



Cite this: *Chem. Sci.*, 2025, **16**, 17885

All publication charges for this article have been paid for by the Royal Society of Chemistry

Received 22nd May 2025
Accepted 11th August 2025

DOI: 10.1039/d5sc03707g
rsc.li/chemical-science

SAGERank: inductive learning of protein–protein interaction from antibody–antigen recognition

Chuance Sun, ^a Xiangyi Li, ^a Honglin Xu, ^a Yike Tang, ^b Ganggang Bai, ^b Yanjing Wang ^a and Buyong Ma ^{a,b}

Predicting Antibody–Antigen (Ab–Ag) docking and structure-based design represent significant long-term and therapeutically important challenges in computational biology. We present SAGERank, a general, configurable deep learning framework for antibody design using Graph Sample and Aggregate Networks. SAGERank successfully predicted the majority of epitopes in a cancer target dataset. In nanobody–antigen structure prediction, SAGERank, coupled with a protein dynamics structure prediction algorithm, slightly outperforms AlphaFold3. Most importantly, our study demonstrates the real potential of inductive deep learning to overcome the small dataset problem in molecular science. The SAGERank models trained for antibody–antigen docking can be used to examine general protein–protein interaction tasks, such as T Cell Receptor-peptide-Major Histocompatibility Complex (TCR-pMHC) recognition, classification of biological *versus* crystal interfaces, and prediction of ternary complexes of molecular glues. In the cases of ranking docking decoys and identifying biological interfaces, SAGERank is competitive with or outperforms state-of-the-art methods.

1 Introduction

The recognition of foreign antigens by antibodies is a crucial step in the immune response, and deciphering antibody–protein antigen recognition is of fundamental and practical significance. The antibody–protein antigen interaction is a subset of general protein–protein interaction (PPI), and both categories share similar principles but antibody–antigen interaction involves distinctly different sequences and structural features from non-antibody PPI.^{1,2} Consequently, general protein–protein docking programs, such as ZDOCK and HADDOCK, often need specialized treatment for their application in antibody–antigen docking. Especially, benchmark datasets or programs for antibody–antigen docking^{3–5} have been developed, including traditional docking methods and recent AI approaches.^{5,6} Overall, two major problems exist for the computational study of antibody–antigen interaction. The first one is the relatively small dataset available and used for antibody–antigen training. For example, the expanded benchmark for antibody–antigen docking has only 67 antibody–antigen cases, and the latest ABAG set just reaches 257.⁵ The second question is a related but more general situation in modern machine learning training: how to obtain meaningful information from small data.⁷ Protein language models may help protein interaction prediction,

for example hot spot prediction with limited data,⁸ but antibody and general protein may have different NLP patterns.^{9–11}

The high expressive power of deep neural networks enables efficient training with a large amount of data.¹² Enlarging the antibody–antigen dataset is certainly the right direction,^{5,13} and some approaches also use a combination of structure modeling and computational docking to create training data sets of antibody–antigen complexes. Still the problem stands: if we can use general protein–protein interaction data to study antibody–antigen recognition or *vice versa*? For the small-data challenges in molecular science, exploration of the latest advances in deep learning algorithms and new methods are needed.¹⁴

In the last few years, deep learning techniques have attracted much attention as a promising alternative to physicochemical based approaches.¹⁵ Compared to docking calculations, deep learning methods have improved the performance by learning the extracted features from protein–ligand complexes.¹⁶ They can automatically extract task-related features directly from data without handcrafted features or rules. Among many deep learning methods, various graph neural networks (GNNs) are especially suitable for questions related to protein structure and protein–protein interactions, as illustrated in a recent study on hierarchical graph neural networks for protein–protein interactions.¹⁷

Here, we tested the generalization of a Graph Sample and Aggregate Network (GraphSAGE) model, SAGERank initially trained for ranking antibody–antigen docking models using only an antibody–antigen dataset. Low-dimensional vector embeddings of nodes in large graphs have proved extremely useful as feature inputs for a wide variety of prediction and

^aEngineering Research Center of Cell & Therapeutic Antibody (MOE), School of Pharmacy, Shanghai Jiao Tong University, Shanghai 200240, China. E-mail: mbuyong@sjtu.edu.cn

^bShanghai Digiwiser Biotechnology, Limited, Shanghai 201203, China



graph analysis tasks.^{18,19} However, most embedding frameworks are inherently transductive and can only generate embeddings for a single fixed graph. These transductive approaches do not efficiently generalize to unseen nodes (e.g., in evolving graphs), and these approaches cannot learn to generalize across different graphs.²⁰ In contrast, GraphSAGE is an inductive framework that leverages node attribute information to efficiently generate representations on previously unseen data. Our work demonstrated that the SAGERank model trained for ranking antibody–antigen docking poses learned general

features of protein–protein interaction and can be directly applied to different tasks including distinguishing antibody–antigen pairing, TCR–pMHC recognition, and non-antibody protein–protein interactions.

2 Materials and methods

2.1. SAGERank architecture of antibody–antigen interfaces

SAGERank is built as a Python 3 package that allows end-to-end training on datasets of 3D antibody–antigen complexes. Fig. 1

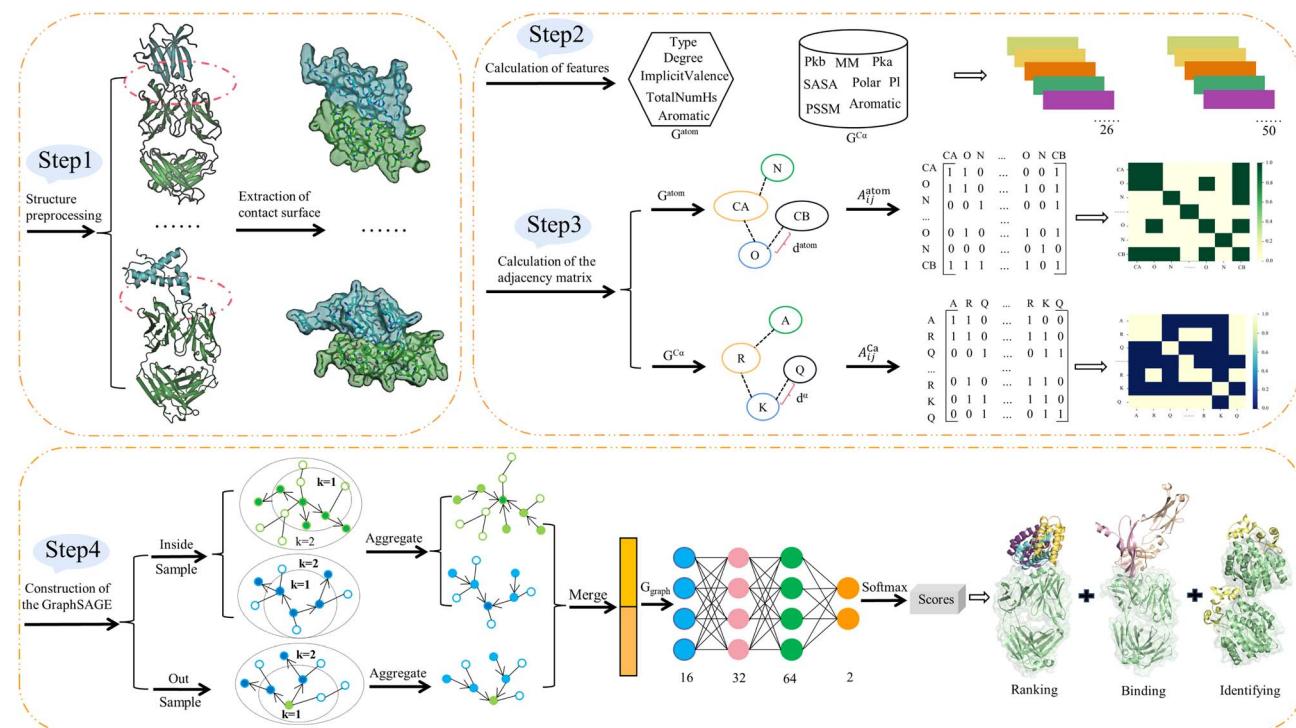


Fig. 1 The framework of SAGERank. step 1: SAGERank extracts the interface region of antibody–antigen complexes; step 2: constructing a graph based on atom nodes (26 features) and residue nodes (50 features); step 3: in order to save memory, we compress the sparse adjacency matrix using COO (coordinate format). The COO contains three arrays that store the row index, column index, and value of all non-zero elements; step 4: the graph representation of the interface region is split into two sub-graphs: the internal graph (including either receptor or ligand) and the external graph (including both receptor and ligand). The internal and external graphs are sequentially passed to 4 consecutive SAGEConv layers and 1 global mean pooling layer. The two final graph representations are merged before applying the softmax function for output.

Table 1 Features computed in the residue network G^{Ca}

Name of the feature	Full name	Description	Dimension
Type	Residue type	One-hot encoded	20
Aliphatic	Residue aliphatic	One-hot encoded	1
Aromatic	Residue aromatic	One-hot encoded	1
Polar neutral	Residue polar neutral	One-hot encoded	1
Charged	Acidic/basic charged	—	2
Weight	Residue weight	—	1
Pka	The negative of the logarithm of the dissociation constant for the -COOH group	—	1
Pkb	The negative of the logarithm of the dissociation constant for the -NH ₃ group	—	1
Isoelectric point	Residue isoelectric point	—	1
Hydrophobicity	Hydrophobicity of the residue (pH = 2)	—	1
Hydrophobicity	Hydrophobicity of the residue (pH = 7)	—	1
PSSM	Position-specific scoring matrix	—	20
SASA	Solvent-accessibility surface area	—	1



Table 2 Features computed in the atomic network G^{atom}

Name of the feature	Full name	Description	Dimension
Type	Atom type	One-hot encoded	10
Degree	Number of atomic connections	One-hot encoded	6
TotalNumHs	Number of hydrogen atoms	One-hot encoded	5
ImplicitValence	Atomic implicit valence	One-hot encoded	6
Aromatic	Atomic aromatic	—	1

shows the architecture of the network. The framework consists of two main parts: one focusing on data pre-processing and featurization and the other on the training, evaluation, and testing of the neural network.

We have considered two types of networks, one based on amino acids represented by Ca coordinates (G^{Ca}) and another based on specific atoms in each amino acid (G^{atom}). Starting from the 3D structures of Ab–Ag complexes, the interface region is identified as a set of residues located within 10.0 Å of any residues of the other antibody or antigen (Fig. 1). Graph G is defined by V (node set), E (set of edges), and A (adjacency matrix). Therefore, the amino acids in the interface region and the atoms that make up the amino acids represent two different types of nodes. The residue-based and atomic node features are listed in Tables 1 and 2, respectively. Overall, 50-bit features and 28-bit features are used to describe the residue nodes and atom nodes, respectively.

The adjacency matrix is constructed using the following procedure. For the amino acid network G^{Ca} with N nodes, the adjacency matrix A^{Ca} has a dimension of $N \times N$ (eqn (1)). Within the ligand (antigen) and receptor (antibody) $A_{ij}^{\text{Ca}} = 0$ if the Euclidean distance d_{ij}^{Ca} between the i -th node and the j -th node is greater than 4.5 Å, and $A_{ij}^{\text{Ca}} = 1$ otherwise. In addition, we also

d_{ij}^{atom} between the i -th node and the j -th node is greater than 8 Å, and $A_{ij}^{\text{atom}} = 0$ otherwise. Intraprotein edges (4.5 Å) capture covalent bonds and close-range contacts, while interprotein edges (10 Å) accommodate larger interface distances typical of antibody–antigen interactions. This differentiation aligns with physicochemical principles and improves model accuracy.

$$A_{ij}^{\text{Ca}} = \begin{cases} 1 & \begin{cases} \text{if } d_{ij}^{\text{Ca}} \leq 10 \text{ Å and } i \in \text{receptor and } j \in \text{ligand} \\ \text{if } d_{ij}^{\text{Ca}} \leq 4.5 \text{ Å and } i, j \in \text{ligand} \\ \text{if } i, j \in \text{CDR loops and } i, j \in \text{receptor} \end{cases} \\ 0 & \begin{cases} \text{if } d_{ij}^{\text{Ca}} > 10 \text{ Å and } i \in \text{receptor and } j \in \text{ligand} \\ \text{if } d_{ij}^{\text{Ca}} > 4.5 \text{ Å and } i, j \in \text{ligand} \\ \text{if } i, j \in \text{non-CDR loops and } i, j \in \text{receptor} \end{cases} \end{cases} \quad (1)$$

Node aggregation and update follow the standard GraphSAGE algorithm. The core steps of GraphSAGE are neighbor sampling and feature aggregation. The forward propagation algorithm for GraphSAGE is as follows:

$$A_{ij}^{\text{atom}} = \begin{cases} 1 & \begin{cases} \text{if } d_{ij}^{\text{atom}} \leq 8 \text{ Å and } i \in \text{receptor and } j \in \text{ligand} \\ \text{if } i \text{ and } j \text{ are connected by a covalent bond or if } i = j \text{ and } i, j \in \text{receptor or } i, j \in \text{ligand} \end{cases} \\ 0 & \begin{cases} \text{if } d_{ij}^{\text{atom}} > 8 \text{ Å and } i \in \text{receptor and } j \in \text{ligand} \\ \text{if } i \text{ and } j \text{ are not connected by a covalent bond or if } i = j \text{ and } i, j \in \text{receptor or } i, j \in \text{ligand} \end{cases} \end{cases} \quad (2)$$

take into account the specific properties of the antibody CDR region in the receptor where $A_{ij}^{\text{Ca}} = 1$ if the i -th node and the j -th node belong to the CDR loops, and $A_{ij}^{\text{Ca}} = 0$ otherwise. Between the receptor and ligand, $A_{ij}^{\text{Ca}} = 1$ if the distance d_{ij}^{Ca} between the i -th node and the j -th node is greater than 10 Å, and $A_{ij}^{\text{Ca}} = 0$ otherwise. For the atomic network G^{atom} with M nodes, the adjacency matrix A^{atom} has a dimension of $M \times M$ (eqn (2)). Within the receptor and ligand where $A_{ij}^{\text{atom}} = 1$ if atom i and atom j are connected by a covalent bond, and $A_{ij}^{\text{atom}} = 0$ otherwise. Between receptor and ligand $A_{ij}^{\text{atom}} = 1$ if the distance

In the algorithm, the first for loop is used to traverse the number of layers, and the second for loop is used to traverse all nodes in the Graph. Sampling is performed on the neighbors of each node v to obtain N_v . Next, the embedding of neighbor nodes are aggregated through AGGREGATE_k to obtain $h_{N(v)}^k$. Then $h_{N(v)}^k$ is spliced with the current embedding h_v^{k-1} of the target node, and assigned to h_v^k after nonlinear transformation, thereby completing an update of the target node v . When the outer loop ($k = 1 \dots K$) ends, the node v will complete the information aggregation of k -order neighbors.



GraphSAGE embedding generation algorithm

Input: Graph G ; input features; depth K ; weight matrices $W^k, \forall k \in \{1, \dots, K\}$; non-linearity σ ;
 Different aggregator functions $\text{AGGREGATE}_k \forall k \in \{1, \dots, K\}$;
 neighborhood function $N: v \rightarrow 2^V$

Output: Vector representations z_v for all $v \in V$

```

1  $h_v^0 \leftarrow X_v, \forall v \in V;$ 
2 for  $k = 1 \dots K$  do
3   for  $v \in V$  do
4      $h_{N(v)}^k = \text{AGGREGATE}_k(\{h_u^{k-1}, \forall u \in N(v)\})$ 
5      $h_v^k = \sigma(W^k \cdot \text{CONCAT}(h_v^{k-1}, h_{N(v)}^k))$ 
6   end
7    $h_v^k = h_v^k / \|h_v^k\|_2, \forall v \in V$ 
8 end
9  $z_v = h_v^k, \forall v \in V$ 

```

2.2. Training hyperparameters and quality metrics

The training hyperparameter settings are shown in Table 3. The entire method is implemented using the Pytorch-geometric deep learning library.

The hit-rate and success-rate are used to evaluate the performance of different scoring functions for ranking docked decoys. The hit-rate is defined as the percentage of near-native (models with $i\text{RMSD} \leq 4 \text{ \AA}$) models in the top-ranked models for a specific complex:

$$\text{Hit rate} = \frac{N(\text{hits})}{M} \quad (3)$$

where $N(\text{hits})$ is the number of near-native models among the top models and M is the total number of near-native models for this case. The hit-rate was calculated for each individual case in our test set and the higher the value, the higher the accuracy of ranking decoys. The success-rate is defined as the percentage of complexes for which at least one near-native model is found in the top K selected models. It is therefore defined as:

$$\text{Success rate} = \frac{N(\text{successful hits})}{K} \quad (4)$$

where $N(\text{successful hits})$ is the number of cases with at least one near-native model among top models, and K is the total number of cases. The commonly used statistical indicators for measuring binary classification problems are accuracy (SI eqn (1)), precision (SI eqn (2)), recall (SI eqn (3)) and F_1 (SI eqn (4)). Among them, TN is true negative, TP is true positive, FN is false negative and FP is false positive. When the number of positive

and negative samples in the dataset is relatively balanced, accuracy can be used as an evaluation indicator, but when the number of positive and negative samples is unbalanced, F_1 values and ROC are more trustworthy. The ROC curve is defined as the fraction of the true positive rate as a function of the fraction of the false positive rate while navigating through the ranking provided by the scoring function. The AUC is the integral of the ROC curve and is equal to 1 for an ideal classifier and 0.5 for a random classifier.

3 Results

3.1. Training and performance of the SAGERank docking model to rank antibody–antigen docking poses

We collected a dataset of 287 different Ab–Ag complexes with sequence identity <95%. We used megadock²¹ to generate a set of docking poses of various RMSDs for all Ab–Ag complexes. In the process of docking, we adopted a semi-flexible docking method, that is, the receptor conformation does not change, and the ligand rotates in the CDR region of the antibody. 20 000 different conformations were generated for each Ab–Ag complex. Therefore, based on interface iRMSD, we divided the decoys of docking into two categories: negative models posed with iRMSD greater than 4 Å are non-native poses, and those with iRMSD less than 4 Å are near-native poses. In the end, 287 different Ab–Ag complexes generated a total of 455 420 docking decoys with a 1 : 3 ratio of positive and negative samples. The conformations were divided into training (80%), validation (10%), and test (10%) sets. In addition, in order to compare with other docking methods, we constructed 10 groups of Ab–Ag complexes and 8 protein–protein complexes with 17 552 and 10 707 docking decoys, respectively.

As can be seen in SI Fig. 1–1, the model with atoms as nodes achieves much better ranking accuracy than that with amino acids as nodes. There are two factors contributing to the different performances. Firstly, the graph core and the interaction between interfaces can be more accurately represented and captured using a larger number of nodes in the atomic network G^{atom} than in the residue-based network G^{Ca} . Secondly,

Table 3 The hyperparameter settings using human experience

Hyperparameter	Setting
Epoch	50
Batchsize	512
Learning rate	0.0003
Optimizer	Adam
Loss function	Cross entropy loss
The layers of GraphSAGE	4



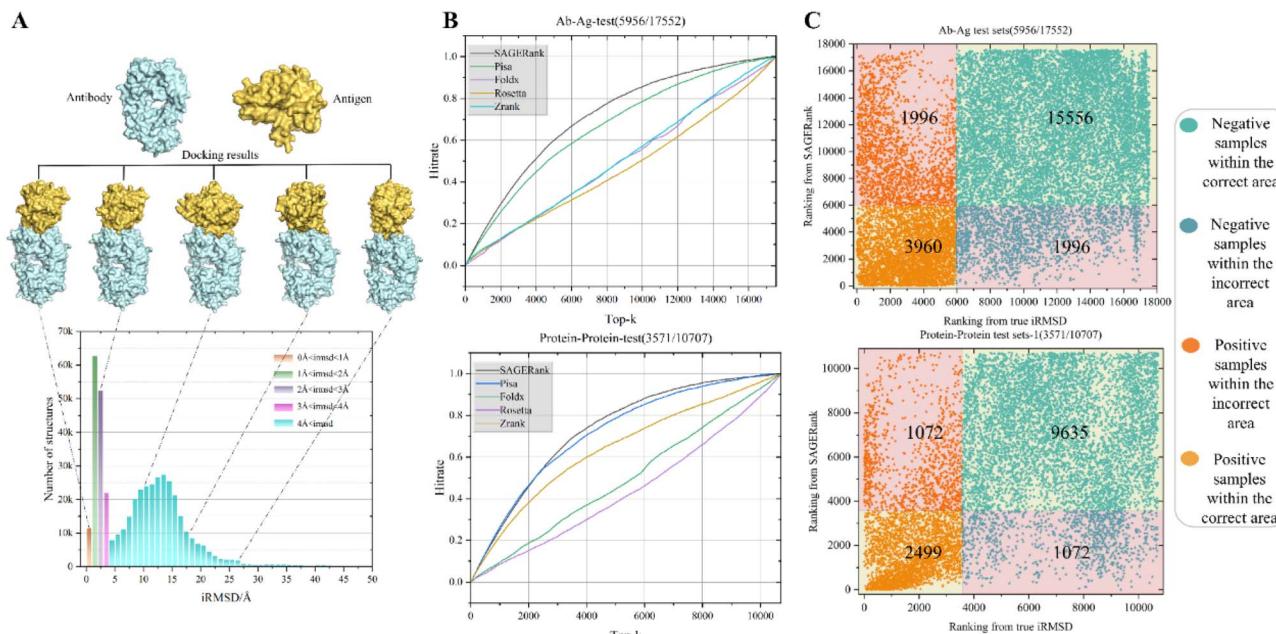


Fig. 2 SAGERank applied to the ranking of docking decoys. (A) iRMSD distribution of all docking decoys in the training set. (B) Performance of five different evaluation methods on the Ab–Ag docking decoy dataset; the number of positive and negative samples was 5956 and 11 596, respectively (top). Performance of five different evaluation methods on the protein–protein docking decoy dataset; the number of positive and negative samples was 3571 and 7136, respectively (bottom). (C) The ranking comparison between SAGERank and true iRMSD for Ab–Ag test sets (top) and protein–protein test sets (bottom).

the atomic network G^{atom} captures essential physicochemical features underlying antibody–antigen recognition.

We then compared the SAGERank docking model with four leading scoring functions in ranking protein–protein docking poses: Zrank,²² Pisa,²³ FoldX,²⁴ and Rosetta²⁵ (Fig. 2B, SI 1-2, 1-3 and Table 1-1). Clearly, the SAGERank docking model exhibits superior performance compared to other methods in the Ab–Ag docking decoys set. A detailed analysis of the success rates of the SAGERank docking model and Pisa in specific cases, such as the 7MLH case (SI Fig. 1-4) and the 2B42 case (SI Fig. 4-3), further validates its outstanding performance. In these instances, when we select the consensus hits from the top 200 poses predicted by both the SAGERank docking model and Pisa, we observe only a minimal number of false positive hits. This finding underscores the reliability and accuracy of our model in identifying true positive docking poses, even when compared to other leading scoring functions.

3.2. Cognate antigen identification: predicting if antibody can bind to an antigen in right epitopes

A dataset with shuffled antibody–antigen pairing was generated using the Megadock as negative antibody–antigen recognition. Firstly, sequence alignment of 287 different antigen–antibody complexes was performed to remove the entries with sequence similarity between antigens >70%, and between antibodies >95%. After that, we obtained 230 Ab–Ag complexes of which 200 complexes were used as the training set and the remaining 30 complexes were used as the test set. Next, the positive samples consisted of decoys with iRMSD in the 3.5 Å range

generated by docking the antibody with its own antigen. The negative sample dataset consists of two parts: one included decoys with iRMSD greater than 8 Å generated by docking the antibody with its own antigen, and the other is the structure generated by docking antibodies with the remaining 199 antigens. In the end, the dataset produced a total of 369 932 complex structures, with a ratio of positive and negative samples of about 1 : 5 (Fig. 3A). Among them, 80% were used as the training set, 10% as the verification set, and 10% as the test set.

Because the proportion of positive and negative samples in the dataset is not balanced, it is more reliable to use the AUC for model evaluation in this case, with AUC reaching 0.82 in the test set (SI Fig. 2-1), slightly smaller than the AUC obtained in the docking pose ranking model. Fig. 3B shows the distribution of SAGERank model2 scores for positive and negative samples in test sets. The average scores for all positive and negative samples are 0.57 and 0.20, respectively. It is obviously seen from Fig. 3C that when the threshold is 0.3, the F_1 score can achieve a maximum value of 0.74. Therefore, choosing a threshold of 0.3 on the dataset for determining whether an antibody–antigen can be recognized maximizes the performance of the SAGERank model2. Fig. 3E shows the confusion matrix of SAGERank model2 on the test dataset with a threshold of 0.3. SAGERank model2 accurately classified 9813 out of 11 307 negative samples and 3685 out of 4642 positive samples.

We further tested the model to examine antibody–antigen complexes for three targets with a large number of structures available, including 222 structures for the SARS-CoV-2 target, 41



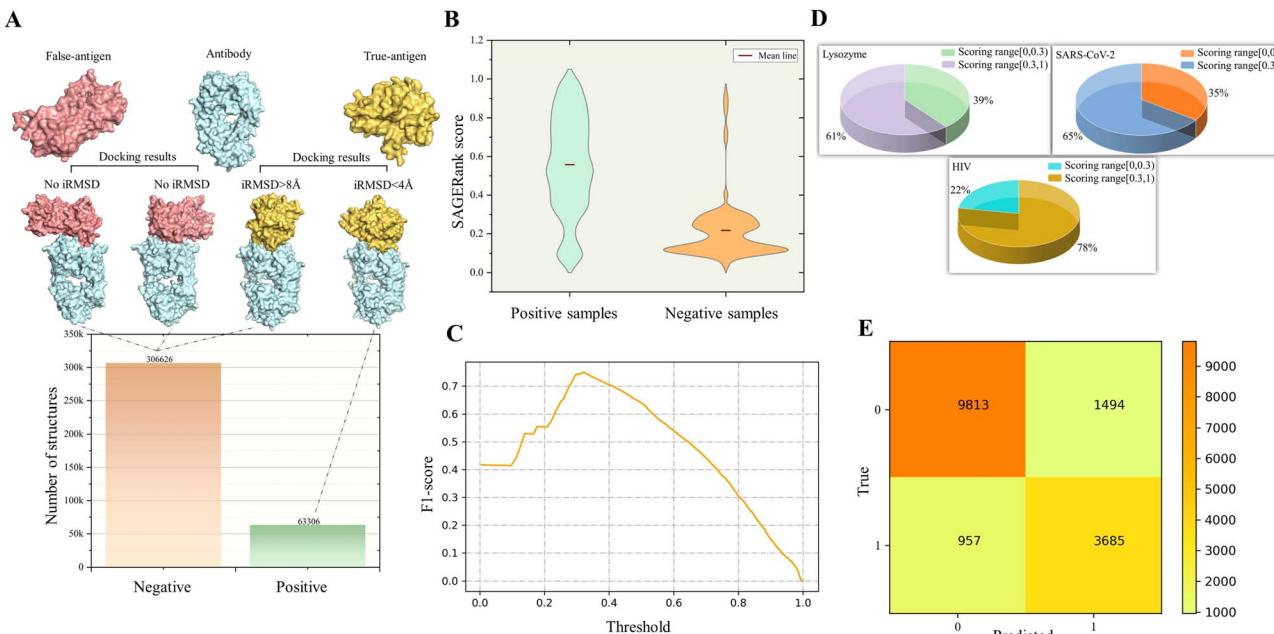


Fig. 3 SAGERank applied for detecting the antibody–antigen binding problem. (A) The number of complexes on the training set where antibodies bind to antigens (positive samples) and complexes where antibodies cannot bind to antigens (negative samples). (B) SAGERank model2's scoring distribution for positive and negative samples on the test set. (C) The F_1 -score curve with threshold variation. (D) Recognition success rate of SAGERank model2 for antibody antigen complexes of SARS-CoV-2, HIV and lysozyme. (E) The confusion matrix of SAGERank model2 on the test dataset.

structures for the lysozyme target, and 367 structures for the HIV target. All these 630 positive samples were chosen to test the distinguishing ability of SAGERank model2. The recognition success rates of the model for three targets HIV, SARS-CoV-2 and Lysozyme are 77.7%, 65.3%, and 61.0%, respectively (Fig. 3D). Overall, the SAGERank model2 achieved a considerable success rate in determining whether antibodies and antigens can bind. It is interesting to compare the performance of the SAGERank docking model trained with antibody–antigen docking poses. The success rates of the docking model for three targets, SARS-CoV-2, HIV, and lysozyme, are 50%, 48.6%, and 70.0%, respectively (SI Fig. 2-3). It is worth noting that the recognition accuracy of the docking model for antigens and antibodies in the lysozyme group is higher than that of the specifically trained SAGERank model2. Finally, we constructed positive protein–protein and pseudo protein–protein binding datasets. It can be seen in SI Fig. 2-4B that the SAGERank docking model still has good distinction accuracy; it correctly classified 153 out of 230 positive protein–protein and 304 out of 438 negative protein–protein bindings.

3.3. Extension of antibody–antigen trained SAGERank to TCR-pMHC recognition

Antibody and TCR share many similarities, but they have distinct different CDR patterns and antigen interactions. Our first test of the inductive learning ability of SAGERank is to check if it can be used to discriminate TCR-pMHC recognition.

We collected 153 pairs of TCR-pMHC1 ternary complexes from the STCRDab database. Unlike typical protein–protein or

antibody–antigen complexes with diverse interfacial morphologies, TCR-pMHC1 ternary complexes not only share similar interfaces but also conserve a diagonal orientation, presenting a further challenge for accurate prediction. To generate negative samples with rational binding patterns, TCRs were randomly picked and aligned to other non-cognate pMHC1 complexes, followed by 1000 steps of OpenMM energy minimization to optimize interface conformations. Here, we employ a contact map format to illustrate the conformational differences at the interfaces of positive and negative samples. Using complex 3bnq (PDB ID) as an illustrative example, it becomes evident that, in comparison to the TCR-pMHC1 complex of negative samples, the TCR-pMHC1 complex of positive samples exhibits a stronger mutual contact between the peptide interface and the TCR interface (Fig. 4A). We compared the performance of the SAGERank docking model with two different types of networks, the atom-based (G^{atom}) features and the amino-acid-based (G^{Ca}) features. The results indicate that the G^{Ca} network better discriminates between homologous and non-homologous TCRs, with ROCAUC and PRAUC values of 0.6467 and 0.6739, respectively (Fig. 4C). The G^{atom} features exhibited a high false positive rate (Fig. 4B) and the difference in the scores between positive and negative samples is more pronounced (Fig. 4D). We attribute this to the highly conserved binding geometry of TCR-pMHC1 interfaces, which leads to nearly identical atomic-level graph representations across different complexes. These structurally similar inputs produce indistinguishable embeddings in the GNN, limiting the model's ability to discriminate binding specificity. In contrast, residue-level (G^{Ca}) features implicitly

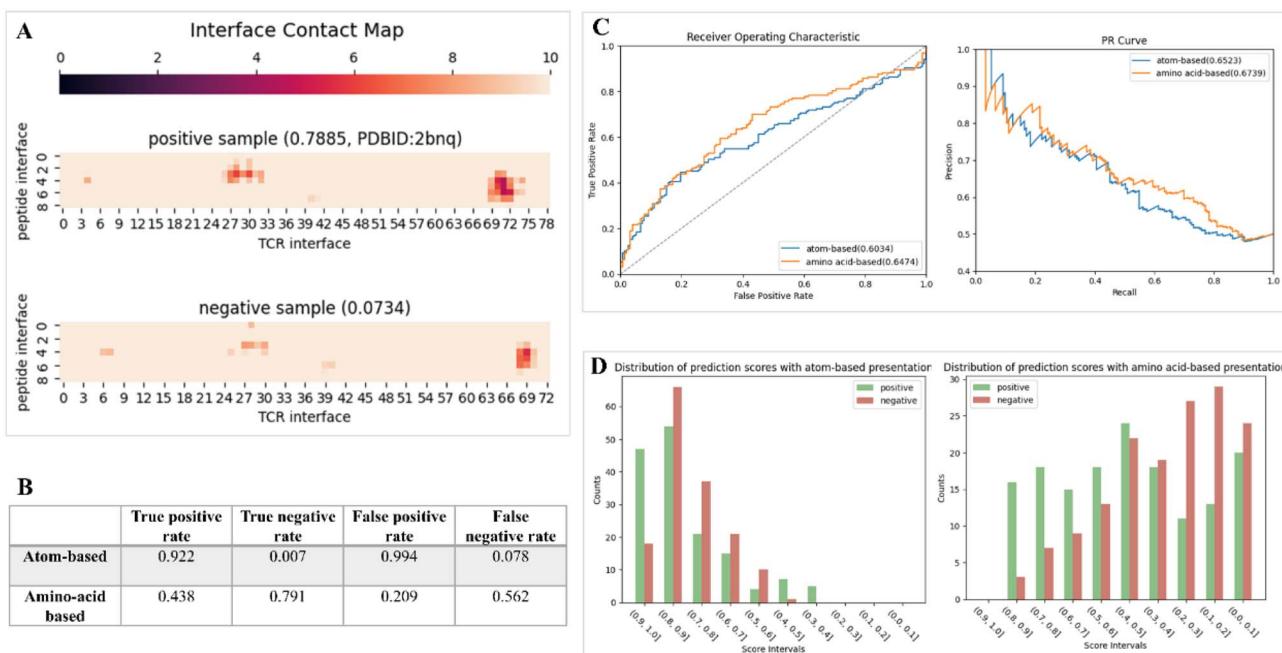


Fig. 4 The identification results for binding patterns of TCR-pMHC1 complexes from the SAGERank docking model. (A) A contact map is utilized to visually depict the disparities between the interfaces of positive and negative samples; (B) comparison of the true positive rate, true negative rate, false positive rate, and false negative rate in atom – amino acid detection; (C) the outcomes of ROCAUC and PRAUC evaluations for various TCR-pMHC1 complexes; (D) the scoring distribution of the SAGERank docking model across diverse TCR-pMHC1 complexes.

incorporate critical sequence information—such as physico-chemical and evolutionary profiles of key peptide and CDR3 residues—effectively complementing the structural data. This observation further suggests that future models combining high-resolution structural features with explicit sequence-based information (*e.g.*, amino acid propensities or positional motifs) could significantly improve TCR specificity prediction. In conclusion, our study demonstrates the potential of the SAGERank docking model, particularly when using amino-acid-based features, in accurately predicting TCR specificity for TCR-pMHC1 ternary complexes. This approach holds promise for accelerating the discovery of neoantigens and improving the efficacy of adoptive immunotherapy for cancer treatment.

3.4. Extension of antibody–antigen trained SAGERank to general protein–protein interaction

While the application of SAGERank to the TCR system could be a natural extension of SAGERank, we then examined if SAGERank can be used to discriminate general protein–protein recognition. Initially, we constructed a small protein–protein docking decoy set consisting of 8 protein–protein complexes and 10 707 docking decoys. Surprisingly, we discovered that the SAGERank docking model demonstrated competitiveness with scoring functions specifically optimized for protein–protein docking, even slightly outperforming Pisa in certain aspects (Fig. 2B, SI 4-1, 4-2, 4-5 and Table 4-1). To further validate the ranking accuracy of the SAGERank docking model on the protein–protein dataset, we expanded the size of the test set to include 80 protein–protein complexes, totaling 62 220 structures. As anticipated, the

SAGERank docking model significantly surpassed the success rates of the other three scoring methods and was on par with PISA. Our results strongly suggest that the SAGERank docking model possesses the capability to accurately rank docking decoys, regardless of whether they are antibody–antigen complexes or protein–protein complexes. This finding demonstrates the versatility and potential of the SAGERank docking model as a robust docking model that can be effectively applied to diverse biological systems.

After discovering the superior accuracy of the SAGERank docking model in protein docking ranking, we intended to employ this model for predicting binding sites and epitopes. The specific operational steps are illustrated in Fig. 5. Initially, two proteins or antibody–antigen pairs undergo unrestricted docking using Megadock (proteins do not specify binding sites and antigens do not specify epitopes), resulting in 2000 docking decoys. Subsequently, these decoys are ranked by the SAGERank docking model, and the top 100 decoys from the ranking are subjected to amino acid contact frequency analysis. By applying a specific threshold, we can determine the binding sites or epitopes of the protein or antigen.

Thus, we selected 10 protein groups and 5 antigen groups for binding site and epitope prediction. As demonstrated in SI Fig. 4-6 and 4-7, which showcase representative examples from our test dataset, the SAGERank docking model achieves high predictive accuracy for both binding sites and epitopes. Through comprehensive analysis of amino acid frequency distributions and three-dimensional structural visualizations, our results reveal that the model successfully identified binding sites in 8 out of 10 protein cases (80% accuracy) and correctly



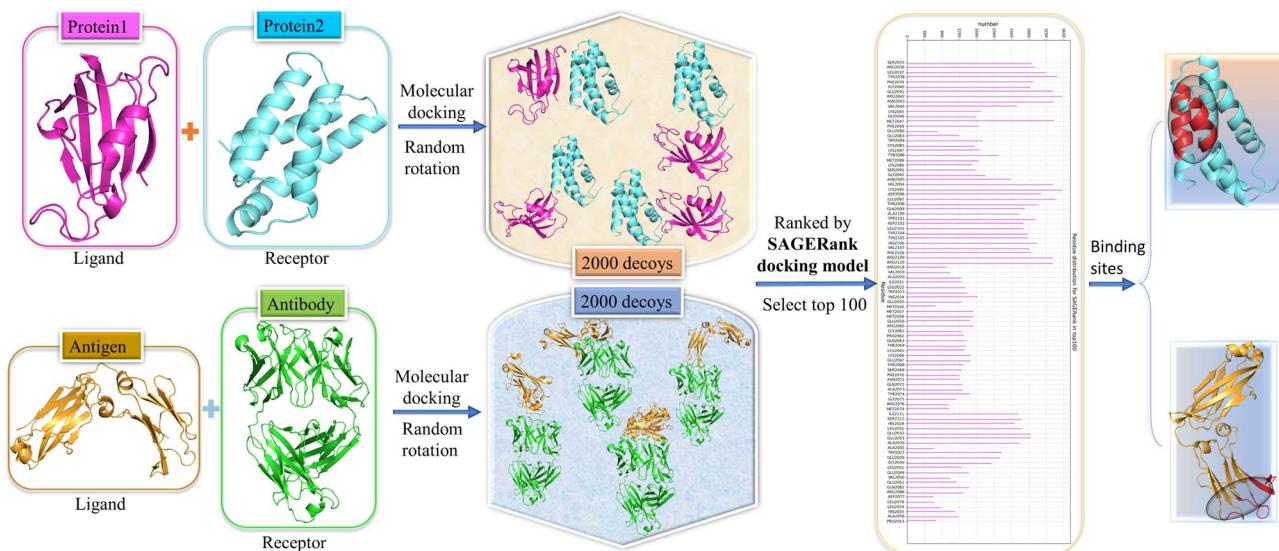


Fig. 5 The schematic diagram of binding sites prediction for protein using the SAGERank docking model.

predicted epitopes in 3 out of 5 antigen cases (60% accuracy). These findings not only validate the model's robust inductive capabilities but also highlight its strong generalization performance, positioning SAGERank as a reliable computational tool for structural immunology and protein interaction studies.

We then examined if SAGERank can discriminate biological protein–protein interfaces and “crystal interfaces” arising from packing in crystals using a commonly used DC dataset containing 90 biological and 71 crystal interface entries. DeepRank,²⁶ PISA,²⁷ and PRODIGY-crystal^{28,29} have shown the highest

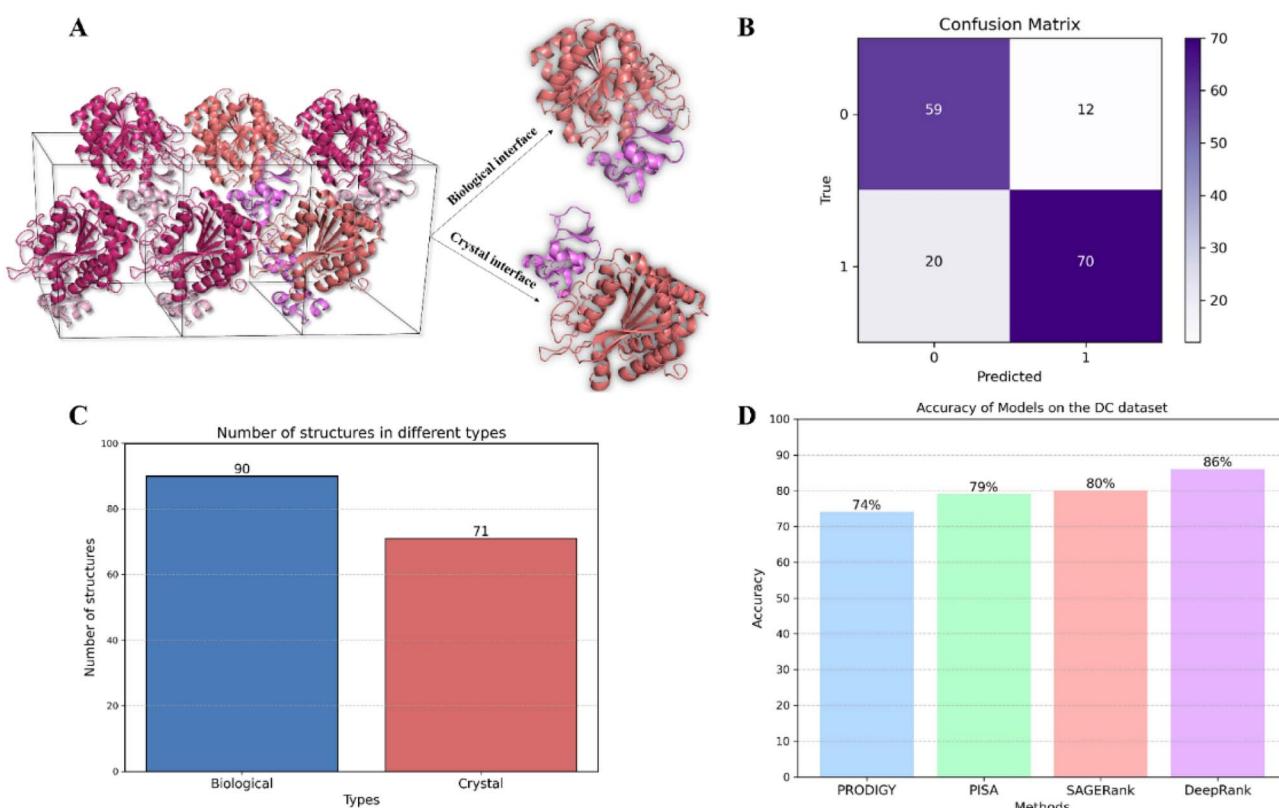


Fig. 6 (A) Schematic representation of the biological interface classification. (B) The confusion matrix of SAGERank on the DC test dataset. (C) The number distribution of biological and crystal interfaces on the DC dataset. (D) Comparison of accuracy of five different methods on DC data sets.



prediction performance in distinguishing crystal interfaces from biological ones in previous studies. The results using SAGERank are reported in Fig. 6 and SI 4-8. Fig. 6A illustrates protein structures with annotations for both biological and crystal interfaces. Fig. 6B shows a confusion matrix from SAGERank prediction, with 0 and 1 indicating crystal interfaces and biological interfaces, respectively. Fig. 6C presents the number of structures in different types within the commonly – used DC dataset, which contains 90 biological interfaces (in blue) and 71 crystal interfaces (in red). Comparative analysis (Fig. 6D) demonstrates that our SAGERank model achieves 80% accuracy in interface differentiation—comparable to PISA (79%) and superior to PRODIGY (74%), although slightly below DeepRank's 86% benchmark, which was specifically trained using the DC dataset.

3.5. Extension of SAGERank to reconstruct ternary complex interfaces in molecular glue

Since we used atoms as nodes in SAGERank, we challenged ourselves to see if SAGERank can be generalized to molecular systems with both proteins and small molecule drugs. We selected molecular glues, which enable the formation of ternary complexes (two protein and a molecular glue). These molecular glues alter the conformation or stability of proteins through non-covalent interactions with their target proteins or other associated proteins, thereby modulating their functions and metabolic processes. Despite significant efforts, progress in selecting a specific compound capable of simultaneously binding two targets to form a ternary complex has been slow.

We chose the molecular glue system FKBP12-rapamycin-FRAP as a paradigmatic case.³⁰ (Fig. 7). We conducted a series of group deletions on rapamycin (the positive compound) to transform it into a compound (negative) incapable of binding two targets simultaneously. The flowchart depicting the overall process is presented in Fig. 7. In the first step 1, we employed the SAGERank docking model to predict the possible binding poses of protein 1 with protein 2, subsequently in step 2,

Autodock-vina was utilized to dock the molecular glue compound (MGC) near protein–protein binding sites, allowing full conformational flexibility of MGC, generating 10 MGC conformations. Next in step 3, Megadock was invoked to dock the complex composed of the protein 1–MGC binary complex with protein 2, generating 2000 ternary complex poses for each set of 10 MGC conformations. Finally in step 4, the ternary complex poses were ranked and screened using the SAGERank docking model to select the possible ternary complex structure. For each set of 10 MGC conformations, we selected the top ten ternary complexes (from 2000) based on the SAGERank docking model's ranking. We then calculated the structural differences between these complexes and the authentic ternary complex, defining those with RMSD values between protein 2 and the docking poses of less than 8 Å as native samples, *i.e.*, the near native ternary complexes.

We counted the total near-native ternary structures in the top rankings (top 10) using SAGERank and Pisa (for comparison). The comprehensive statistical outcomes are presented in Table 4, with additional details available in SI Fig. 5-1. It is evident that the SAGERank docking model demonstrates a remarkable ability to distinguish between ternary complexes formed by positive and negative compounds. There are 10 rapamycin conformations generated using the positive compound, and we selected top 10 ternary complexes ranked by SAGERank and Pisa, leading to 100 ternary complex conformations selected for comparison. SAGERank correctly identified 30 positive structures, and mis-classified 2 negative complexes as positive. In comparison, Pisa only identified 10 positive structures, and mis-classified 6 negative complexes as positive. Thus, utilizing the SAGERank docking model, we accurately reproduced the structural configuration of this three-member complex. This underscores the superiority of the SAGERank docking model in accurately recognizing the interfaces formed between small molecular compounds and proteins, suggesting its possible application in the screening of molecular glue systems. The results not only enhance our understanding of molecular glue mechanisms but also hold promise for future drug design and discovery efforts.

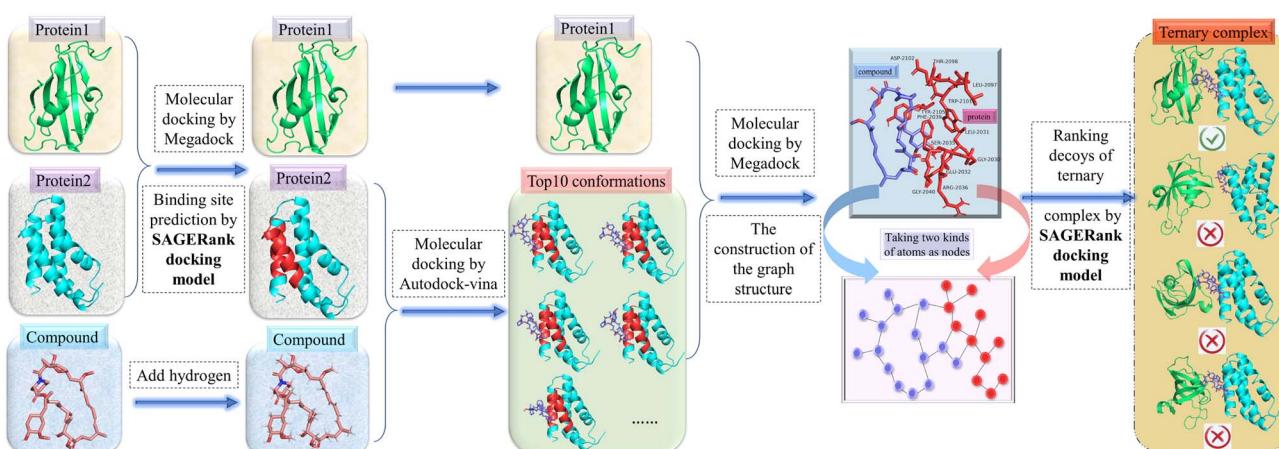
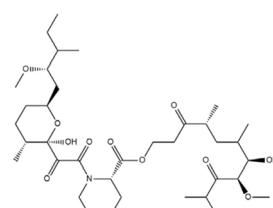
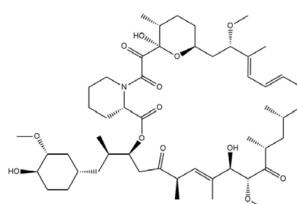


Fig. 7 Reproduction of the molecular glue system (FKBP12-rapamycin-FRAP) by the SAGERank docking model.



Table 4 Ranking statistics results of the SAGERank docking model for ternary complexes produced by positive and negative compounds

Positive compound		Negative compound			
Compound conformations (native/total)	The number of native ternary complexes in top10 by the SAGERank docking model	The number of native ternary complexes in top10 by Pisa	Compound conformations (native/total)	The number of native ternary complexes in top10 by the SAGERank docking model	The number of native ternary complexes in top10 by Pisa
Conf1(16/2000)	2	1	Conf1(43/2000)	0	0
Conf2(21/2000)	6	2	Conf2(32/2000)	1	0
Conf3(14/2000)	1	2	Conf3(40/2000)	0	1
Conf4(20/2000)	2	1	Conf4(38/2000)	0	0
Conf5(19/2000)	2	1	Conf5(31/2000)	0	0
Conf6(6/2000)	3	1	Conf6(14/2000)	0	0
Conf7(16/2000)	4	1	Conf7(11/2000)	0	1
Conf8(34/2000)	2	1	Conf8(48/2000)	0	3
Conf9(17/2000)	5	0	Conf9(22/2000)	0	0
Conf10(10/2000)	3	0	Conf10(28/2000)	1	1
All	30	10	All	2	6



3.6. Examination of SAGERank to obtain biological insights into cancer targets recognition

To examine if the small-dataset-trained SAGERank docking model can be used to obtain biological insights into epitope identification, we curated a specialized benchmark dataset containing 66 well-characterized cancer targets. Each target in this dataset was annotated with experimentally validated epitope information extracted from the Immune Epitope Database (IEDB; <https://www.iedb.org/>).³¹ Following preprocessing of the structural profiles of these 66 cancer targets and their corresponding antibodies, the SAGERank docking model was employed to predict the specific epitope amino acids. The exhaustive details and predictive outcomes pertaining to these cancer targets are presented in SI Table 1. We consider the SAGERank docking model to be predictively accurate if it is capable of predicting more than half of the total number of amino acids comprising the epitope of each target. Building upon this foundation, the model successfully predicted the epitopes for 39 out of the 66 cancer target groups, translating to a success rate approaching 60%. This substantial achievement underscores the model's heightened sensitivity in recognizing epitopes across diverse targets, firmly establishing its credibility and potential as a reliable tool in epitope prediction.

Next, we delved deeper into our analysis by selecting two targets for closer analysis. One of these is the versatile GTP-binding protein, which orchestrates a myriad of cellular processes, ranging from protein biosynthesis to intricate

intracellular membrane trafficking. In recent times, the functions and underlying mechanisms of GTP-binding proteins have significantly broadened, with the small GTP-binding protein Ras garnering particular attention due to its pivotal role in neoplastic transformation.^{32,33} Leveraging X-ray crystallography, the crystal structure of the RAS-anti-RAS single domain complex (PDB ID: 2UZI) was obtained.³² As depicted in Fig. 8A, the genuine epitope of RAS encompasses approximately 18 amino acids. Notably, the position of the epitope predicted by the SAGERank docking model aligns almost perfectly with the actual epitope, achieving an impressive coincidence rate of 88.9%. Furthermore, we conducted a comprehensive evaluation of the top 50 complex structures prioritized by our model, as shown in Fig. 8B. The analysis revealed that 37 of these structures closely resemble the native conformation, while 13 deviate from it.

Another cancer target under investigation is Plasminogen Activator Inhibitor-1 (PAI-1), a prominent member of the serine-protease inhibitor superfamily, commonly referred to as serpins. PAI-1 serves as the primary inhibitor for both tissue-type and urokinase-type plasminogen activators, enzymes crucial for activating plasminogen.³⁴ Recent research underscores PAI-1's central role in various age-related subclinical (such as inflammation, atherosclerosis, and insulin resistance) and clinical conditions (including obesity, comorbidities, and Werner syndrome). Structurally, PAI-1 is a single-chain glycoprotein composed of 379 amino acids, with a molecular weight of approximately 48 kDa. As the primary inhibitor of t-PA and u-PA, PAI-1 inactivates these plasminogen activators by cleaving



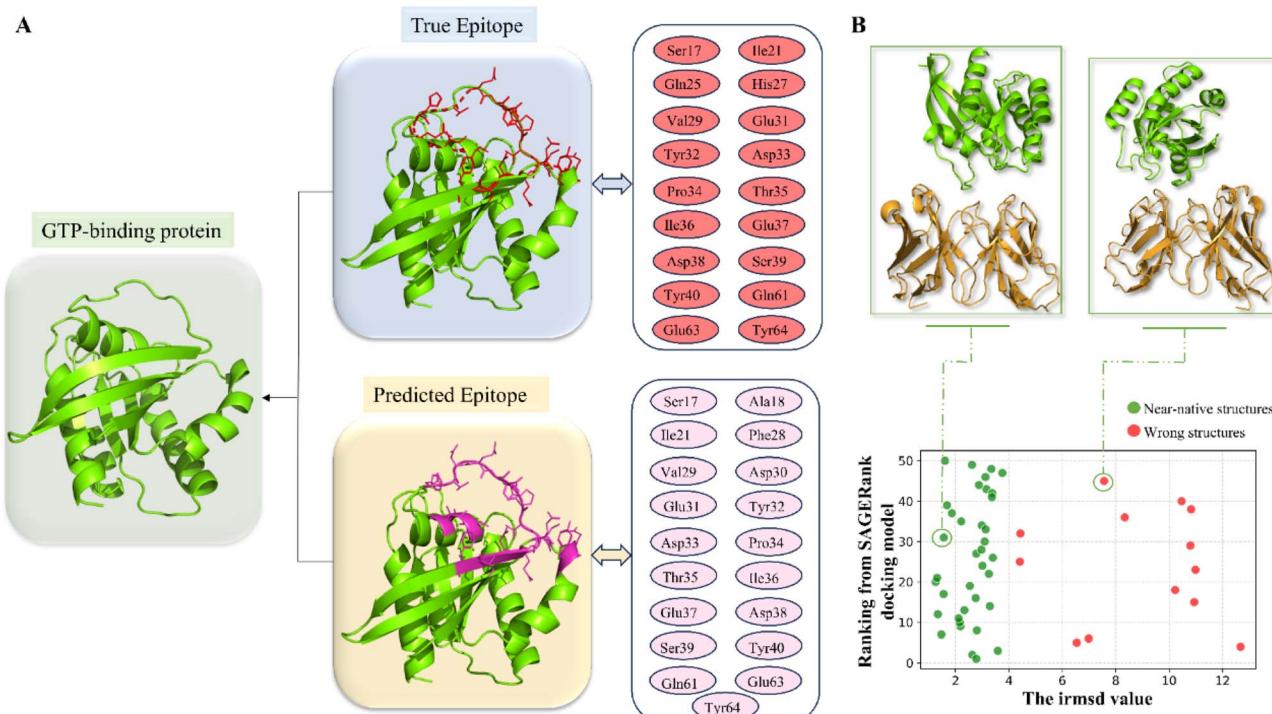


Fig. 8 (A) The results of epitope amino acids predicted by the model; (B) the relationship between the ranking of each complex by the model and iRMSD.

a specific Arg–Val peptide bond within their protease domain.³⁵ Intriguingly, PAI-1 possesses the capacity to form complex structures with two distinct types of nanoantibodies, suggesting

the presence of at least two antigen epitopes. Fig. 9A showcases the predicted epitope of PAI-1 (the prediction result for the other epitope can be found in SI Table 1), encompassing 12

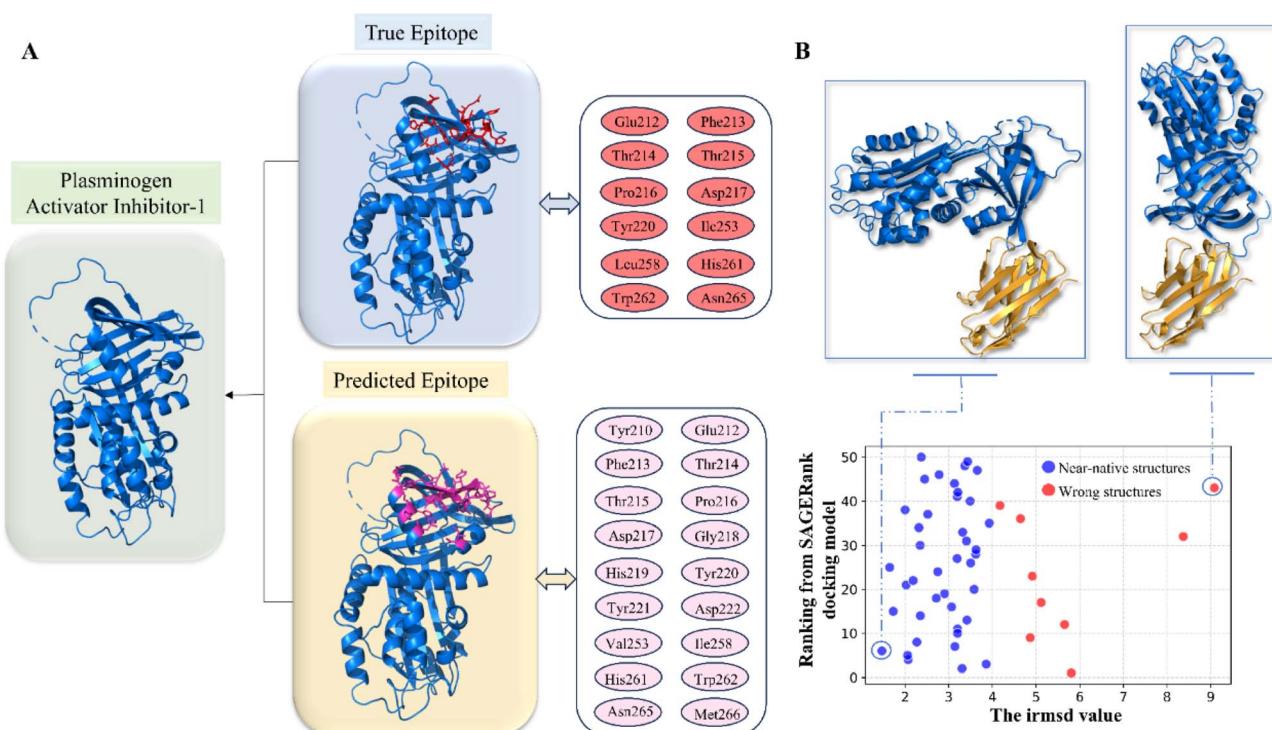


Fig. 9 (A) The results of epitope amino acids predicted by the model; (B) the relationship between the ranking of each complex by the model and iRMSD.

amino acids. Remarkably, the prediction aligns perfectly with the actual epitope, achieving a coincidence rate of 100%. When assessing the top 50 complex structures ranked by our model, we find that 41 resemble the native structure closely, while 9 deviate. Notably, even among the incorrect structures, the iRMSD remains below 10 Å, demonstrating the model's remarkable capacity to discern between positive and negative samples. These findings further validate the SAGERank docking model's accuracy and effectiveness in epitope prediction.

3.7. Comparison of SAGERank + deepconformer with alphafold3 in prediction of nanobody–protein recognition

AlphaFold 3 (AF3) has achieved remarkable success in predicting protein complex structures, but it suffers from a critical limitation: its ipTM performance metric, which is closely linked to AF3_score, fails to fully reflect actual structural accuracy, as prior evaluations have shown that high ipTM scores can mask significant conformational errors while low scores may coexist with reasonably accurate structures, given ipTM's focus on interface regions rather than global structural fidelity.³⁶ Our evaluation examined the performance of two scoring methods, SAGERank and AF3_score, across 43 protein complexes. For each complex, DeepConformer (the generative AI algorithm model for predicting dynamic three-dimensional protein structures from amino acid sequences³⁷) generated 40 predicted conformations, which were then ranked by both scoring methods. We analyzed the top four conformations from each method by computing their interface RMSD (iRMSD) values against the reference structure, with the minimum iRMSD value used for comparative assessment (Table 5).

The overall performance metrics revealed interesting insights. While SAGERank showed a marginally better mean iRMSD (6.62 Å vs. 6.88 Å), AF3_score demonstrated superior median performance (5.32 Å vs. 5.75 Å), suggesting that AF3_score generally yields better predictions but is more susceptible to outliers. Notably, AF3_score achieved near-native predictions (iRMSD ≤ 2.0 Å) in 27.9% of cases, slightly outperforming SAGERank's 25.6% success rate.

A detailed examination of individual cases revealed distinct strengths and weaknesses for each method. AF3_score exhibited remarkable predictive capabilities for certain complexes, such as 8sk5 (improving from 15.284 Å to 2.681 Å) and 7x2j (12.05 Å to 4.654 Å), demonstrating its potential to identify highly accurate conformations missed by SAGERank. However, this method also showed significant inconsistencies, with notable failures in complexes like 7sqp (1.261 Å to 11.214 Å) and 7uia (5.065 Å to 11.331 Å). In contrast, SAGERank displayed more consistent performance, with only one case (8oud) exceeding 20 Å iRMSD.

The strong positive correlation ($r > 0.7$) between the iRMSD values from both methods indicates substantial agreement in their assessment of conformation quality. This correlation suggests that both scoring functions recognize similar structural features important for near-native docking predictions, with each method typically identifying comparable low-iRMSD conformations. These findings suggest that while AF3_score

Table 5 Comparison of minimum iRMSD in protein structure prediction between SAGERank and AF3 Score

No	pdb_code	min_irmsd_sagerank1	min_irmsd_af3_score
0	7nxx	7.314	7.756
1	7q6c	10.678	9.923
2	7sp8	1.534	1.478
3	7sqp	1.261	11.214
4	7srk	1.366	9.763
5	7uia	5.065	11.331
6	7unz	6.94	9.327
7	7usv	7.054	9.201
8	7vke	5.748	3.755
9	7vnb	14.66	14.613
10	7voa	9.798	10.108
11	7wn0	6.759	6.585
12	7wn1	7.43	3.818
13	7x2j	12.05	4.654
14	7x2l	1.083	0.708
15	7x2m	14.155	13.73
16	7x7e	9.89	10.548
17	7z1c	2.455	4.038
18	7z1x	12.887	12.485
19	7zkz	4.457	10.298
20	7zml	12.706	3.123
21	7zmm	1.116	0.652
22	7zw1	7.34	6.999
23	8b17	0.781	0.646
24	8b8i	1.194	0.728
25	8be2	1.023	0.624
26	8c3l	0.952	1.118
27	8ce4	12.368	10.417
28	8dly	14.274	9.857
29	8dqu	5.333	5.317
30	8e0e	2.008	2.584
31	8eln	0.736	0.649
32	8emz	8.607	4.053
33	8en3	4.103	4.186
34	8gni	1.522	1.493
35	8h3x	1.374	1.29
36	8h3y	11.851	15.079
37	8h5u	18.321	16.717
38	8hxq	4.308	4.877
39	8ido	9.005	14.181
40	8ont	5.184	7.891
41	8oud	21.177	21.478
42	8pyr	0.772	0.502
43	8qf5	1.172	1.158
44	8sk5	15.284	2.681

can provide superior predictions in specific cases, SAGERank offers more reliable performance overall. The complementary strengths of these methods highlight the potential value of a hybrid approach that could leverage the best features of both scoring functions to further improve docking accuracy (Fig. 10).

4 Discussion and conclusion

High-resolution structures of Ab–Ag complexes are necessary for understanding the mechanisms of Ab–Ag interactions, analyzing mutations, and modulating binding affinity.³⁸ The large gap



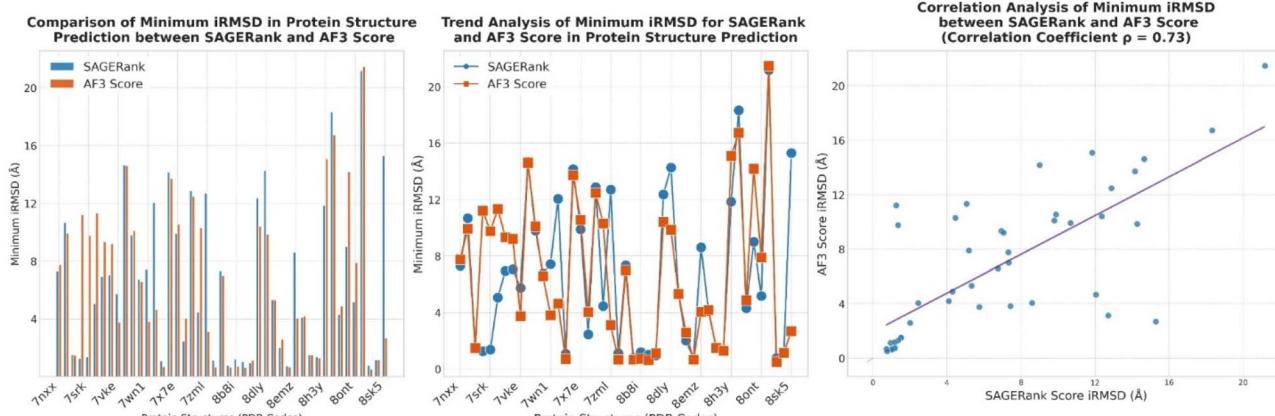


Fig. 10 Comparison of SAGERank + deepconformer with alphaFold3 in prediction of nanobody–protein recognition.

between the number of experimentally determined complex structures and the available sequences of pairs of Ab–Ag complexes underscores the challenges, time required, and cost of experimental approaches.³⁹ AlphaFold antibody–antigen modeling has only 30–50% success,⁴⁰ and other complementary methods are needed to support antibody engineering. One approach is to initially predict antibody and antigen structures separately and dock them.⁴¹ Although substantial improvements have been made in protein docking, selecting near-native models from a large number of produced models is still challenging.⁴² Various protein docking studies in recent years have taken advantages of deep learning to improve their performances.^{43,44} Challenges for docking algorithms include side chain and backbone conformational changes between unbound and bound structures, large search spaces, and the inability to capture key energetic features in grid-based and other rapidly computable functions, leading to false positive models among top-ranked models or lack of any near-native models within large sets of predicted models.^{45,46} Although substantial improvements have been made in protein docking, selecting near-native models out of many generated poses is still challenging.^{42,46,47}

Two major factors are responsible for the above difficulties. Firstly, although the principles of protein–protein interactions have been actively investigated during last two decades,² we also increasingly realize the complexity of these interactions. Many PTMs for example, phosphorylation and glycans, exist and modulate protein–protein interfaces and interactions. Protein structures and conformations are dynamic,^{48–51} which adds a hidden dynamical element to protein–protein interaction. The dynamics is especially important for antibody–antigen recognition due to the highly flexible CDR regions in antibodies. Secondly, the traditional transductive learning limited the application of proteins with unusual sequence features and those with only small data available, which is the second major difficulty that need to be solved for the development of biological drugs.

Here we developed the SAGERank model to predict the structure of antibody–antigen complexes, using an antibody–antigen complex dataset. Then we thoroughly examined the

generalization of SAGERank to broad protein–protein interaction prediction tasks. Overall, our model has demonstrated excellent performance. First, in the application of ranking docking decoys, the SAGERank docking model outperformed major existing scoring functions. SAGERank successfully predicted majority epitopes in a cancer target dataset. In nanobody–antigen structure prediction, SAGERank coupled with a protein dynamics structure prediction algorithm slightly outperformed AlphaFold3.

Most importantly, our study demonstrated the real potential of inductive deep learning, coupled with atomic interaction features, to overcome the small-dataset problem in molecular science. A natural graph network with atoms as nodes can be formed at the interface between antibody–antigen complexes and protein–protein complexes to accurately capture the fundamental physicochemical features of amino acid interactions. As a result, even though SAGERank was trained using only a small antibody–antigen complex dataset, it can be extended to predict general protein–protein interaction problems, such as TCR–pMHC recognition, classification of biological *versus* crystal interfaces, and prediction of ternary complexes of molecular glues.

In summary, we have designed a reliable and efficient deep learning framework for accelerating research based on antibody–antigen 3D structures, with the potential to expand to general protein–protein interaction. In the future, we will examine models trained with larger datasets to fully explore SAGERank's potential.

Author contributions

Chuance Sun: writing – original draft, methodology, investigation, data curation; Xiangyi Li: investigation, data curation; Honglin Xu: investigation; Yike Tang: data curation; Ganggang Bai: investigation; Yanjing Wang: supervision, funding acquisition; Buyong Ma: writing – review & editing, supervision, funding acquisition.

Conflicts of interest

There are no conflicts to declare.



Data availability

The data and code are available at <https://github.com/sunchuanu/SAGERank>.

The supplementary information file contains additional parameters of model training, including learning rate settings at different stages, and more experimental data such as detailed performance metrics of the model on various validation data-sets. See DOI: <https://doi.org/10.1039/d5sc03707g>.

Acknowledgements

B. Ma thanks support from the Natural Science Foundation of China (Grant No. 32171246) and Shanghai Municipal Government Science Innovation grant 21JC1403700. Y. Wang thanks support from the grants from the National Natural Science Foundation of China (No. 32200531), the Joint Research Funds for Medical and Engineering and Scientific Research at Shanghai Jiao Tong University (YG2022QN114 and YG2022QN082), and Startup Fund for Young Faculty at SJTU (SFYF at SJTU).

References

- 1 M. Wang, *et al.*, Local and global anatomy of antibody-protein antigen recognition, *J. Mol. Recognit.*, 2018, **31**(5), 14.
- 2 O. Keskin, *et al.*, Principles of protein-protein interactions: What are the preferred ways for proteins to interact?, *Chem. Rev.*, 2008, **108**(4), 1225–1244.
- 3 J. D. Guest, *et al.*, An expanded benchmark for Antibody-Antigen docking and affinity prediction reveals insights into antibody recognition determinants, *Structure*, 2021, **29**(6), 606–621.
- 4 C. Ye, W. Hu and B. Gaeta, Prediction of Antibody-Antigen Binding via Machine Learning: Development of Data Sets and Evaluation of Methods, *JMIR Bioinform. Biotechnol.*, 2022, **3**(1), e29404.
- 5 N. Zhao, *et al.*, ABAG-docking benchmark: a non-redundant structure benchmark dataset for Antibody-Antigen computational docking, *Briefings Bioinf.*, 2024, **25**(2), 48.
- 6 A. Davila, *et al.*, AbAdapt: an adaptive approach to predicting Antibody-Antigen complex structures from sequence, *Bioinform. Adv.*, 2022, **2**(1), vbac015.
- 7 B. Dou, *et al.*, Machine Learning Methods for Small Data Challenges in Molecular Science, *Chem. Rev.*, 2023, **123**(13), 8736–8780.
- 8 K. Sargsyan and C. Lim, Using protein language models for protein interaction hot spot prediction with limited data, *BMC Bioinf.*, 2024, **25**(1), 115.
- 9 H. Wang, *et al.*, AbImmPred: An immunogenicity prediction method for therapeutic antibodies using AntiBERTy-based sequence features, *PLoS One*, 2024, **19**(2), e0296737.
- 10 X. Jing, *et al.*, Single-sequence protein structure prediction by integrating protein language models, *Proc. Natl. Acad. Sci. U. S. A.*, 2024, **121**(13), e2308788121.
- 11 H. Jing, *et al.*, Accurate prediction of antibody function and structure using bio-inspired antibody language model, *Briefings Bioinf.*, 2024, **25**(4), 245.
- 12 J. Lim, *et al.*, Predicting Drug-Target Interaction Using a Novel Graph Neural Network with 3D Structure-Embedded Graph Representation, *J. Chem. Inf. Model.*, 2019, **59**(9), 3981–3988.
- 13 Y. Huang, Z. D. Zhang and Y. Zhou, AbAgIntPre: A deep learning method for predicting Antibody-Antigen interactions based on sequence information, *Front. Immunol.*, 2022, **13**, 1053617.
- 14 J. Jiang, *et al.*, A review of machine learning methods for imbalanced data challenges in chemistry, *Chem. Sci.*, 2025, **16**(18), 7637–7658.
- 15 Y. LeCun, Y. Bengio and G. Hinton, Deep learning, *Nature*, 2015, **521**(7553), 436–444.
- 16 S. L. Kinnings, *et al.*, A Machine Learning-Based Method To Improve Docking Scoring Functions and Its Application to Drug Repurposing, *J. Chem. Inf. Model.*, 2011, 1195–1197.
- 17 Z. Q. Gao, *et al.*, Hierarchical graph learning for protein-protein interaction, *Nat. Commun.*, 2023, **14**(1), 1093.
- 18 S. Cao, W. Lu, and Q. Xu, GraRep: Learning Graph Representations with Global Structural Information, in *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, Association for Computing Machinery, Melbourne, Australia, 2015, pp. 891–900.
- 19 A. Grover and J. Leskovec, *node2vec: Scalable Feature Learning for Networks*, Kdd, 2016, vol. 2016, pp. 855–864.
- 20 W. L. Hamilton, R. Ying and J. Leskovec, Inductive Representation Learning on Large Graphs, *Advances in neural information processing systems*, 2017, vol. 30(11), pp. 1025–1035.
- 21 M. Ohue, *et al.*, MEGADOCK 4.0: an ultra-high-performance protein-protein docking software for heterogeneous supercomputers, *Bioinformatics*, 2014, **30**(22), 3281–3283.
- 22 B. Pierce and Z. Weng, ZRANK: Reranking protein docking predictions with an optimized energy function, *Proteins: Struct., Funct., Bioinf.*, 2007, **67**(4), 1078–1086.
- 23 S. Viswanath, D. V. S. Ravikant and R. Elber, Improving ranking of models for protein complexes with side chain modeling and atomic potentials, *Proteins: Struct., Funct., Bioinf.*, 2013, **81**(4), 592–606.
- 24 J. Delgado, *et al.*, FoldX 5.0: working with RNA, small molecules and a new graphical interface, *Bioinformatics*, 2019, **35**(20), 4168–4169.
- 25 S. Lyskov and J. J. Gray, The RosettaDock server for local protein-protein docking, *Nucleic Acids Res.*, 2008, **36**(suppl_2), W233–W238.
- 26 N. Renaud, *et al.*, DeepRank: a deep learning framework for data mining 3D protein-protein interfaces, *Nat. Commun.*, 2021, **12**(1), 7068.
- 27 E. Krissinel and K. Henrick, Inference of Macromolecular Assemblies from Crystalline State, *J. Mol. Biol.*, 2007, **372**(3), 774–797.



28 B. Jiménez-García, *et al.*, PRODIGY-crystal: a web-tool for classification of biological interfaces in protein complexes, *Bioinformatics*, 2019, **35**(22), 4821–4823.

29 K. Elez, A. M. Bonvin and A. Vangone, Distinguishing crystallographic from biological interfaces in protein complexes: role of intermolecular contacts and energetics for classification, *BMC Bioinf.*, 2018, **19**(suppl 15), 438.

30 R. C. E. Deutscher, *et al.*, Discovery of fully synthetic FKBP12-mTOR molecular glues, *Chem. Sci.*, 2025, **16**(10), 4256–4263.

31 R. Vita, *et al.*, The Immune Epitope Database (IEDB): 2018 update, *Nucleic Acids Res.*, 2019, **47**(D1), D339–D343.

32 T. Tanaka, R. L. Williams and T. H. Rabbits, Tumour prevention by a single antibody domain targeting the interaction of signal transduction proteins with RAS, *EMBO J.*, 2007, **26**(13), 3250–3259.

33 T. Schweins and A. Wittinghofer, GTP-binding proteins. Structures, interactions and relationships, *Curr. Biol.*, 1994, **4**(6), 547–550.

34 M. Cesari, M. Pahor and R. A. Incalzi, Plasminogen activator inhibitor-1 (PAI-1): a key factor linking fibrinolysis and age-related subclinical and clinical conditions, *Cardiovasc. Ther.*, 2010, **28**(5), e72–e91.

35 M. Sillen, *et al.*, Molecular mechanism of two nanobodies that inhibit PAI-1 activity reveals a modulation at distinct stages of the PAI-1/plasminogen activator interaction, *J. Thromb. Haemostasis*, 2020, **18**(3), 681–692.

36 J. Wee and G. W. Wei, Evaluation of AlphaFold 3's Protein-Protein Complexes for Predicting Binding Free Energy Changes upon Mutation, *J. Chem. Inf. Model.*, 2024, **64**(16), 6676–6683.

37 Y. Tang, *et al.*, Deep learning of protein energy landscape and conformational dynamics from experimental structures in PDB, *bioRxiv*, 2024, 600251.

38 J. J. Gray, High-resolution protein–protein docking, *Curr. Opin. Struct. Biol.*, 2006, **16**(2), 183–193.

39 A. Sircar, J. J. Gray and T. Kortemme, SnugDock: Paratope Structural Optimization during Antibody-Antigen Docking Compensates for Errors in Antibody Homology Models, *PLoS Comput. Biol.*, 2010, **6**(1), e1000644.

40 R. Yin and B. G. Pierce, Evaluation of AlphaFold Antibody-Antigen modeling with implications for improving predictive accuracy, *Protein Sci.*, 2024, **33**(1), e4865.

41 K. M. McCoy, M. E. Ackerman and G. Grigoryan, A comparison of Antibody-Antigen complex sequence-to-structure prediction methods and their systematic biases, *Protein Sci.*, 2024, **33**(9), e5127.

42 X. Wang, *et al.*, Protein docking model evaluation by 3D deep convolutional neural networks, *Bioinformatics*, 2020, **36**(7), 2113–2118.

43 H. Cai, *et al.*, CarsiDock: a deep learning paradigm for accurate protein-ligand docking and screening based on large-scale pre-training, *Chem. Sci.*, 2024, **15**(4), 1449–1471.

44 J. Zhu, *et al.*, DiffBindFR: an SE(3) equivariant network for flexible protein-ligand docking, *Chem. Sci.*, 2024, **15**(21), 7926–7942.

45 R. Yin, *et al.*, Benchmarking AlphaFold for protein complex modeling reveals accuracy determinants, Cold Spring Harbor Laboratory, 2021.

46 M. Buttenschoen, G. M. Morris and C. M. Deane, PoseBusters: AI-based docking methods fail to generate physically valid poses or generalise to novel sequences, *Chem. Sci.*, 2024, **15**(9), 3130–3139.

47 D. E. B. Gomes, *et al.*, Integrating Dynamic Network Analysis with AI for Enhanced Epitope Prediction in PD-L1:Affibody Interactions, *J. Am. Chem. Soc.*, 2024, **146**(34), 23842–23853.

48 G. H. Wei, *et al.*, Protein Ensembles: How Does Nature Harness Thermodynamic Fluctuations for Life? The Diverse Functional Roles of Conformational Ensembles in the Cell, *Chem. Rev.*, 2016, **116**(11), 6516–6551.

49 R. Nussinov and B. Y. Ma, Protein dynamics and conformational selection in bidirectional signal transduction, *BMC Biol.*, 2012, **10**, 5.

50 B. Y. Ma, *et al.*, Multiple diverse ligands binding at a single protein site: A matter of pre-existing populations, *Protein Sci.*, 2002, **11**(2), 184–197.

51 B. Y. Ma, *et al.*, Folding funnels and binding mechanisms, *Protein Eng.*, 1999, **12**(9), 713–720.

