases, and drug abuse, for example, combating synthetic opioid intoxication [1] and blocking abused drugs [2]. Moreover, antibodies play an important role in fighting infectious viruses, especially current SARS-CoV-2 [3]. Antibodies not only can be applied in COVID-19 testing [4], but also can neutralize the SARS-CoV-2 virus through targeting the spike protein of SARS-CoV-2 [3]. Additionally, antibodies have well-established expressing and purifying methods that are highly attractive which enable antibodies to be one of the critical agents in therapy [5]. Therefore, antibody therapy has attracted more and more attention in the past years. The number of approved antibodies has steady risen (**Figure 2**) and its market becomes larger [6]. Researchers have been working on further improvement of the antibody's properties, such as affinity, stability [7], or decreasing the immunogenicity through antibody rational design. Due to these factors, the development of novel antibody design techniques is essential.

1.3 Current Status of Antibody Design

Antibody design can be classified into redesign and de novo design [8]. Antibody redesign means modifying the structure based on a known antibody structure and its sequence. Since the process of redesign is based on the existing antibodies, especially FDA-approved antibodies, it is easier to modify antibodies properties without causing severe immunogenicity and the antibodies can be applied to clinical usage much faster, which can reduce time and cost. Antibody redesign

[8] heavily relies on a native protein or an interface and mainly focuses on engineering the surface or interior of the antibody-antigen complex to acquire better affinity or stability of the complex [9, 10]. In vivo affinity maturation is an essential method to get better properties of an antibody, which can increase its affinity several-fold by somatic hypermutation (SHM) during an immune response [11]. SHM usually occurs in the Complementarity-determining regions (CDR) of the immunoglobulin genes and has a rapid mutation rate [12].

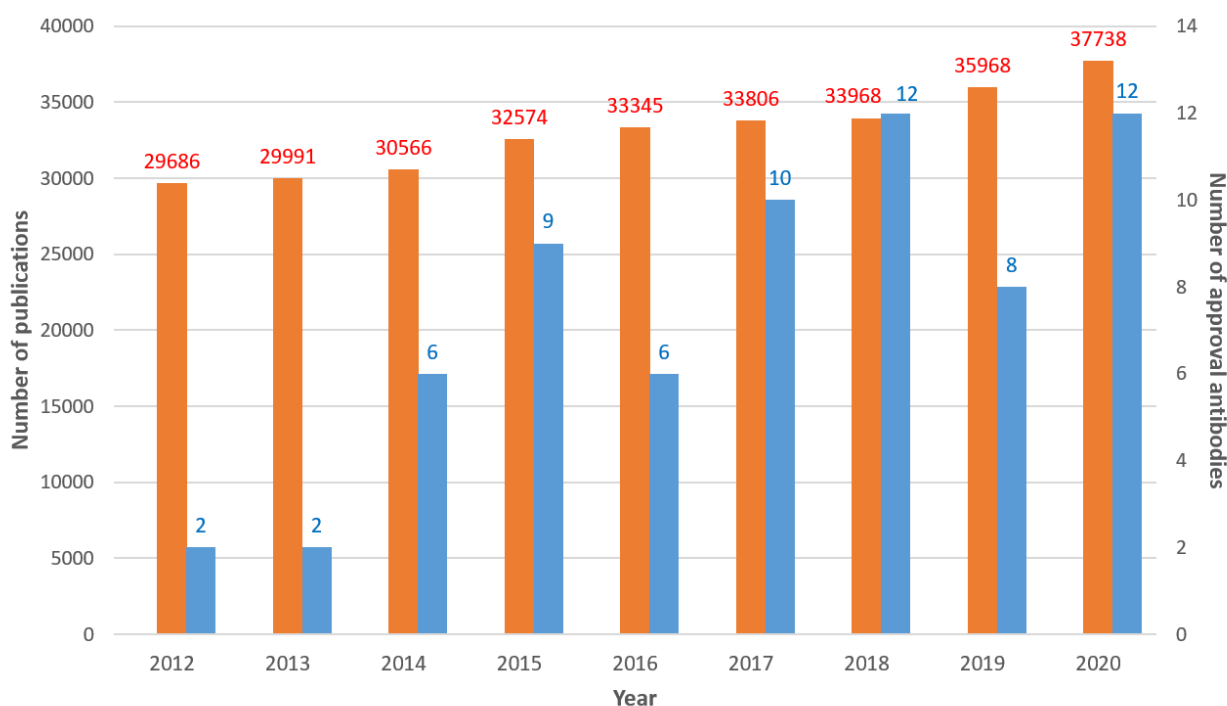


Figure 2. The number of approval antibodies and relative publications each year.

Publications number data retrieved from the Web of Science through using key word ‘antibody,’ then the function ‘Results Analysis.’ Number of approval antibodies data generated from FDA website. The blue bars are the number of FDA approval antibodies. The orange bars are the number of publications regarding antibodies.

However, there are still many limitations on antibody empirical rational design. For example, using an X-ray to obtain antibody structure and screening the libraries [13] to optimize

the antibody affinity is time-consuming and expensive. Random mutagenesis requires a heavy workload. Thus, it is essential to apply alternative strategies to aid in antibody design.

In recent years, computational methods have been used extensively in drug discovery and design, which give rise to the field called computer-aided drug design (CADD). CADD has lots of advantages, such as cost reduction by decreasing the number of failures and promoting the success rate of the test [14]. As a result, CADD is utilized in drugs development, such as CB2-agonist complex generation [15] and SARS-CoV-2 drugs exploration [16]. CADD also plays a key role in protein drug design by establishing biological databases such as EMBL [17] (<https://www.ebi.ac.uk/>), setting up sequence searching programs such as BLAST [18], developing computational modeling analysis programs such as SWISS-MODEL [19] (<https://swissmodel.expasy.org/>), and initiating ligand selectivity program DOCK [20]. Among them, artificial intelligence becomes more and more popular in recent years. Its techniques like, machine learning, deep learning have been applied in antibodies discovery and design, due to its ability to rationalize and actions that have chance of achieving a specific goal. Computational techniques generate testable hypotheses to guide experiments or validation by experiments. In this case, it is suitable to use computational techniques to assist with antibody rational design, as described in the review by Kuroda and co-workers [21] (**Figure 3**).

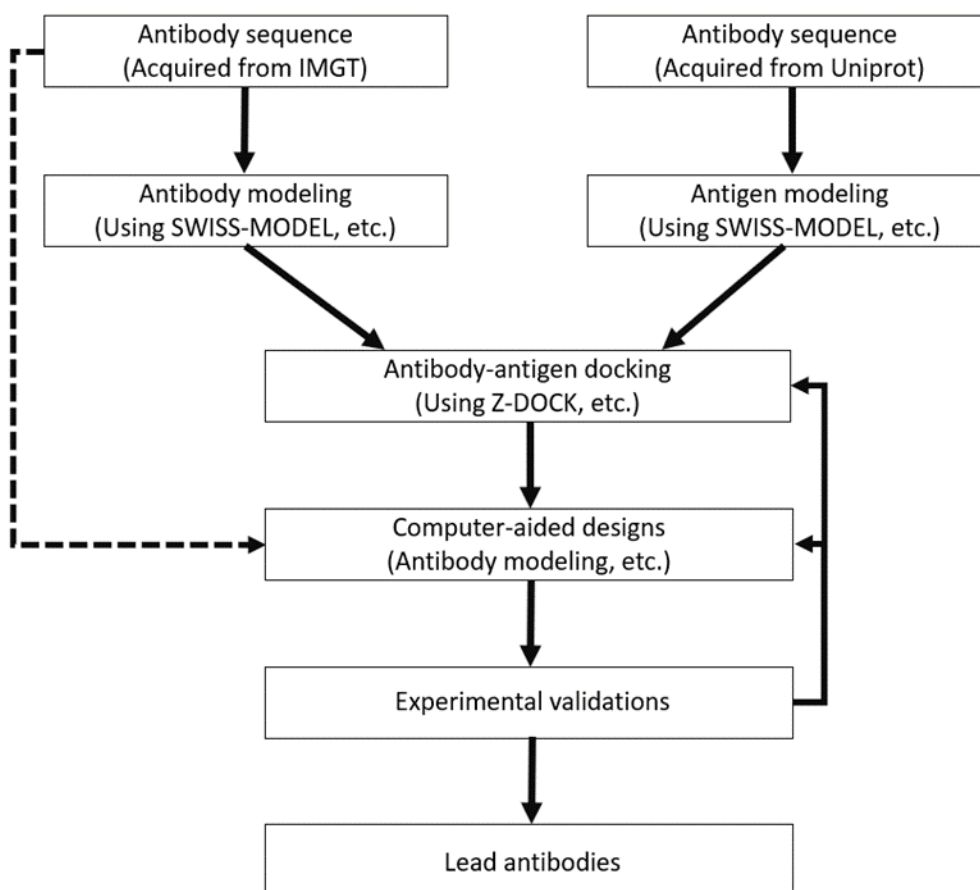


Figure 3. Computational antibody design process.

Step 1: 3D homology modeling of antibody and antigen based on their sequences. Step 2: Construction of the antibody-antigen complex by molecular docking. Step 3: Validation of the antibody-antigen complex via the experiment.

1.4 The Usage of Computational Antibody Design in Current Diseases

Computational antibody design with its advantage is applied in many current diseases drug development. For the current COVID-19 crisis, Desautels *et al.* [22] developed a computational-experimental platform which generated antibody targets to the SARS-CoV-2 receptor binding domain (RBD) based on published neutralizing SARS-CoV-1 antibodies. The computational-

experimental platform combined data-driven (machine learning), theory-driven, and experiment-driven approaches and leveraged the strength of each approach while minimizing their limitations. Desautels *et al.* used a machine learning module to propose mutations to the SARS-CoV-1 antibody (M396) and applied FoldX [23] to calculate the binding free energy of mutated antibodies. They also performed additional free energy calculations through Rosetta and molecular dynamics simulation, and finally selected 20 antibodies. According to the result, the 20 antibodies ranged from baseline free energy of -48.1 kcal/mol to -82.0 kcal/mol. The baseline energy of M396 combined with SARS-CoV-1 was -52.2 kcal/mol. This result suggested that the predicted antibody mutants may bind the SARS-CoV-2 RBD.

To obtain high-efficiency therapeutic antibodies to address acquired immune deficiency syndrome (AIDS), Farhadi *et al.* [24] used computational methods to improve the affinity of Ibalizumab, an CD4 antibody targeting HIV-1. They first employed PyMOL to mutate six amino acids located on the VH chain of Ibalizumab which contains key residues for interacting with CD4, along with other 19 interacting residues. Then, they applied Chimera 1.10 software [25] to prepare CD4 and Ibalizumab. Ibalizumab and CD4 docking simulations were performed through ZDOCK server while the docking results were evaluated using Hex docking software [26]. They analyzed hydrogen bonds using LigPlot tool [27]. Based on the above calculations, they selected seven potential mutants among 720 designed antibodies. Five selected antibodies showed higher energy of interactions with CD4 and six of them had more H-bonds with CD4 compared to Ibalizumab. The computational prediction results suggested that the designed antibodies may have higher potential targeting HIV-1.

1.5 Common Techniques in Computational Antibody Design

1.5.1 Antibody Modeling

Antibody modeling is the key process of antibody design, which uses amino acid sequences from the antibodies to predict its structures [28]. With antibody modeling techniques, we can rationally modify antibody molecules, for example, redesigning smaller antibody fragments, improving their stability or affinity for the antigens [29], and engineering fusion proteins to enhance their specificity of the cellular targets [30]. Antibody modeling techniques can also help to identify epitopes of the antigen-binding sites. Moreover, this technique gives us a deeper understanding of the antibody-antigen recognition mechanism, which is essential in the humanization of animal-derived antibodies [31]. As mentioned in the introduction, the variable regions are the antigen-binding sites of most antibodies. Therefore, antibody modeling focuses on the structure of the variable regions. However, the CDRs inside the variable regions are the core challenge for computational antibody modeling due to variability across domains. Despite the high sequence variability of CDRs, the “canonical structures” of CDR made up of H1, H2, L1, L2, and L3 loops in the main chain structure are very similar [32]. The combined mode of residues which determines canonical structures have already been discovered. Hence, the researchers can predict the canonical structure accurately by aligning an unknown structure to the known structures in the database or base on the sequences alone [33]. The H3 loop is more unique than other loops within the main chain structure. The conformation of the H3 loop cannot find a suitable catalog in the canonical structure due to its highly variable sequence, length, and structure. For these reasons, CDR loops, especially the H3 loop, are the core challenge in antibody modeling.

The general method to model immunoglobulin is aligning the unknown sequences of each variable domain of antibodies to the known structures. Then, using the template domains of known structures to decide the framework regions of each domain. The prediction of the H3 loop could be made after examining whether the target sequences are similar to the hypervariable loops. The side chain is minimized at the end of the process. Long *et al.* [34] applied a machine learning approach (Gradient Boosting Machine [GBM]) to model the antibody. They applied both the GBM method and RosettaAntibody in triplicate 10-fold cross-validation of non-H3 CDR loops modeling using data from the cluster annotated antibody database PyIgClassify [35]. Their modeling results were compared after applying the methods. According to the result, GBM approach ($83.4\% \pm 0.11\%$) had higher accuracy than RosettaAntibody ($79.0\% \pm 0.23\%$). Also, the GBM method could reduce the errors in misclassifications of specific cluster membership. Machine learning approaches are becoming increasingly relevant to antibody modeling, providing advantages for data integration and simplification of feature selection.

Several researchers have contributed to the modeling process of the H3 loop. Zhu *et al.* [36] used program Prime, an ab initio method-based tool, to predict the structure of the H3 loop through energy calculations and conformational sampling of the program. They predicted 53 antibodies' H3 loop, and 70% of root means square deviation (RMSD) of cases they sampled were below 2.0 Å. Kuroda *et al.* [37] proposed new H3-rules based on the rules proposed by Shirai *et al.*, which renewed the signals for base identification and the relationship of total length and base forms. Yamashita *et al.* [38] developed a fully automated web service (Kotai Antibody Builder: https://bio.tools/kotai_antibody_builder) based on the new H3-rules and tested 11 antibodies' H3 loop. Unlike other web services, Kotai Antibody Builder was able to refine the model by using fragment assembly and then performed knowledge-based scorings. The result showed that the

Kotai Antibody Builder was better than other web services like Prediction of Immunoglobulin Structure (PIGS) (http://www.biocomputing.it/pigs/index_single). Furthermore, the average RMSD of Kotai Antibody Builder was the same as the semi-automated predictors and is easier to operate. Weitzner *et al.* [39] used the RosettaAntibody algorithm to predict short (8-10 residues) and long (11-14 residues) CDR H3 loops. The result showed that the average RMSDs of short H3 loops were 1.66 ± 0.96 Å and the long loops were 2.77 ± 1.16 Å, which meant RosettaAntibody could produce models with high accuracy. Interestingly, Ruffolo *et al.* [40] used DeepH3, a deep residual neural network, to predict orientations and inter-residue distances of antibodies based on the light and heavy chain sequences. In order to differentiate the decoy structures from RosettaAntibody, the output of DeepH3 is converted to geometric potential. Ruffolo *et al.* retrieved data from SAbDab as the training dataset and collected data from PylgClassify database as a benchmark dataset. The result showed that DeepH3 picked worse-, same-, and better-RMSD structure (measured by RMSD from the experimental CDR H3 loop structure) for 6, 13, and 30 of 49 targets from benchmark dataset, with an average RMSD improvement of 0.40 Å.

1.5.2 Affinity Maturation

In vivo antibody affinity maturation is the process that increases antibody affinity by Tfh cell (Follicular B helper T cells) activation during an immune response [11]. After repeating exposures to the same antigens, the antibody will increase its affinity by somatic hypermutation (SHM), which occurs on the surface of germinal center B cells and is selected by Tfh cells. The affinity of antibody in the secondary response increases several-fold compared to a primary response. SHM usually occurs in the CDR of the immunoglobulin genes which has a rapid mutation rate [41]. Although the exact mechanism of the SHM is not clear, the researchers still

modify the antibodies in vitro using the antibody engineering techniques, such as random mutations and phage display. In vitro affinity maturation is based on mutation and selection, and aims to optimize antibodies with desirable properties [42]. To improve the success rate of mutation, it is essential to identify the configuration of the antibody complexes or the paratopes inside the antibodies. Hence, combining computational approaches with display technologies is a wise method to enhance the efficiency of mutation [43].

Based on the single-chain antibodies (scFvs) 2D10 and its antigen DVFYPYPYASGS, Poosarla *et al.* [44] combined computational framework named Optimal Method for Antibody Variable region Engineering (OptMAVEN) [45] (<https://github.com/maranasgroup>) with molecular dynamics (MD) simulation to conduct de novo design of scFvs-2D10. They used MD simulation to refine scFvs-2D10 which ensures the stability of the antibody-antigen complex. Later *in silico* affinity maturation is performed to accumulate potent mutations to the refined antibodies. Also, interaction energies (IEs) and complex were used as the stability metric to select useful mutations. The result showed that three of five mutations (scFvs-1, 2, and 4; $K_d=8.9, 14.4,$ and 23.8 nM) exhibited a higher binding affinity than the scFvs-2D10 ($KD=3,7$ nM). The result supported the potency of OptMAVEN in *silico* antibody affinity maturation. Currently, Cannon *et al.* [46] developed a novel computational method with the help of alanine scanning mutations and combination with *in silico* complex to drive *in silico* affinity maturation when the crystal structure of antibody-antigen complex was not available and the wild-type antibody had strong K_d in the nanomolar range. The murine protein CCL20 and its binder AB1 acted as an example since they had strong K_d in the nanomolar range. Cannon *et al.* first used Antibody-Builder [47] to generate antibody structure and obtained antigen's nuclear magnetic resonance (NMR) structure from the protein data bank. Docking between antigen and antibody was performed by ZDOCK

(<http://zdock.umassmed.edu/>), and RDOCK [48] was used to refine the result. Subsequently, they conducted experimental site-directed mutagenesis and *in silico* alanine scanning simultaneously. The results were combined to identify important residues involved in antibody-antigen binding which could refine the docking of the complex. Under the guidance of results gained previously, they redocked the antibody to antigen, then transmitted the results to *in silico* affinity maturation. The mutagenesis was performed by three *in silico* algorithms, Discovery Studio 2016, Schrödinger Biologics Suite 2016–3 (<https://www.schrodinger.com/glide>), and Rosetta separately. They integrated the results from three algorithms and obtained 20 potential variants. After the ELISAs (enzyme-linked immunosorbent assay) experiment, two of the mutants showed improvement in binding dissociation constant when compared with AB1. Cannon *et al.* also conducted an *in vitro* study which discovered the two mutants with improved K_d property and with a 2-fold and 4-fold decrease in IC_{50} . These results suggested that these novel computational methods could improve the affinity maturation of antibody design.

In addition to computational methodology, machine learning methods were also applied to improve the antibody affinity. Liu *et al.* [49] presented a high-capacity machine learning approach, named Ens-Grad, which can design CDR of human IgG without the known antigen structure through sufficiently modeling the biophysics of antibody-target interaction. Liu *et al.* used ranibizumab as an example and attempted to develop a new CDR-H3 sequence with better affinity for ranibizumab. They optimized the antibody sequence by employing a gradient-based optimization framework from an ensemble of neural networks in Ens-Grad. The result indicated that the sequences designed by machine learning method had equivalent or superior affinity for ranibizumab than the candidate from different phage display panning experiments. The best result from machine learning method had an EC_{50} of 0.29nM.

1.6 Aim

Currently, most of the antibody design tools are separately and there are no comprehensive antibody design protocols which can give instruction for users from the beginning to the end. It requires users to spend a lot of time to figure out a complete workflow of antibody design. Besides, there are still some challenges regarding computational methods that need to be addressed. In some circumstances, the 3D structures of antibodies are unavailable which limits the binding information of an antibody-antigen complex. Another challenge is the flexibility of antigens, some loops may be invisible, while others may appear. In addition, most of the antibody design web servers are commercial, which means that the users need to pay for the design work.

Hence, we combined powerful antibody engineering protocols together to propose a free *in silico* antibody design protocol-IsAb to address these challenges and provide detail instruction for the users from the beginning to the end. IsAb mainly focus on antibody redesign which modifies the antibodies properties based on existing antibodies. Thus, the operation of IsAb can facilitate the design of antibodies and potentially accelerate the development process.

The first step of our protocol is to use RosettaAntibody to solve the problem of unavailable 3D structures of antibodies. Then, RosettaRelax is applied to minimize the energy of protein structures by making the input conformations closer to the bound state and increase the accuracy of docking. To address the lack of binding information, the protocol will suggest performing two-step docking, including the global and local dockings. After acquiring the binding conformation, alanine scanning is applied to predict the hotspots (or key residues) of the antibody. Finally, computational affinity maturation is conducted to improve the properties of the existing antibody. To give step-by-step instruction, we went through this protocol using cemiplimab design as an example.

2.0 MATERIAL AND METHODS

2.1 Generating and Preparing the Structures

3D structures of an antigen or an antibody can be downloaded from the Protein Data Bank (<https://www.rcsb.org/>). Then PyMol can be used to extract the 3D structures of an antigen from its antigen-antibody (not the target antibody) complex. For antibodies whose 3D structures were unavailable, their Fv sequence was obtained from IMGT (<http://www.imgt.org/>) and submitted to ROSIE (<https://rosie.graylab.jhu.edu/>) to construct antibody 3D models. The RosettaAntibody function [50-54] in ROSIE was used to model the 3D structure of the antibody. For the modeling results, we chose the “Grafted-Relaxed-Model” provided by the web server, in which the “Grafted-Relaxed-Model” was relaxed during the modeling process. For structures of the antigen, we applied a Rosetta relax protocol (“relax.static.linuxgccrelease”) for preparations (**Appendix Figure 1**). Finally, we selected the lowest score decoy among the ten output results as the relaxed structure.

2.2 Antibody-Antigen Global Docking

We submitted the antibody Fv sequence to an antibody paratopes prediction database developed by Robin *et al.* [55] to predict the possible paratopes on the antibody (**Figure 4**). We chose the residues whose $\Delta\Delta G$ was larger than 1kcal/mol as possible paratopes. The global docking for an antibody-antigen was performed by ClusPro (<https://cluspro.bu.edu/login.php>), in

which the “Antibody Mode” in ClusPro [56-61] was selected for antibody docking. In the “Attraction and Repulsion” section, we input the possible paratopes predicted by the database developed by Robin *et al.* to the “attraction” box to increase the docking accuracy.

Enter your VH and VL sequences in the boxes below using IMGT numbering

Using IMPG/DomainGapAlign <http://www.imgt.org/3Dstructure-DB/cgi/DomainGapAlign.cgi>

Your VH (IMGT)
 EVQLLESQG.VLVQPGGSLRLSCAASGFTF...SNFGMTWVRQAPGKLEWVSGISGG..GRDTYFADSVK.GRF
 TISRDNSEKNTLYLQMNSLKGEDTAVYYCVKQGN...IYFDYWGQGLTVTVSS

Your VL (IMGT)
 DIQMTQSPSSLSASVGDSTITTCRASLSI.....NTFLNWWYQQKPKAPNLLIYAA.....SSLHGGVP.SRF
 SGSG...SGTDFTLTIRTLQPEDFATYYCQSSN...TPFTFGPGTVVDFR.

Convert VH to AHo & align with DG values

Convert VL to AHo & align with DG values

	A	CZ	DA	DB	DC	DD	DE	DF	DG	DH	DI	DJ	DK	DL	DM	DN	DO	DP
1	AHo	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119
2	VH_Frequency	####	####	####	####	99%	97%	91%	91%	81%	72%	56%	36%	21%	13%	6%	3%	2%
3	VH_mean_Energy	0.00	0.00	0.00	0.00	0.01	0.50	0.96	0.71	0.97	1.02	0.79	0.61	0.61	0.66	0.89	0.06	0.28
4	aa hot propensity	0.33	0.66	0.66	0.40	0.33	0.22	0.74	0.00	0.25	0.59	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	Your VH	V	Y	Y	C	V	K	W	G	N	I

Figure 4. Predicting possible paratopes from database.

The top figure is the input surface of the database. The figure below is the output result of the database.

2.3 Antibody-Antigen Local Docking

SnugDock [62] function (https://rosie.graylab.jhu.edu/snug_dock) on ROSIE was used to refine the possible binding poses generated from global docking. Then we chose “thorough mode” to perform the local docking and accepted the docking result based on whether the local docking formed a docking funnel. Once the docking was completed, we picked the lowest I_{sc} decoy. The

antibody sequence was compared to other antibodies to discover possible binding poses of the antibody-antigen complex. The sequence alignment was performed by EMBL “Clustal Omega” server (<https://www.ebi.ac.uk/services>).

2.4 Computational Alanine Scanning and Affinity Maturation

Rosetta alanine scanning protocol (AlaScan.xml, <https://github.com/Kortemme-Lab/ddg/>) was used to predict the possible hotspots (or key residues) on the antibody (**Appendix Figure 2**). The interface cutoff value was 5 Å. We excluded residues with $\Delta\Delta G$ lower than 1 kcal/mol for the result.

The computational affinity maturation was performed by affinity maturation protocol generated from Dr. Jens Meiler’s lab website. We first used “define_interface.py” to prepare a residue file (resfile) (**Appendix Figure 3**). Resfile is used to define which interface residues should be mutated. Then, we applied affinity maturation protocol (design.xml, <http://www.meilerlab.org/index.php/rosetta-tutorials>) to modify the antibody (**Appendix Figure 4**). Next, the scripts in Rosetta common were applied to analyze the results. The script of “compare_design_to_control.py” was utilized to analyze different metrics comparisons between the control and the design group (**Appendix Figure 5**). The residue energy breakdown of the control and the design group was calculated by the script of “per_residue_energies” and “residue_energy_breakdown” (**Appendix Figures 6, 7**). Lastly, we calculated the residues energy changing during the mutation and made the energy changing plot.

2.5 Molecular Dynamics (MD) Simulation and Molecular Mechanics/ Generalized Born Surface Area (MM/GBSA) Calculation

Four systems of cemiplimab-PD-1 with mutation at cemiplimab (H: G100A, N101G, I102T, and Y103A) were used to perform molecular dynamics (MD) simulation. These systems were put in a 0.15M NaCl solution with a cubic water box. The same force fields or parameters [63-65] described in our previous publications [66-69] were applied to both the cemiplimab and PD-1 receptors and water molecules (TIP3P model). MD simulation was carried out using the AMBER18 [70] software package. The MD system was first relaxed by a set of minimizations that removed possible steric clashes. There were three phases for the subsequent NPT (constant particle number, pressure, and temperature) MD simulations: the relaxation phase (0.2 nanoseconds for each temperature from 50 to 250 K at a step of 50 K), the equilibrium phase (5 nanoseconds, 298 K), and the sampling phase (50 nanoseconds). Integration of the equations of motion was conducted at a time step of 1 fs for the relaxation phase and 2 fs for the equilibrium and sampling phases.

100 snapshots were evenly selected from the sampling phase (40-50 ns) for MM/GBSA binding free energy decomposition analysis. For each MD snapshot, the molecular mechanical (MM) energy (EMM) and the MM/GBSA solvation free energy were calculated without further minimization [71, 72]. The interaction energies between each residue and its ligand were calculated with the solvent effect being taken into account using a MM/GBSA solvation model [73].

3.0 RESULTS

3.1 Workflow of Antibody Design Protocol

The proposed procedure of *in silico* antibody design protocol is described in **Figure 5**. Briefly, the *first step* is to obtain the structures of both antibody and antigen, which can be downloaded from the Protein Data Bank. Users who already have the antibody's structures can jump directly to the *third step*. Otherwise, users need to retrieve the antibody sequence from IMGT. *Step 2*, the RosettaAntibody web server is used for antibody modeling. The RosettaAntibody will help users generate the antibody structure through its sequence and output a relaxed structure. *Step 3*, RosettaRelax protocol is applied to refine the structures of antibody and antigen. RosettaRelax uses rosetta force-field to make simple all-atom refinement of structures, driving a starting structure towards its native state and improving docking accuracy. *Step 4*, the antibody is docked to the antigen. If the binding information between the antibody and the antigen is available, the user can jump to the *fifth step*. With limited structural information, the antibody and the antigen structures can be submitted to ClusPro to perform global docking. The ten most populated clusters will be outputted as the results, which give users several possible binding poses of the antibody-antigen complex. *Step 5*, based on the global docking results, a local refined docking is carried out by SnugDock, which allows flexibility of interfacial side chains and CDR loops. SnugDock refines the possible binding poses and outputs the final antibody-antigen complex. *Step 6*, based on the refined antibody-antigen complex, alanine scanning is performed: (1) it first mutates the residues on antibody and antigen interface to alanine, then (2) it calculates the residues' energy change during the mutation to find out the hotspots on the antibody, which can facilitate the following

antibody design or give insight into future studies. *Step 7*, antibody affinity maturation protocol is applied to design the antibody. Based on the Rosetta scoring function, the antibody affinity maturation protocol will lead to a result of the best mutated antibody that has better affinity and stability than the original one. All the code (relaxation, alanine scanning, affinity maturation, define interface, compare control to design, and residue calculation) used in IsAb protocol was modified by us to suit the specific project.

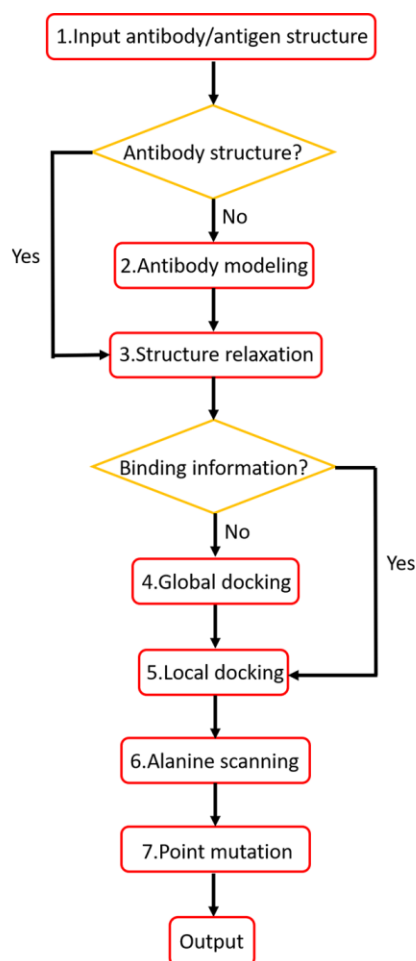


Figure 5. The workflow of the IsAb.

Generation of the structure from protein data bank and modeling the antibody structure by RosettaAntibody. Relaxing antibody and antigen structures and performing two steps docking. Global and

local dockings were performed by ClusPro and SnugDock. After generating final binding pose, Rosetta alanine scanning was utilized to predict hotspots on the antibody. Finally, *in silico* affinity maturation was used to design the antibody and improve its properties.

In the following section, we select several key steps to illustrate the detailed procedures in our protocol.

3.1.1 Generating and Preparing the Structures

Before performing antibody docking, we need to have structures of both antigen and antibody. In this step, we can obtain the 3D structure of an antigen from the Protein Data Bank. To generate the 3D structure of an antibody, we need to find out its sequence from the IMGT database.

After retrieving the sequence, we used RosettaAntibody to construct the Fv region of the antibody. RosettaAntibody models the antibody structure in two main steps. First, the BLAST-based method is used to search homologous templates for framework regions and CDR loops for the input sequences. Then, RosettaAntibody inserts the template CDRs onto template frameworks and optimizes side chains of all residues in the model. In the second step, RosettaAntibody performs a low-resolution and a high-resolution phase. In the low-resolution phase, the candidate conformation of the CDR H3 loop will be found by cyclic coordinate descent and fragment assembly [74]. The high-resolution phase will repeatedly optimize and minimize the side chains [75]. Also, the orientation of heavy and light chains and backbone torsion angles of CDR will be modified. After the second step, it will generate one thousand potential structures.

According to experimental results [76] of the protein-protein interaction, the protein-protein docking always experiences backbone conformational changes to obtain minimum-energy conformation in the correct binding site [77, 78]. Therefore, there are several angstroms of root-mean-square deviation (RMSD) of their bound structures compared with the unbound state (the unbound state is easy to access, so users may use the unbound state as input structure in most of the time). Moreover, the protein docking accuracy relies on the quality of input structures. The closer the starting structure is to the bound state, the higher will be the likelihood of success will be [79]. Therefore, we implemented Rosetta relax protocol (FastRelax) [80] to lower the energy of the starting structure and made the antibody or antigen structures closer to their native state.

FastRelax is composed of several rounds of energy minimization in torsion angle space and all atoms are repacking while ramping up the van der Waals repulsive weights. This can deal with clashes caused by large repulsive forces and avoid the protein unfolding. In the repacking step, side-chain conformations are randomly chosen from a rotamers library by Monte Carlo simulated annealing run [81].

3.1.2 Antibody-Antigen Global Docking

Before antibody-antigen docking, the binding information of the antibody and antigen is essential. The best way to access this information is to perform global docking. Global docking assumes a spherically general structure of proteins and rotates the ligand protein around the receptor protein. It randomizes the starting position of unbound proteins in every run and outputs possible binding poses of the antibody-antigen complex.

ClusPro is a protein docking server that can conduct the antibody-antigen docking. At the beginning of the antibody-antigen docking, a Fast Fourier Transform (FFT) correlation approach

[82] based docking program PIPER [83] is used to conduct a rigid-body docking step. During rigid-body docking, the antigen moves around the antibody, and their interaction energy is calculated by FFTs. These results will provide 1000 lowest energy docked structures. Specifically, to address the problem of asymmetry between two sides of the antibody and the antigen interface [84], ClusPro integrates ADARS, which is the asymmetric DARS-type [85] potential in the energy function of PIPER. After docking, ClusPro calculates the pairwise interface root mean square and uses it to cluster retained conformations. Lastly, the CHARMM potential is used to minimize the energy of structures by removing the steric crash and to output the structures at the centers of the ten most populated clusters [86].

3.1.3 Antibody-Antigen Local Docking

To refine the results from global docking, local docking is performed. RosettaDock [87] is one of the best local docking methods, which consists of a Monte Carlo-based multi-scale docking algorithm. The algorithm allows both intramolecular and interfacial flexibility in the antibody during docking with high accuracy. Built on RosettaDock, SnugDock is a specific algorithm for antibody docking, and it incorporates the sampling characteristic of RosettaAntibody.

SnugDock runs local docking in two stages: a low-resolution stage and a high-resolution stage. During the low-resolution stage, SnugDock adds to RosettaDock by additionally perturbing and minimizing the CDR H2 and H3 loops. In the high-resolution phase, the algorithm will randomly choose one trail among the five in a move set with indicated frequencies. The move set is composed of 1,2) rigid body minimization and transformation of VL, VH, or antibody-antigen orientation; 3,4) minimization and perturbation of CDR H2 and H3 loops; and 5) gradient-based minimization of the CDRs H1, L1, L2, and L3 backbones. Before each minimization and Monte

Carlo decoy acceptance decision, the high-resolution stage will pack sidechains. After that, SnugDock will sort decoys by interface score and output docked complex.

To evaluate docking performance and structural accuracy, we followed the measurement rules defined by Rosetta and used its metrics [88]. Due to the energetic differences within small backbone changes during local docking, it is difficult to capture total energy accurately. Interface score (intermolecular energy/ I_{sc}) [89] can provide the best measurement of the structure as the metric to analyze the result of local docking. I_{rmsd} is defined as interface root mean square deviation of heavy atoms in the interface residues between the reference structure (input structure) and result structure, which describes the atomic accuracy of the interface between the two structures. According to the Critical Assessment of Protein Interactions (CAPRI) [90], a decoy with $I_{rmsd} < 1.0 \text{ \AA}$ is defined as high quality, $1.0 \text{ \AA} < I_{rmsd} < 2.0 \text{ \AA}$ is defined as medium quality, and $2.0 \text{ \AA} < I_{rmsd} < 4 \text{ \AA}$ is defined as acceptable quality. Thus, the decoy with an I_{rmsd} value lower than 4 \AA will be referred to as the “near-native” structure, which means that the decoy has high structural accuracy. The most robust evaluation of success in a docking result is the presence of a “docking funnel”, in which the “near-native” decoy (low I_{rmsd}) has lower energy (low I_{sc}) than non-native decoys [80]. In this case, we created an “ I_{sc} versus I_{rmsd} ” plot to count the number of decoys with an I_{rmsd} value lower than 4 \AA among the top five I_{sc} (N_5) and identified a docking simulation which has a docking funnel if it had $N_5 \geq 3$.

3.1.4 Computational Alanine Scanning

To predict energetically important residues [91] in antibody-antigen interfaces, computational alanine scanning protocol [92] is a great choice. Alanine scanning protocol automatically defines interface residues of the complex and conducts *in silico* alanine mutations

on the interface. After mutation, the protocol uses free energy function to calculate the wild type (original complex) and mutated complex's ΔG energy separately. Then the difference of ΔG between wild type and mutated complex is calculated. Finally, the protocol outputs the predicted changes in binding free energy ($\Delta\Delta G_{\text{binding}}$) for all alanine mutations. Since we were under the help of alanine scanning protocol developed by Kortemme *et al.*, according to their definition, residues whose $\Delta\Delta G_{\text{binding}}$ was greater than 1 kcal/mol could be the hotspots [92]. Their experiments proved that the protocol they developed and the criteria they defined were reliable in accurately predicting hotspots. Therefore, we followed the $\Delta\Delta G$ definition set up by Kortemme *et al.* According to the classification, hotspots and neutral residues were defined as $\Delta\Delta G_{\text{binding}}$ more or less than 1 kcal/mol during the alanine mutation, respectively [93].

3.1.5 *In silico* Affinity Maturation

Computational antibody affinity maturation protocol can theoretically increase the affinity of an antibody to an antigen. First, an *in silico* affinity maturation method needs to be assigned to a mutated site on the antibody-antigen complex. Then, a rotamer library will be used to explore 20 amino acids at each position [94]. The backbone will experience a small change [95] during sidechain sampling. The sampling step will generate different types of mutations, and the mutations are scored by the Rosetta score function *REF2015* [96]. The output structure decision is based on the Metropolis criterion to output the best-mutated structure, which has better affinity and stability among other mutations.

In the following section, to test the reliability of IsAb, we applied it to redesign anti-lysozyme monoclonal antibody D44.1 and compared our designed results with the previously reported data.

3.2 Antibody D44.1 Redesign

3.2.1 D44.1 Structure and Binding Pose Generation

The D44.1 sequence was generated from the IMGT database (sequence number: 1mlc), and the RosettaAntibody function was used to obtain its 3D structure. The RMSD value between modeling D44.1 and crystal one (PDB: 1mlc) was 0.342 Å. A lysozyme structure was generated from the protein data bank (PDB: 1mlc). After minimizing the lysozyme structure through RosettaRelax, D44.1 and lysozyme structures were submitted to ClusPro. The paratopes were predicted through the database developed by Robin *et al.* Combining literature research [97] and the Immune Epitope Database (IEDB) epitope prediction result [98], we chose residues 41-51 and 66-71 of lysozyme as our epitopes. Both possible paratopes and epitopes were entered into ClusPro to increase the docking accuracy. After sequence comparison with other lysozyme antibodies, we noticed that the D44.1 heavy chain and light chain sequence shared 84.5% and 58.1% similarity with HyHEL-5 sequence, respectively. Therefore, we speculated that the binding pose of D44.1-

lysozyme may be similar to HyHEL-5-lysozyme (**Figure 6**). We chose the pose that

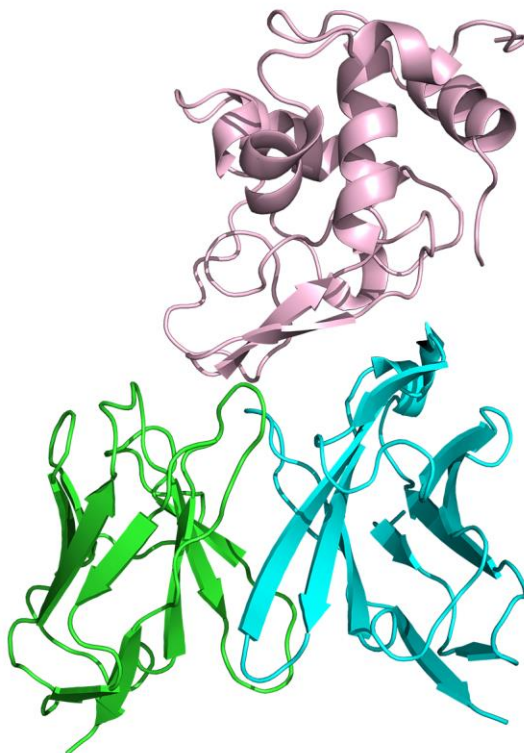


Figure 6. The crystal structure of HyHEL-5-lysozyme complex.

The structure is retrieved from PDB 1bql. Light chain colored with cyan, heavy chain colored with green, and lysozyme colored with pink.

was similar to HyHEL-5-lysozyme complex (PDB: 1bql) among the top ten global docking results. Then, SnugDock was applied to refine the decoys from global docking. A docking funnel was formed in the local docking results, which meant local docking was successful (**Figure 7**). The lowest I_{sc} decoy among the top 5 I_{sc} results was picked as the final docking pose of the D44.1-lysozyme complex. The RMSD value of predicted D44.1-lysozyme and its crystal structure (PDB:

1mlc) was 0.967 Å, indicating that the predicted binding pose could be utilized in the later step (Figure 8, Appendix Table 1).

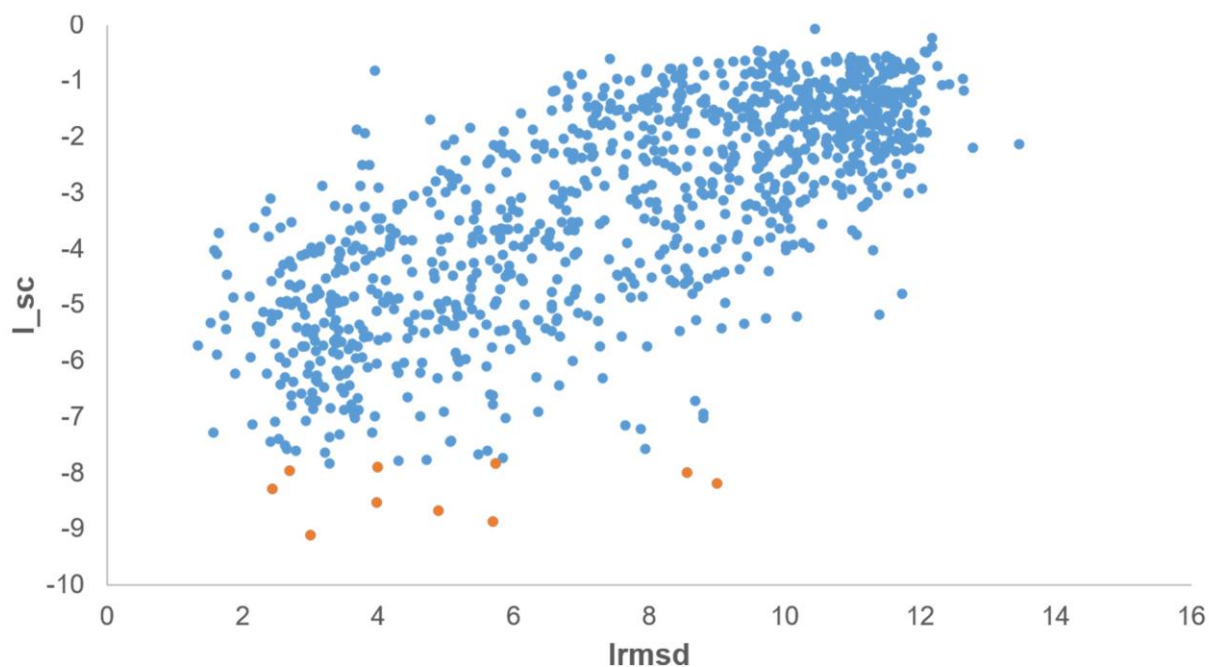


Figure 7. The I_{sc} vs Irmsd plot of D44.1-lysozyme local docking.

The orange dots are the top10 I_{sc} decoys. The local docking result of D44.1-lysozyme complex, five decoys

Irmsd among top10 I_{sc} are smaller than 4 Å and forms docking funnel.

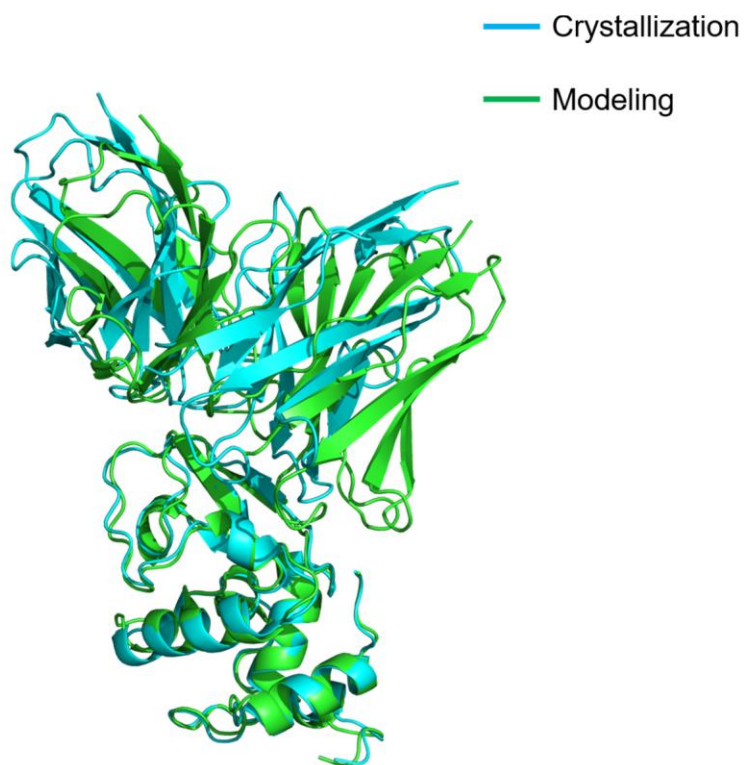


Figure 8. Superimposition of modeling and crystallization D44.1-lysozyme structures.

The structure in cyan is crystal D44.1-lysozyme (PDB: 1mlc). The structure in green is modeling D44.1-lysozyme complex.

3.2.2 D44.1 Affinity Maturation

Once we got the D44.1-lysozyme structure, we applied Rosetta affinity maturation protocol to design the D44.1 and compared the result with Lippow *et al.*'s works [99]. Lippow *et al.* had proposed 9 potential CDR residues point mutations, and the experiment had proved that 6 of them could increase D44.1-lysozyme complex binding affinity. Since IsAb focused on interface residues and not all the CDR residues located on the interface, the residues we mutated were little different

from Lippow *et al.* For our results, 5 of the positions we mutated were the same as Lippow *et al.* Among the 5 positions, 3 of them (L: N92A, H: T28D, T31A) have the same mutation with the reported data. Within three mutations, T31A has been proved by the experiment that it could not enhance binding affinity. In our prediction, N92A and T28D both could lower the total score of the complex. Meanwhile, N92A could increase the affinity and stability of the structure (both binding density and energy were 1.01 folds lower than the original one), which was the same as experimental results. Moreover, N92A was the best mutation validated by the experiment. We performed double mutation at positions L: N32, N92. The IsAb affinity maturation result was L: N32A, N92A, which had one mutation that was the same as Lippow *et al.*'s work (N32G, N92A) (**Appendix Table 2**). From **Figure 9a**, we found that because the score of most design models were close to each other, the average score value of the design group was lower than that of the control group, indicating that the structure of the design group may be more energetics-favorable than that of the control group. The binding density was negatively correlated with binding affinity [100]. Based on **Figure 9b**, we noticed that the average binding density value of the design group was lower than that of the controlled, which specified that the designed complex may have better binding affinity than the controlled. From **Figure 9c**, we could see that the average binding energy of the control group was higher than that of the design group, which could draw the conclusion that the design group had better stability. The outcome that IsAb designed D44.1 had better affinity was consistent with previously reported data. The work we have done indicated that the IsAb protocol could overlap with better antibody designs and showed the reliability of our protocol.

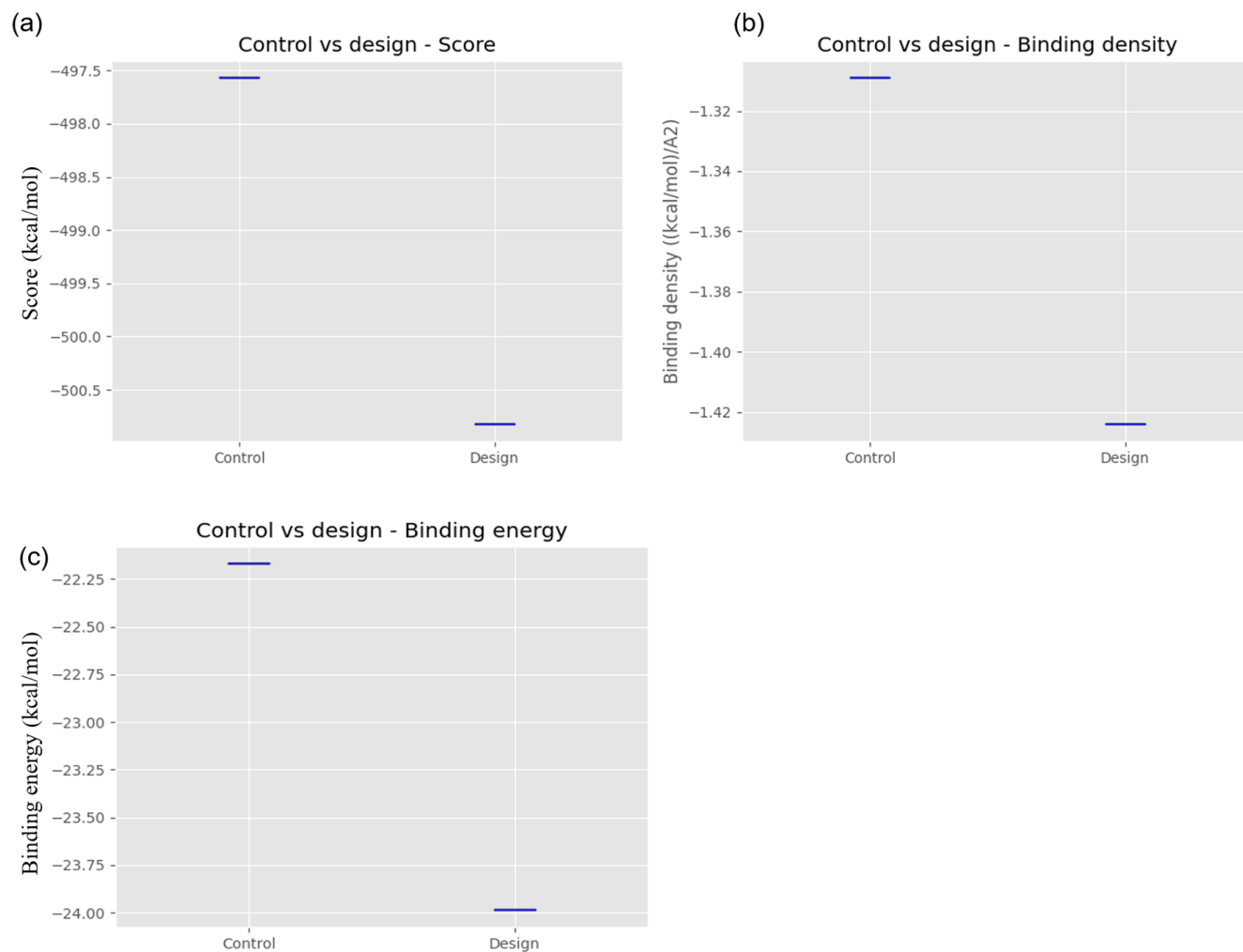


Figure 9. Boxplot of comparison of three metrics between antibody D44.1 control and design groups.

(a) Comparison of total score between D44.1 control and design groups. (b) Comparison of binding density between D44.1 control and design groups. (c) Comparison of binding energy between D44.1 control and design groups.

3.3 Case Study: Cemiplimab Design Procedure

Cemiplimab is an FDA proved PD-1 checkpoint inhibitor (**Figure 10**) used to treat cutaneous squamous cell carcinoma (cSCC). However, cemiplimab does not have a published 3D crystal or cryo-EM structure. Moreover, the structures of PD-1 resolved in different conditions reveal substantial flexibility (**Figure 11**). According to research, PD-1 C'D, N, and FG loops may experience huge changes during interactions with different antibodies. C'D and N loop are too flexible to be resolved. For example, when binding with pembrolizumab [101] (**Figure 12a**), the PD-1 N loop is missing, and the C'D loop dominantly affects the binding affinity of the complex. However, when binding with nivolumab [102] (**Figure 12b**), the C'D loop becomes invisible, and N loop dominates the binding affinity with nivolumab. The C'D and N loops are both invisible when binding with toripalimab [103] (**Figure 12c**). On the other hand, most literature about cemiplimab focuses on clinical data, which means that the interaction information is limited. Therefore, the cemiplimab example is under the extreme situation where the antibody has no structure and no binding information.

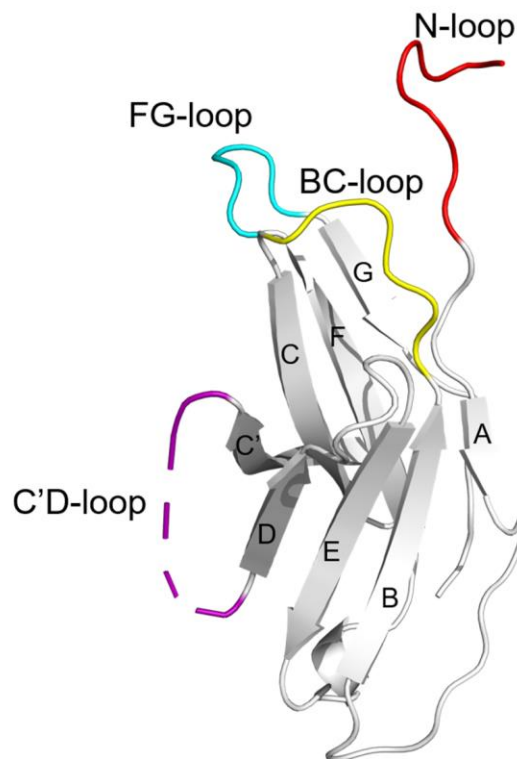


Figure 10. The crystal structure of PD-1 (extracted from PDB:6jbt).

The colored loops are the important loops on PD-1. N loop marked as red, FG loop marked as cyan, BC loop marked as yellow, C'D loop marked as purple.

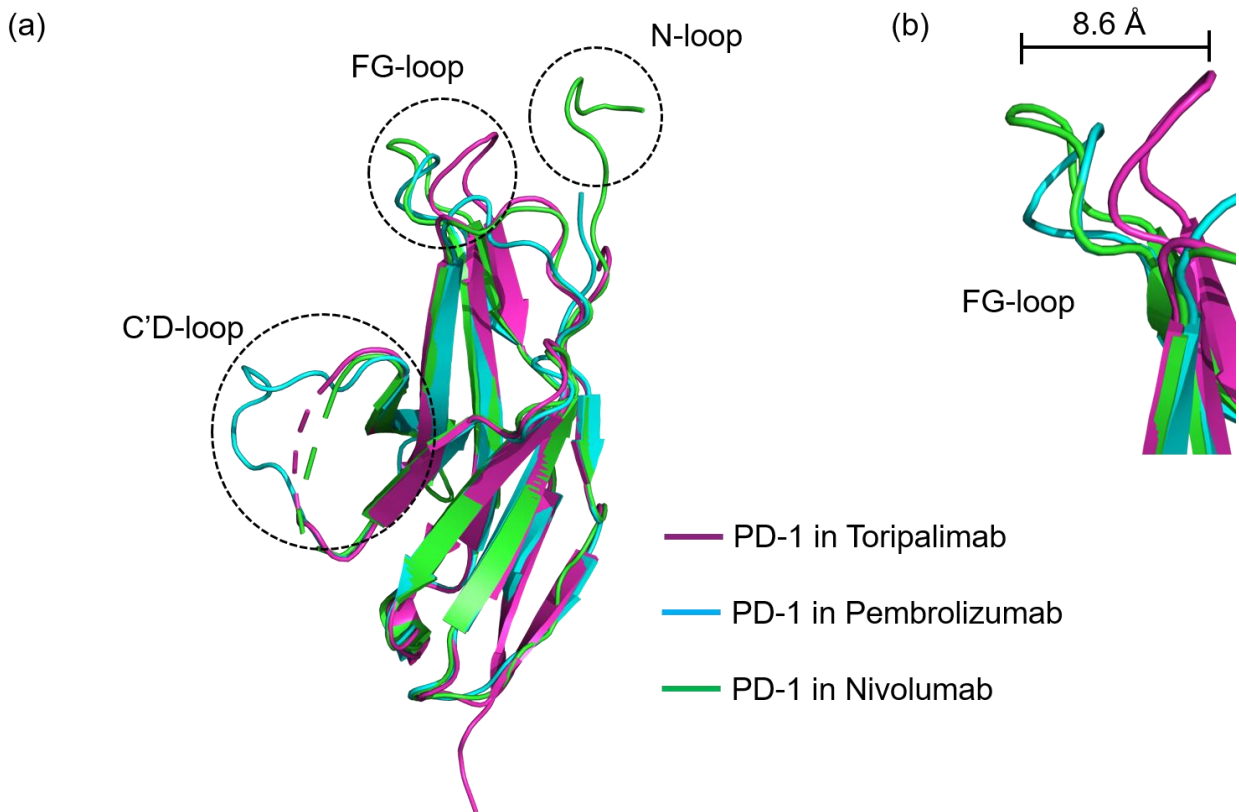


Figure 11. Superimposition of PD-1 structures from three PD-1/antibody complexes.

(a) PD-1/toripalimab (purple), PD-1/pembrolizumab (cyan), and PD-1/nivolumab (green). When binding to different antibodies, some PD-1 C'D and N loops are invisible. (b) Comparison of the PD-1 FG loop from the complex structures. The PD-1 FG loop shifts 8.6 Å on the binding to toripalimab, pembrolizumab, or nivolumab.

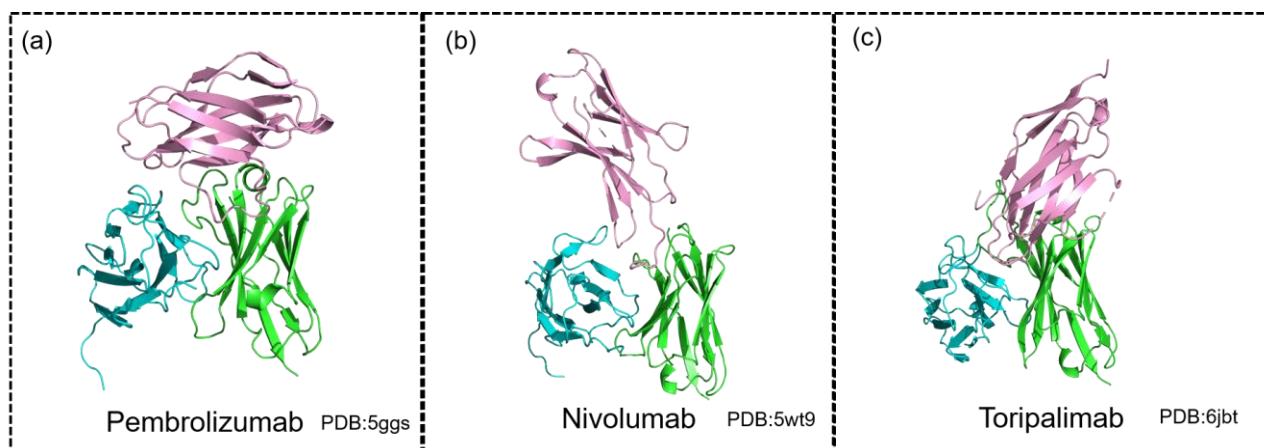


Figure 12. Three crystal structures of pembrolizumab, nivolumab, and toripalimab complex with PD-1.

Light chain colored with cyan, heavy chain colored with green, and PD-1 colored with pink. (a) Crystal structure of pembrolizumab/PD-1 complex. (b) Crystal structure of nivolumab/PD-1 complex. (c) Crystal structure of toripalimab/PD-1 complex.

3.3.1 Generating and Preparing the PD-1 and Cemiplimab Structures

For the construction of the cemiplimab-PD1 complex, there is not much information about the binding between cemiplimab and PD-1. First, we collected the crystal structures of PD-1-antibody complex from the protein data bank and generated three representative PD-1 structures, named **PD1v1**, **PD1v2**, and **PD1v4** (**Figure 13**), by extracting the structures from their complexes (PDB: 5wt9, 5ggs, and 6jbt). We then docked these three PD-1s to cemiplimab, respectively. Taken docking convenience and mechanism of antigen-antibody binding into consideration, we used the antibody Fv region to dock with PD-1s. PyMol was used to extract the crystal structure of PD-1 from its complex. Cemiplimab's Fv sequence was obtained from IMGT and submitted to ROSIE. The RosettaAntibody function was used to model the 3D structure of cemiplimab. For the modeling results, we chose the "Grafted-Relaxed-Model" provided by the web server, in which

the “Grafted-Relaxed-Model” was relaxed during the modeling process. We applied Rosetta relax protocol (“relax.static.linuxgccrelease”) for the preparation of the antigen structure. We selected the lowest score decoy among the ten output results as the relaxed structure.

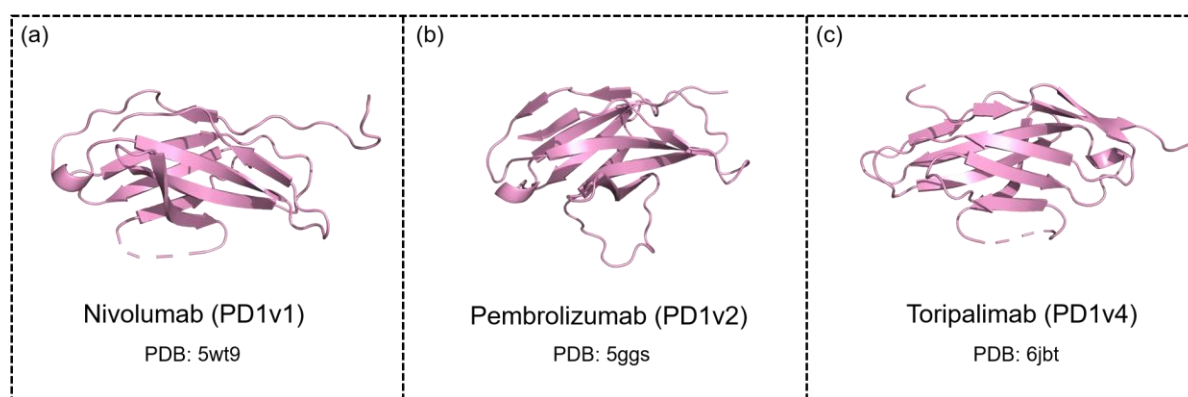


Figure 13. Three most representative PD-1s structures.

(a) PD1v1 structure was extracted from PD-1/nivolumab complex. (b) PD1v2 structure was extracted from PD-1/pembrolizumab complex. (c) PD1v4 structure was extracted from PD-1/toripalimab complex. RMSD value between PD1v1 and PD1v2 is 0.445 Å. RMSD value between PD1v1 and PD1v4 is 0.395 Å. RMSD value between PD1v2 and PD1v4 is 0.393 Å.

If the crystal structures of proteins are available, the web server can directly relax the crystal structure and skip the modeling step. In some situations, antigens like PD-1 experience huge changes during the interaction with antibodies (structure missing or appearing). In this case, users can do the literature search to conclude the representative structure types of the antigen, then input these structures to docking, and determine the conformation by analyzing the results in later steps. For convenience, we carried out the modeling and docking by the web server in this case study. The RosettaAntibody and SnugDock functions can also be conducted on Rosetta software.

If Rosetta software is used, it can implement an ensemble of the ten lowest-scoring structures obtained from antibody modeling as the input. For protein relaxation, KIC protocol [104] or molecular dynamic (MD) simulation [105] is also an alteration method. Except Rosetta, there are some other databases that can provide antibodies information like sequence and structure data or help users to model a 3D structure of antibodies. For example, the abYsis [106] and abYbank [107] functions in Antibodies database can provide antibodies information for the users and help them to analyze the sequence information. At the same time, users can use abYmod function in the database to generate an accurate 3D structure of antibodies. The other powerful database is SAbDab [108, 109] which contains antibodies data and modeling function. All these databases can help the user to obtain basic resources of antibodies engineering.

3.3.2 Cemiplimab-PD-1 Global Docking

Since there is no experimental data for Cemiplimab-PD-1 complex, we first used global docking to generate the epitopes information and possible binding poses. To increase the accuracy of global docking, we submitted cemiplimab sequence into the paratopes prediction database developed by Robin *et al.* [55] to predict the possible paratopes on the antibody. From the results shown in the paratope prediction database, we chose the residues whose energy contribution was larger than 1kcal/mol as possible paratopes. Then, we input three possible PD-1s structures and predicted paratopes in ClusPro to perform global docking. The three criteria from the literature concluded that the PD-1 FG loop would form hydrogen bonds with antibodies, the C'D or N loop would form hydrogen bonds with antibodies when they appeared, and the antibodies structure would not experience a significantly change during docking. Following these criteria, we selected possible binding poses from the top 10 score decoys. Each docking trait will select one possible

binding pose as the local docking input. A total of three results were selected and named as **models1, 2, and 3 (Figure 14)**, respectively.

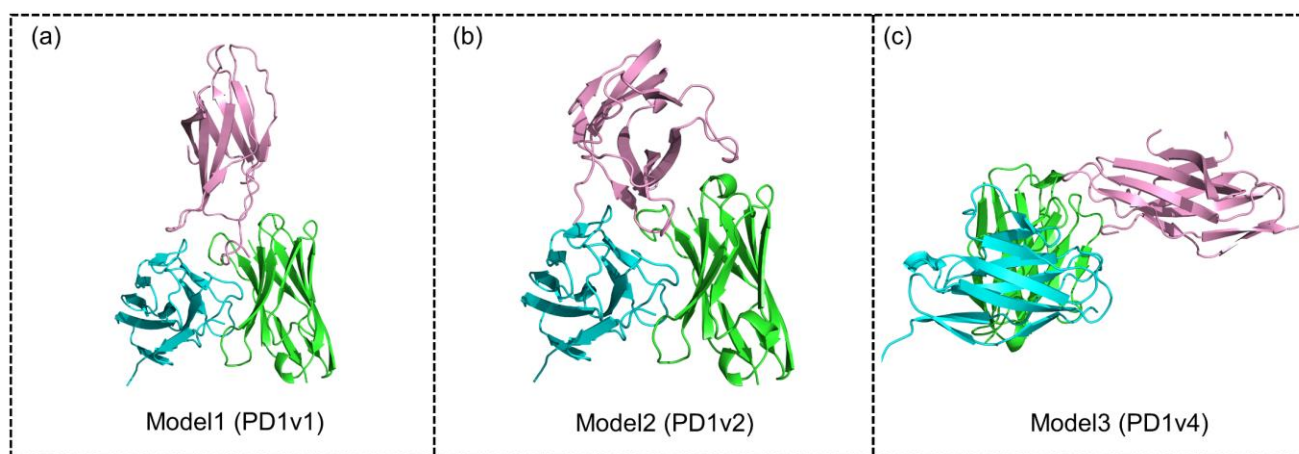


Figure 14. Three possible binding poses generated from global docking.

(a) Model1 is from PD1v1 structure docking result. (b) Model2 is from PD1v2 structure docking result.

(c) Model3 is from PD1v4 structure docking result.

Global docking is used when binding information of the antibody-antigen is not clear. If the paratopes and the epitopes are already known, the starting pose for local docking can be built by PyMol. The information of paratopes and epitopes can be acquired from experiments, such as point mutation, epitopes mapping, and alanine scanning. For virtual methods, Parapred [110], Antibody i-patch [111], the webserver like Immune Epitope Database and Analysis Resources (IEDB), and Robetta [112] are available methods to predict paratopes and epitopes. Although the

information of paratopes and epitopes is not essential for global docking, and global docking can be performed without this information, it can improve the docking accuracy if provided in the docking process. The more accurate the prediction of paratopes and epitopes is, the better-quality the local docking starting conformation will have. In the cemiplimab case, instead of only relying on the docking score, we used the criteria concluded from the literature to filter global docking results, which was more precise in picking the right conformation for the complex. The screening criteria can be obtained from the reporting experiments such as key residues. The binding rules of an antibody-antigen complex can be obtained by comparing the binding structure of the antigen with different antibodies if there is no experimental data available. In some situations, the antigen may be a novel one and does not have an existing antibody. In this case, we suggest users choose the top rank complex as the global docking result.

3.3.3 Cemiplimab-PD-1 Local Docking

To refine the results from global docking, the SnugDock function in ROSIE web server was used to conduct local docking. The criteria to accept the docking result was that local docking formed a docking funnel ($N_5 > 3$). If the docking was successful and the lowest I_{sc} decoy met the three criteria mentioned in the global docking step, then the lowest I_{sc} decoy will be the final local docking result. Following these rules, **models1** and **2** both had three decoys among the five top I_{sc} decoys that were better than acceptable quality (**Figure 15a, b**). However, N_5 of **model3** was smaller than three, meaning that the local docking of **model3** failed. To sum up, local dockings of **models1** and **2** were successful, and we filtered one possible binding pose from the results of **models1** and **2**, respectively (**Appendix Tables 3, 4**).

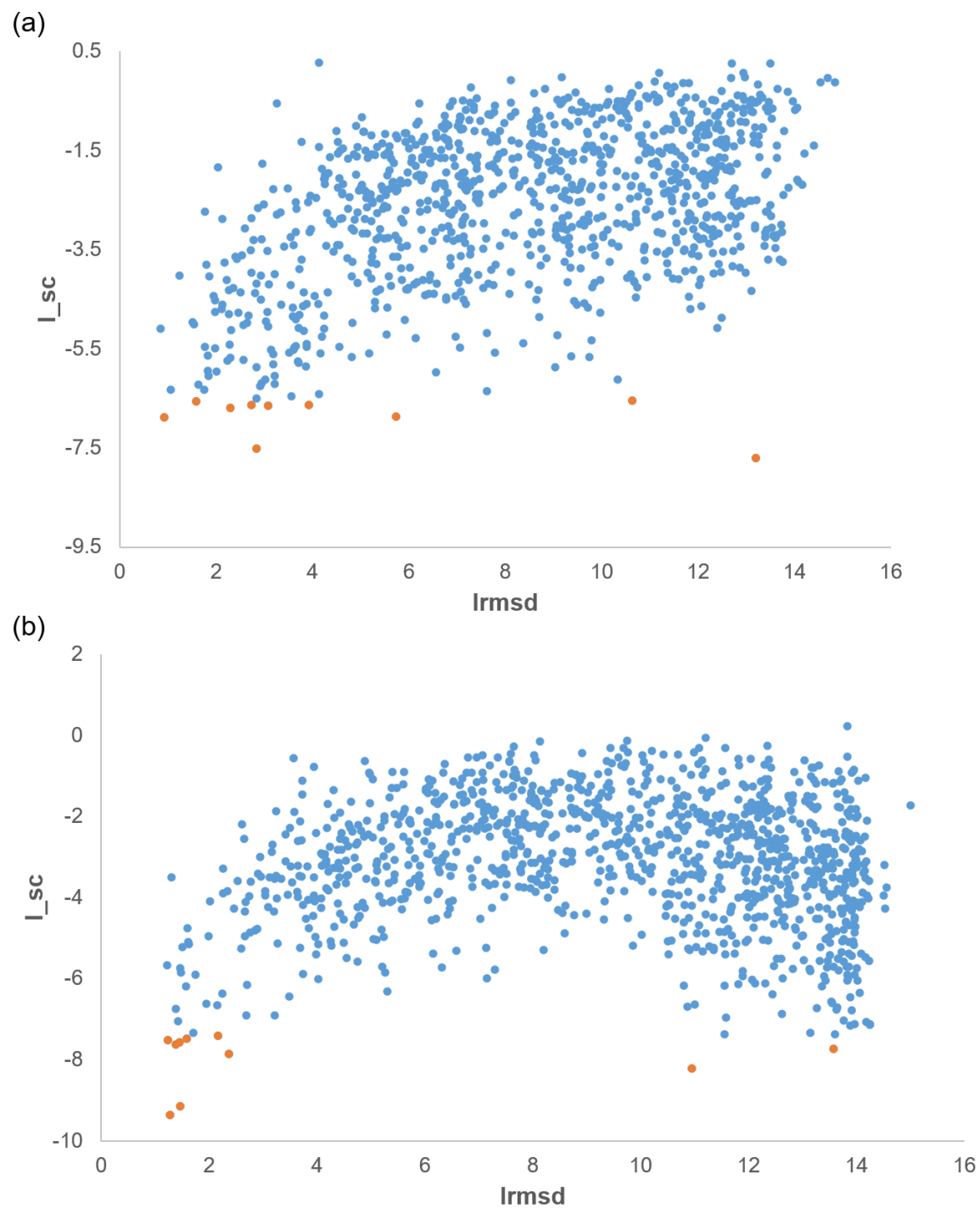


Figure 15. The I_{sc} vs $Irmsd$ plot of models1 and 2 local docking.

The orange dots are the top10 I_{sc} decoys. (a) The local docking result of model1, seven decoys Irmsd among top10 I_{sc} are smaller than 4 Å and forms docking funnel. (b) The local docking result of model2, eight decoys Irmsd among top10 I_{sc} are smaller than 4 Å and forms docking funnel.

We compared the sequence of cemiplimab with published PD-1 antibodies structure. The result showed that the sequence of cemiplimab shared a high similarity to nivolumab (PDB: 5wt9), including 64% similarity on light chain and 76% similarity on heavy chain, especially its HCDR3 which shared a higher sequence similarity than with the other PD-1 antibodies (**Appendix Table 5**). According to the report, heavy chain has a higher binding frequency and contributes higher energy to the binding with antigen [55]. **Model1** binding pose was similar with nivolumab-PD-1 binding pose. Therefore, we chose **Model1** as our final binding pose.

Structural analysis revealed that PD-1 BC, N, FG, DE loops were involved in the interaction with cemiplimab. The complex interface included 9 hydrogen bonds. Unlike binding to nivolumab, PD-1 N loop did not act as a dominating role in binding with cemiplimab. Instead, the BC loop contributed lots of hydrogen bonds in the interaction with cemiplimab (3 out of 9). The residues of the BC loop (N58, T59, S60) formed three hydrogen bonds with cemiplimab HCDR1 (N31) and HCDR3 (N101) (**Figure 16**). Although the N loop did not act as a dominant role in binding, it still contributed many hydrogen bonds (3 out of 9). One residues of N loop (R30) formed three hydrogen bonds with heavy chain (CDR2 D57, and framework region Y59) (**Figure 16**). **Figure 16** shows that heavy chain N31 and N101 formed two hydrogen bonds with PD-1, respectively. From the calculation of each residue energy contribution, we knew that the residues Y59, D57, and N31 from heavy chain made a large energy contribution (-3.697 kcal/mol, -2.721 kcal/mol, -2.118 kcal/mol) to the cemiplimab-PD-1 complex. Also, the hydrogen bonds between Y59, D57, N31 on cemiplimab and R30, N58/N102 on PD-1 were strong, with a distance of 2.6

Å, 2.9/3.0 Å, and 3.4/3.0 Å, respectively. In addition, PD-1 FG and DE loops contributed to the remaining hydrogen bonds. FG loop (S127) formed one hydrogen bond with HCDR3 (N101), and DE loop formed two hydrogen bonds with HCDR1 (N31, T28). On the other hand, P28 on N loop and R104 on BC loop made strong hydrophobic interactions (-1.946 kcal/mol and -2.152 kcal/mol) with LCDR3 and HCDR1 (**Figure 17**).

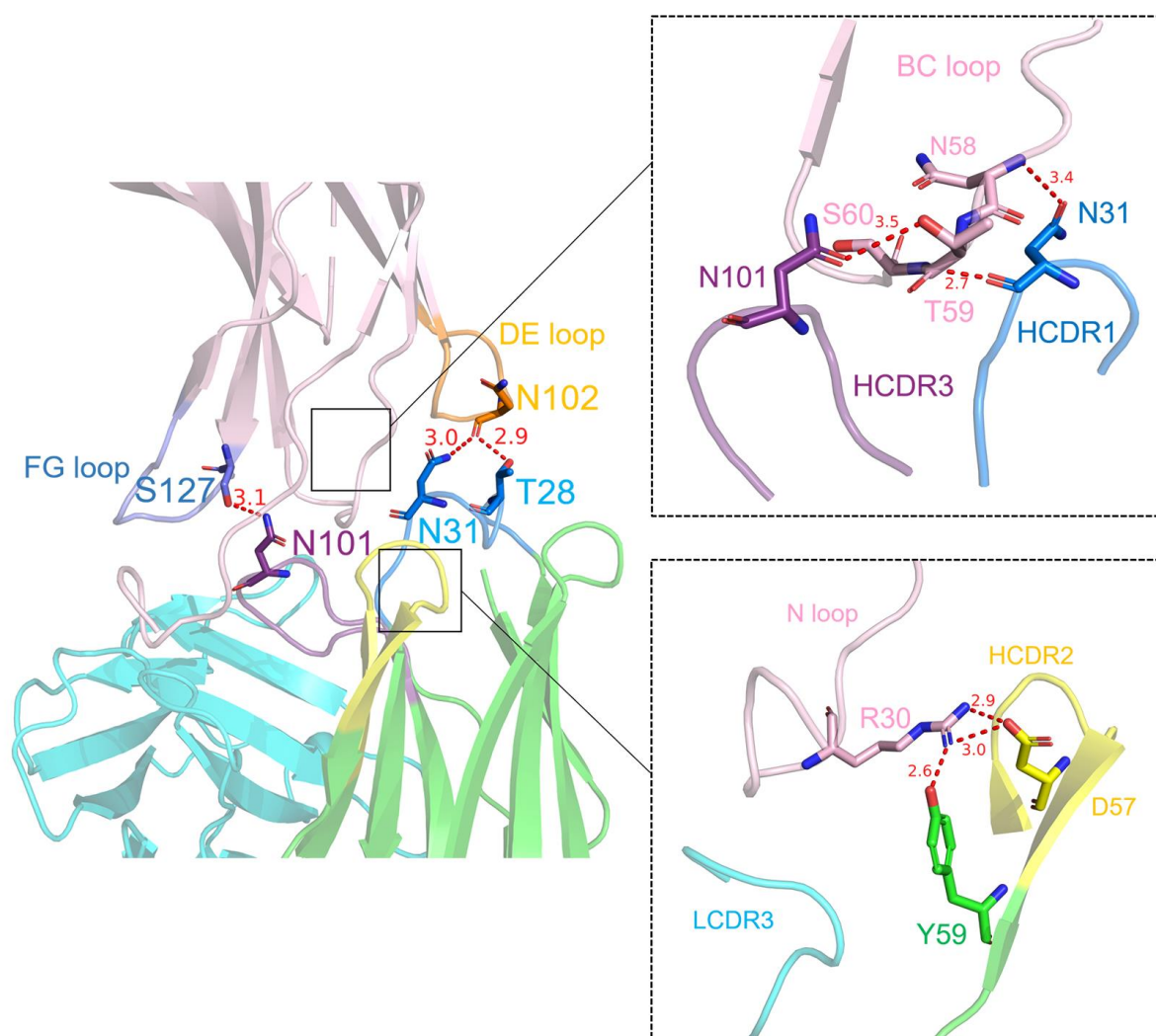


Figure 16. Detailed hydrogen bonds in cemiplimab/PD-1 complex.

Red dash line is hydrogen bond. The square frames show the detailed interactions between N, BC loops and cemiplimab.

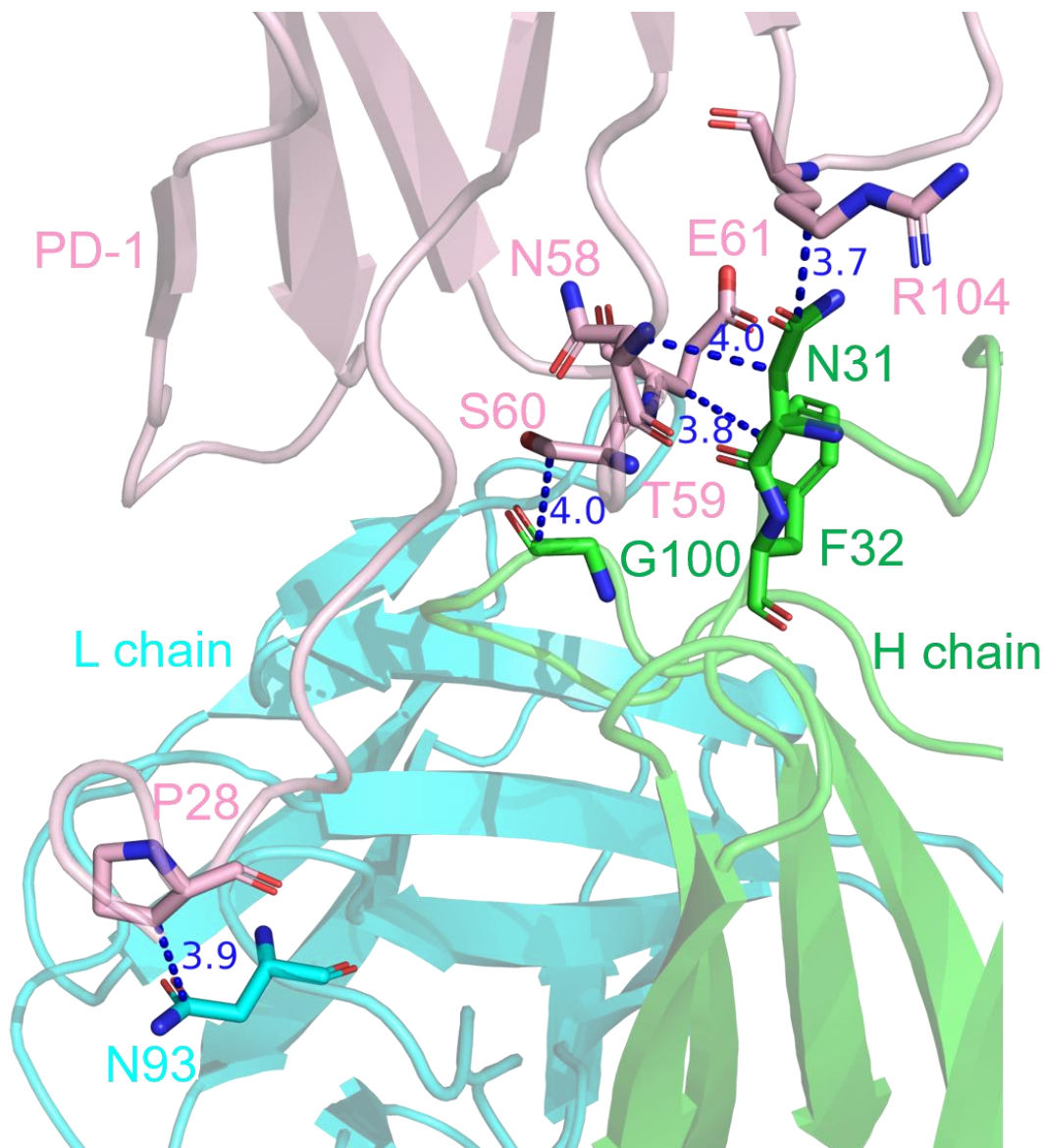


Figure 17. Hydrophobic interaction in cemiplimab/PD-1 complex.

Blue dash line is hydrophobic interaction (C-C). The distances between PD-1 P28/LCDR N93 and PD-1 R104/HCDR N31 are short.

There are several sources that can be used as inputs for local docking, such as the crystal structure of proteins, the ensemble of the ten lowest scores of RosettaAntibody homology models, or the Web Antibody Modeling (WAM) model [113]. The crystal structure may be the best one as

an input among these sources. No matter what methods are used to get the input structure, the closer the input structure is to the native one, the better the docking result is expected. For docking result filtering in the cemiplimab example, sequence alignment assisted in the prediction of possible final binding poses of the complex. Under some situations, there may be various types of docking results (the antigen binds to different locations on antibodies, or the antigen binds to the same location with different binding poses). In this case, the sequence alignment may be a choice to determine the final pose. The target antibody interaction pose may be similar to other antibodies that share high similarity in sequence. Besides, as we mentioned previously, the heavy chain plays an important role in an antibody-antigen interaction. So, we suggest that sequence similarity of heavy chain can be the priority during the sequence alignment, especially the alignment in the CDR region.

3.3.4 Computational Alanine Scanning for Cemiplimab

Computational alanine scanning was used to predict hotspots on cemiplimab/PD-1 interface. The interface cutoff value was defined as 5 Å. After alanine scanning, we obtained $\Delta\Delta G_{\text{binding}}$ of several interface residues. According to the classification, hotspots and neutral residues were defined as $\Delta\Delta G_{\text{binding}}$ more or less than 1 kcal/mol during the alanine mutation, respectively [93]. Therefore, we found two hotspots, ASN101 and ILE102, located on HCDR3, and their $\Delta\Delta G_{\text{binding}}$ were 1.799 and 1.552 kcal/mol, respectively. Mutation to Ala made N101 lose two hydrogen bonds, which may be responsible for the large binding energy change. On the contrary, I102 did not form any hydrogen bond with PD-1 but still accounted for the great change of the energy. This means I102 may have contributed to the stability of the complex.

3.3.5 *In silico* Affinity Maturation for Cemiplimab

Based on the knowledge of hotspots on cemiplimab, the *in silico* mutation was performed by following the protocol of antibody affinity maturation. We first tried to mutate the interface residues to 20 proteinogenic amino acids. 17 residues on the cemiplimab had been mutated, including 11 from the heavy chain and 6 from the light chain. To view the results clearly, we calculated the score changing of mutated residues. **Figure 18** shows that the mutations of Y103A, I102T, G100D, and F27S had the highest changes of each residue score among the heavy chain.

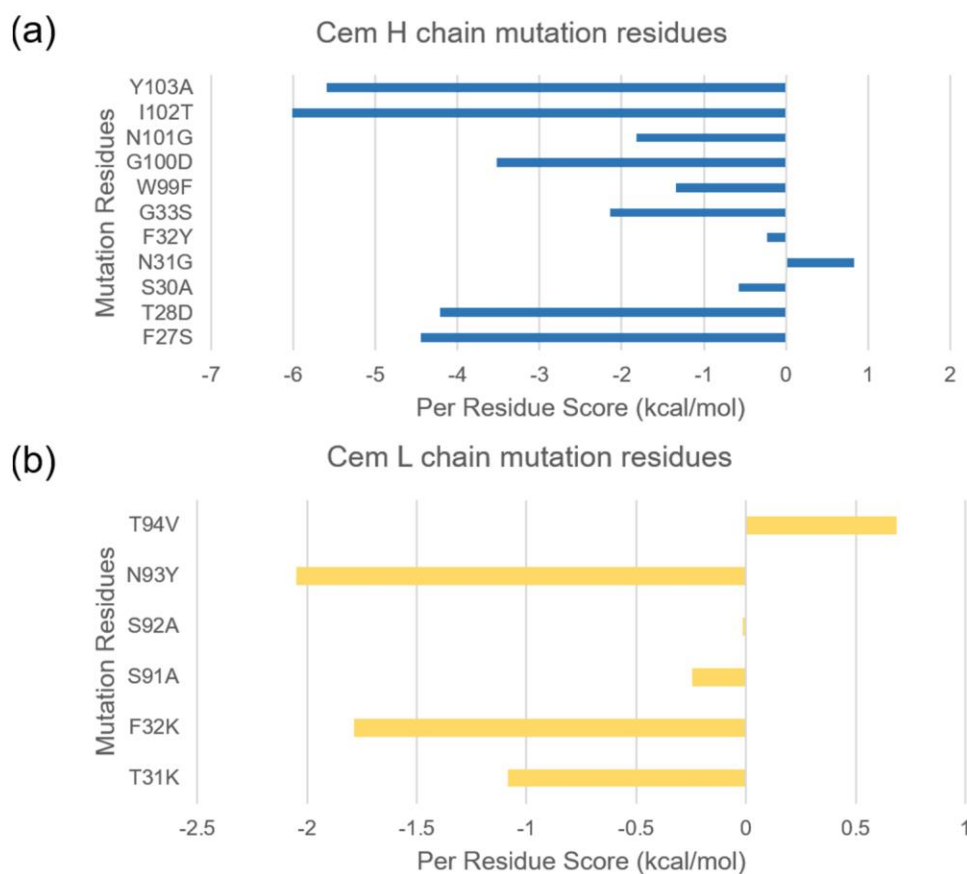


Figure 18. Score changing of mutated residues on antibody during affinity maturation.

(a) The score changing of mutated residues on cemiplimab heavy chain. (b) The score changing of mutated residues on cemiplimab light chain.

On the other hand, the mutations of N93Y and F32K had the highest changes of per residue score on the light chain. Interestingly, most of the mutation residues were located on the complementarity determining region of cemiplimab. This result was consistent with the phenomenon that major mutagenesis during in vivo somatic hypermutation mainly accumulates in the CDR and the framework regions [114]. We chose residues that had the highest score changing to perform the point mutation. Because the HCDR3 region played an important role in binding, the residues on H3 formed major hydrogen bonds, and their energy changing was high, we decided to perform point mutation for residues H: Y103, I102, N101, and G100. The residues were mutated on H chain at cemiplimab: G100A, N101G, I102T, and Y103A, and the results predicted by our protocol showed that both mutations had a better total score, binding affinity, and stability than the original one. To validate these predictions, we conducted the molecular dynamics (MD) simulation and the MM/GBSA calculations for these four mutations. The RMSD of MD simulation result was 2.8 Å. **Figure 19** shows that three mutations including G100A (total binding energy: -173.29 kcal/mol), N101G (total binding energy: -197.49 kcal/mol), and Y103A (total binding energy: -195.27 kcal/mol), improved the binding energy of cemiplimab (total binding energy: -159.68 kcal/mol), while I102T (total binding energy: -97.15 kcal/mol) decreased the total binding energy of cemiplimab. Interestingly, our results showed that these three systems had better

electrostatic properties than that of the original cemiplimab.

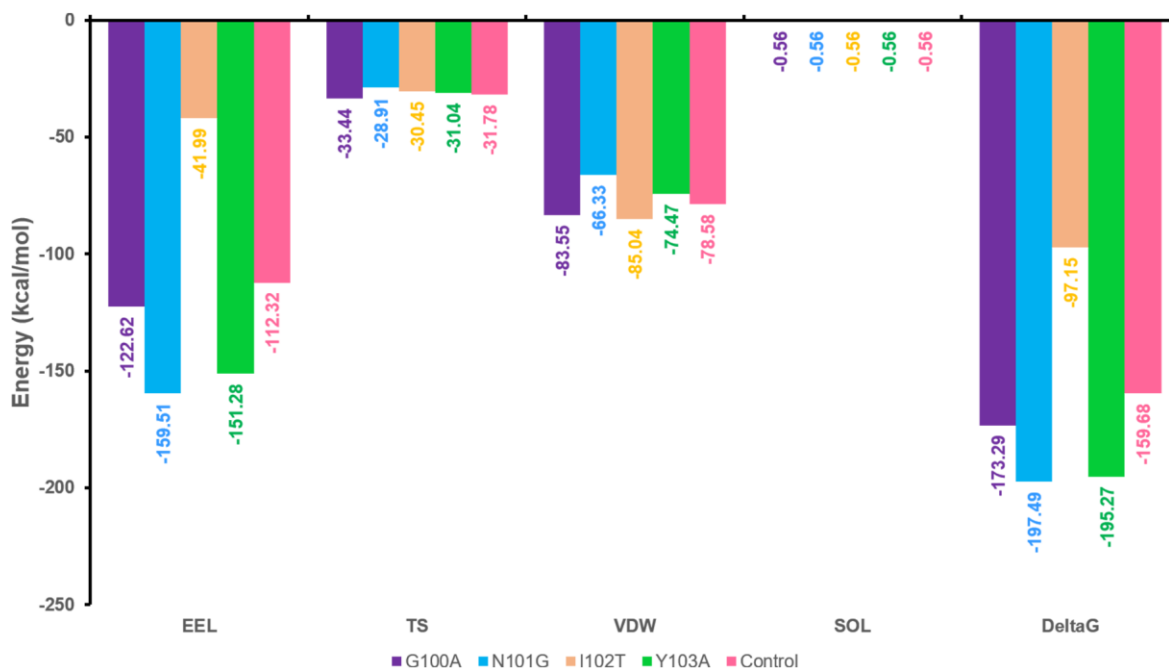


Figure 19. Comparison of MD simulation results between four point mutations and control cemiplimab.

The binding energy (ΔG) and other energy terms are compared between four point mutations G100A (purple), N101G (cyan), I102T (light yellow), Y103A (green), and control (pink).

For the N101G and Y103A, their thermodynamic and van der Waals energy were lower than that of the control group. But G100A had higher conformational entropy change and van der Waals energy than that of the controlled, which meant G100A mutation allowed cemiplimab to form strong van der Waals interaction with PD-1. From the structural analysis, we found that the N101G

mutation lost two hydrogen bonds (**Figure 20**). The overlap of IsAb prediction result from the MD simulation proved that our protocol was reliable.

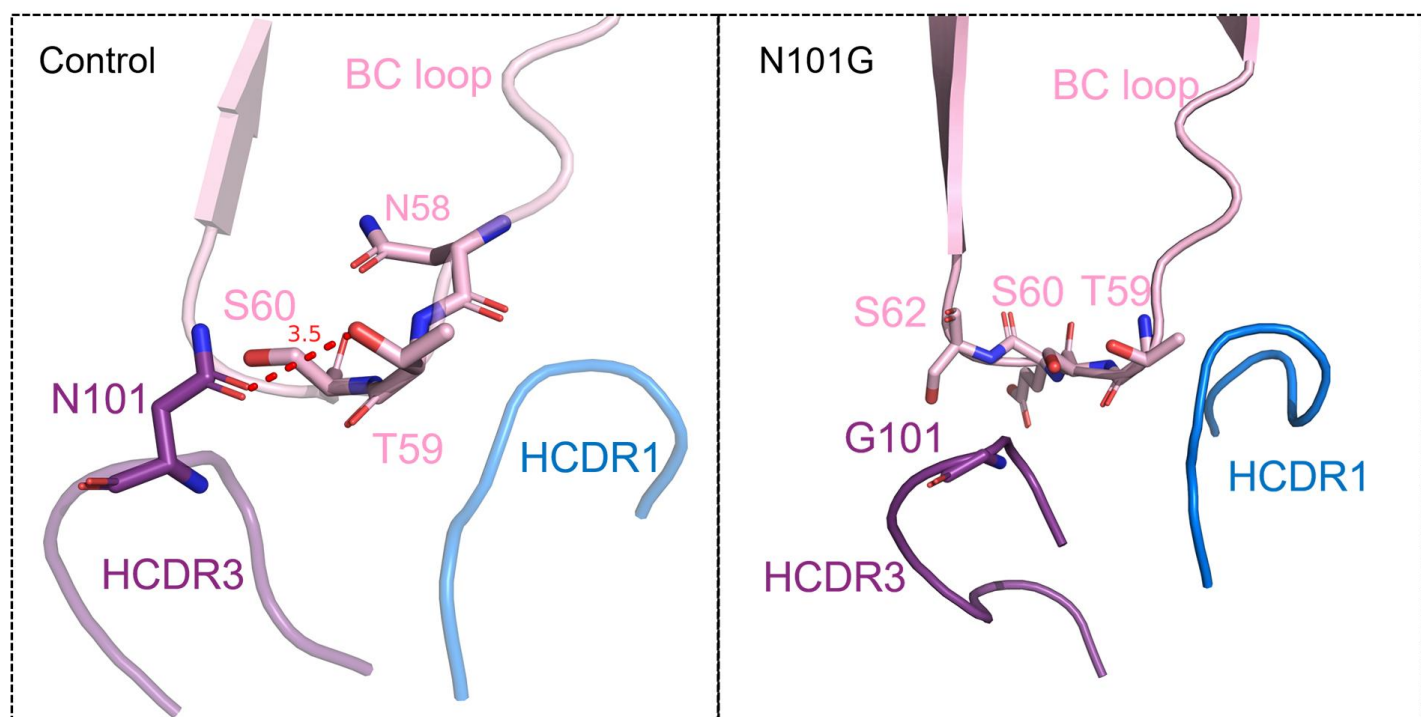


Figure 20. Comparison of hydrogen bonds between the control and the design groups.
Red dash line is hydrogen bond. The hydrogen bonds mainly change in mutation N101G.

4.0 CONCLUSION

In the present work, we developed an *in silico* antibody design protocol, IsAb. The protocol first used RosettaAntibody to model the 3D structure of antibodies. Then, global docking was used to generate the possible binding poses when lacking binding information. Subsequently, SnugDock was applied to refine the antibody-antigen complex and generate a final binding pose. In addition, alanine scanning was used to predict the hotspots on the antibody, which will facilitate affinity maturation and future study. Finally, computational affinity maturation was used to improve affinity and stability of the antibody.

To validate our protocol, we applied IsAb to redesign antibody D44.1 and compared the prediction result with reported data. According to the comparison, IsAb could overlap the better designs with experimental data, which increased the reliability of IsAb.

In the cemiplimab design case study, we used RosettaAntibody to generate the 3D structure of cemiplimab. Due to the lack of binding information of the cemiplimab-PD-1 complex, two steps docking was applied to predict possible binding poses of the complex. With the binding pose, alanine scanning was used to predict the hotspots on the antibody. Based on the prediction hotspots, we conducted *in silico* affinity maturation to modify the antibody. According to our prediction and MD simulation results, the designed cemiplimab theoretically had better affinity and stability than the original one. Thus, IsAb protocol like other computational tools including our Virus-CKB [16] and MCCS [115] algorithms may be used to redesign FDA-approved antiviral antibodies for the potential treatment of COVID-19.

In comparing with other software programs that were reported to design antibodies, IsAb has its advantage. For example, OptMAVEN-2.0 developed by Chowdhury *et al.* [116] is a de novo

design of antibody variable regions. They obtained six antibody parts from the Modular Antibody Parts (MAPs) database to create framework antibody and cluster a set of MAPs parts with an antigen position that will not cause steric clash with the framework antibody. Then, they output the lowest interaction energy in each cluster as the results. Two subsequent software in the antibody design area were Rosetta Antibody Design (RAbD) by Adolf-Bryfogle *et al.* [117] and AbDesign by Lapidoth *et al.* [118] These two methods are similar, they both cluster antibodies sequence and structural fragments. At the same time, they applied Rosetta to implement docking, repacked the sidechain to optimize the structure of the antibody-antigen complex, and built new antibodies that have new backbones and sequences. However, AbDesign and RAbD both require existing antibody-antigen complexes. In most of the cases, the antibody-antigen complex may not be available, which limits the usage of these software. IsAb provides solutions to common problems during antibody design, such as the unavailability of a 3D structure of antibodies or binding information. Different from previous methods, IsAb focuses on antibody redesign that mainly optimizes the antibody sequence. Since the whole antibody structure does not change significantly, it is less likely for the designed antibodies to cause immunogenicity, which they can be progressed to clinical trial faster. In addition, unlike other protocols, IsAb provides a comprehensive tutorial for the antibody design from the beginning to the end, gives future insight for the antibody engineering, and allows the users from other research areas to easily use it to facilitate their research.

However, our protocol cannot automatically conduct antibody design, and the affinity maturation protocol can be improved to solve the problem of trade-off between affinity and stability in a better way. In the future, we will address these challenges and refine our protocol.

We also want to integrate the antibodies data and IsAb protocol to create a one-stop webserver and design antibodies automatically for the input antigens.

APPENDIX

Appendix Table 1 D44.1 local docking top10 I_sc decoys results

No.	Irmsd	I_sc
1	3.002	-9.1
2	5.688	-8.87
3	4.895	-8.666
4	3.974	-8.529
5	2.44	-8.286
6	8.999	-8.176
7	8.56	-7.983
8	2.691	-7.952
9	3.998	-7.89
10	5.731	-7.832

Appendix Table 2 Comparison of D44.1 redesign results

	IsAb	Lippow <i>et al.</i>
Single point mutation	H: T28D T31A S57L L: N32A N92A	H: T28D T31A/V S57V/A T58D K65D L: N32G N92A
Double combination mutation	(L32)A (L92)A	(L32)G (L92)A



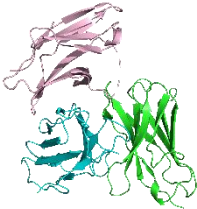


Appendix Table 3 Model1 local docking top10 I_sc decoys results

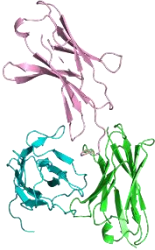


No.	Irmsd	I_sc
1	13.202	-7.704
2	2.843	-7.506
3	0.921	-6.884
4	5.735	-6.861
5	2.292	-6.696
6	3.079	-6.648
7	3.917	-6.638
8	2.738	-6.628
9	1.579	-6.558
10	10.627	-6.551

Appendix Table 4 Model2 local docking top10 I_sc decoys results

No.	Irmsd	I_sc
1	1.279	-9.359
2	1.459	-9.154
3	10.944	-8.207
4	2.364	-7.852
5	13.57	-7.727
6	1.38	-7.616
7	1.446	-7.576
8	1.229	-7.51
9	1.578	-7.488
10	2.165	-7.401

Appendix Table 5 PD-1 antibodies and their sequence alignment with cemiplimab

Antibody name	PDB ID	Structure	Light chain similarity with Cemiplimab	Heavy chain similarity with Cemiplimab
GY 14	6J14		56%	47%
GY 15	6J15		56%	48%
Tislelizumab	7BXA		56%	52%
mAb059c	6K0Y		74%	53%
MW11-h317	6JJP		57%	77%

Nivolumab	5WT9		64%	76%
Pembrolizumab	5GGS		63%	52%
Toripalimab	6JBT		56%	51%

```

./relax.static.linuxgccrelease \
-database /mnt/f/rosetta_bin_linux_2020.08.61146_bundle/main/database \
-s PD1v2.pdb \
-in:file:fullatom \
-out:pdb \
-relax:thorough \
-relax:constrain_relax_to_start_coords \
-relax:ramp_constraints false \
-ex1 \
-use_input_sc -flip_HNQ \
-no_optH false \
-nstruct 10

```

Appendix Figure 1. The script of Rosetta relax.

```

./rosetta_scripts.static.linuxgccrelease \
-s proteins_0887.pdb \
-use_input_sc \
-nstruct 10 \
-jd2:ntrials 1 \
-database /mnt/f/rosetta_bin_linux_2020.25.61318_bundle/main/database \
-parser:protocol AlaScan.xml \
-parser:view \
-out:overwrite

```

Appendix Figure 2. The script of Rosetta alanine scanning.

```

run define_interface.py \
-side1 LH \
-side2 G \
-design-side 1 \
-repack \
-output proteins_0887 proteins_0887.pdb

```

Appendix Figure 3. The script of defining antibody-antigen interface.

```

./rosetta_scripts.static.linuxgccrelease \
-linmem_ig 100 \
-use_input_sc \
-ex1 \
-nstruct 10 \
-s proteins_0887.pdb \
-parser:protocol design.xml \
-out:suffix _design \
-scorefile proteins_0887.fasc

```

Appendix Figure 4. The script of Rosetta affinity maturation.

```

run compare_design_to_control.py proteins_0887_control.fasc proteins_0887.fasc

```

Appendix Figure 5. The script of comparing controlled and designed antibody.

```
./per_residue_energies.static.linuxgccrelease \  
-database /mnt/f/rosetta_bin_linux_2020.25.61318_bundle/main/database \  
-s proteins_0887_design_0003.pdb \  
-use_input_sc
```

Appendix Figure 6. The script of per residue energy analysis.

```
./residue_energy_breakdown.static.linuxgccrelease \  
-database /mnt/f/rosetta_bin_linux_2020.25.61318_bundle/main/database \  
-s proteins_0887_control_0008.pdb
```

Appendix Figure 7. The script of residue energy breakdown analysis.

BIBLIOGRAPHY

1. Smith, L.C., et al., *Monoclonal Antibodies for Combating Synthetic Opioid Intoxication*. Journal of the American Chemical Society, 2019. **141**(26): p. 10489-10503.
2. Shen, X. and T.R. Kosten, *Immunotherapy for drug abuse*. CNS & neurological disorders drug targets, 2011. **10**(8): p. 876-879.
3. Chen, J., et al., *Review of COVID-19 Antibody Therapies*. Annual Review of Biophysics, 2020.
4. Jacofsky, D., E.M. Jacofsky, and M. Jacofsky, *Understanding Antibody Testing for COVID-19*. The Journal of Arthroplasty, 2020. **35**(7, Supplement): p. S74-S81.
5. Reichert, J.M., *Which are the antibodies to watch in 2013?* mAbs, 2013. **5**(1): p. 1-4.
6. Lu, R.-M., et al., *Development of therapeutic antibodies for the treatment of diseases*. Journal of Biomedical Science, 2020. **27**(1): p. 1.
7. Tiller, K.E. and P.M. Tessier, *Advances in Antibody Design*. Annual review of biomedical engineering, 2015. **17**: p. 191-216.
8. Samish, I., et al., *Theoretical and Computational Protein Design*. Annual Review of Physical Chemistry, 2011. **62**(1): p. 129-149.
9. Dahiyat, B.I., D.B. Gordon, and S.L. Mayo, *Automated design of the surface positions of protein helices*. Protein science : a publication of the Protein Society, 1997. **6**(6): p. 1333-1337.
10. Su, A. and S.L. Mayo, *Coupling backbone flexibility and amino acid sequence selection in protein design*. Protein science : a publication of the Protein Society, 1997. **6**(8): p. 1701-1707.
11. Victora, G.D. and M.C. Nussenzweig, *Germinal Centers*. Annual Review of Immunology, 2012. **30**(1): p. 429-457.
12. Diaz, M. and P. Casali, *Somatic immunoglobulin hypermutation*. Current opinion in immunology, 2002. **14**(2): p. 235-240.
13. Chames, P., et al., *Therapeutic antibodies: successes, limitations and hopes for the future*. British journal of pharmacology, 2009. **157**(2): p. 220-233.
14. Song, C.M., S.J. Lim, and J.C. Tong, *Recent advances in computer-aided drug design*. Brief Bioinform, 2009. **10**(5): p. 579-91.
15. Xing, C., et al., *Cryo-EM Structure of the Human Cannabinoid Receptor CB2-Gi Signaling Complex*. Cell, 2020. **180**(4): p. 645-654.e13.
16. Feng, Z., et al., *Virus-CKB: an integrated bioinformatics platform and analysis resource for COVID-19 research*. Briefings in Bioinformatics, 2020.
17. O'Donovan, C., et al., *High-quality protein knowledge resource: SWISS-PROT and TrEMBL*. Briefings in Bioinformatics, 2002. **3**(3): p. 275-284.
18. Altschul, S.F., et al., *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs*. Nucleic acids research, 1997. **25**(17): p. 3389-3402.
19. Schwede, T., et al., *SWISS-MODEL: an automated protein homology-modeling server*. Nucleic Acids Research, 2003. **31**(13): p. 3381-3385.

20. Kellenberger, E., et al., *Comparative evaluation of eight docking tools for docking and virtual screening accuracy*. Proteins: Structure, Function, and Bioinformatics, 2004. **57**(2): p. 225-242.
21. Kuroda, D., et al., *Computer-aided antibody design*. Protein Engineering, Design and Selection, 2012. **25**(10): p. 507-522.
22. Desautels, T., et al., *Rapid &in silico& design of antibodies targeting SARS-CoV-2 using machine learning and supercomputing*. bioRxiv, 2020: p. 2020.04.03.024885.
23. Schymkowitz, J., et al., *The FoldX web server: an online force field*. Nucleic acids research, 2005. **33**(Web Server issue): p. W382-W388.
24. Farhadi, T., A. Fakharian, and S.M. Hashemian, *Affinity Improvement of a Humanized Antiviral Antibody by Structure-Based Computational Design*. International Journal of Peptide Research and Therapeutics, 2019. **25**(1): p. 181-186.
25. Pettersen, E.F., et al., *UCSF Chimera—A visualization system for exploratory research and analysis*. Journal of Computational Chemistry, 2004. **25**(13): p. 1605-1612.
26. Pierce, K., et al., *Preference for geometric patterns early in life as a risk factor for autism*. Archives of general psychiatry, 2011. **68**(1): p. 101-109.
27. Laskowski, R.A. and M.B. Swindells, *LigPlot+: Multiple Ligand–Protein Interaction Diagrams for Drug Discovery*. Journal of Chemical Information and Modeling, 2011. **51**(10): p. 2778-2786.
28. Marcatili, P., et al., *Antibody modeling using the Prediction of ImmunoGlobulin Structure (PIGS) web server*. Nature Protocols, 2014. **9**(12): p. 2771-2783.
29. Winter, G. and C. Milstein, *Man-made antibodies*. Nature, 1991. **349**(6307): p. 293-299.
30. Morea, V., A.M. Lesk, and A. Tramontano, *Antibody Modeling: Implications for Engineering and Design*. Methods, 2000. **20**(3): p. 267-279.
31. Steipe, B., et al., *Sequence Statistics Reliably Predict Stabilizing Mutations in a Protein Domain*. Journal of Molecular Biology, 1994. **240**(3): p. 188-192.
32. Chothia, C. and A.M. Lesk, *Canonical structures for the hypervariable regions of immunoglobulins*. Journal of Molecular Biology, 1987. **196**(4): p. 901-917.
33. Chothia, C., et al., *The predicted structure of immunoglobulin D1.3 and its comparison with the crystal structure*. Science, 1986. **233**(4765): p. 755.
34. Long, X., J.R. Jeliakov, and J.J. Gray, *Non-H3 CDR template selection in antibody modeling through machine learning*. PeerJ, 2019. **7**: p. e6179-e6179.
35. Adolf-Bryfogle, J., et al., *PyIgClassify: a database of antibody CDR structural classifications*. Nucleic Acids Res, 2015. **43**(Database issue): p. D432-8.
36. Zhu, K. and T. Day, *Ab initio structure prediction of the antibody hypervariable H3 loop*. Proteins: Structure, Function, and Bioinformatics, 2013. **81**(6): p. 1081-1089.
37. Kuroda, D., et al., *Structural classification of CDR-H3 revisited: A lesson in antibody modeling*. Proteins: Structure, Function, and Bioinformatics, 2008. **73**(3): p. 608-620.
38. Yamashita, K., et al., *Kotai Antibody Builder: automated high-resolution structural modeling of antibodies*. Bioinformatics, 2014. **30**(22): p. 3279-3280.
39. Weitzner, B.D., et al., *Blind prediction performance of RosettaAntibody 3.0: grafting, relaxation, kinematic loop modeling, and full CDR optimization*. Proteins, 2014. **82**(8): p. 1611-1623.
40. Ruffolo, J.A., et al., *Geometric potentials from deep learning improve prediction of CDR H3 loop structures*. Bioinformatics, 2020. **36**(Supplement_1): p. i268-i275.

41. Teng, G. and F.N. Papavasiliou, *Immunoglobulin Somatic Hypermutation*. Annual Review of Genetics, 2007. **41**(1): p. 107-120.
42. Finlay, W.J.J. and J.C. Almagro, *Natural and man-made V-gene repertoires for antibody discovery*. Frontiers in immunology, 2012. **3**: p. 342-342.
43. Guntas, G., C. Purbeck, and B. Kuhlman, *Engineering a protein–protein interface using a computationally designed library*. Proceedings of the National Academy of Sciences, 2010. **107**(45): p. 19296.
44. Poosarla, V.G., et al., *Computational de novo design of antibodies binding to a peptide with high affinity*. Biotechnology and bioengineering, 2017. **114**(6): p. 1331-1342.
45. Li, T., R.J. Pantazes, and C.D. Maranas, *OptMAVEN – A New Framework for the de novo Design of Antibody Variable Region Models Targeting Specific Antigen Epitopes*. PLOS ONE, 2014. **9**(8): p. e105954.
46. Cannon, D.A., et al., *Experimentally guided computational antibody affinity maturation with de novo docking, modelling and rational design*. PLOS Computational Biology, 2019. **15**(5): p. e1006980.
47. Leem, J., et al., *ABodyBuilder: Automated antibody structure prediction with data-driven accuracy estimation*. mAbs, 2016. **8**(7): p. 1259-1268.
48. Dunbar, J., et al., *SAbPred: a structure-based antibody prediction server*. Nucleic acids research, 2016. **44**(W1): p. W474-W478.
49. Liu, G., et al., *Antibody complementarity determining region design using high-capacity machine learning*. Bioinformatics, 2020. **36**(7): p. 2126-2133.
50. Marze, N.A., S. Lyskov, and J.J. Gray, *Improved prediction of antibody VL-VH orientation*. Protein Eng Des Sel, 2016. **29**(10): p. 409-418.
51. Weitzner, B.D. and J.J. Gray, *Accurate Structure Prediction of CDR H3 Loops Enabled by a Novel Structure-Based C-Terminal Constraint*. J Immunol, 2017. **198**(1): p. 505-515.
52. Weitzner, B.D., et al., *Modeling and docking of antibody structures with Rosetta*. Nature Protocols, 2017. **12**(2): p. 401-416.
53. Lyskov, S., et al., *Serverification of Molecular Modeling Applications: The Rosetta Online Server That Includes Everyone (ROSIE)*. PLOS ONE, 2013. **8**(5): p. e63906.
54. Sivasubramanian, A., et al., *Toward high-resolution homology modeling of antibody Fv regions and application to antibody-antigen docking*. Proteins, 2009. **74**(2): p. 497-514.
55. Robin, G., et al., *Restricted Diversity of Antigen Binding Residues of Antibodies Revealed by Computational Alanine Scanning of 227 Antibody–Antigen Complexes*. Journal of Molecular Biology, 2014. **426**(22): p. 3729-3743.
56. Desta, I.T., et al., *Performance and Its Limits in Rigid Body Protein-Protein Docking*. Structure, 2020. **28**(9): p. 1071-1081.e3.
57. Vajda, S., et al., *New additions to the ClusPro server motivated by CAPRI*. Proteins, 2017. **85**(3): p. 435-444.
58. Kozakov, D., et al., *The ClusPro web server for protein-protein docking*. Nat Protoc, 2017. **12**(2): p. 255-278.
59. Kozakov, D., et al., *How good is automated protein docking?* Proteins, 2013. **81**(12): p. 2159-66.
60. Brenke, R., et al., *Application of asymmetric statistical potentials to antibody-protein docking*. Bioinformatics, 2012. **28**(20): p. 2608-14.
61. Comeau, S.R., et al., *ClusPro: an automated docking and discrimination method for the prediction of protein complexes*. Bioinformatics, 2004. **20**(1): p. 45-50.

62. Sircar, A. and J.J. Gray, *SnugDock: Paratope Structural Optimization during Antibody-Antigen Docking Compensates for Errors in Antibody Homology Models*. PLOS Computational Biology, 2010. **6**(1): p. e1000644.
63. Dickson, C.J., et al., *Lipid14: The Amber Lipid Force Field*. Journal of Chemical Theory and Computation, 2014. **10**(2): p. 865-879.
64. Maier, J.A., et al., *ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB*. Journal of Chemical Theory and Computation, 2015. **11**(8): p. 3696-3713.
65. Wang, J., et al., *Development and testing of a general amber force field*. J Comput Chem, 2004. **25**(9): p. 1157-74.
66. Cheng, J., et al., *Computational Systems Pharmacology-Target Mapping for Fentanyl-Laced Cocaine Overdose*. ACS Chem Neurosci, 2019. **10**(8): p. 3486-3499.
67. Wu, N., et al., *Insight of Captagon Abuse by Chemogenomics Knowledgebase-guided Systems Pharmacology Target Mapping Analyses*. Scientific Reports, 2019. **9**(1): p. 2268.
68. Chen, M., et al., *DAKB-GPCRs: An Integrated Computational Platform for Drug Abuse Related GPCRs*. J Chem Inf Model, 2019. **59**(4): p. 1283-1289.
69. Wang, Y., et al., *An insight into paracetamol and its metabolites using molecular docking and molecular dynamics simulation*. J Mol Model, 2018. **24**(9): p. 243.
70. Case, D., et al., *Amber 2018*. 2018.
71. Hou, T., et al., *Assessing the Performance of the MM/PBSA and MM/GBSA Methods. I. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations*. Journal of Chemical Information and Modeling, 2011. **51**(1): p. 69-82.
72. Hou, T., et al., *Assessing the performance of the molecular mechanics/Poisson Boltzmann surface area and molecular mechanics/generalized Born surface area methods. II. The accuracy of ranking poses generated from docking*. J Comput Chem, 2011. **32**(5): p. 866-77.
73. Hawkins, G.D., C.J. Cramer, and D.G. Truhlar, *Parametrized Models of Aqueous Free Energies of Solvation Based on Pairwise Descreening of Solute Atomic Charges from a Dielectric Medium*. The Journal of Physical Chemistry, 1996. **100**(51): p. 19824-19839.
74. Canutescu, A.A. and R.L. Dunbrack, Jr., *Cyclic coordinate descent: A robotics algorithm for protein loop closure*. Protein science : a publication of the Protein Society, 2003. **12**(5): p. 963-972.
75. Wang, C., O. Schueler-Furman, and D. Baker, *Improved side-chain modeling for protein-protein docking*. Protein science : a publication of the Protein Society, 2005. **14**(5): p. 1328-1339.
76. Hwang, H., et al., *Protein-protein docking benchmark version 4.0*. Proteins, 2010. **78**(15): p. 3111-3114.
77. Zacharias, M., *Accounting for conformational changes during protein-protein docking*. Curr Opin Struct Biol, 2010. **20**(2): p. 180-6.
78. Bonvin, A.M., *Flexible protein-protein docking*. Curr Opin Struct Biol, 2006. **16**(2): p. 194-200.
79. Kuroda, D. and J.J. Gray, *Pushing the Backbone in Protein-Protein Docking*. Structure, 2016. **24**(10): p. 1821-1829.
80. Tyka, M.D., et al., *Alternate states of proteins revealed by detailed energy landscape mapping*. J Mol Biol, 2011. **405**(2): p. 607-18.

81. Bower, M.J., F.E. Cohen, and R.L. Dunbrack, Jr., *Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: a new homology modeling tool*. J Mol Biol, 1997. **267**(5): p. 1268-82.
82. Katchalski-Katzir, E., et al., *Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques*. Proceedings of the National Academy of Sciences of the United States of America, 1992. **89**(6): p. 2195-2199.
83. Kozakov, D., et al., *PIPER: an FFT-based protein docking program with pairwise potentials*. Proteins, 2006. **65**(2): p. 392-406.
84. Lo Conte, L., C. Chothia, and J. Janin, *The atomic structure of protein-protein recognition sites*. J Mol Biol, 1999. **285**(5): p. 2177-98.
85. Chuang, G.Y., et al., *DARS (Decoys As the Reference State) potentials for protein-protein docking*. Biophys J, 2008. **95**(9): p. 4217-27.
86. Kozakov, D., et al., *The ClusPro web server for protein-protein docking*. Nature Protocols, 2017. **12**(2): p. 255-278.
87. Lyskov, S. and J.J. Gray, *The RosettaDock server for local protein-protein docking*. Nucleic Acids Res, 2008. **36**(Web Server issue): p. W233-8.
88. Chaudhury, S., et al., *Benchmarking and Analysis of Protein Docking Performance in Rosetta v3.2*. PLOS ONE, 2011. **6**(8): p. e22477.
89. Wang, C., P. Bradley, and D. Baker, *Protein-protein docking with backbone flexibility*. J Mol Biol, 2007. **373**(2): p. 503-19.
90. Méndez, R., et al., *Assessment of blind predictions of protein-protein interactions: current status of docking methods*. Proteins, 2003. **52**(1): p. 51-67.
91. Clackson, T. and J.A. Wells, *A hot spot of binding energy in a hormone-receptor interface*. Science, 1995. **267**(5196): p. 383-6.
92. Kortemme, T., D.E. Kim, and D. Baker, *Computational Alanine Scanning of Protein-Protein Interfaces*. Science's STKE, 2004. **2004**(219): p. pl2.
93. Kortemme, T. and D. Baker, *A simple physical model for binding energy hot spots in protein-protein complexes*. Proceedings of the National Academy of Sciences, 2002. **99**(22): p. 14116.
94. Dunbrack, R.L., Jr., *Rotamer libraries in the 21st century*. Curr Opin Struct Biol, 2002. **12**(4): p. 431-40.
95. Ollikainen, N., et al., *Chapter Four - Flexible Backbone Sampling Methods to Model and Design Protein Alternative Conformations*, in *Methods in Enzymology*, A.E. Keating, Editor. 2013, Academic Press. p. 61-85.
96. Shringari, S., et al., *Rosetta Custom Score Functions Accurately Predict $\Delta\Delta G$ of Mutations at Protein-Protein Interfaces Using Machine Learning*. bioRxiv, 2020: p. 2020.03.17.996066.
97. Darsley, M.J. and A.R. Rees, *Three distinct epitopes within the loop region of hen egg lysozyme defined with monoclonal antibodies*. The EMBO journal, 1985. **4**(2): p. 383-392.
98. Ponomarenko, J.V. and P.E. Bourne, *Antibody-protein interactions: benchmark datasets and prediction tools evaluation*. BMC Struct Biol, 2007. **7**: p. 64.
99. Lippow, S.M., K.D. Wittrup, and B. Tidor, *Computational design of antibody-affinity improvement beyond in vivo maturation*. Nature Biotechnology, 2007. **25**(10): p. 1171-1176.

100. Chen, J., N. Sawyer, and L. Regan, *Protein-protein interactions: general trends in the relationship between binding affinity and interfacial buried surface area*. Protein Sci, 2013. **22**(4): p. 510-5.
101. Horita, S., et al., *High-resolution crystal structure of the therapeutic antibody pembrolizumab bound to the human PD-1*. Sci Rep, 2016. **6**: p. 35297.
102. Tan, S., et al., *An unexpected N-terminal loop in PD-1 dominates binding by nivolumab*. Nature Communications, 2017. **8**(1): p. 14369.
103. Liu, H., et al., *Glycosylation-independent binding of monoclonal antibody toripalimab to FG loop of PD-1 for tumor immune checkpoint therapy*. MAbs, 2019. **11**(4): p. 681-690.
104. Mandell, D.J., E.A. Coutsiadis, and T. Kortemme, *Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling*. Nature Methods, 2009. **6**(8): p. 551-552.
105. Phillips, J.C., et al., *Scalable molecular dynamics with NAMD*. J Comput Chem, 2005. **26**(16): p. 1781-802.
106. Swindells, M.B., et al., *abYsis: Integrated Antibody Sequence and Structure-Management, Analysis, and Prediction*. J Mol Biol, 2017. **429**(3): p. 356-364.
107. Ferdous, S. and A.C.R. Martin, *AbDb: antibody structure database-a database of PDB-derived antibody structures*. Database : the journal of biological databases and curation, 2018. **2018**: p. bay040.
108. Dunbar, J., et al., *SAbDab: the structural antibody database*. Nucleic Acids Research, 2014. **42**(D1): p. D1140-D1146.
109. Raybould, M.I.J., et al., *Thera-SAbDab: the Therapeutic Structural Antibody Database*. Nucleic Acids Research, 2020. **48**(D1): p. D383-D388.
110. Liberis, E., et al., *Parapred: antibody paratope prediction using convolutional and recurrent neural networks*. Bioinformatics, 2018. **34**(17): p. 2944-2950.
111. Krawczyk, K., et al., *Antibody i-Patch prediction of the antibody binding site improves rigid local antibody-antigen docking*. Protein Eng Des Sel, 2013. **26**(10): p. 621-9.
112. Kim, D.E., D. Chivian, and D. Baker, *Protein structure prediction and analysis using the Robetta server*. Nucleic Acids Res, 2004. **32**(Web Server issue): p. W526-31.
113. Whitelegg, N.R.J. and A.R. Rees, *WAM: an improved algorithm for modelling antibodies on the WEB*. Protein Engineering, Design and Selection, 2000. **13**(12): p. 819-824.
114. Julian, M.C., et al., *Efficient affinity maturation of antibody variable domains requires co-selection of compensatory mutations to maintain thermodynamic stability*. Scientific Reports, 2017. **7**(1): p. 45259.
115. Feng, Z., et al., *MCCS: a novel recognition pattern-based method for fast track discovery of anti-SARS-CoV-2 drugs*. Brief Bioinform, 2020.
116. Chowdhury, R., M.F. Allan, and C.D. Maranas, *OptMAVEEn-2.0: De novo Design of Variable Antibody Regions against Targeted Antigen Epitopes*. Antibodies, 2018. **7**(3).
117. Adolf-Bryfogle, J., et al., *RosettaAntibodyDesign (RAbD): A general framework for computational antibody design*. PLOS Computational Biology, 2018. **14**(4): p. e1006112.
118. Lapidoth, G.D., et al., *AbDesign: An algorithm for combinatorial backbone design guided by natural conformations and sequences*. Proteins, 2015. **83**(8): p. 1385-406.