

PPAP: A Protein–protein Affinity Predictor Incorporating Interfacial Contact-Aware Attention

Jie Qian, Lin Yang, Zhen Duan, Renxiao Wang,* and Yifei Qi*



Cite This: *J. Chem. Inf. Model.* 2025, 65, 9987–9998



Read Online

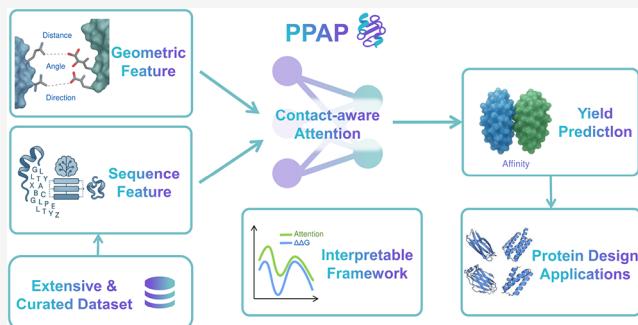
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: Protein–protein interactions (PPIs) play fundamental roles in biological processes and therapeutic development. Accurately predicting PPI binding affinity is critical for understanding interaction mechanisms and guiding protein engineering. Recent advances in structure prediction like AlphaFold have enabled accurate modeling of protein–protein complexes, creating new opportunities for structure-based affinity prediction. However, existing methods predominantly rely on sequence information and fail to fully exploit structural insights at interaction interfaces. To address this gap, we propose PPAP, a novel deep learning framework that integrates structural features with sequence representations through an interfacial contact-aware attention mechanism. Our model demonstrated superior prediction performance across all evaluated data sets, outperforming strong sequence-based large language models on the internal test ($R = 0.540$, MAE = 1.546). On the external test set, our model achieved a higher Pearson correlation coefficient ($R = 0.63$) than all benchmarked models. In protein binder design, we further demonstrate that incorporating our model’s prediction can enhance enrichment by up to 10-fold in comparison to the metrics based on AlphaFold-Multimer prediction. Given its robust performance, PPAP holds promise as a valuable tool not only for protein design but also for a wide range of protein interaction-related applications.



INTRODUCTION

Protein–protein interactions (PPIs) are of central importance in numerous physiological processes, both at the cellular level and in the context of the entire organism.^{1,2} These interactions play critical roles in regulating cellular functions such as signal transduction, gene expression, and cellular trafficking.^{3–6} The dynamic nature of PPIs is essential for maintaining homeostasis and responding to environmental stimuli.^{7–10} Furthermore, PPI networks are fundamental in various biological mechanisms, including immune responses, metabolic pathways, and development.¹¹ Given their relevance, understanding PPIs has become a cornerstone in drug discovery and disease modeling.¹²

Currently, the evaluation of PPIs can be conducted through both qualitative and quantitative approaches, with quantitative assessment playing a more crucial role in drug development.¹² Given the importance of precise measurements in understanding and optimizing PPIs, researchers often rely on experimental assays to determine protein–protein binding affinity. However, these methods are relatively costly and time-consuming, involving complex and troublesome procedures. Furthermore, the purification and preparation of certain proteins present significant challenges, limiting their availability for experimental studies.¹³ In contrast, *in silico* methods for predicting PPI affinities have garnered significant attention in recent years due to the high efficiency and low cost.

The existing *in silico* methods can be mainly divided into two categories: (1) Sequence-based approaches leverage evolutionary relationships and semantic patterns in amino acid sequences, frequently employing machine learning models combined with transformer-based architectures to infer binding strengths.^{14–16} (2) Structure-based methods exploit 3D complex geometries through interfacial contact analysis and surface property calculations, where machine learning methods and deep learning methods are commonly applied to structural informations.^{17–22} Building upon the advancements in the success of AlphaFold in predicting protein structures with remarkable accuracy,^{23–25} the prediction of protein–protein complex structures has gained significant reliability. This has made the use of complex structures as a basis for PPI prediction a promising and increasingly favored approach, as it allows for more precise modeling of the interactions between protein partners. Recent advancements in structure-based models, such as GeoPPI²⁶ and binding-ddg-predictor,²⁷ have

Received: June 16, 2025

Revised: September 11, 2025

Accepted: September 11, 2025

Published: September 19, 2025



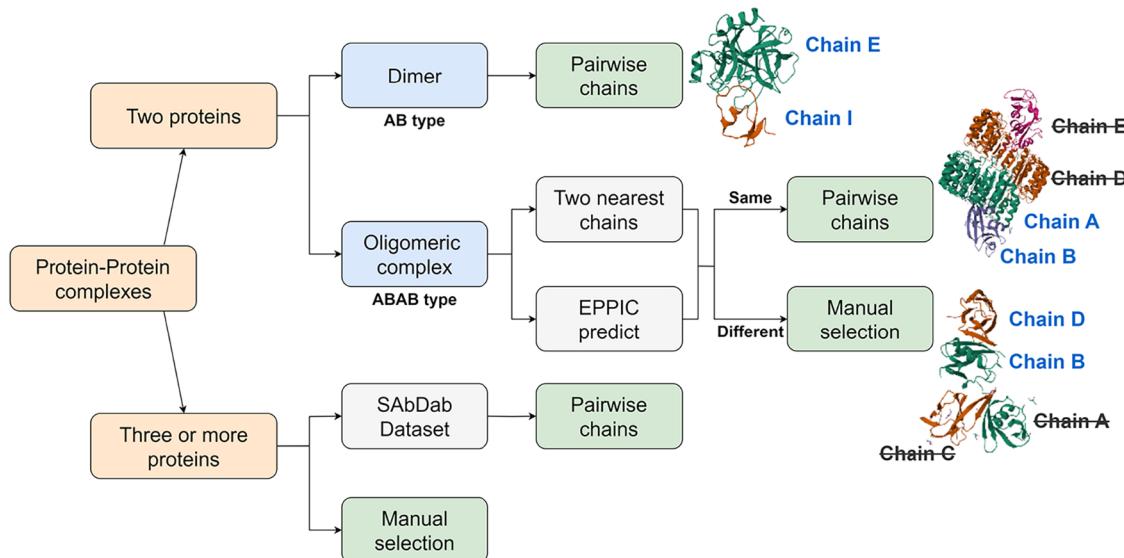


Figure 1. Workflow of data set processing. The figure illustrates the selection process for protein–protein complexes, categorized into dimers, oligomeric complexes, and complexes containing three or more proteins. Different strategies, including pairwise chain selection, EPPIC prediction, and manual selection, are applied based on complex type. The right side shows structural examples, with final selected chains highlighted in blue.

effectively utilized geometric representations and deep learning to predict changes in binding affinity upon mutations. Similarly, PLMGraph-Inter²⁸ integrates protein language models with geometric graphs for interprotein contact prediction.

However, current methods for predicting protein–protein affinity have several limitations. First, although structure-based affinity prediction models provide some predictive capabilities, they fail to adequately capture the complex interactions at the protein–protein interface, which limits their predictive accuracy. For example, ProAffinity-GNN relies solely on atomic distances and ESM-2 sequence embeddings in its residue graph, overlooking critical geometric information such as amino acid side-chain orientations and relative normal angles.²² Similarly, PRODIGY employs a traditional interface residue contact statistical model to estimate binding free energy but fails to integrate the informational representations offered by language models.¹⁹ Second, while the PDBbind data set provides binding affinity data alongside corresponding PDB files,²⁹ the presence of redundant chains in these PDB files that are unrelated to the binding affinity compromises data quality. Current approaches to address this issue have notable limitations: some studies restrict their analysis to PDB files containing only two chains, thereby excluding a substantial portion of valuable data; others manually curate data sets comprising two or three chains, yet still omit PPI data involving more than three chains. Moreover, these manual curation processes lack standardization, resulting in inconsistent data quality across studies.^{17,18,30}

To overcome these challenges, we utilized the protein–protein complex data in PDBbind and developed a standardized pipeline to identify the chain components involved in the direct pairwise interactions, generating 3,135 uniquely defined protein–protein pairs with binding affinities. This meticulously curated resource, representing a 50% increase over prior data sets, establishes new standards for affinity data sets of PPIs while addressing the data scarcity problem in the field.^{21,22} Based on our curated data set, we propose “PPAP” (Protein–protein Affinity Predictor), an innovative structure-

based deep learning model designed to advance the prediction of protein–protein in interactions. Our model integrates protein sequences and structures with interfacial contact-based attention, combining protein language models with geometric structure to accurately predict protein–protein binding affinity. PPAP achieves state-of-the-art performance across various benchmark data sets, where it reduces the error in predicted K_d values by approximately 60%, with a mean absolute error (MAE) reduction of 0.55 kcal/mol in binding free energy compared to the best existing model. Additionally, our model shows promising potential in the design of protein binder, offering valuable insights for optimizing protein–protein interactions and enhancing the accuracy of design processes.

METHODS

Data Processing. To identify chain components directly engaged in protein–protein binding, we employed an integrated approach incorporating interprotein distance analysis, EPPIC predictions,³¹ and manual curation on the protein–protein interaction data set in PDBbind version v2021. Each protein–protein complex, classified by composition into two-protein and multiprotein complexes, underwent separate processing pipelines. For protein complexes, if the complex is a dimer, the receptor and ligand were directly assigned. For multimeric complexes, assignment was based on interchain distances and EPPIC predictions. If both methods yielded consistent results, the assignment was accepted; otherwise, manual curation was performed. For complexes involving multiple proteins, antibody-related complexes were assigned according to the SAbDab database, while other cases were manually curated (Figure 1). The details are elaborated below.

Simple complexes composed of two chains (e.g., chains A and B) could be directly treated as a receptor and a ligand (e.g., receptor A and ligand B). However, for complexes containing three or more chains (e.g., ABAB-type complexes), two distinct approaches were employed to cross-verify the results.

First, the pairs were enumerated by selecting one chain from each of the two proteins and exhaustively evaluating all possible pairwise matches. The closest chain pair was identified by the number of carbon atoms in their main chains within 8.5 Å. This criterion assumed that the chains with the highest number of main-chain carbon atoms within 8.5 Å are most likely to interact.

Second, the EPPIC (Evolutionary Protein–protein Interface Classifier) method was employed. EPPIC differentiated between biologically relevant protein–protein interaction interfaces and those formed by crystal packing, therefore facilitated the identification of the interface within the complex that most closely resembles a natural interaction, i.e., the one with the highest confidence score. Using EPPIC, a detailed analysis of various interface types within the protein complex was performed, which allowed for the determination of receptor–ligand interaction chains.

If the candidate protein pairs from the first step were absent in the EPPIC prediction results, it indicated inconsistency between the two methods. In such cases, the following steps were taken for further analysis. The original literature was visually inspected, and manual selection was conducted to identify receptor–ligand pairs corresponding to natural protein–protein interaction interfaces. If the distinction between a natural interface and one caused by crystal packing could not be concluded from the PDB file, a thorough review of the relevant literature was undertaken to determine the interacting receptor and ligand. If a definitive conclusion could not be reached after reviewing the original literature, or if the complex presented unresolved issues, the data were excluded from the final data set of interaction protein pairs.

A significant proportion of complexes composed of three or more proteins were antigen–antibody complexes. For these complexes, the SAbDab database³² was utilized to confirm the receptor and ligand chain information. SAbDab (The Structural Antibody Database) is a specialized resource for collecting and organizing structural antibody information. It compiles antibody structures from the Protein Data Bank (PDB)³³ and standardizes them using automated methods, providing uniform formats and annotations. For complexes not included in SAbDab, manual curation was performed to identify the receptor–ligand chains.

By employing these methods, we ensured the accuracy and reliability of the identified protein–protein interaction pairs. Cross-validation of these methods enabled more accurate identification and confirmation of interacting protein pairs in complexes, providing a robust foundation for subsequent research. Moreover, we compared our data processing pipeline with that of PPB-affinity,³⁴ a recently published, comprehensive data set that focuses on protein–protein binding affinity. This comparison, detailed in the Supporting Information (Figure S1), highlights key differences in our approach, which we believe results in a more refined data set for model training.

The final data set contained a total of 3135 entries, which include different affinity test values such as K_d , K_b , and IC_{50} . All binding affinity data were converted to ΔG (kcal/mol), where $T = 298.15$ K and $R = 0.00199$ kcal/(mol·K). During the calculation of ΔG , the original inequalities in the data were transformed into equalities.

To build the training and test sets, we clustered the data to prevent data leakage of similar data. Here, referring to the clustering method for complexes in AlphaFold3 (AF3),²³ all chains in the complexes were clustered first, and if the

sequence similarity of any chain between two complexes was more than 40%, the two complexes were treated as a cluster. Complexes containing peptide chains shorter than 16 amino acids were excluded from the clustering process. In total, our clustering resulted in 1971 clusters. Among them, 1461 clusters contain only a single member, while 510 clusters contain multiple members. To maximize data retention, our data set partitioning strategy incorporated all data within each cluster.

The PDBbind S90 test set, initially introduced in the PPI-affinity study, has been widely adopted as an external benchmark in recent investigations.^{20–22} To maintain the independence of this test set, we excluded all entries overlapping with PDBbind S90 from our data set. The remaining 3046 data entries were then partitioned into training, validation, and test sets in a 7:1:2 ratio (Table S1). This test set, comprising 598 entities, is hereafter referred to as the internal test set. To ensure robustness, the training and validation sets were randomly resampled five times, and both the PPAP model and the sequence-based model were trained on each split. The final results represent the mean performance across all splits.

To assess the generalizability of various models when confronted with novel protein–protein complexes, we constructed the test set S34 from the newly released PDBbind v2021 PPI data in the internal test set. Furthermore, to guarantee that the test set does not exhibit high similarity to our training data, the selection of this test set was guided by the following criteria: (1) The data is exclusively sourced from the new entities in PDBbind v2021 release, thereby eliminating potential overlap between S34 and the training data used in prior models. (2) None of the entries in S34 share $>= 40\%$ similarity with any data from previous clustering analyses, thereby preventing data leakage.

Graph Representation of Protein–Protein Complex. *Node Features.* We employed Evolutionary Scale Modeling-2 (ESM2–3B), a deep learning-based pretrained model designed to capture evolutionary information and functional features of protein sequences, for node embedding. Protein sequences were extracted from PDB files and processed with ESM2–3B, yielding a 2560-dimensional representation for each node.

Edge Features. We calculated the C_β distances between different residues to establish edges, with intrachain residue pairs set at 10 Å and interchain pairs at 12 Å for edge relationship establishment. These thresholds were empirically determined and biologically justified: the shorter intrachain distance accounts for covalent bonding and tighter packing within a single chain, while the extended interchain distance captures longer-range interactions between chains.^{35–37}

For the interaction edges between molecules, we constructed an equivariant edge feature vector comprising the following elements

$$\mathbf{e}_{ij} = \text{concat}(d_{ij}, R_{ij}, D_{ij}, \alpha), \mathbf{e}_{ij} \in \mathbb{R}^{505} \quad (1)$$

where d_{ij} is the distance feature, R_{ij} is the angle feature, D_{ij} is the direction feature, and α is the chain information feature which will be described in detail below.

Distance Features. For two sets of atom coordinates A and B

$$d_{ij} = \text{RBF}(\|A - B\|) \quad (2)$$

where $A \in \{N_i, C_{\alpha i}, C_i, O_i, C_{\beta i}, \dots\}$, $B \in \{N_j, C_{\alpha j}, C_j, O_j, C_{\beta j}, \dots\}$, RBF represents the radial basis function.³⁸ d_{\min} and d_{\max}

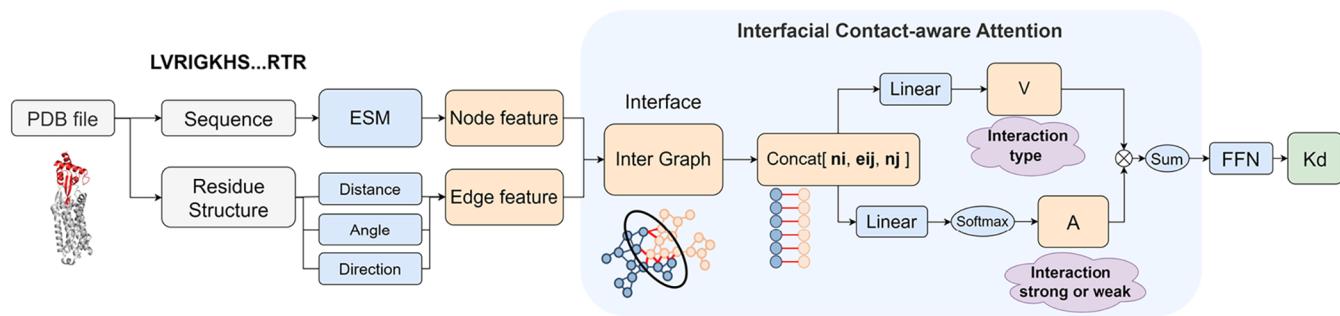


Figure 2. Flowchart of PPAP framework. The model extracts sequence features from ESM and processes interaction interface information through a decoder. Features from protein complex interactions, including edge and node attributes, are concatenated and passed through a linear layer, followed by an interfacial contact-aware attention mechanism that determines interaction type and strength. The graph representation is then processed through a feedforward network (FFN) to predict the affinity.

define the range of distances over which the RBF function operates. Here, the default range is set from 0 to 20. The parameter governing the number of Gaussian basis functions (kernels) employed in the radial basis function expansion determined their spatial distribution within this interval. In the current implementation, a set of 16 uniformly distributed kernels is automatically generated across the predefined interval. We computed distances between all atoms in the two residues. In order to ensure that different residue pairs form edges of the same dimension, we selected distances between the five backbone atoms in each residue, resulting in a total of 25 distances. However, since backbone atoms may not interact, we included the five shortest distances between all atoms in residues i and j to capture potential interactions, resulting in a total of 30 distances. Eventually, we obtained the distance characterization $d_{ij} \in \mathbb{R}^{(30 \times 16)}$. For glycine, which lacks C_β atoms, we used the MP-NeRF method to compute the C_β atom coordinates.³⁹ We utilized pdbfixer to complete the missing backbone atoms.⁴⁰ For any remaining issues that pdbfixer could not resolve, we manually completed the missing backbone atoms.

Angle Features. R_{ij} is obtained by multiplying the rotation matrices of the two residues corresponding to the edge.

$$R_{ij} = R_j^{-1} R_i, R_i, R_j \in \mathbb{R}^{3 \times 3} \quad (3)$$

Here, we obtain rotation matrix R_p, R_j based on the Gram-Schmidt orthogonalization process.⁴¹

$$v_1 = C - C_\alpha \quad (4)$$

$$v_2 = N - C_\alpha \quad (5)$$

$$e_1 = \frac{v_1}{\|v_1\|} \quad (6)$$

$$u_2 = v_2 - (e_1 \cdot v_2)e_1 \quad (7)$$

$$e_2 = \frac{u_2}{\|u_2\|} \quad (8)$$

$$e_3 = e_1 \times e_2 \quad (9)$$

$$R = \text{concat}(e_1, e_2, e_3) \quad (10)$$

Direction Features. Direction features D_{ij} were defined using unit vectors between the C_α of the starting residue and the five backbone atoms of the target residue. We multiply this

vector by R_i^T (obtained from angle features) to ensure its rotational invariance.

$$D_{ij} = \text{concat}\left(R_i^T \frac{C_{ai} - N_j}{\|C_{ai} - N_j\|}, R_i^T \frac{C_{ai} - O_j}{\|C_{ai} - O_j\|}, \dots, R_i^T \frac{C_{ai} - C_j}{\|C_{ai} - C_j\|}\right) \quad (11)$$

Chain Information. Edges within same chains were labeled as 0, and interchain edges as 1.

PPAP Model. Considering the close relationship between interactions at the protein binding interface, as well as its physicochemical properties and binding affinity, we propose an interfacial contact-aware attention decoder design approach (Figure 2).⁴² Specifically, node features are extracted using the protein language model ESM2–3B to capture high-dimensional representations of individual residues.¹⁶

Drawing inspiration from physics-informed statistical models of interfacial residue, we introduce a contact-aware attention mechanism.¹⁹ This approach, termed interfacial attention, integrates ESM embeddings with edge features to effectively model the relationships between potentially contacting amino acid residues.

Specifically, to construct the edge features, we first compute all pairwise atomic distances between interfacial residues, subsequently selecting the five shortest distances to characterize potential interactions. Building upon this, we further incorporate angular and directional features. These edge features serve as a key input to the attention module, wherein the model learns multiheaded attention weights (A) via a Multi-Layer Perceptron (MLP), rather than through the separate computation of query (Q) and key (K) matrices (Figure 2). The mechanism is defined as follows

$$w_{ij} = \text{Att MLP}(\text{concat}(h_i, e_{ij}, h_j)) \quad (12)$$

$$a_{ij} = \frac{\exp w_{ij}}{\sum_{ij \in \mathcal{N}} \exp w_{ij}} \quad (13)$$

$$v_{ij} = \text{value MLP}(\text{concat}(h_i, e_{ij}, h_j)) \quad (14)$$

$$\mathcal{G} = \sum_{ij \in \mathcal{N}} a_{ij} v_{ij} \quad (15)$$

where \mathcal{N} represents the set of edge index at the interface and \mathcal{G} denotes the model's output. This design is intended to assign higher attention weights to residue pairs that are in close

spatial proximity and exhibit physical contact. Finally, a feed-forward network (FFN) layer and a fully connected layer yield the prediction values.

Model Training. For both the PPAP and sequence-based models, we used the Adam optimizer (learning rate: 3×10^{-5} for PPAP, 1×10^{-4} for the sequence-based model) with a batch size of 1.⁴³ Training efficiency was optimized using an early stopping strategy based on validation set performance, whereby training was terminated if no improvement was observed over 10 epochs. A dropout layer (with a rate of 0.1) is applied subsequent to the attention weight calculation to mitigate model overfitting. Concurrently, a Batch Normalization layer has been incorporated into the edge feature projection module. All trainable parameters within our custom graph neural network are initialized via the Xavier uniform distribution, culminating in a total model size of approximately 18 M parameters, excluding the precomputed ESM embeddings. The objective for both models was to minimize the mean squared error (MSE) loss.

Evaluation Metrics. To evaluate the performance of our model in predicting binding free energy changes (ΔG), we used three standard metrics: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Pearson's Correlation Coefficient (R).

Mean Absolute Error (MAE) (kcal/mol): Measures the average absolute difference between predicted and experimental ΔG values.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|$$

where \hat{y}_i is the predicted ΔG , y_i is the experimental ΔG , and n is the number of samples.

Root Mean Square Error (RMSE) (kcal/mol): Quantifies the square root of the average squared differences between predicted and experimental ΔG values.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}$$

where \hat{y}_i is the predicted ΔG , y_i is the experimental ΔG , and n is the number of samples.

Pearson's Correlation Coefficient (R): Measures the linear correlation between predicted and experimental ΔG values.

$$R = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2} \sum_{i=1}^n (y_i - \bar{y})^2}$$

where \bar{y} and \bar{y} are the means of predicted and experimental ΔG values, respectively.

All ΔG values were consistently reported in kcal/mol, the standard unit for binding free energy measurements, ensuring direct comparability with existing benchmarks. This unit choice eliminates the need for additional conversions, as ΔG maintains a well-defined thermodynamic relationship with dissociation constants ($\Delta G = RT \ln K_d$). The combined use of these metrics enables a multidimensional and standardized assessment of model performance.

RESULTS AND DISCUSSION

Performance on Internal Test Set. To evaluate the accuracy and generalizability of our model, we conducted

comparative tests across different models on the internal test set ($N = 598$). Our model achieved a correlation coefficient of $R = 0.540$ and a mean absolute error (MAE) of 1.546 kcal/mol (Table 1). For comparison, we evaluated several models,

Table 1. Performance Comparison of Different Models on the Internal Test Set

model	pearson $R \uparrow$	RMSE \downarrow	MAE \downarrow
ESM-1v	0.536 ± 0.031	<u>2.094 ± 0.122</u>	1.627 ± 0.079
ESM2-3B	0.532 ± 0.030	2.101 ± 0.125	1.629 ± 0.081
ESM2-15B	0.533 ± 0.034	2.101 ± 0.127	1.621 ± 0.086
ESM3-open ^a	0.453 ± 0.062	2.306 ± 0.087	1.772 ± 0.063
Saprot	<u>0.537 ± 0.028</u>	2.107 ± 0.108	<u>1.619 ± 0.074</u>
Masif	0.303 ± 0.009	2.277 ± 0.027	1.820 ± 0.041
Δ SASA	0.050 ± 0.002	/	/
PPAP	0.540 ± 0.032	1.998 ± 0.027	1.546 ± 0.018

^aThe model achieved convergence in only three out of five data partitions, with the final results being calculated based on these three convergent training runs.

including powerful protein sequence-based large language models ESM-1v, ESM2 (3B, 15B)^{16,44} and ESM3-open,⁴⁵ the structure-aware Saprot model,⁴⁶ and Masif,⁴⁷ which incorporates protein surface properties. In addition, the change in solvent-accessible surface area (Δ SASA), which quantifies surface area changes upon protein–protein binding, was used it as a naïve baseline. Our model demonstrates superior performance compared to sequence-based models such as ESM2 and ESM3-open, highlighting the critical role of incorporating geometric information and our interfacial contact-aware attention mechanism (Table 1). This integration not only enhances the model's ability to capture spatial dependencies but also significantly improves prediction accuracy, underscoring the importance of leveraging both structural and interfacial features in protein–protein interaction studies. Interestingly, while the protein language model Saprot achieved higher accuracy in Table 1, we observed that replacing ESM2-3B with Saprot as a feature encoder did not yield improved performance in our model. We believe this can be attributed to the unique advantage provided by the geometric information integrated into the PPAP model. This additional geometric context may compensate for the structural information that is inherently limited in ESM2-3B, thus enhancing its performance over Saprot in our final model.

Performance on Test Sets S34 and PDBbind S90. We then compared our model with recent PPI affinity prediction models on the test set S34. The comprehensive results are presented in Table 2, demonstrating our model's superior performance across multiple evaluation metrics. While all other models, including ProAffinity-GNN,²² PPI-Affinity,²⁰ AAIN_Predictor,¹⁸ and PRODIGY,¹⁹ showed limited predictive capability on the test set S34, our model achieved an root-mean-square error (RMSE) of 1.77 kcal/mol and an MAE of 1.42 kcal/mol, significantly lower than existing methods. Notably, our model achieved a significant reduction of 0.67 kcal/mol in the MAE metric compared to the current best-performing model in affinity prediction. This improvement corresponds to an approximately 70% decrease in the error of predicted K_d values, representing a substantial advancement in the field.

Table 2. Performance Comparison of Different Models on the Test Set S34 ($N = 34$)

model	pearson R↑	RMSE↓	MAE↓
ProAffinity-GNN	0.28	3.16	2.26
PPI-Affinity	0.11	2.77	2.22
AAIN_Predictor	0.19	2.66	2.24
PRODIGY	0.23	2.66	2.09
ΔSASA	-0.04	/	/
AF3_ranking_score	-0.03	/	/
AF3_ipTM	0.00	/	/
AF3_pTM	-0.22	/	/
AF3_pae_interact	-0.18	/	/
PPAP	0.32 ± 0.09	1.77 ± 0.05	1.42 ± 0.02

Recent studies have indicated that the confidence scores provided by AlphaFold, such as the ranking score and ipTM, are associated with changes in protein–protein binding affinity upon mutations.⁴⁸ Building upon these findings, we hypothesized that these metrics might serve as potential indicators of protein–protein complex binding affinity. To investigate this hypothesis, we conducted a comprehensive correlation analysis between AlphaFold3’s confidence metrics (ipTM, pTM, ranking score, and pAE_interaction) and binding affinity using the test set S34 (Table 2). AlphaFold3’s confidence scores correlated poorly with affinity. The poor performance of ΔSASA and AlphaFold3’s confidence metrics on test set S34 highlights the complexity of protein–protein binding affinity prediction and the need for more sophisticated approaches. Our results demonstrate that the model represents a significant advancement in protein–protein interaction prediction, achieving robust performance in binding affinity prediction on the independent test set S34.

In general, experimentally determined protein structures serve as the primary benchmark for high-resolution protein structural analysis. However, when protein structures are unavailable—such as in certain protein design tasks or other scenarios—researchers increasingly utilize AlphaFold3 (AF3)-predicted structures as alternatives. We investigated whether employing AF3-predicted structures as input to our model would yield results comparable to those of PDB structures, and whether potential inaccuracies in AF3 predictions would adversely affect the model’s output.

We employed AF3 to predict structures for test set S34 and used these as input to our model. The resulting predictions were systematically compared against those generated using experimentally determined structures (Figure S2). We observed a direct correlation: the greater the structural deviation between AF3 predictions and PDB structures (quantified by higher Root-Mean-Square Deviation, RMSD), the larger the discrepancy between the model’s predictions based on each structure type. Therefore, we recommend utilizing high-accuracy AF3 structures for affinity prediction to minimize the impact of structural uncertainties on prediction reliability, when experimental structures are unavailable.

In addition, we also evaluated the performance of various models on the PDBbind S90 test set ($N = 90$, Table 3). Compared to five other models, our method achieved the highest Pearson correlation coefficient ($R = 0.63$), outperforming the others (R ranging from 0.28 to 0.59). In terms of MAE, our model reached 1.58 kcal/mol, comparable to ProAffinity-GNN (1.55 kcal/mol) but lower than those of the remaining four methods (ranging from 1.75 to 2.51 kcal/

Table 3. Performance Comparison of Different Models on the PDBbind S90 Test Set ($N = 90$)

model	pearson R↑	MAE↓
ProAffinity-GNN ^a	0.59	1.55
ProBAN ^b	0.55	1.75
PPI-Affinity ^a	0.50	1.80
PRODIGY ^a	0.31	2.51
ISLAND ^a	0.28	2.18
AF3_ranking_score	0.27	/
AF3_ipTM	0.32	/
AF3_pTM	0.49	/
AF3_pae_interact	0.33	/
ΔSASA	0.08	/
PPAP	0.63 ± 0.01	1.58 ± 0.03

^aPerformance data was sourced from the ProAffinity-GNN study.²²

^bProBAN was evaluated on a subset of 82 structures from PDBbind S90. Performance data was sourced from the ProBAN study.²¹

mol). Among the AlphaFold3-derived metrics, pTM exhibits the strongest correlation with binding affinity ($R = 0.49$), outperforming both ranking score ($R = 0.27$) and pAE_interaction ($R = 0.33$), indicating their potential applicability in affinity prediction tasks. However, this result contrasts sharply with that of the test set S34, further underscoring the difficulty and challenge of S34. The ΔSASA metric performs the least effectively, with a Pearson R of only 0.08. These results highlight the competitive performance of PPAP in predicting binding affinities, particularly in terms of correlation strength, while maintaining comparable accuracy in MAE to the top-performing models.

To evaluate the contribution of individual components within the interfacial attention mechanism, we performed ablation studies. As described in the Supporting Information, these studies revealed that the complete model achieved the lowest error. Removing individual components, particularly the interfacial attention module, resulted in an increase in mean absolute error (MAE), confirming their collective importance for the model’s high predictive accuracy and stability.

Retraining and Benchmarking on a Legacy Data set. To objectively evaluate the performance of our method and eliminate potential bias from the new data in PDBbind v2021, we conducted a comparative analysis using the data set from the ProBAN study.²¹ This study first manually annotates chain components directly involved in pairwise interactions within the PDBbind v2020 database. We therefore retrained our model on this data set and evaluated its performance on three distinct test sets: PDBbind S90 (Table 4), a subset of the structure-based benchmark for protein–protein binding affinity used in the ProAffinity-GNN study (Test set S79, Table 4),⁴⁹ and a curated data set from new entities in PDBbind v2021 with 40% similarity reduction to the training data (Test set S115, Table 4). Following ProBAN’s data partitioning protocol, test set data were excluded prior to partitioning the remaining data set into training and test subsets at an 8:2 ratio.

As shown in Table 4, comparative evaluations demonstrate that our model consistently achieves state-of-the-art (SOTA) performance across the first two test sets. Specifically, on the PDBbind S90 data set, PPAP attains the highest Pearson correlation coefficient ($R = 0.629$) and the lowest mean absolute error (MAE = 1.54 kcal/mol), substantially outperforming traditional methods such as PRODIGY ($R = 0.306$,

Table 4. Performance Comparison of Different Models on the Test Set PDBbind S90 ($N = 90$), S79 ($N = 79$) and S115 ($N = 115$)

method	PDBbind S90		test set S79		test set S115	
	$R \uparrow$	MAE \downarrow	$R \uparrow$	MAE \downarrow	$R \uparrow$	MAE \downarrow
PRODIGY ^a	0.306	2.51	<u>0.735</u>	<u>1.43</u>	0.089	2.55
DFIRE ^a	0.095	25.37	0.602	4.64	/	/
CP_PIE ^a	0.095	7.59	-0.517	8.80	/	/
ISLAND ^a	0.276	2.18	0.378	2.10	/	/
PPI-Affinity ^a	0.495	1.80	0.616	1.82	/	/
ProBAN ^b	0.550	1.75	/	/	/	/
ProAffinity-GNN ^a	<u>0.594</u>	<u>1.55</u>	0.733	1.51	0.494	2.20
ESM2-15B	0.532	1.74	0.712	1.55	0.515	1.52
ESM2-3B	0.546	1.69	0.669	1.63	<u>0.544</u>	1.49
ESM-1v	0.569	1.69	0.726	1.45	0.518	1.48
Saprot	0.571	1.71	0.679	1.59	0.547	1.52
ESM3-open	0.494	1.70	0.667	1.62	0.506	1.58
AF3_ranking_score	0.270	/	0.247	/	-0.028	/
Δ SASA	0.080	/	0.440	/	-0.058	/
PPAP ^c	0.629	1.54	0.793	1.36	0.540	<u>1.49</u>

^aPerformance data was sourced from the ProAffinity-GNN study.²² ^bProBAN was evaluated on a subset of 82 structures from PDBbind S90. Performance data was sourced from the ProBAN study.²¹ ^cThe model is trained based on ESM-1v.

MAE = 2.51 kcal/mol) and deep learning-based approaches like ProAffinity-GNN ($R = 0.594$, MAE = 1.55 kcal/mol). On the S79 test set, PPAP similarly achieves superior results ($R = 0.793$, MAE = 1.36 kcal/mol), surpassing all previously reported models and highlighting its strong generalization across data sets. Although some sequence-based models perform comparably to PPAP on the test set S115, their performance degrades significantly on older benchmarks, suggesting limited predictive capability for similar data. Additionally, when evaluated on the test set S115 comprising previously unseen data, PPAP reduces the MAE by 0.71 kcal/mol compared to two leading open-source models (PRODIGY¹⁹ and ProAffinity-GNN²²), as shown in Table 4. Notably, PPAP is among the few models that exhibit robust and stable performance across all three independent test sets (S90, S79, and S115), underscoring its reliability and practical utility across diverse protein–ligand binding scenarios. These findings suggest that the performance enhancement stems primarily from the model structure, rather than resulting from the incorporation of the PDBbind v2021 data set.

Performance on SKEMPI Test Set. To assess the generalization capability of our method for predicting the effects of protein mutations, we evaluated its performance using a subset of the SKEMPI 2.0 database.⁵¹ We utilized a previously established subset comprising 26 wild-type and 151 mutant entries.⁵² We excluded protein entries that appeared in both the PDBbind and SKEMPI data sets, along with their mutant variants, to prevent data leakage. After this filtering, we curated a final test set consisting of 25 wild-type and 145 mutant protein–protein complexes. Full structural models for the mutant complexes were generated using the AlphaFold3. A scatter plot comparing the predicted versus experimental ΔG values for the test set is shown in Figure S3. Subsequently, we evaluated the performance of our PPAP model (Table 5), which also benchmarks against other state-of-the-art methods for protein–protein binding affinity prediction, with their reported performance on the SKEMPI database or its subsets.

PPAP demonstrated promising performance on data sets containing mutant protein–protein complexes, despite not being explicitly trained on such systems. This underscores the

Table 5. Performance Comparison of Different Models on the SKEMPI Test Set

method	performance	specific for mutation
PPIformer ⁵³	$R = 0.46^a$	yes
GeoPPI ²⁶	$R = 0.52^{a,c}$	yes
binding-ddg-predictor ²⁷	$R = 0.73^a$	yes
MT-TopLap ⁵⁴	$R = 0.88^{a,c}$	yes
SSIPe ⁵⁵	$R = 0.61^a$	yes
PPI-Affinity ⁵²	$R = 0.77^{b,c}$	no
ProAffinity-GNN ⁵⁶	$R = 0.73^{b,c}$	no
PPAP	$R = 0.80^{b,c}$	no

^aPerformance is evaluated on respective test sets. Note that these test sets are different from the test set used in this study. R represents the correlation with $\Delta\Delta G$. Values may be biased due to varying deduplication protocols across studies. ^bTest sets are derived from the PPI-Affinity study and evaluated after deduplication based on respective training sets. R represents the correlation with ΔG . ^cData sourced from a previous study.⁵⁶

model's inherent robustness and generalization capability across diverse protein interactions, extending even to scenarios absent from its training data. It is crucial to note that while some alternative methods may superficially exhibit superior metrics, their reported performance is likely inflated due to inadequate assessment of data similarity within the test set.

Furthermore, we explored predicting relative changes in binding free energy ($\Delta\Delta G$) for 145 mutant data points. Acknowledging the task's complexity and that PPAP was not specifically designed for this purpose, the Mean Absolute Error (MAE) between experimental and predicted $\Delta\Delta G$ values was 1.440 kcal/mol. This represents a substantial improvement over the previously reported MAE of 1.964 kcal/mol for ProAffinity-GNN, a protein affinity prediction model also not trained on mutant data. While this result highlights the potential for further optimization in this specific prediction domain, it demonstrates PPAP's strong baseline capability even for this challenging application.

Application to Protein Binder Design. In protein binder design pipelines, after generating an extensive library of

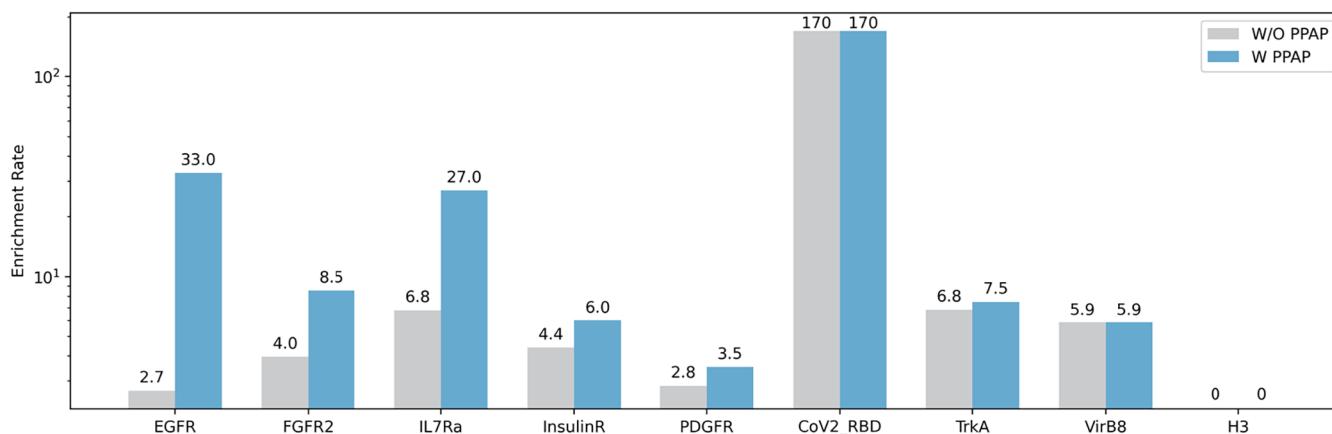


Figure 3. Comparison of enrichment rates between filtering methods. The enrichment rates for nine protein targets (EGFR, FGFR2, IL7Ra, InsulinR, PDGFR, CoV2_RBD, TrkA, VirB8 and H3) were compared between methods with PPAP (W PPAP) and without PPAP (W/O PPAP). Enrichment rate values were calculated based on the data collected by Cao et al.⁵⁷

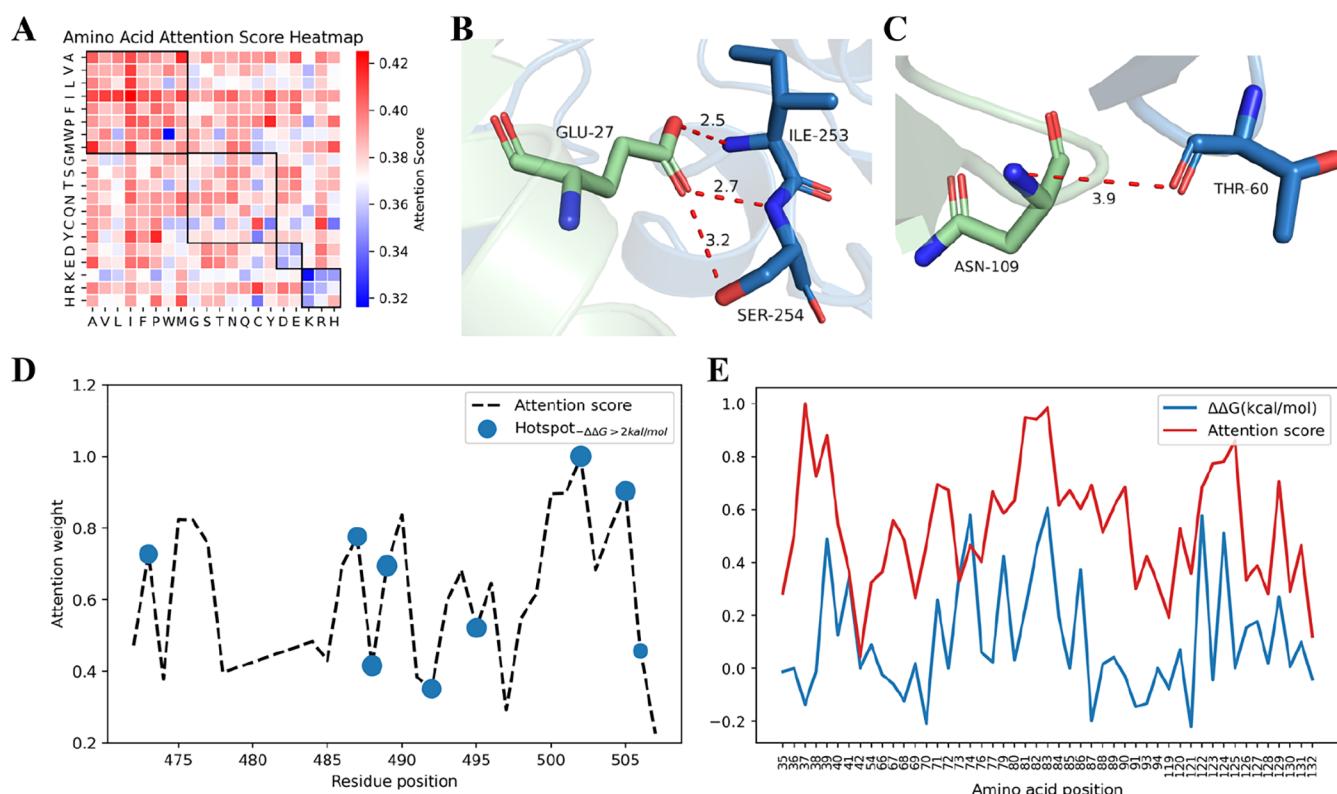


Figure 4. Decoding interaction patterns through attention mechanisms and interface residue analysis. (A) Attention score of amino-acid pairs on protein–protein interfaces. The attention scores was obtained by normalization and averaging of attention scores across different complexes in internal test set. The square boxes from upper left to bottom right in the figure represent hydrophobic–hydrophobic, polar–polar, acidic–acidic, basic–basic interactions. (B, C) Interactions of the hotspot residues. (B) Streptococcal protein G and IgG complex (PDB ID: 1FCF). C Acute leukemia-associated transcription factors (AML1) and core-binding factor (CBFbeta) complex (PDB ID: 1E50). The numbers between two atoms on the diagram represent the distance between them, measured in Å. (D) Diagram of hotspot residues and normalized attention scores at the SARS-CoV-2 RBD interface. The blue dots represent hotspot residues with a free energy change greater than 2 kcal/mol before and after alanine mutation. The size of the dots indicating the magnitude of the free energy change. (E) Correlation between normalized attention scores and binding affinity loss. The plot shows the normalized attention scores (green line) and binding affinity loss ($\Delta\Delta G$ in kcal/mol, blue line) for each amino acid position in the complex (PDB ID: 1OHZ). The observed correlation suggests that the model may help identify interfacial residues whose alanine mutations could influence binding affinity.

candidate sequences, researchers routinely employ AlphaFold-Multimer for complex structure prediction to validate successful target engagement and binding capability of the designed proteins.^{25,50} In previous studies, structures predicted by AlphaFold were screened *in silico* using the criteria of

pAE_interaction < 10 and complex RMSD < 5 Å,⁵⁰ which achieved a substantially greater success rate in success rate during *in vitro* verification. We reasoned that applying PPAP for affinity prediction after complex structure prediction may enhance the success rate of binder design.

To this end, we employed protein binder design data from Cao et al.⁵⁷ to compare the enrichment rate of two distinct screening strategies: (1) without PPAP, which involved filtering based solely on the AlphaFold–criterion, and (2) with PPAP, which entailed an additional PPAP screening step following the AlphaFold–criterion. The PPAP model was trained using the PDBbind v2021 data set, and the top-performing model from multiple random splits was subsequently selected for screening based on its test set performance. Two strategies were explored: proportional selection and fixed-number selection (Tables S2 and S3). Ultimately, the optimal screening method was determined to be the selection of the top 200 protein sequences predicted by our model, which showed improved enrichment rates for 6 of the 9 targets (Figure 3). Remarkably, the enrichment for EGFR improved by nearly 10-fold, while IL7Ra exhibited a 4-fold increase, demonstrating the effectiveness of model-guided selection. For three targets (CoV2_RBD, VirB8 and H3), the number of candidate sequences was already fewer than 200 prior to our screening process. Therefore, these targets were not subjected to our screening step. PPAP exhibits strong potential in protein binder design and provides a novel and effective strategy to enhance experimental success rates.

Mining the Key Elements at the Protein–Protein Binding Interface. To better understand the influence of residue pair interactions on protein affinity, we visualized the attention scores corresponding to these interactions. By averaging and visualizing the attention scores corresponding to residue pairs among the 20 amino acids in the internal test set, we aimed to identify patterns that reveal how specific amino acid interactions contribute to overall affinity. The results demonstrate that hydrophobic interactions have higher attentions scores and therefore contribute more to the predicted binding affinity compared to polar interactions, as illustrated in the upper left corner of Figure 4A.

Amino acids with similar electric charges exhibited lower attention scores, whereas interactions between oppositely charged residues had higher attention. This matches the frequency of occurrence of residues at the interface as reported in a previous study.⁵⁸ Studies have noted that cysteine (C) is rarely present at the interface and has a low probability of becoming a hotspot residue,⁵⁹ which aligns with our attention scores. Most residue pairs involving cysteine display low attention scores. Additionally, study found that isoleucine (I) is more likely to serve as a hotspot amino acid at the interface compared to leucine (L), despite their similar chemical properties as isomers.⁵⁹ The attention scores in our model also support this observation. These findings demonstrate that the model has, to a certain extent, learned physicochemical principles underlying protein–protein interactions, thereby enhancing its predictive accuracy.

To further assess whether the model has captured the physical interactions within PPIs, we investigated the relationship between the attention distribution in the model and the hotspots in PPIs. Streptococcal protein G and IgG complex (PDB ID: 1FCC) has a 4.9 kcal/mol decrease in affinity after mutation to alanine at position 27 of streptococcal protein G, which is the largest known mutation-induced decrease in binding affinity for this complex reported in the SKEMPI database.⁵¹ In accordance with this, the attention scores corresponding to the amino acid at position 27 are also the first and second among all scores (Table S4). As shown in Figure 4B, the GLU27 residue in streptococcal protein G interacts

with the isoleucine and serine of the IgG and potentially makes a large contribution to the affinity.

Similarly, acute leukemia-associated transcription factors (AML1) and core-binding factor (CBFbeta) complex (PDB ID: 1ES0) has a 2.4 kcal/mol decrease in binding affinity after mutation to alanine at position 109 of AML1.⁶⁰ It is the largest known mutation-induced decrease in binding affinity for this complex reported in the SKEMPI database (Table S5). According to the attention score of our model, the attention score corresponding to the amino acid at position 109 ranks first among all scores. As shown in Figure 4C, the ASN109 residue of the AML1 interacts with the THR60 residue of the CBFbeta and makes a large contribution to the affinity.

These examples illustrate that our model has learned to a certain extent to identify key residue pairs at the interface that significantly contribute to binding affinity. This ability is essential for accurately predicting the binding affinity of protein–protein interactions.

Alanine scanning determines the contribution of specific residues to the stability of protein complexes by mutating each amino acid residue in a protein to alanine one by one and then determining the effect of these mutations on the stability of the protein complex.⁶¹ This approach reveals the key residues that mediate molecular recognition and quantitatively assesses the contribution of each site to the strength of the interaction by comparing the difference in binding capacity between wild-type and alanine mutants. The attention score of each residue also represents its contribution to the interaction strength. We analyzed two independent alanine scanning studies with accompanying crystal structure data to systematically investigate potential associations between attention score and free energy changes ($\Delta\Delta G$) observed in the experimental scans.^{62,63}

Due to the 12-Å cutoff distance used in graph construction between residue pairs, each residue at the protein interface establishes numerous edges with adjacent residues, among which only a few actually represent true interactions. Therefore, we selected the highest attention score for each residue as a representative attention score for visualization. We calculated the normalized attention scores between ACE2 and the RBD of SARS-CoV-2 (PDB ID: 8DF5).⁶³

As shown in Figure 4D, the residues with high attention scores in our model align well with actual hotspot residues. However, some hotspot residues fall within regions of low attention score. We find that two hotspot amino acids located in these low-attention regions, specifically residue 488, form a disulfide bond within the chain that stabilizes the RBD structure, thus contributing significantly to affinity; similarly, amino acid 492 is oriented inward, forming hydrogen bonds that stabilize the RBD structure. Our model predominantly focuses on interprotein interactions, which may explain its limited capacity to capture the contributions of interfacial residues to protein stability internally.

Additionally, we visualized the attention scores for Cohesin-Dockerin complex from the cellulosome of *Clostridium thermocellum* (PDB ID: 1OHZ) (Figure 4E).^{62,64} The attention weights assigned by the model partially mirrored the trend of experimentally measured affinity loss upon alanine scanning mutagenesis, suggesting that the model may have implicitly learned to identify interfacial residues critical for binding affinity.

CONCLUSIONS

Quantitative prediction of the binding affinity of a protein–protein complex represents a significant challenge in the field of protein engineering and drug discovery. To reach this goal, the two major obstacles are the lack of sufficient training data and certain limitations in the existing computational methods. In response, we have compiled the largest data set to date of protein–protein complex structures with experimental binding affinity data. The binding affinity data of each complex was carefully curated from the public literature, and the relevant interacting units in each complex were explicated annotated. On this basis, we have developed PPAP, a new deep-learning model that leverages protein language models and the interfacial contact-aware attention mechanism. It effectively integrates both sequence and structural features and utilizes an elaborated decoder to capture the important elements at the protein–protein binding interface while still keeping the model simple. PPAP achieved superior performance across several benchmark data sets, showcasing the model’s reasonable generalization ability and consistency. Furthermore, the model’s interpretability aligns with previous experimental findings, while its application in protein design shows promising potential. This approach establishes a novel computational paradigm for the precise modulation of protein–protein interactions, offering promising potential for advancing protein design and engineering.

ASSOCIATED CONTENT

Data Availability Statement

The complete data set, including PDB IDs for train and test sets, chain annotations for receptor and ligand proteins, raw and processed structural data of protein–protein complexes, and binding affinity data, is publicly available on the PDBbind+ Web site (<https://www.pdbbind-plus.org.cn>). Access to the data set requires free user registration on the PDBbind+. The source code, including scripts for data processing and model implementation, is available on our GitHub repository (<https://github.com/TEKHOO/PPAP>). Additionally, the model checkpoint is hosted on Hugging Face (<https://huggingface.co/qj666/PPAP/tree/main>).

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jcim.5c01390>.

Comparison of our data processing pipeline with that of PPB-affinity, description of the ablation study, examples of different receptor and ligand chain assignments in our approach and PPB-affinity (Figure S1), correlation between AF3 structural deviation and model prediction discrepancy (Figure S2), performance of PPAP on the mutation test set (Figure S3), Data set Partitioning (Table S1), comparison of enrichment rates under different screening strategies (Table S2–3), normalized attention weight for Interprotein residues (Table S4–5), detailed ablation study results (Table S6) ([PDF](#))

AUTHOR INFORMATION

Corresponding Authors

Renxiao Wang – Department of Medicinal Chemistry, School of Pharmacy, Fudan University, Shanghai 201203, People’s Republic of China; orcid.org/0000-0003-0485-0259; Email: wangrx@fudan.edu.cn

Yifei Qi – Department of Medicinal Chemistry, School of Pharmacy, Fudan University, Shanghai 201203, People’s Republic of China; orcid.org/0000-0003-2853-7910; Email: yfqi@fudan.edu.cn

Authors

Jie Qian – Department of Medicinal Chemistry, School of Pharmacy, Fudan University, Shanghai 201203, People’s Republic of China

Lin Yang – Department of Medicinal Chemistry, School of Pharmacy, Fudan University, Shanghai 201203, People’s Republic of China

Zhen Duan – Shanghai TopScience Co., Ltd., Shanghai 200436, People’s Republic of China

Complete contact information is available at: <https://pubs.acs.org/10.1021/acs.jcim.5c01390>

Author Contributions

J.Q., R.W. and Y.Q.: Conceived the research project. J.Q., L.Y. and Y.Q.: Designed and trained the model. J.Q.: Performed data analysis and wrote the paper. Z.D.: Provided support for maintaining the PDBbind+ Web site and data set. All authors read and commented on the paper.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This study was financially supported by the National Natural Science Foundation of China (Grants No. 22033001, 22373020 to Y.Q. and 22537001, 82173739 to R.W.) and the Shanghai Municipal Science and Technology Commission (Grant No. 25JS2830200 to R.W.). The computations in this research were performed using the CFFF platform of Fudan University.

REFERENCES

- (1) Rabbani, G.; Baig, M. H.; Ahmad, K.; Choi, I. Protein-Protein Interactions and Their Role in Various Diseases and Their Prediction Techniques. *Curr. Protein Pept. Sci.* **2018**, *19* (10), 948–957.
- (2) Soleymani, F.; Paquet, E.; Viktor, H.; Michalowski, W.; Spinello, D. Protein–Protein Interaction Prediction with Deep Learning: A Comprehensive Review. *Comput. Struct. Biotechnol. J.* **2022**, *20*, 5316–5341.
- (3) Ramos, R. H.; De Oliveira Lage Ferreira, C.; Simao, A. Human Protein–Protein Interaction Networks: A Topological Comparison Review. *Heliyon* **2024**, *10* (5), No. e27278.
- (4) Peng, R.; Deng, M. Mapping the Protein–Protein Interactome in the Tumor Immune Microenvironment. *Antibody Ther.* **2023**, *6* (4), 311–321.
- (5) Farooq, Q. U. A.; Shaukat, Z.; Aiman, S.; Li, C.-H. Protein-Protein Interactions: Methods, Databases, and Applications in Virus-Host Study. *WJV* **2021**, *10* (6), 288–300.
- (6) Lage, K. Protein–Protein Interactions and Genetic Diseases: The Interactome. *Biochim. Biophys. Acta, Mol. Basis Dis.* **2014**, *1842* (10), 1971–1980.
- (7) Leregian, A.; Marsden, H. S.; Palù, G. Protein–Protein Interactions as Targets for Antiviral Chemotherapy. *Rev. Med. Virol.* **2002**, *12* (4), 239–262.
- (8) Ryan, D.; Matthews, J. Protein–Protein Interactions in Human Disease. *Curr. Opin. Struct. Biol.* **2005**, *15* (4), 441–446.
- (9) Peng, X.; Wang, J.; Peng, W.; Wu, F.-X.; Pan, Y. Protein–Protein Interactions: Detection, Reliability Assessment and Applications. *Brief Bioinform.* **2016**, *18*, 798–819.

- (10) Gromiha, M. M.; Yugandhar, K.; Jemimah, S. Protein–Protein Interactions: Scoring Schemes and Binding Affinity. *Curr. Opin. Struct. Biol.* **2017**, *44*, 31–38.
- (11) Akbarzadeh, S.; Coşkun, Ö.; Günçer, B. Studying Protein–Protein Interactions: Latest and Most Popular Approaches. *J. Struct. Biol.* **2024**, *216* (4), No. 108118.
- (12) Elhabashy, H.; Merino, F.; Alva, V.; Kohlbacher, O.; Lupas, A. N. Exploring Protein–Protein Interactions at the Proteome Level. *Structure* **2022**, *30* (4), 462–475.
- (13) Zhou, M.; Li, Q.; Wang, R. Current Experimental Methods for Characterizing Protein–Protein Interactions. *ChemMedChem* **2016**, *11* (8), 738–756.
- (14) Zheng, F.; Jiang, X.; Wen, Y.; Yang, Y.; Li, M. Systematic Investigation of Machine Learning on Limited Data: A Study on Predicting Protein–Protein Binding Strength. *Comput. Struct. Biotechnol. J.* **2024**, *23*, 460–472.
- (15) Abbasi, W. A.; Yaseen, A.; Hassan, F. U.; Andleeb, S.; Minhas, F. U. A. A. ISLAND: In-Silico Proteins Binding Affinity Prediction Using Sequence Information. *BioData Min.* **2020**, *13* (1), No. 20.
- (16) Lin, Z.; Akin, H.; Rao, R.; Hie, B.; Zhu, Z.; Lu, W.; Smetanin, N.; Verkuil, R.; Kabeli, O.; Shmueli, Y.; dos Santos Costa, A.; Fazel-Zarandi, M.; Sercu, T.; Candido, S.; Rives, A. Evolutionary-Scale Prediction of Atomic-Level Protein Structure with a Language Model. *Science* **2023**, *379* (6637), 1123–1130.
- (17) Yue, Y.; Li, S.; Cheng, Y.; Wang, L.; Hou, T.; Zhu, Z.; He, S. Integration of Molecular Coarse-Grained Model into Geometric Representation Learning Framework for Protein–Protein Complex Property Prediction. *Nat. Commun.* **2024**, *15* (1), No. 9629.
- (18) Yi, C. H.; Taylor, M. L.; Ziebarth, J.; Wang, Y. Predictive Models and Impact of Interfacial Contacts and Amino Acids on Protein–Protein Binding Affinity. *ACS Omega* **2024**, *9* (3), 3454–3468.
- (19) Xue, L. C.; Rodrigues, J. P.; Kastritis, P. L.; Bonvin, A. M.; Vangone, A. PRODIGY: A Web Server for Predicting the Binding Affinity of Protein–Protein Complexes. *Bioinformatics* **2016**, *32* (23), 3676–3678.
- (20) Romero-Molina, S.; Ruiz-Blanco, Y. B.; Mieres-Perez, J.; Harms, M.; Münch, J.; Ehrmann, M.; Sanchez-Garcia, E. PPI-Affinity: A Web Tool for the Prediction and Optimization of Protein–Peptide and Protein–Protein Binding Affinity. *J. Proteome Res.* **2022**, *21* (8), 1829–1841.
- (21) Bogdanova, E. A.; Novoseletsky, V. N. PROBAN: Neural Network Algorithm for Predicting Binding Affinity in Protein–Protein Complexes. *Proteins* **2024**, *92* (9), 1127–1136.
- (22) Zhou, Z.; Yin, Y.; Han, H.; Jia, Y.; Koh, J. H.; Kong, A. W.-K.; Mu, Y. ProAffinity-GNN: A Novel Approach to Structure-Based Protein–Protein Binding Affinity Prediction via a Curated Dataset and Graph Neural Networks. *J. Chem. Inf. Model.* **2024**, *64*, 8796–8808, DOI: [10.1021/acs.jcim.4c01850](https://doi.org/10.1021/acs.jcim.4c01850).
- (23) Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A. J.; Bambrick, J.; Bodenstein, S. W.; Evans, D. A.; Hung, C.-C.; O'Neill, M.; Reiman, D.; Tunyasuvunakool, K.; Wu, Z.; Žemgulytė, A.; Arvaniti, E.; Beattie, C.; Bertolli, O.; Bridgland, A.; Cherepanov, A.; Congreve, M.; Cowen-Rivers, A. I.; Cowie, A.; Figurnov, M.; Fuchs, F. B.; Gladman, H.; Jain, R.; Khan, Y. A.; Low, C. M. R.; Perlin, K.; Potapenko, A.; Savvy, P.; Singh, S.; Stecula, A.; Thillaisundaram, A.; Tong, C.; Yakneen, S.; Zhong, E. D.; Zielinski, M.; Žídek, A.; Bapst, V.; Kohli, P.; Jaderberg, M.; Hassabis, D.; Jumper, J. M. Accurate Structure Prediction of Biomolecular Interactions with AlphaFold 3. *Nature* **2024**, *630* (8016), 493–500.
- (24) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589.
- (25) Evans, R.; O'Neill, M.; Pritzel, A.; Antropova, N.; Senior, A.; Green, T.; Žídek, A.; Bates, R.; Blackwell, S.; Yim, J.; Ronneberger, O.; Bodenstein, S.; Zielinski, M.; Bridgland, A.; Potapenko, A.; Cowie, A.; Tunyasuvunakool, K.; Jain, R.; Clancy, E.; Kohli, P.; Jumper, J.; Hassabis, D. Protein Complex Prediction with AlphaFold-Multimer *bioRxiv* **2021**.
- (26) Liu, X.; Luo, Y.; Li, P.; Song, S.; Peng, J. Deep Geometric Representations for Modeling Effects of Mutations on Protein–Protein Binding Affinity. *PLoS Comput. Biol.* **2021**, *17* (8), No. e1009284.
- (27) Shan, S.; Luo, S.; Yang, Z.; Hong, J.; Su, Y.; Ding, F.; Fu, L.; Li, C.; Chen, P.; Ma, J.; Shi, X.; Zhang, Q.; Berger, B.; Zhang, L.; Peng, J. Deep Learning Guided Optimization of Human Antibody against SARS-CoV-2 Variants with Broad Neutralization. *Proc. Natl. Acad. Sci. U.S.A.* **2022**, *119* (11), No. e2122954119.
- (28) Si, Y.; Yan, C. Protein Language Model-Embedded Geometric Graphs Power Inter-Protein Contact Prediction. *eLife* **2024**, *12*, No. RP92184.
- (29) Wang, R.; Fang, X.; Lu, Y.; Wang, S. The PDBbind Database: Collection of Binding Affinities for Protein–Ligand Complexes with Known Three-Dimensional Structures. *J. Med. Chem.* **2004**, *47* (12), 2977–2980.
- (30) Nikam, R.; Yugandhar, K.; Gromiha, M. M. Deep Learning-Based Method for Predicting and Classifying the Binding Affinity of Protein–Protein Complexes. *Biochim. Biophys. Acta, Proteins Proteomics* **2023**, *1871* (6), No. 140948.
- (31) Duarte, J. M.; Srebnik, A.; Schärer, M. A.; Capitani, G. Protein Interface Classification by Evolutionary Analysis. *BMC Bioinf.* **2012**, *13* (1), No. 334.
- (32) Raybould, M. I. J.; Marks, C.; Lewis, A. P.; Shi, J.; Bujotzek, A.; Taddese, B.; Deane, C. M. Thera-SAbDab: The Therapeutic Structural Antibody Database. *Nucleic Acids Res.* **2020**, *48* (D1), D383–D388.
- (33) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235–242.
- (34) Liu, H.; Chen, P.; Zhai, X.; Huo, K.-G.; Zhou, S.; Han, L.; Fan, G. PPB-Affinity: Protein–Protein Binding Affinity Dataset for AI-Based Protein Drug Discovery. *Sci. Data* **2024**, *11*, No. 1316, DOI: [10.1038/s41597-024-03997-4](https://doi.org/10.1038/s41597-024-03997-4).
- (35) Faisal, F. E.; Newaz, K.; Chaney, J. L.; Li, J.; Emrich, S. J.; Clark, P. L.; Milenković, T. GRAFENE: Graphlet-Based Alignment-Free Network Approach Integrates 3D Structural and Sequence (Residue Order) Data to Improve Protein Structural Comparison. *Sci. Rep.* **2017**, *7* (1), No. 14890.
- (36) Chiang, Y.; Hui, W.-H.; Chang, S.-W. Encoding Protein Dynamic Information in Graph Representation for Functional Residue Identification. *Cell Rep. Phys. Sci.* **2022**, *3* (7), No. 100975.
- (37) Li, B.; Ming, D. GATSol, an Enhanced Predictor of Protein Solubility through the Synergy of 3D Structure Graph and Large Language Modeling. *BMC Bioinf.* **2024**, *25* (1), No. 204.
- (38) Park, J.; Sandberg, I. W. Universal Approximation Using Radial-Basis-Function Networks. *Neural Comput.* **1991**, *3* (2), 246–257.
- (39) Aldeano, E. A. MP-NeRF: A Massively Parallel Method for Accelerating Protein Structure Reconstruction from Internal Coordinates. *J. Comput. Chem.* **2021**, *43*, 74–78, DOI: [10.1002/jcc.26768](https://doi.org/10.1002/jcc.26768).
- (40) Swails pdbfixer; GitHub. <https://github.com/openmm/pdbfixer>.
- (41) Pursell, L.; Trimble, S. Y. Gram–Schmidt Orthogonalization by Gauss Elimination. *Am. Math. Mon.* **1991**, *98* (6), 544–549.
- (42) Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; Polosukhin, I. *Attention Is All You Need*; Advances in Neural Information Processing Systems 30; NIPS, 2023.
- (43) Kingma, D. P.; Ba, J. Adam: A Method for Stochastic Optimization. 2017. arXiv:1412.6980. arXiv.org e-Printarchive. <https://doi.org/10.48550/arXiv.1412.6980>.
- (44) Meier, J.; Rao, R.; Verkuil, R.; Liu, J.; Sercu, T.; Rives, A. *Language Models Enable Zero-Shot Prediction of the Effects of Mutations*

- on Protein Function; Advances in Neural Information Processing Systems 34; NIPS, 2021.
- (45) Hayes, T.; Rao, R.; Akin, H.; Sofroniew, N. J.; Oktay, D.; Lin, Z.; Verkuil, R.; Tran, V. Q.; Deaton, J.; Wiggert, M.; Badkundri, R.; Shafkat, I.; Gong, J.; Derry, A.; Molina, R. S.; Thomas, N.; Khan, Y. A.; Mishra, C.; Kim, C.; Bartie, L. J.; Nemeth, M.; Hsu, P. D.; Sercu, T.; Candido, S.; Rives, A. Simulating 500 Million Years of Evolution with a Language Model. *Science* **2025**, *387*, No. eads0018.
- (46) Su, J.; Han, C.; Zhou, Y.; Shan, J.; Zhou, X.; Yuan, F. SaProt: Protein Language Modeling with Structure-Aware Vocabulary. 2023.
- (47) Gainza, P.; Sverrisson, F.; Monti, F.; Rodolà, E.; Boscaini, D.; Bronstein, M. M.; Correia, B. E. Deciphering Interaction Fingerprints from Protein Molecular Surfaces Using Geometric Deep Learning. *Nat. Methods* **2020**, *17* (2), 184–192.
- (48) Lu, W.; Zhang, J.; Rao, J.; Zhang, Z.; Zheng, S. AlphaFold3, a Secret Sauce for Predicting Mutational Effects on Protein-Protein Interactions *bioRxiv* 2024.
- (49) Kastritis, P. L.; Moal, I. H.; Hwang, H.; Weng, Z.; Bates, P. A.; Bonvin, A. M. J. J.; Janin, J. A Structure-based Benchmark for Protein–Protein Binding Affinity. *Protein Sci.* **2011**, *20* (3), 482–491.
- (50) Bennett, N. R.; Coventry, B.; Goreshnik, I.; Huang, B.; Allen, A.; Vafeados, D.; Peng, Y. P.; Dauparas, J.; Baek, M.; Stewart, L.; DiMaio, F.; De Munck, S.; Savvides, S. N.; Baker, D. Improving de Novo Protein Binder Design with Deep Learning. *Nat. Commun.* **2023**, *14* (1), No. 2625.
- (51) Jankauskaite, J.; Jiménez-García, B.; Dapkusas, J.; Fernández-Recio, J.; Moal, I. H. SKEMPI 2.0: An Updated Benchmark of Changes in Protein–Protein Binding Energy, Kinetics and Thermodynamics upon Mutation. *Bioinformatics* **2019**, *35* (3), 462–469.
- (52) Luo, S.; Su, Y.; Wu, Z.; Su, C.; Peng, J.; Ma, J. Rotamer Density Estimator Is an Unsupervised Learner of the Effect of Mutations on Protein–Protein Interaction *bioRxiv* 2023.
- (53) Bushuiev, A.; Bushuiev, R.; Kouba, P.; Filkin, A.; Gabrielova, M.; Gabriel, M.; Sedlar, J.; Pluskal, T.; Damborsky, J.; Mazurenko, S.; Sivic, J. Learning to Design Protein–Protein Interactions with Enhanced Generalization. 2024. arXiv:2310.18515. arXiv.org e-Printarchive. <https://doi.org/10.48550/arXiv.2310.18515>.
- (54) Wee, J.; Chen, J.; Xia, K.; Wei, G.-W. Integration of Persistent Laplacian and Pre-Trained Transformer for Protein Solubility Changes upon Mutation. *Comput. Biol. Med.* **2024**, *169*, No. 107918.
- (55) Huang, X.; Zheng, W.; Pearce, R.; Zhang, Y. SSIPe: Accurately Estimating Protein–Protein Binding Affinity Change upon Mutations Using Evolutionary Profiles in Combination with an Optimized Physical Energy Function. *Bioinformatics* **2020**, *36* (8), 2429–2437.
- (56) Zhou, Z.; Yin, Y.; Han, H.; Jia, Y.; Koh, J. H.; Kong, A. W.-K.; Mu, Y. ProAffinity-GNN: A Novel Approach to Structure-Based Protein–Protein Binding Affinity Prediction via a Curated Data Set and Graph Neural Networks. *J. Chem. Inf. Model.* **2024**, *64* (23), 8796–8808.
- (57) Cao, L.; Coventry, B.; Goreshnik, I.; Huang, B.; Sheffler, W.; Park, J. S.; Jude, K. M.; Marković, I.; Kadam, R. U.; Verschueren, K. H. G.; Verstraete, K.; Walsh, S. T. R.; Bennett, N.; Phal, A.; Yang, A.; Kozodoy, L.; DeWitt, M.; Picton, L.; Miller, L.; Strauch, E.-M.; DeBouver, N. D.; Pires, A.; Bera, A. K.; Halabiya, S.; Hammerson, B.; Yang, W.; Bernard, S.; Stewart, L.; Wilson, I. A.; Ruohola-Baker, H.; Schlessinger, J.; Lee, S.; Savvides, S. N.; Garcia, K. C.; Baker, D. Design of Protein-Binding Proteins from the Target Structure Alone. *Nature* **2022**, *605* (7910), 551–560.
- (58) Ansari, S.; Helms, V. Statistical Analysis of Predominantly Transient Protein–Protein Interfaces. *Proteins* **2005**, *61* (2), 344–355.
- (59) Bogan, A. A.; Thorn, K. S. Anatomy of Hot Spots in Protein Interfaces. *J. Mol. Biol.* **1998**, *280* (1), 1–9.
- (60) Warren, A. J. Structural Basis for the Heterodimeric Interaction between the Acute Leukaemia-Associated Transcription Factors AML1 and CBFbeta. *EMBO J.* **2000**, *19* (12), 3004–3015.
- (61) Weiss, G. A.; Watanabe, C.; Zhong, A.; Goddard, A.; Sidhu, S. Rapid Mapping of Protein Functional Epitopes by Combinatorial Alanine Scanning. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97* (16), 8950–8954.
- (62) Kowalsky, C. A.; Whitehead, T. A. Determination of Binding Affinity upon Mutation for Type I Dockerin–Cohesin Complexes from *C. Listerium Thermocellum* and *C. Listerium cellulolyticum* Using Deep Sequencing. *Proteins* **2016**, *84* (12), 1914–1928.
- (63) Starr, T. N.; Greaney, A. J.; Hannon, W. W.; Loes, A. N.; Hauser, K.; Dillen, J. R.; Ferri, E.; Farrell, A. G.; Dadonaite, B.; McCallum, M.; Matreyek, K. A.; Corti, D.; Veesler, D.; Snell, G.; Bloom, J. D. Shifting Mutational Constraints in the SARS-CoV-2 Receptor-Binding Domain during Viral Evolution. *Science* **2022**, *377* (6604), 420–424.
- (64) Carvalho, A. L.; Dias, F. M. V.; Prates, J. A. M.; Nagy, T.; Gilbert, H. J.; Davies, G. J.; Ferreira, L. M. A.; Romão, M. J.; Fontes, C. M. G. A. Cellulosome Assembly Revealed by the Crystal Structure of the Cohesin–Dockerin Complex. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100* (24), 13809–13814.



CAS BIOFINDER DISCOVERY PLATFORM™

BRIDGE BIOLOGY AND CHEMISTRY FOR FASTER ANSWERS

Analyze target relationships,
compound effects, and disease
pathways

Explore the platform

