TITLE

Tianjian Liang, Hui Chen, Jiayi Yuan, Chen Jiang, Zhiwei Feng\*, Xiang-Qun Xie\*

\* To whom correspondence should be addressed. Zhiwei Feng: Tel: +1 (412) 419 4896; Email: [zhf11@pitt.edu](mailto:zhf11@pitt.edu); Correspondence may also be addressed to Xiang-Qun Xie. Tel: +1 (832) 755 1268; Email: [xix15@pitt.edu](mailto:xix15@pitt.edu)

**Abstract**

The design of the therapeutic antibody has attracted a large amount of attention over the 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 significant advances in the past years, there are still some challenges that need to be solved, such as lack of antibody structure data, flexibility of antigen structure, not much detailed antibody design protocol, and no free available antibody design web servers. In the present work, we elaborated on an *in-silico* antibody design protocol for users to perform computer-aided antibody design easily. First, Rosetta web server will be applied to generate the 3D structure of query antibody if there was no structural information available. Then, two-step docking will be used to identify the binding pose of antibody-antigen when the binding information of antibody and antigen is unknown: ClusPro is used first to conduct the global docking, and SnugDock is applied for the local docking. Sequentially, based on the predicted binding pose(s), *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 the antibody to increase its affinity and stability theoretically. To illustrate how this protocol works, we used cemiplimab, a PD-1 checkpoint inhibitor antibody, as an example to show a step-by-step tutorial.

**Introduction**

Antibodies are widely used in the treatment of cancers, autoimmune diseases, and inflammatory diseases due to their high affinity as well as specificity to a wide variety of macromolecules. Moreover, antibodies are simple to express and purify, which makes the antibody design more and more attractive, and the number of approved antibodies has steadily risen.

Among antibody design, antibodies redesign is one of the most important methods (we focus on redesign here). The antibodies redesign1 heavily relies on a native protein or an interface, in which redesign mainly focuses on engineering the surface or interior of the antibody-antigen complex to acquire better affinity or stability of the complex2, 3. For antibody design, *in vivo* affinity maturation is one of the essential methods to get better properties of antibodies, which can increases antibody affinity several-fold by somatic hypermutation (SHM) during an immune response4. SHM usually occurs in the Complementarity-determining regions (CDR) of the immunoglobulin genes, which has a rapid mutation rate5. The experimental methods for the antibody design require either the crystal structure or the cryo-EM structure of the antibody-antigen complex. After obtaining the 3D complex, random mutations or phage display is preformed6. However, all these methods are time-consuming and require a heavy workload. Therefore, computational methods may be alternative strategies to aid in antibody design.

Computational antibody design methods have been improved and indicated several principles in protein architecture7, 8. The work by Li *et al*.9 pointed out that the mutation positions always locate at the surrounding of the binding interface. Additionally, the mutated residues must be the amino acids that frequently appear in the recognition site of antibody during affinity maturation. Li *et al*. combined the evolutionary information with computational simulation techniques to improve the affinity of trastuzumab, the human epidermal growth factor receptor 2 (HER2) antibody. Kiyoshi *et al*.10 combined thermodynamic analysis with computational calculations to improve the affinity of antibody 11K2. Based on the single-chain antibodies (scFvs) 2D10 and its antigen DVFYPYPYASGS, Poosarla *et al*.11 combined computational framework named Optimal Method for Antibody Variable region Engineering (OptMAVEn)12 with MD simulation to conduct de novo design of scFvs-2D10.

However, there are still some challenges that computational methods need to address. In some circumstances, the 3D structural of antibody is unavailable. Therefore, the binding information of antibody-antigen complex may be limited. Another challenge is, when interacting with antibodies some antigens may undergo significant conformational changes, such as some loops may be invisible, and others may appear. Currently, most of the antibody design web servers are commercial, which means that the users need to pay for the design work.

Hence, we established an *in-silico* antibody design protocol to address these challenges and to guide the undemanding antibody design. The protocol will first use RosettaAnitbody to solve the situation when the 3D structure of antibody is not available. Then, RosettaRelax is applied to minimize the energy of protein structures to make the input conformations closer to the bound state and make the docking more accurate. To address the lack of binding information, the proposed protocol performs two-step docking, including the global and local dockings. After acquiring the binding conformation, alanine scanning will be applied to predict the hotspots on the antibody. Finally, the protocol will conduct a computational affinity maturation to improve the properties of existing antibody. To give step-by-step instruction, we went through this protocol by using cemiplimab design as an example.

**Material and Method**

**Generating and Preparing the Structures.** We downloaded 3D structures of antigen or antibody from Protein Data Bank (<https://www.rcsb.org/>). If the 3D structure is antigen-antibody (not the target antibody) complex, we then use PyMol to extract the 3D structure of antigen. Antibody Fv sequence was obtained from IMGT (<http://www.imgt.org/>) and submitted to ROSIE (<https://rosie.graylab.jhu.edu/>). RosettaAntibody was used to model the 3D structure of antibody. For the modeling results, we chose the “Grafted-Relaxed-Model” provided by the web server, which “Grafted-Relaxed-Model” was relaxed during the modeling process. For the structures of antigen, we applied Rosetta relax protocol (“relax.static.linuxgccrelease”) for preparations. We selected the lowest score decoy among the ten output results as the relaxed structure.

**Antibody-Antigen Global Docking.** We submitted the antibody Fv sequence to the antibody paratopes prediction database developed by Robin *et al.*13 to predict the possible paratopes on the antibody. We chose the residues whose ΔΔG is larger than 1kcal/mol as possible paratopes. The global docking for antibody-antigen was performed by ClusPro (<https://cluspro.bu.edu/login.php>), in which we selected the “Antibody Mode” to specific 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.

**Antibody-Antigen Local Docking.** SnugDock function (<https://rosie.graylab.jhu.edu/snug_dock>) on ROSIE was used to refine the possible binding poses generated from global docking. We chose “thorough mode” to perform the local docking and accepted the docking result based on whether the local docking formed a docking funnel (N5 > 3). Once the docking completed, we picked the lowest I\_sc decoy. The sequence alignment was performed by EMBL “Clustal Omega” server (<https://www.ebi.ac.uk/services>).

**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. The interface cutoff value was 5 Å. For the result, we excluded residues whose ΔΔG is lower than 1 kcal/mol.

The *in-silico* affinity maturation was performed by the affinity maturation protocol generated from Dr. Jens Meiler lab website. We first used “define\_interface.py” to prepare a residue file (resfile). Resfile is used to define what kinds of amino acids will the interface residues be mutated. Then we applied affinity maturation protocol (design.xml, <http://www.meilerlab.org/index.php/rosetta-tutorials>) to modify cemiplimab. The script of “compare\_design\_to\_control.py” will be utilized to analyze the different metrics comparison between the control and design group. The residue energy breakdown of control and design groups was calculated by the script of “PerResidueEnergies.py”. Then, we calculated the residues energy changing during the mutation and made the energy changing plot.

**Result**

**Workflow of Antibody Design Protocol**

The proposed procedure of *in-silico* antibody design protocol is described in **Figure 1**. Briefly, the *first step* is to obtain the structure of antibody and antigen. Users can download antigen and antibody structures from Protein Data Bank (<https://www.rcsb.org/>). For users who already obtain antibody structures can directly jump to the third step. Otherwise, users need to retrieve the antibody sequence from IMGT (<http://www.imgt.org/>). *Step 2*, RosettaAntibody web server (<https://rosie.graylab.jhu.edu/>) is used for antibody modeling. The RosettaAntibody14 will help users to get the antibody structure through its sequence and output a relaxed structure. *Step 3*, RosettaRelax protocol is applied to refine antibody and antigen structures. The RosettaRelax uses rosetta force-field to make simple all-atom refinement of structure, driving a starting structure towards its native state and improve docking accuracy. *Step 4*, antibody is docked to the antigen. If the binding information of antibody and antigen is available, the user can jump to the fifth step. Without much structural information, the antibody and antigen structures can be submitted to the ClusPro15 (<https://cluspro.bu.edu/login.php>) to perform global docking. Ten most populated clusters will output as the results, which can give users several possible binding poses for the antibody-antigen complex. *Step 5*, based on the global docking results, a local refined docking is carried out by SnugDock16 (<https://rosie.graylab.jhu.edu/snug_dock>), 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 changing 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 output the best mutated antibody that has better affinity and stability than the original one.

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

***Step (1-3)*:** **Generating and Preparing the Structures**

Before performing antibody docking, we need to have structures for both antigen and antibody. In this work, we can obtain the 3D structure of antigen from the Protein Data Bank. To generate the 3D structure of 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 antibody structure in two main steps. First, BLAST-based method is used to search homologous templates for framework regions and CDR loops for the input sequences. Then, the 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. The low-resolution phase will find candidate conformation of CDR H3 loop by cyclic coordinate descent and fragment assembly17. The high-resolution phase will repeatedly optimize and minimize the side chains18. 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 the experimental result19, during the protein-protein interaction, there are several angstroms of root-mean-square deviation (RMSD) of their structure compared with the unbound state. To observe minimum-energy conformation in the correct binding site, the protein-protein docking always experiences backbone conformational change20, 21. Besides, the protein docking accuracy relies on the quality of input structures. The closer starting structure is to the bound state, the more likely to success22. Therefore, we implemented Rosetta relax protocol (FastRelax)23 to lower the energy of the starting structure.

FastRelax is composed of several rounds of energy minimization in torsion angle space and all atom 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 chosen from a rotamers library randomly by Monte Carlo simulated annealing run24.

***Step 4*: Antibody-Antigen Global Docking**

Before antibody-antigen docking, the binding information of antibody and antigen are 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 the unbound proteins in every run and output possible binding poses of the antibody-antigen complex.

ClusPro (<https://cluspro.bu.edu/login.php>) is a protein docking server that can conduct the antibody-antigen docking. In the beginning, a Fast Fourier Transform (FFT) correlation approach25 based docking program PIPER26 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 antibody and antigen interface27, ClusPro integrates ADARS, which is the asymmetric DARS-type28 potential in the energy function of PIPER. After docking, ClusPro calculates the pairwise interface root mean square bases on it to cluster retained conformations. Lastly, the CHARMM potential is used to minimize the energy of structures by removing the steric crash and output the structures at the centers of the 10 most populated clusters29.

***Step 5*:** **Antibody-Antigen Local Docking**

To refine the results from global docking, local docking is performed. RosettaDock30 is one of the best local docking methods, which consist of Monte Carlo-based multi-scale docking algorithm. The algorithm allows both intramolecular and interfacial flexibility in the antibody during docking with high accuracy. SnugDock is a specific algorithm for antibody docking, it is built on RosettaDock and incorporates the sampling characteristic of RosettaAntibody.

SnugDock runs local docking in two stages: low-resolution and high-resolution. 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 move set with indicated frequencies. The move set 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 sidechain. After that, SnugDock will sort decoys by interface score and output docked complex.

Due to the energetic differences within small backbone changes in local docking that are difficult to capture accurately. Interface score (intermolecular energy)31, which can provide the best measurement of the structure as the metric is used to analyze the result of local docking. In addition, to match the I\_sc, I\_rmsd is applied to measure interface root mean square deviation of heavy atoms in the interface residues between the reference structure and result structure. According to the Critical Assessment of Protein Interactions (CAPRI), a ‘docking funnel’ can be used as a metric to measure success in a local docking. From CAPRI definition, a decoy with I\_rmsd<1.0 Å is defined as high quality, 1.0 Å<I\_rmsd<2.0 Å is defined as medium quality, and 2.0 Å< I\_rmsd< 4 Å is defined as acceptable quality. A docking funnel means that three of the top five decoys’ I\_rmsd need to be smaller than 4.0 Å (N5>=3).

***Step 6*: Computational Alanine Scanning**

To predict energetically important residues32 in antibody-antigen interfaces, computational alanine scanning protocol33 is a great choice. Alanine scanning protocol will automatically define interface residues of complex and conduct *in-silico* alanine mutations on the interfaces. After mutation, the protocol uses free energy function to calculate the wild type (original complex) and mutates 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 (ΔΔGbinding) for all alanine mutations. According to the classification, hotspots and neutral residues were defined as ΔΔG more or less than 1 kcal/mol during the alanine mutation, respectively34.

***Step 7*: *In silico* Affinity maturation**

Computational antibody affinity maturation protocol can theoretically increase the affinity of antibody to the antigen. Firstly, *in-silico* affinity maturation method needs to be assigned a mutated site on the antibody-antigen complex. Then, rotamer library will be used to explore 20 amino acids at each position35. The backbone will experience a small change36 during side chain sampling. The sampling step will generate different types of mutations and scoring by Rosetta score function *REF2015*37. The output structure decision is based on the Metropolis criterion and outputs the best mutated structure, which has better affinity and stability among other mutations.

In the following section, we will apply cemiplimab as an example to show the application of our proposed antibody design protocol.

**Case Study: Cemiplimab Design Procedure**

Cemiplimab is an FDA proved PD-1 checkpoint inhibitor (**Figure 2**) used to treat cutaneous squamous cell carcinoma (cSCC). However, cemiplimab does not have published 3D crystal or cryo-EM structure. Moreover, the structures of PD-1 resolved in different conditions show large flexibility (**Figure 3)**. According to research, PD-1 C’D, N, and FG loops may experience huge changes during interactions with different antibodies. For example, when binding with pembrolizumab38 (**Figure 4a)**, PD-1 N loop is missing, and C’D loop dominantly affects the binding affinity of the complex. However, when binding with nivolumab39 (**Figure 4b**), 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 toripalimab40 (**Figure 4c)**. On the other hand, most literature about cemiplimab focuses on clinical data, which means that the interaction information is limit. Therefore, the cemiplimab example is under the extreme situation that the antibody has no structure and no binding information.

***Step (1-3)*: Generating and Preparing the PD-1 and Cemiplimab Structures**

For the construction of cemiplimab-PD1 complex, there is not much information about the binding between cemiplimab and PD-1. Based on the literature research, we generated three representative PD-1 structures. First, we collected the crystal structure of PD-1-antibody complex from the protein data bank (<https://www.rcsb.org/>) and generated three representative PD-1 structures, named **PD1v1**, **PD1v2**, and **PD1v4** (**Figure 5**), by extracting the structure from its complex (PDB: 5wt9, 5ggs, and 6jbt). We then docked these three PD-1 into cemiplimab, respectively. For convenience and consideration of the mechanism of antigen-antibody binding, we used the antibody Fv region to dock with PD-1. Pymol was used to extract the crystal structure of PD-1 from its complex. Cemiplimab’s Fv sequence was obtained from IMGT (<http://www.imgt.org/>) and submitted to ROSIE (<https://rosie.graylab.jhu.edu/>). RosettaAntibody function was used to model the cemiplimab 3D structure. For the modeling results, we chose the “Grafted-Relaxed-Model” provided by the web server, which “Grafted-Relaxed-Model” was relaxed during the modeling process. Therefore, we applied Rosetta relax protocol (“relax.static.linuxgccrelease”) for antigen structure preparation. We selected the lowest score decoy among the ten output results as the relaxed structure.

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 interaction with antibody (structure missing or appear). In this case, users can do the literature search to conclude the representation 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 web server in the cemiplimab example. In addition, the RosettaAntibody and SnugDock function can also be ran on Rosetta software. If Rosetta software was used, it can implement an ensemble of ten lowest-scoring structures obtained from antibody modeling as the input. For the protein relaxation, KIC protocol41 or molecular dynamic (MD) simulation42 is also an alteration method.

***Step 4*: Cemiplimab-PD-1 Global Docking.**

Since there are 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’s sequence into the paratopes prediction database developed by Robin *et al.*13 to predict the possible paratopes on the antibody. We chose the residues whose energy contribution shown in the paratopes prediction database is larger than 1kcal/mol as possible paratopes. Then, we input the predicted paratopes in ClusPro to perform the global docking. The three criteria from the literature conclude that FG loop will form hydrogen bonds with antibodies, C’D or N loop will form hydrogen bonds with antibodies when they appear, and the antibody structure will not experience a significant change during docking. Following these criteria, we selected possible binding poses from the top10 score decoys. Each docking trait will select one possible binding poses as the local docking input. A total of three results were selected and named as **model1**, **2**, and **3** (**Figure 6**), respectively.

Global docking is used when binding information of antibody-antigen is not clear. If the paratopes and the epitopes are already known, the starting pose for local docking can be built by using Pymol. The information of paratopes and epitopes can acquire from experiments, such as point mutation, epitopes mapping, and alanine scanning. For virtual methods, Parapred43, Antibody i-patch44, and the database we used in the cemiplimab example are available methods to predict the paratopes. But, the information of paratopes and epitopes is not essential for global docking. The global docking can perform without this information. However, the quality of starting pose of the local docking significantly affects the final result. The more accurate paratopes and epitopes prediction are, the better-quality starting conformation will be.

In the cemiplimab case, instead of only relying on the docking score, we used the criteria concluded from the literature to filter the 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 antibody-antigen can obtain by comparing the binding structure of antigen with different antibodies if there are no experimental data available. In some situations, the antigen may be a novel target and does not have an existing antibody. In this case, we suggest the users to choose the top rank complex as the global docking result.

***Step 5*:Cemiplimab-PD-1 Local Docking**

To refine the results from global docking, SnugDock function on the ROSIE web server (<https://rosie.graylab.jhu.edu/snug_dock>) was used to conduct the local docking. The criteria to accept docking result is based on whether the local docking formed a docking funnel (N5 >3). If the docking was successful and the lowest I\_sc decoy met the three criteria mentioned in global docking step, then the lowest I\_sc decoy will be the final local docking result. With these rules, **model1** and **2** both had three decoys among five top I\_sc decoys that were better than acceptable quality (**Figure 7a, b)**. However, N5 of **model3** was smaller than three, meaning that the local docking of **model3** has failed. To sum up, local dockings of **model1** and **2** were successful, and we filtered one possible binding pose from the results of **model1** and **2**, respectively.

We compared the sequence of **model1** and **2** with published structure PD-1 antibodies. The result showed the sequence of **model1** shared higher similarity to nivolumab, including 64% similarity on light chain and 76% similarity on heavy chain. According to the report, heavy chain has a higher binding frequency and contributes higher energy to the binding with antigen13. 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 12 hydrogen bonds and hydrophobic interactions. Unlike binding to nivolumab, PD-1 N loop did not act as a dominating role in binding with cemiplimab. Instead, BC loop contributed the majority of hydrogen bonds in the interaction with cemiplimab (5 out of 12). The residues of BC loop (N58, T59, S60, S62) formed five hydrogen bonds with cemiplimab HCDR1 (N31) and HCDR3 (N97, Y99) (**Figure 8**). Although N loop did not act as a dominant role in binding, it still contributed many hydrogen bonds (4 out of 12). Two residues of N loop (D29, R30) formed four hydrogen bonds with LCDR3 (S92) and heavy chain (CDR2 D56, and framework region Y58) (**Figure 8**). **Figure 8** shows that heavy chain N31 and N97 formed three and two hydrogen bonds with PD-1, respectively. From the calculation of per residue energy contribution, we knew that the residues Y58, D56, 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 Y58, D56, N31 on cemiplimab and R30, N58/S60/N102 on PD-1 were strong, with 2.6 Å, 2.9 Å, and 3.4/2.7/3.0 Å. In addition, PD-1 FG and DE loops contributed to the remaining hydrogen bonds. FG loop (S127) formed one hydrogen bond with HCDR3 (N97), 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 a strong hydrophobic interaction (-1.946 kcal/mol and -2.152 kcal/mol) with LCDR3 and HCDR1 (**Figure 9**).

There are several sources that can be used as input of local docking, such as the crystal structure of proteins, the ensemble of ten lowest-score RosettaAntibody homology models, or the Web Antibody Modeling (WAM) model45. 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 will be. 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 the antibody 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 with it, especially the antibody that binds to the same antigen as the target antibody. Besides, as we mentioned previously, the heavy chain plays an important role in antibody-antigen interaction. So, we suggest that heavy chain sequence similarity can be the priority during the sequence alignment, especially the alignment in CDR region.

***Step 6*: 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 ΔΔG of serval interface residues. According to the classification, hotspots and neutral residues were defined as ΔΔG more or less than 1 kcal/mol during the alanine mutation respectively34. Therefore, we found two hotspots, ASN97 and ILE98, located on the HCDR3 and their ΔΔG were 1.799 and 1.552 kcal/mol, respectively. Mutation to Ala made N97 lost three hydrogen bonds, which may responsible to the large binding energy change. On the contrary, I98 did not form any hydrogen bond with PD-1 but account for the great change of the energy. This meant I98 may contribute to the stability of the complex.

***Step 7*: *In silico* Affinity maturation for Cemiplimab**

Based on the knowledge of hotspots on cemiplimab, the *in-silico* mutation was performed by the antibody affinity maturation protocol. We 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. Comparing the result, we constructed a box plot to distinguish the difference of score, binding density, and binding energy between control group (original cemiplimab) and design group (after affinity maturation cemiplimab). From **Figure 10a**, we found that the score of design group was lower than the control group, indicating the structure of design group may be more energy favorable than that of control group. The binding density was negatively correlated with binding affinity46. Based on **Figure 10b,** we noticed that the binding density of design group was lower than the control, which specified that the designed complex may have a better binding affinity than the control. From **Figure 10c,** we concluded that the binding energy of control group was higher than the design, which could draw a conclusion that the design group had better stability.

In order to make further investigation, we calculated per residue score changing of mutated residues. **Figure 11** shows that the mutation of Y99A, I98T, and F27S had the highest per residue score among the heavy chain. On the other hand, the mutation of N93Y and F32K had the highest per residue score on 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 regions47. Combining the residues energy breakdown with structure analysis, we observed that heavy chain Tyr99 mutated to Ala and Phe27 mutated to Ser, and their fa\_dun energy had decrease significantly, which indicated that the rotamer may be more native-like. On the heavy chain, the mutation of Ile98 to Thr increased its fa\_rep, attributing to the shorter side chain of Thr than that of Ile. The shorter side chain increased the distance from other residues to decrease the repulsive energy. The mutation of Phe32 to Lys on light chain decreased the fa\_elec. Since Lys32 formed electrostatic attraction with Asp29 and decreased the energy (**Figure 12**), these mutations had reduced the energy of the complex and made it under a more favorable state.

**Key Points:**

* Antibody design protocol is a method that combines several antibody engineering techniques to provide a general and clear instruction aid in the antibody design.
* Antibody design protocol distinguishes itself from the normal antibody engineering methods for its characteristic that frees for the users, gives whole steps instruction for users, and provides a solution to address some of harsh situations in antibody design.
* The application of antibody design protocol in cemiplimab engineering shows a clear workflow and detailed solution to solve some of the problems, which gives insight into future antibody design.

**Conclusion**

In the present work, we have developed an antibody design protocol. The protocol used RosettaAntibody to model the 3D structure of antibodies. Global docking was used to generate possible binding poses when lacking binding information. Subsequently, SnugDock was applied to refine the antibody-antigen complex and generate a final binding pose. To facilitate affinity maturation and future study, alanine scanning was used to predict the hotspots on antibody. Finally, computational affinity maturation was used to improve antibody’s affinity and stability.

In the cemiplimab design case study, we used RosettaAntibody to generate the 3D structure of cemiplimab. Due to the lack of binding information of cemiplimab-PD-1 complex, two steps docking was applied to predict possible binding poses of the complex. With the complex binding pose, alanine scanning was used to predict the hotspots on antibody. Based on the prediction hotspots, *in-silico* affinity maturation was used to modify the antibody. According to the result, designed cemiplimab theoretically had better affinity and stability than the original one. The finding indicated by antibody design protocol can provide insights into the antibody engineering and cemiplimab further study.

**Legend of Figures**

**Figure 1. The workflow of the experiment.** First generate and prepare antibody/antigen structures. Then perform global docking and refine possible binding poses by local docking. After generating final binding pose, conduct alanine scanning to predict hotspots on antibody. Finally, perform *in silico* affinity maturation to design antibody and improve its properties.

**Figure 2. The crystal structure of PD-1 (extract 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 3. 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 4. 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.

**Figure 5. Three most representative PD-1 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.

**Figure 6. Three possible binding poses generate 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.

**Figure 7. The I\_sc vs Irmsd plot of model1 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 form docking funnel. (b) The local docking result of model2, eight decoys Irmsd among top10 I\_sc are smaller than 4 Å and form docking funnel.

**Figure 8. Detailed hydrogen bonds in cemiplimab/PD-1 complex.** Red dash line is hydrogen bond. The square frames show the detailed interaction between N, BC loops and cemiplimab.

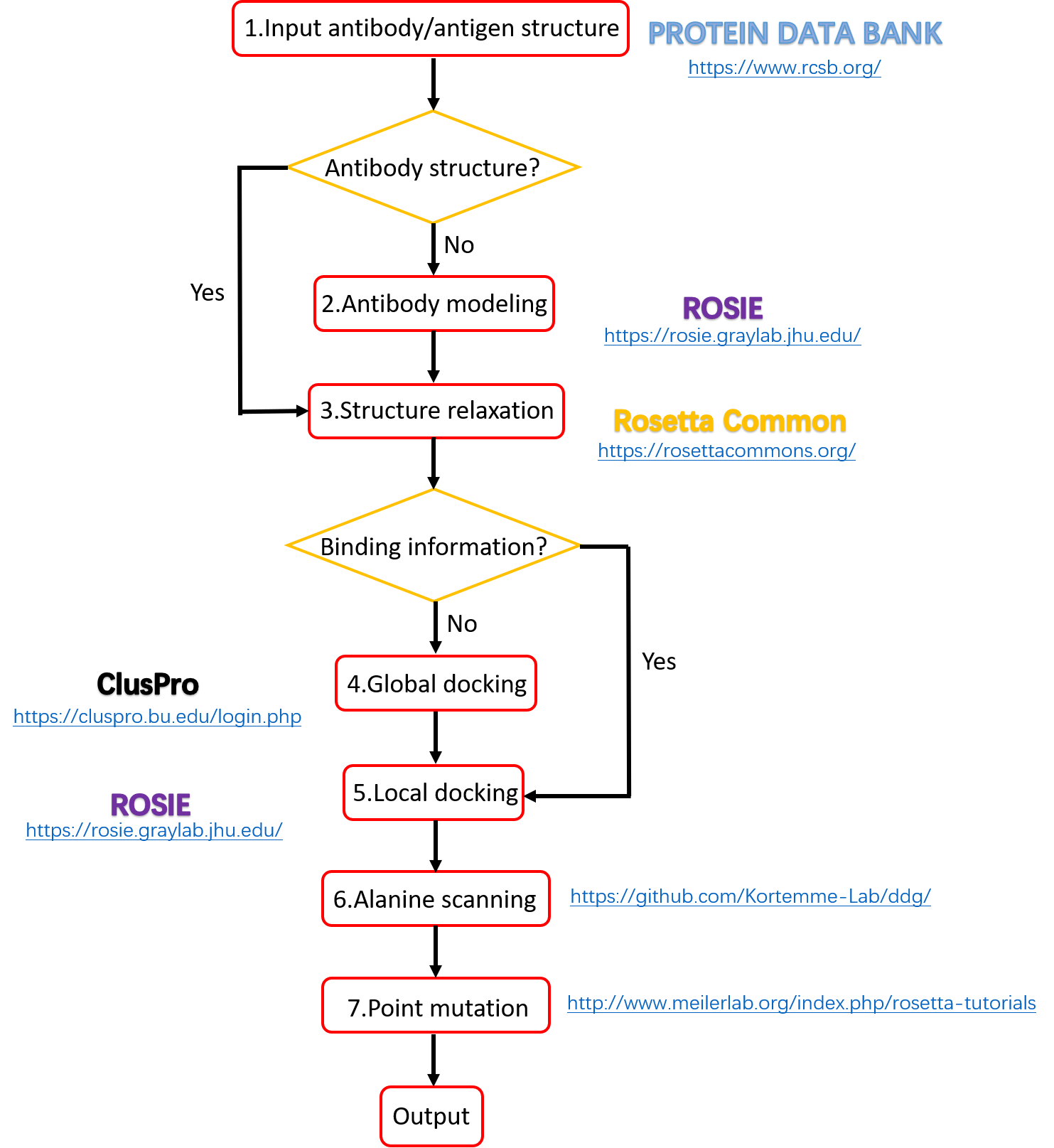
**Figure 9. Hydrophobic interaction in cemiplimab/PD-1 complex.** Blue dash line is hydrophobic interaction. The distances between PD-1 P28/LCDR N93 and PD-1 R104/HCDR N31 are short.

**Figure 10. Boxplot of comparison of three metrics between control and design groups.** (a) Comparison of total score between control and design groups. (b) Comparison of binding affinity between control and design groups. (c) Comparison of binding energy between control and design groups.

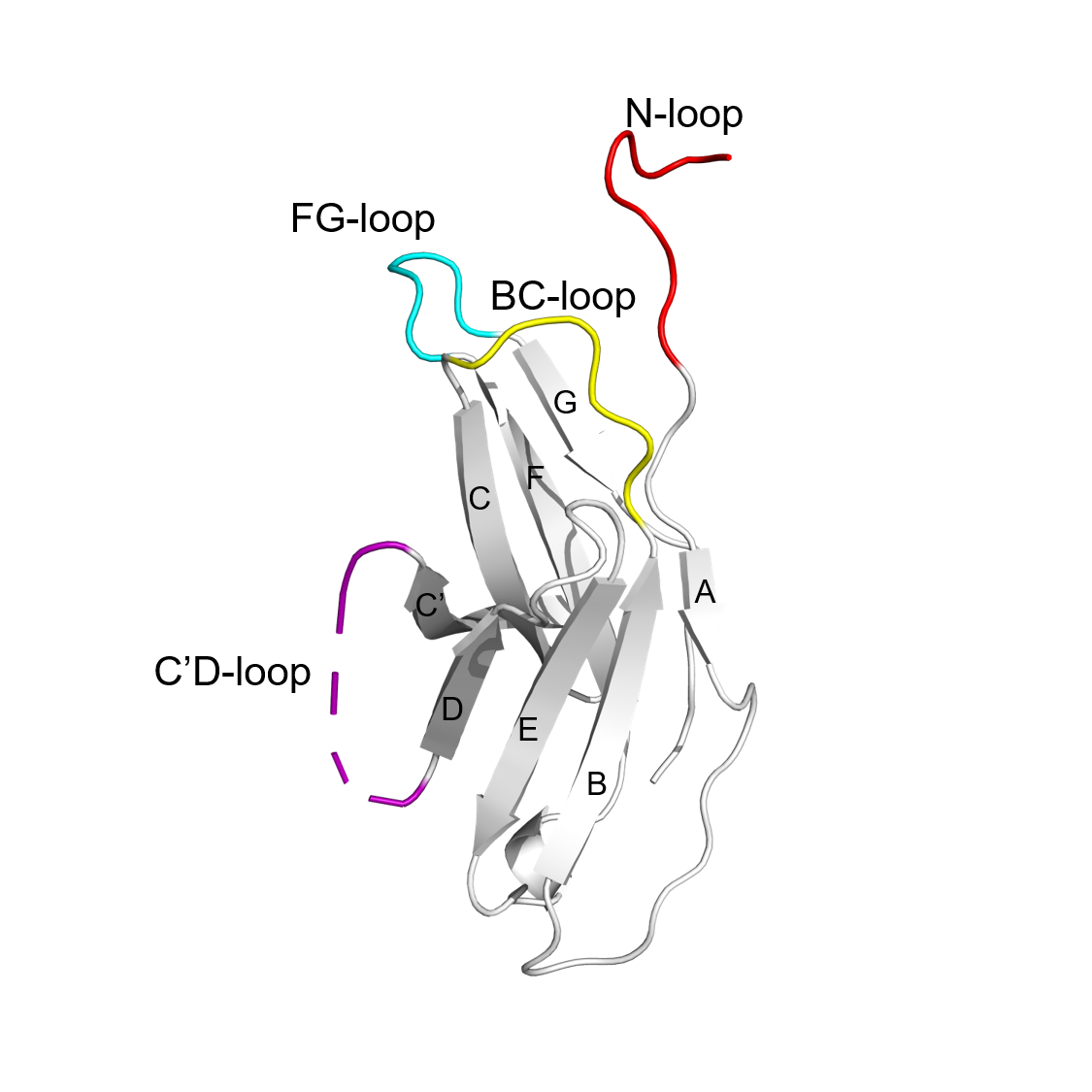
**Figure 11. Score changing of mutation residues on antibody during affinity maturation.** (a) The score changing of mutation residues on cemiplimab heavy chain. (b) The score changing of mutation residues on cemiplimab light chain.

**Figure 12. Important residues changing during affinity mutation.** Light chain in cyan, heavy chain in green, and PD-1 in pink. (a) Control group antibody (original cemiplimab). Important residues marked in purple. (b) Design group antibody (cemiplimab after affinity maturation). Important residues marked in orange.

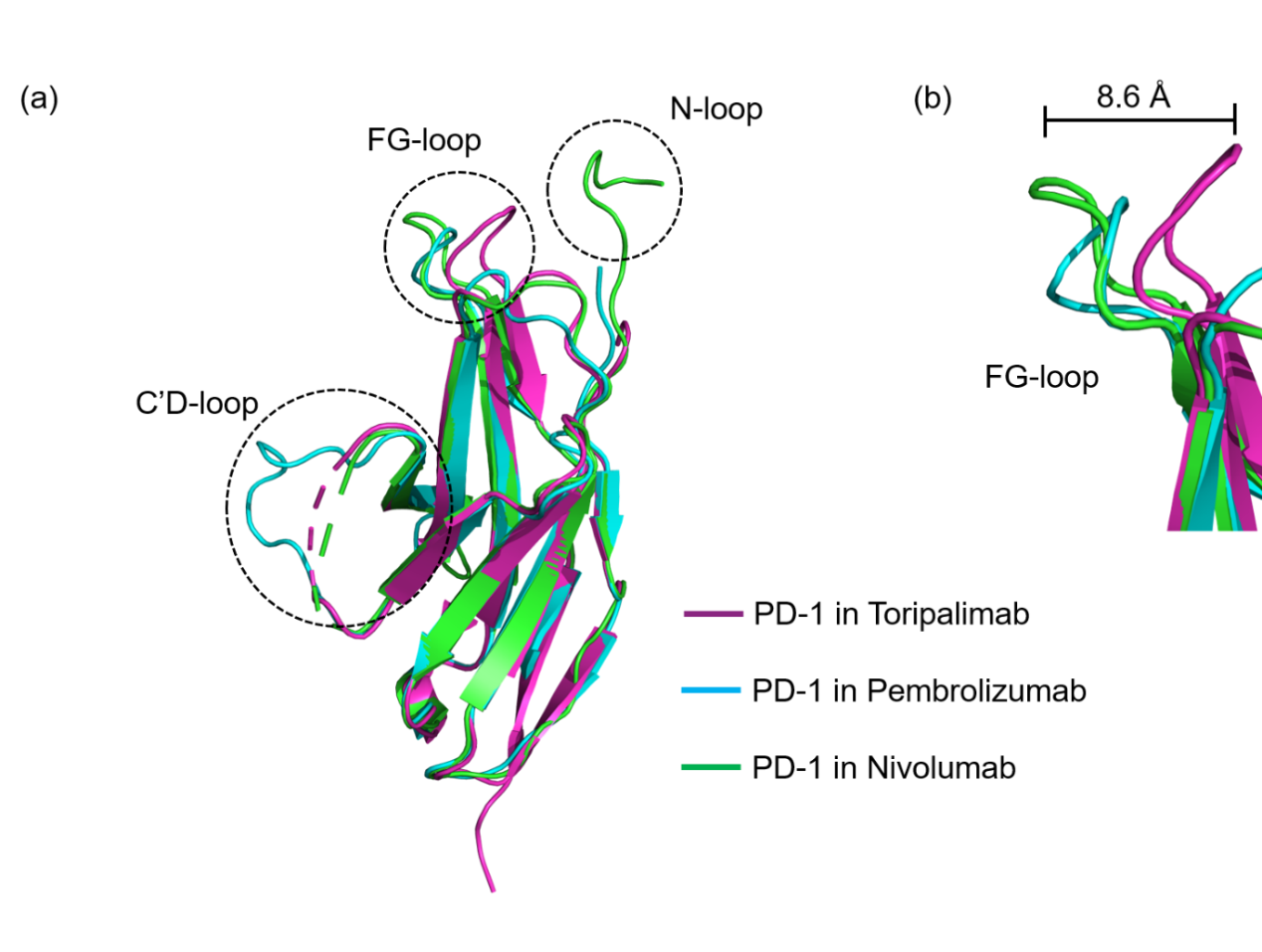
**Figure 1**



**Figure 2**

****

**Figure 3**

****

**Figure 4**

**A close up of a logo

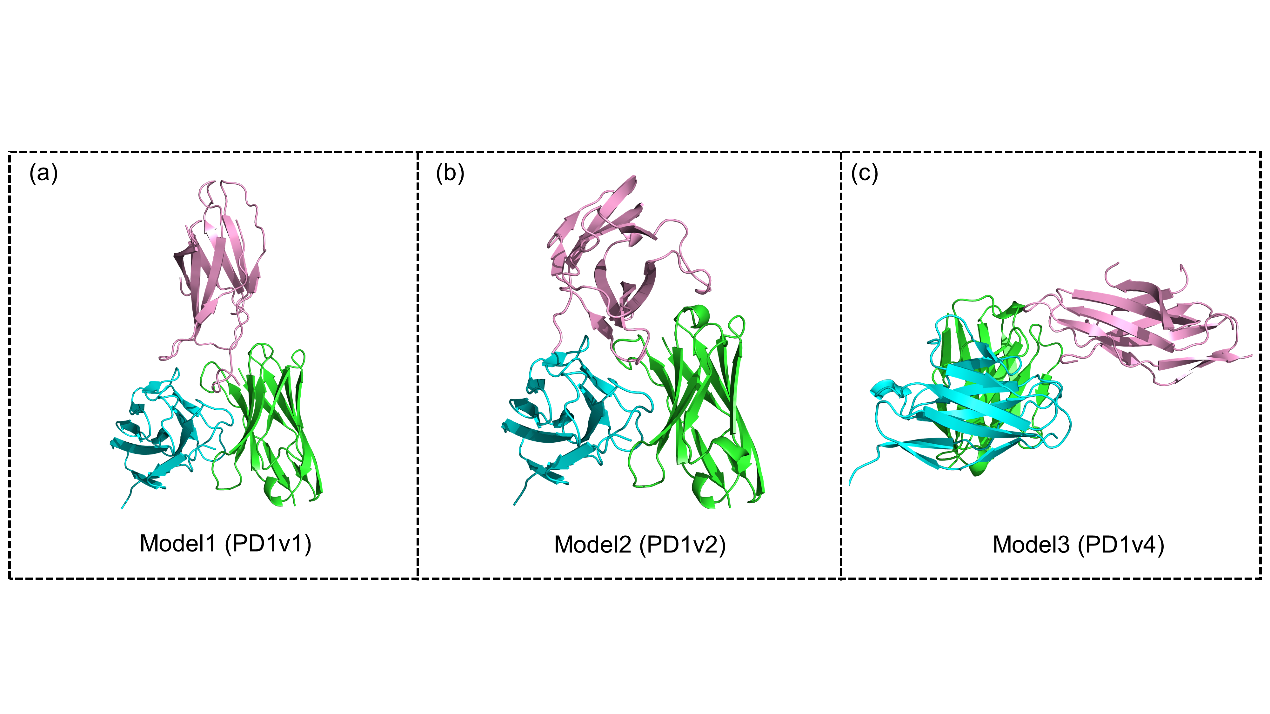
Description automatically generated**

**Figure 5**

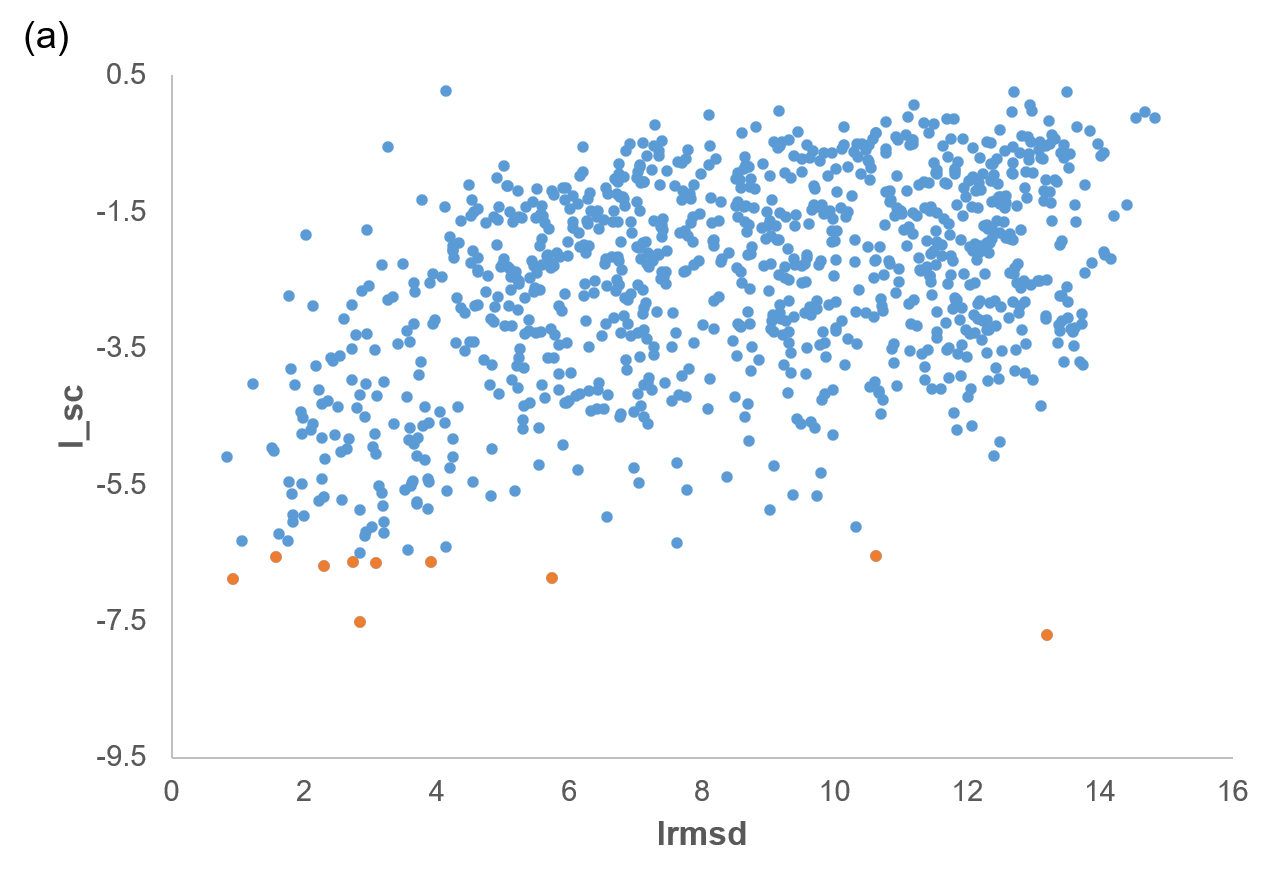
**A picture containing indoor, sitting, table, cat

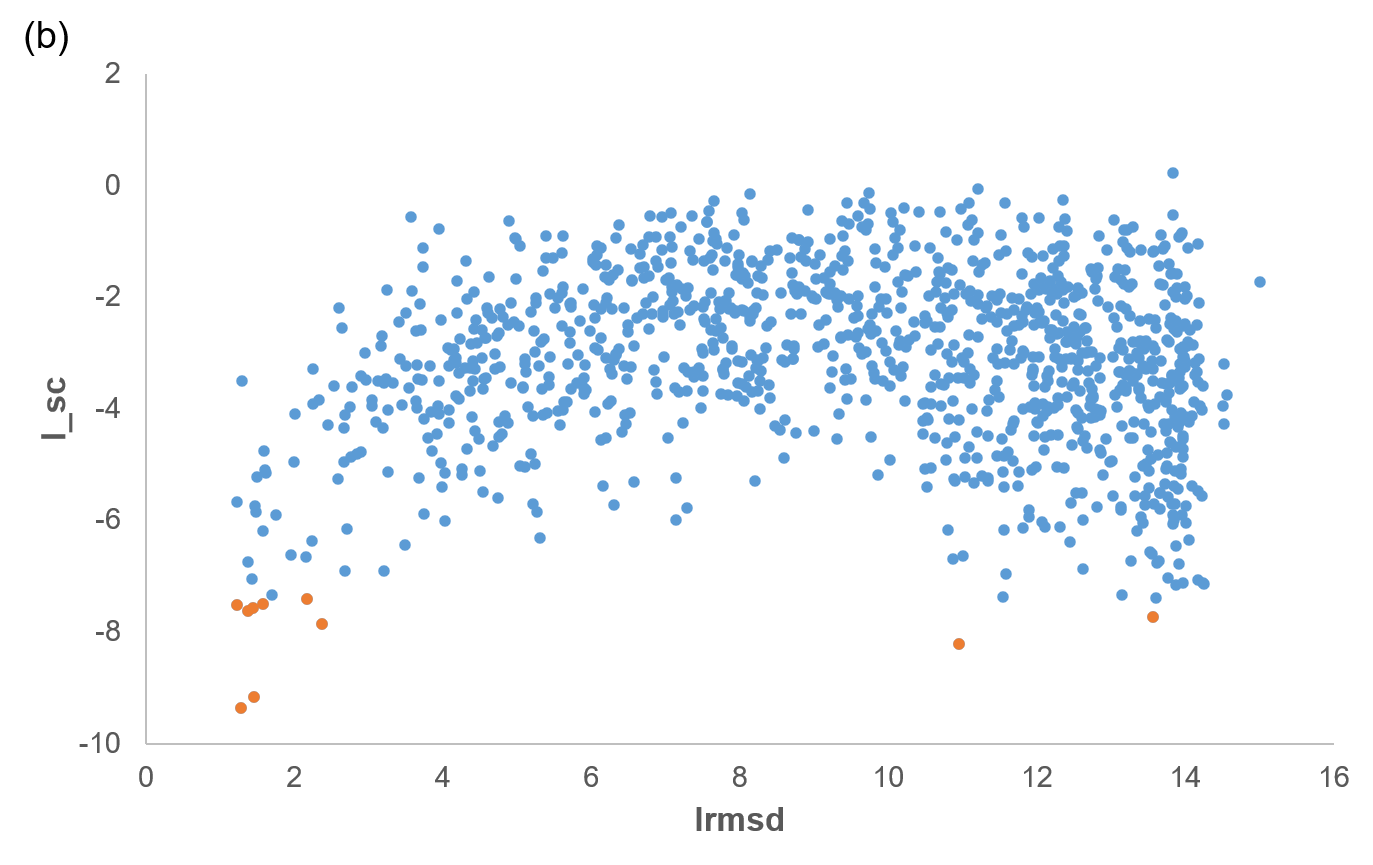
Description automatically generated**

**Figure 6**

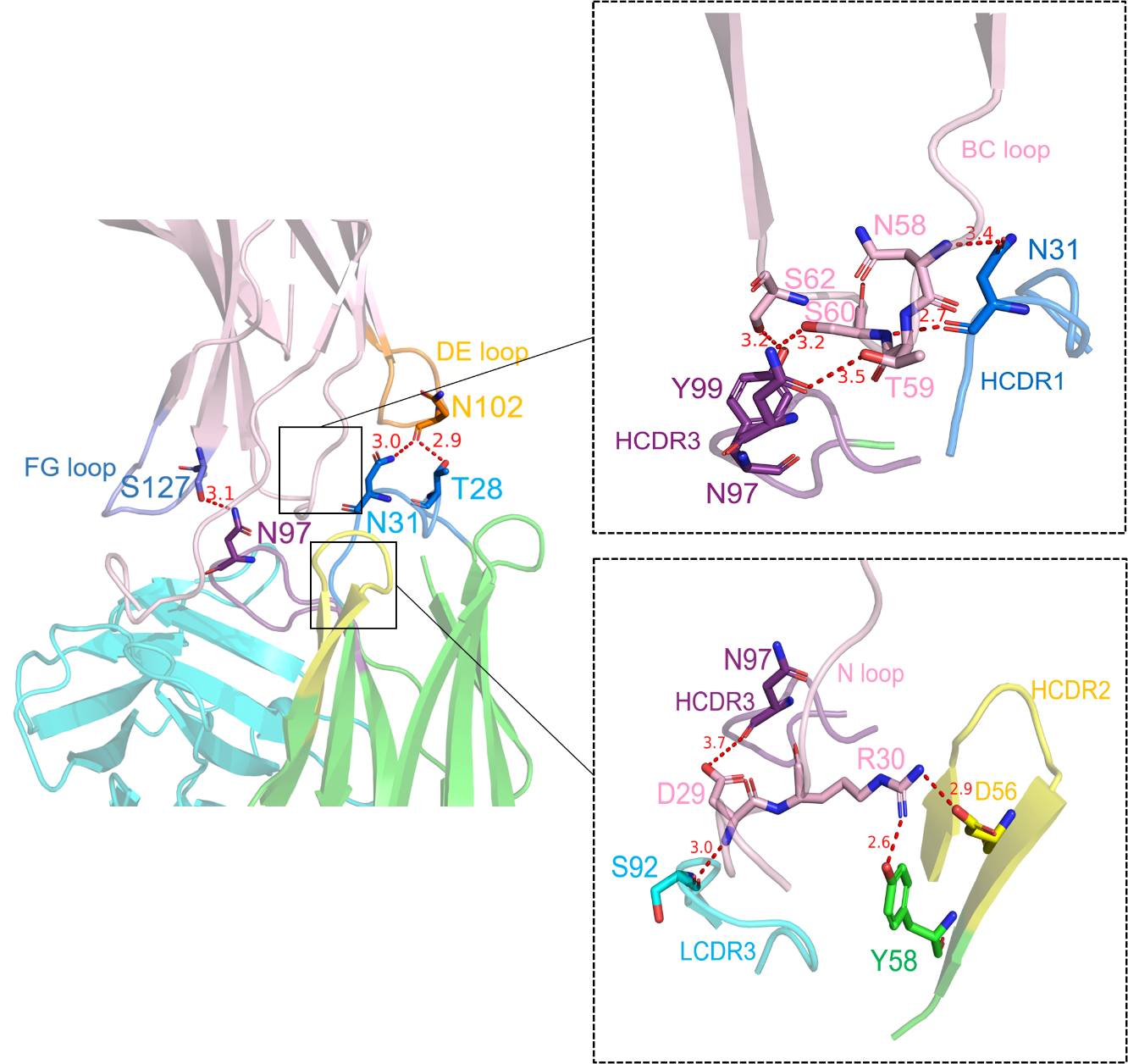
****

**Figure 7**

****

****

**Figure 8**



**Figure 9**

**A picture containing text

Description automatically generated**

**Figure 10**

**A screenshot of a social media post

Description automatically generated**

**A screenshot of a cell phone

Description automatically generated**

**A screenshot of a cell phone

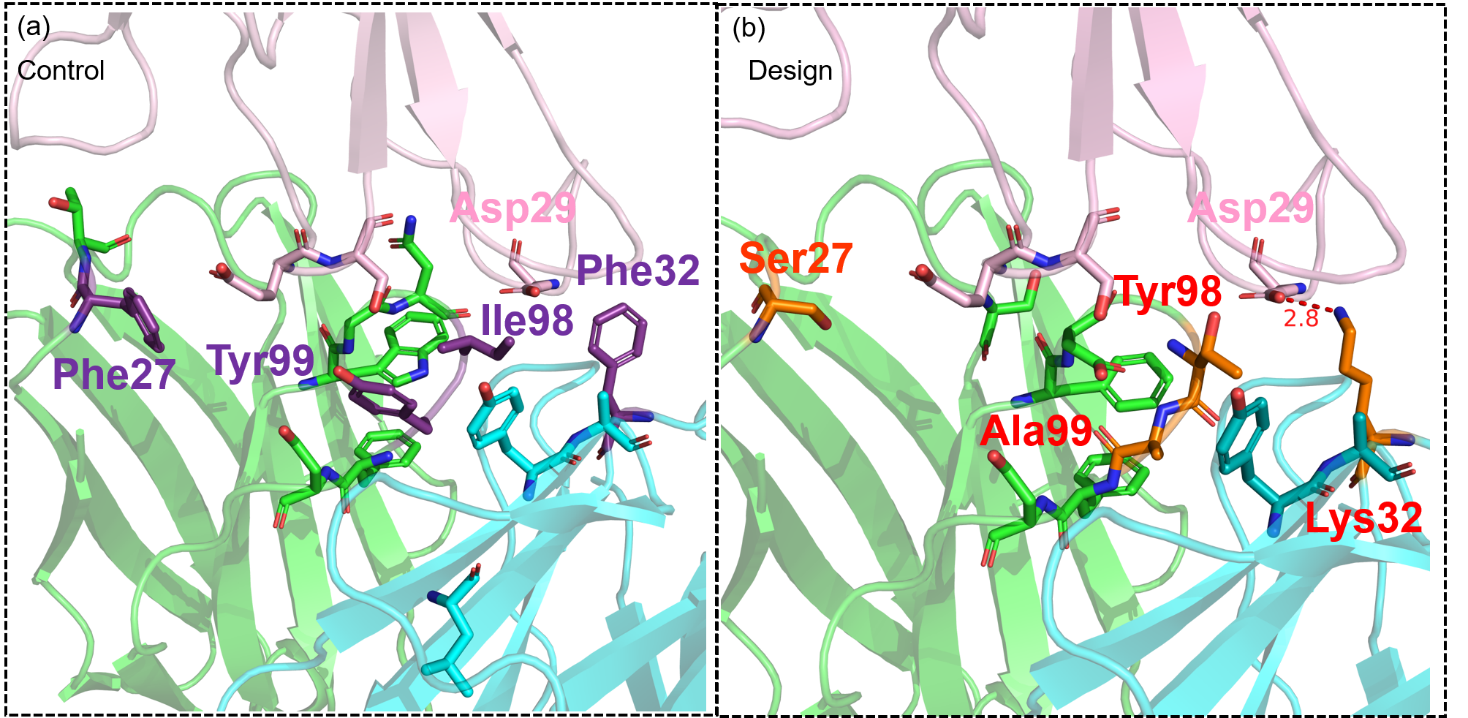
Description automatically generated**

**Figure 10**

**A screenshot of a cell phone

Description automatically generated**

**Figure 12**



**Reference**

1. Samish I, MacDermaid CM, Perez-Aguilar JM*, et al.* Theoretical and Computational Protein Design. *Annual Review of Physical Chemistry* **62**, 129-149 (2011).

2. Dahiyat BI, Benjamin Gordon D, Mayo SL. Automated design of the surface positions of protein helices. *Protein Science* **6**, 1333-1337 (1997).

3. Su A, Mayo SL. Coupling backbone flexibility and amino acid sequence selection in protein design. *Protein Science* **6**, 1701-1707 (1997).

4. Victora GD, Nussenzweig MC. Germinal centers. *Annu Rev Immunol* **30**, 429-457 (2012).

5. Diaz M, Casali P. Somatic immunoglobulin hypermutation. *Curr Opin Immunol* **14**, 235-240 (2002).

6. Babor M, Mandell DJ, Kortemme T. Assessment of flexible backbone protein design methods for sequence library prediction in the therapeutic antibody Herceptin–HER2 interface. *Protein Science* **20**, 1082-1089 (2011).

7. Koga N, Tatsumi-Koga R, Liu G*, et al.* Principles for designing ideal protein structures. *Nature* **491**, 222-227 (2012).

8. Marcos E, Basanta B, Chidyausiku TM*, et al.* Principles for designing proteins with cavities formed by curved β sheets. *Science* **355**, 201 (2017).

9. Li B, Zhao L, Wang C*, et al.* The protein-protein interface evolution acts in a similar way to antibody affinity maturation. *J Biol Chem* **285**, 3865-3871 (2010).

10. Kiyoshi M, Caaveiro JM, Miura E*, et al.* Affinity improvement of a therapeutic antibody by structure-based computational design: generation of electrostatic interactions in the transition state stabilizes the antibody-antigen complex. *PLoS One* **9**, e87099 (2014).

11. Poosarla VG, Li T, Goh BC*, et al.* Computational de novo design of antibodies binding to a peptide with high affinity. *Biotechnol Bioeng* **114**, 1331-1342 (2017).

12. Li T, Pantazes RJ, Maranas CD. OptMAVEn – A New Framework for the de novo Design of Antibody Variable Region Models Targeting Specific Antigen Epitopes. *PLOS ONE* **9**, e105954 (2014).

13. Robin G, Sato Y, Desplancq D*, et al.* Restricted Diversity of Antigen Binding Residues of Antibodies Revealed by Computational Alanine Scanning of 227 Antibody–Antigen Complexes. *Journal of Molecular Biology* **426**, 3729-3743 (2014).

14. Sivasubramanian A, Sircar A, Chaudhury S*, et al.* Toward high-resolution homology modeling of antibody Fv regions and application to antibody-antigen docking. *Proteins* **74**, 497-514 (2009).

15. Comeau SR, Gatchell DW, Vajda S*, et al.* ClusPro: an automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics* **20**, 45-50 (2004).

16. Sircar A, Gray JJ. SnugDock: Paratope Structural Optimization during Antibody-Antigen Docking Compensates for Errors in Antibody Homology Models. *PLOS Computational Biology* **6**, e1000644 (2010).

17. Canutescu AA, Dunbrack RL, Jr. Cyclic coordinate descent: A robotics algorithm for protein loop closure. *Protein Sci* **12**, 963-972 (2003).

18. Wang C, Schueler-Furman O, Baker D. Improved side-chain modeling for protein-protein docking. *Protein Sci* **14**, 1328-1339 (2005).

19. Hwang H, Vreven T, Janin J*, et al.* Protein–protein docking benchmark version 4.0. *Proteins: Structure, Function, and Bioinformatics* **78**, 3111-3114 (2010).

20. Zacharias M. Accounting for conformational changes during protein–protein docking. *Current Opinion in Structural Biology* **20**, 180-186 (2010).

21. Bonvin AMJJ. Flexible protein–protein docking. *Current Opinion in Structural Biology* **16**, 194-200 (2006).

22. Kuroda D, Gray JJ. Pushing the Backbone in Protein-Protein Docking. *Structure* **24**, 1821-1829 (2016).

23. Tyka MD, Keedy DA, André I*, et al.* Alternate states of proteins revealed by detailed energy landscape mapping. *J Mol Biol* **405**, 607-618 (2011).

24. Bower MJ, Cohen FE, Dunbrack RL, Jr. Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: a new homology modeling tool. *J Mol Biol* **267**, 1268-1282 (1997).

25. Katchalski-Katzir E, Shariv I, Eisenstein M*, et al.* Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. *Proc Natl Acad Sci U S A* **89**, 2195-2199 (1992).

26. Kozakov D, Brenke R, Comeau SR*, et al.* PIPER: an FFT-based protein docking program with pairwise potentials. *Proteins* **65**, 392-406 (2006).

27. Conte LL, Chothia C, Janin J. The atomic structure of protein-protein recognition sites11Edited by A. R. Fersht. *Journal of Molecular Biology* **285**, 2177-2198 (1999).

28. Chuang G-Y, Kozakov D, Brenke R*, et al.* DARS (Decoys As the Reference State) potentials for protein-protein docking. *Biophys J* **95**, 4217-4227 (2008).

29. Kozakov D, Hall DR, Xia B*, et al.* The ClusPro web server for protein–protein docking. *Nature Protocols* **12**, 255-278 (2017).

30. Lyskov S, Gray JJ. The RosettaDock server for local protein-protein docking. *Nucleic Acids Res* **36**, W233-W238 (2008).

31. Wang C, Bradley P, Baker D. Protein-protein docking with backbone flexibility. *J Mol Biol* **373**, 503-519 (2007).

32. Clackson T, Wells JA. A hot spot of binding energy in a hormone-receptor interface. *Science* **267**, 383-386 (1995).

33. Kortemme T, Kim DE, Baker D. Computational Alanine Scanning of Protein-Protein Interfaces. *Science&#039;s STKE* **2004**, pl2 (2004).

34. Kortemme T, Baker D. A simple physical model for binding energy hot spots in protein–protein complexes. *Proceedings of the National Academy of Sciences* **99**, 14116 (2002).

35. Dunbrack RL, Jr. Rotamer libraries in the 21st century. *Curr Opin Struct Biol* **12**, 431-440 (2002).

36. Ollikainen N, Smith CA, Fraser JS*, et al.* Chapter Four - Flexible Backbone Sampling Methods to Model and Design Protein Alternative Conformations. In: *Methods in Enzymology* (ed Keating AE). Academic Press (2013).

37. Shringari S, Giannakoulias S, Ferrie JJ*, et al.* Rosetta Custom Score Functions Accurately Predict ΔΔG of Mutations at Protein-Protein Interfaces Using Machine Learning. *bioRxiv*, 2020.2003.2017.996066 (2020).

38. Horita S, Nomura Y, Sato Y*, et al.* High-resolution crystal structure of the therapeutic antibody pembrolizumab bound to the human PD-1. *Sci Rep* **6**, 35297 (2016).

39. Tan S, Zhang H, Chai Y*, et al.* An unexpected N-terminal loop in PD-1 dominates binding by nivolumab. *Nature Communications* **8**, 14369 (2017).

40. Liu H, Guo L, Zhang J*, et al.* Glycosylation-independent binding of monoclonal antibody toripalimab to FG loop of PD-1 for tumor immune checkpoint therapy. *MAbs* **11**, 681-690 (2019).

41. Mandell DJ, Coutsias EA, Kortemme T. Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. *Nature Methods* **6**, 551-552 (2009).

42. Phillips JC, Braun R, Wang W*, et al.* Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **26**, 1781-1802 (2005).

43. Liberis E, Veličković P, Sormanni P*, et al.* Parapred: antibody paratope prediction using convolutional and recurrent neural networks. *Bioinformatics* **34**, 2944-2950 (2018).

44. Krawczyk K, Baker T, Shi J*, et al.* Antibody i-Patch prediction of the antibody binding site improves rigid local antibody–antigen docking. *Protein Engineering, Design and Selection* **26**, 621-629 (2013).

45. Whitelegg NRJ, Rees AR. WAM: an improved algorithm for modelling antibodies on the WEB. *Protein Engineering, Design and Selection* **13**, 819-824 (2000).

46. Chen J, Sawyer N, Regan L. Protein-protein interactions: general trends in the relationship between binding affinity and interfacial buried surface area. *Protein Sci* **22**, 510-515 (2013).

47. Julian MC, Li L, Garde S*, et al.* Efficient affinity maturation of antibody variable domains requires co-selection of compensatory mutations to maintain thermodynamic stability. *Scientific Reports* **7**, 45259 (2017).