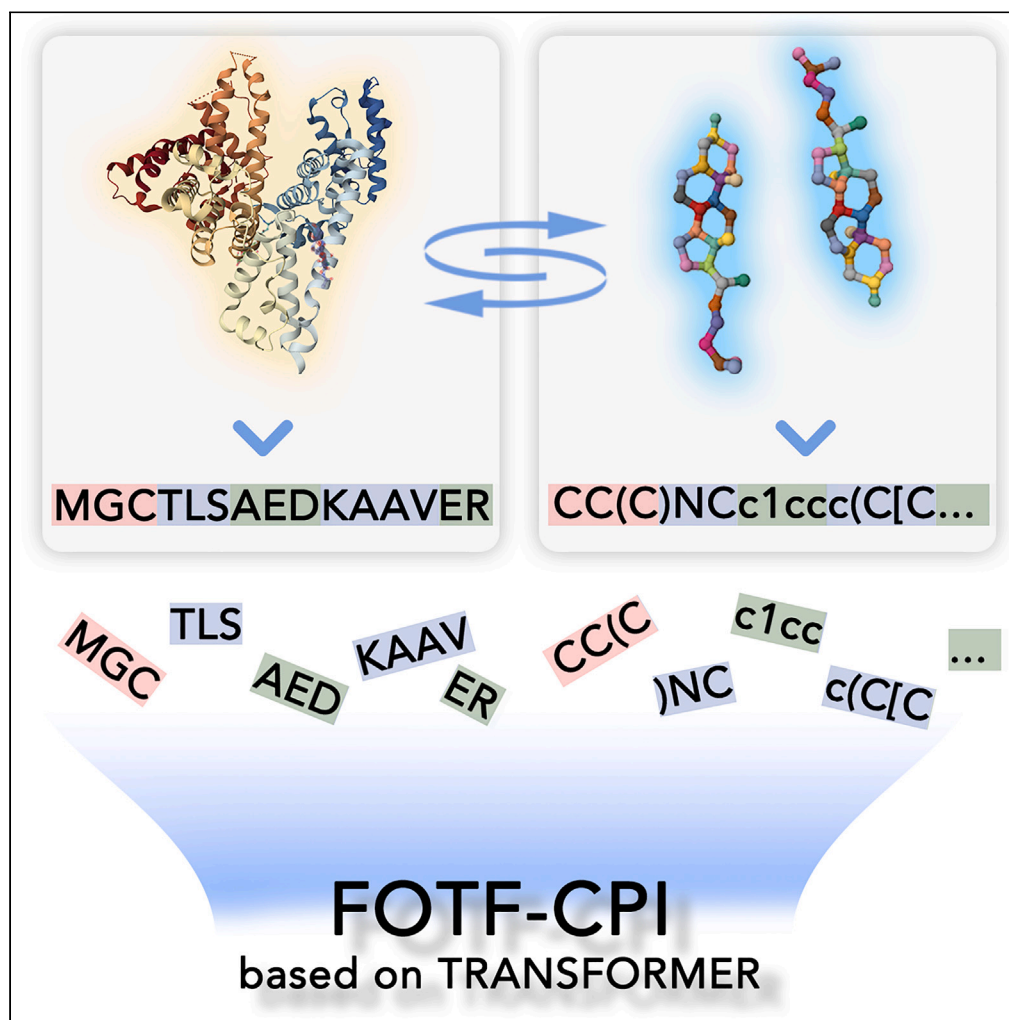


## Article

## FOTF-CPI: A compound-protein interaction prediction transformer based on the fusion of optimal transport fragments



Zeyu Yin, Yu Chen,  
Yajie Hao, Sanjeevi  
Pandiyani, Jinsong  
Shao, Li Wang

wangli@ntu.edu.cn

#### Highlights

A deep learning model  
FOTF-CPI is constructed to  
predict binding affinity

Better affinity prediction  
results are achieved than  
existing methods

An optimal transmission-  
based Fragments slicing  
method is designed

A feature fusion attention  
mechanism designed for  
CPI

Yin et al., iScience 27, 108756  
January 19, 2024 © 2023 The  
Authors.  
[https://doi.org/10.1016/  
j.isci.2023.108756](https://doi.org/10.1016/j.isci.2023.108756)

## Article

## FOTF-CPI: A compound-protein interaction prediction transformer based on the fusion of optimal transport fragments

Zeyu Yin,<sup>1</sup> Yu Chen,<sup>1</sup> Yajie Hao,<sup>1</sup> Sanjeevi Pandiyan,<sup>2</sup> Jinsong Shao,<sup>1</sup> and Li Wang<sup>1,2,3,\*</sup>

## SUMMARY

Compound-protein interaction (CPI) affinity prediction plays an important role in reducing the cost and time of drug discovery. However, the interpretability of how fragments function in CPI is impacted by the fact that current methods ignore the affinity relationships between fragments of compounds and fragments of proteins in CPI modeling. This article introduces an improved Transformer called FOTF-CPI (a Fusion of Optimal Transport Fragments compound-protein interaction prediction model). We use an optimal transport-based fragmentation approach to improve the model's understanding of compound and protein sequences. Additionally, a fused attention mechanism is employed, which combines the features of fragments to capture full affinity information. This fused attention redistributes higher attention scores to fragments with higher affinity. Experimental results show FOTF-CPI achieves an average 2% higher performance than other models on all three datasets. Furthermore, the visualization confirms the potential of FOTF-CPI for drug discovery applications.

## INTRODUCTION

Achieving pharmaceutical industry standards requires cooperative technological developments across numerous research fields for drug discovery.<sup>1,2</sup> Predicting compound-protein interactions (CPI) - encompassing drug-target interactions, drug-target affinity, interaction sites, and biological activity - is crucial to drug development.<sup>3-6</sup> CPI involves time- and labor-intensive compound selection, design, and optimization based on disease-specific target proteins.<sup>7-9</sup> Recently, artificial intelligence has been applied to CPI tasks to identify active compounds before clinical trials.<sup>10,11</sup>

Machine learning (ML) techniques are widely used for CPI, significantly improving clinical trial success rates.<sup>12-14</sup> ML-based approaches utilize techniques such as kernel functions, tree classifiers, and neural networks to predict compound-protein affinity (CPA).<sup>15-17</sup> Deep learning has greatly impacted CPI, as these techniques effectively analyze compound and protein 3D structures.<sup>18-21</sup> However, obtaining quality 3D protein structures experimentally remains challenging, limiting available 3D dataset sizes.<sup>22</sup> While AlphaFold2 helps augment limited 3D data, affinity prediction using 3D datasets still faces high computational costs. Thus, large-scale CPI predictions using 3D structures are hindered by data availability and computational expense. In fact, structure-aware methods have made significant progress in terms of inference speed,<sup>23,24</sup> enabling them to achieve similar computation time as sequence-based methods in predicting affinity. Although these methods have achieved high inference speed, training affinity prediction models based on structure encounters challenges due to the complexity of 3D or graph-structured data, which inevitably requires higher computational resources compared to sequence-based methods.<sup>25</sup> These limitations can impact the practical application of affinity prediction in real-world scenarios and pose obstacles to optimizing data for different contexts.

Sequence-based approaches overcome limited structural data and molecular docking costs. Using sequence data (e.g., SMILES, amino acid sequences) significantly reduces prediction costs.<sup>26</sup> However, these deep learning models lack molecular insights explaining compound-protein interactions. Viewing compound and protein sequences as fragments provides perspective, as CPI derives from fragment interactions.<sup>27,28</sup> Capturing fragment sequence semantics can improve affinity prediction. Current methods overlook fragment interactions, losing information on how partial interactions combine into overall compound-protein affinity.

Fragmenting compounds and proteins can address neural networks ignoring molecular relationships. Fragmentation method strongly impacts model performance, similar to word segmentation in NLP.<sup>29-31</sup> Classification models can consider target independent features and simulate fragment interactive features from a biological lens. However, current fragment-based approaches do not focus on inter-fragment interactions or fragment influence on complete sequences, instead concentrating on fragment generation and global structure.<sup>28,32</sup> Lacking interpretability, these methods cannot distinguish fragment affinities. Attention mechanisms have been combined with compound and

<sup>1</sup>School of Information Science and Technology, Nantong University, Nantong 226001, China

<sup>2</sup>Research Center for Intelligent Information Technology, Nantong University, Nantong 226001, China

<sup>3</sup>Lead contact

\*Correspondence: wangli@ntu.edu.cn

<https://doi.org/10.1016/j.isci.2023.108756>



**Table 1. The statistics of datasets**

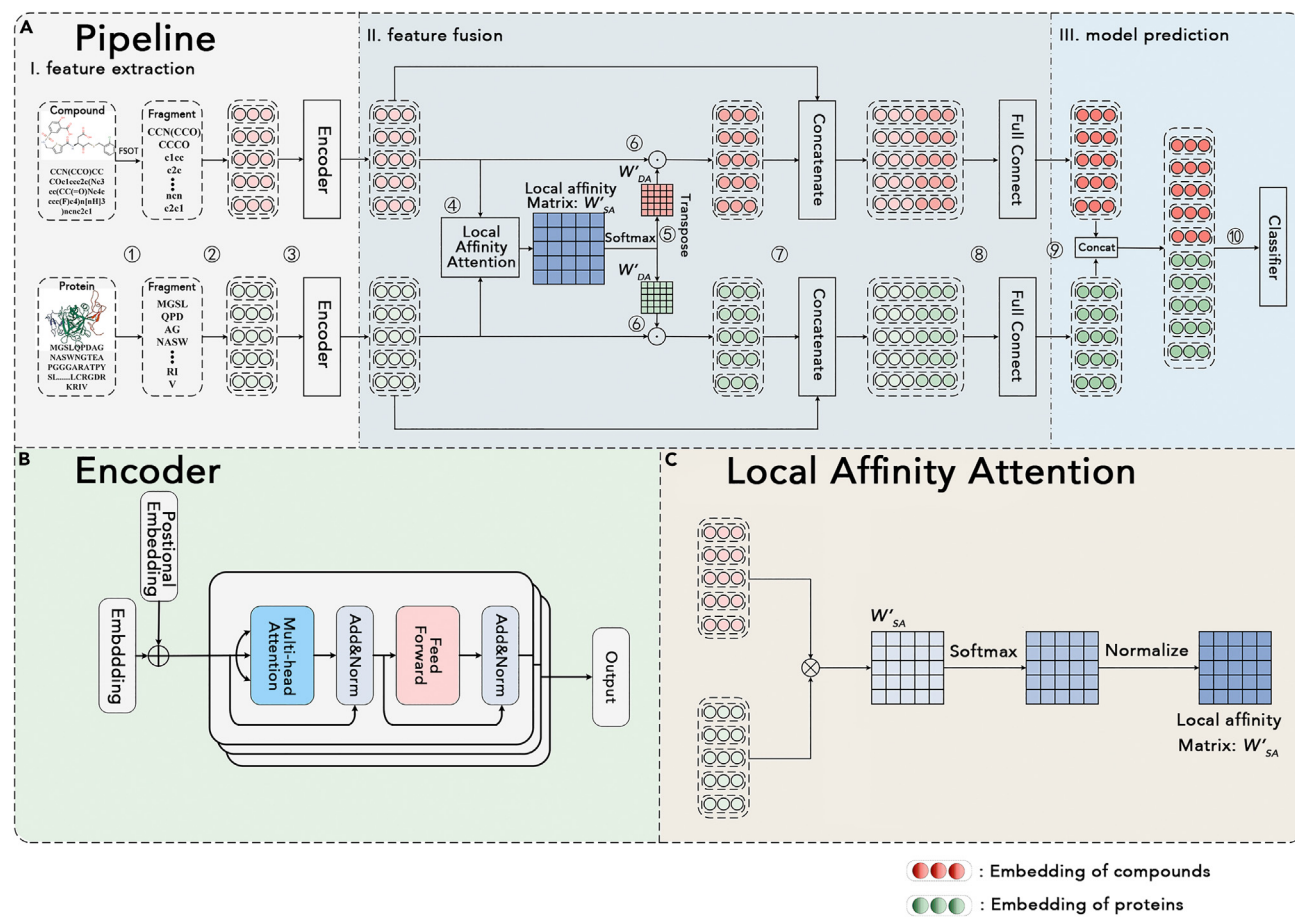
Dataset	Drugs	Proteins	Pos Interactions (train/test)	Neg Interactions (train/test)
BindingDB	10,665	1,413	7,333/1,833	18,748/5,313
Davis	68	379	1,205/301	7,678/1,909
Biosnap	4,510	2,181	10,993/2,748	10,993/2,748
DUD-E	22,886	102	18,116/4,529	1,378,996/281,429

protein structure information to improve model interpretability, paralleling NLP uses.<sup>33,34</sup> However, current attention mechanisms fail to explain compound-protein fragment links.

To address these CPI issues, we propose a transformer called FOTF-CPI (a Fusion of Optimal Transport Fragments compound-protein interaction prediction model), studying fragment interactions to determine overall affinity for faster, lower-cost drug discovery.

## RESULTS

In this section, we report the performance of FOTF-CPI on datasets including BindingDB, Davis, Biosnap, and DUD-E, which are described as targets in the Table 1 and STAR methods section. We first compared the performance of the FOTF-CPI with other baseline models which are described as targets in the STAR methods section. Second, we reported on the AUC and PRC performance of our model on BindingDB dataset. Third, we compared different fragment slicing methods to evaluate our slicing method FOST (Fragment Slicing based on Optimal Transport). Fourth, we assessed the role of each of the proposed modules on affinity prediction by ablation experiments. Fifth, we validate the generalization ability of FOTF-CPI in affinity prediction through cross dataset experiments. Six, we evaluate the performance of



**Figure 1. The pipeline of FOTF-CPI**

(A) This is the complete pipeline we proposed for the affinity prediction of proteins and compounds using FSOT.

(B) This is the process of encoding the segmented fragments into embeddings.

(C) This is the process of seeking local interactions between fragments of proteins and compounds.

**Table 2. The results of the BindingDB dataset**

Model	AUC (SD)	PRC (SD)	Sensitivity (SD)	Specificity (SD)	F1 (SD)	Cost (h)
DeepDTA	0.901 (0.007)	0.810 (0.006)	0.780 (0.024)	0.905 (0.017)	0.757 (0.014)	12.65
TransformerCPI	0.910 (0.007)	0.788 (0.011)	0.736 (0.014)	0.890 (0.012)	0.731 (0.005)	5.44
Moltrans	0.903 (0.002)	0.806 (0.007)	0.762 (0.013)	0.908 (0.007)	0.752 (0.004)	1.20
ML-DTI	0.902 (0.007)	0.785 (0.011)	0.753 (0.007)	0.851 (0.005)	0.763 (0.003)	2.32
IIFDTI	0.917 (0.003)	0.793 (0.004)	0.817 (0.011)	0.883 (0.013)	0.745 (0.003)	18.9
FOTF-CPI	<b>0.929</b> (0.003)	<b>0.834</b> (0.002)	<b>0.822</b> (0.006)	<b>0.924</b> (0.012)	<b>0.789</b> (0.005)	0.98

The best results are shown in bold font.

sequence-based FOTF-CPI with other 3D structure-based methods on 3D structured datasets, confirming the value of FOTF-CPI in reducing computational resources and improving affinity prediction. Last, we used molecular docking techniques to visualize some of our experimental results to further improve the interpretability of the models.

### Fusion of optimal transport fragments compound-protein interaction prediction model

FOTF-CPI conducts CPI prediction by integrating compound and protein fragment information features. The FOTF-CPI pipeline contains three main components: I. feature extraction based on FSOT, II. feature fusion, and III. model prediction. First, we slice the fragments of compounds and proteins by FSOT and extracted their features using transformers. Then, the local affinity attention method fuses the features of molecules and fragments as a whole. Finally, the fused features are fed into the classifier to get the prediction results of CPI. An overview of our proposed framework is shown in Figure 1.

### Affinity experiments on multiple datasets

50-fold cross-validation experiments were conducted for affinity prediction on the three datasets, using an 8:2 train-test split. Tables 2, 3, and 4 present FOTF-CPI results on BindingDB, Davis and Biosnap compared to baseline methods.

Table 2 presents a comparison study of the BindingDB dataset based on AUC, PRC, sensitivity, specificity, and F1. Notably, the FOTF-CPI proposed in this research obtained the highest scores on each of the metrics. Table 3 provides a comparative study of the Davis dataset. Except for specificity, which ML-DTI performs better on, FOTF-CPI earns the highest score values on the other four metrics. Similarly, Table 4 provides the experimental results for the Biosnap dataset. The experimental findings demonstrate that FOTF-CPI, except sensitivity, which performs worse than ML-DTI, produces the best score outcomes on the other four metrics of this dataset.

Although FOTF-CPI has slightly lower specificity and sensitivity on the Davis and Biosnap datasets alone, it is still the best model among several models when both specificity and sensitivity metrics are considered together. The DeepDTA, TransformerCPI, Moltrans, ML-DTI, and IIFDTI models concentrate on interactions between proteins and compounds as entire sequences, while the interactions between protein and compound fragments are not examined as essential components. The FOTF-CPI model prioritizes the collection of sequence characteristics of molecular fragments and then mines them for information links. As a result, more feature information is acquired and the performance of CPI is enhanced.

### AUC and PRC of BindingDB

FOTF-CPI and baseline model robustness were evaluated on BindingDB. Each model has an arbitrary prediction bias. Figure 2 plots ROC and PR curves using true CPI labels and model predictions. Baseline AUC and PRC curves were more erratic compared to FOTF-CPI. Thus, FOTF-CPI demonstrated superior robustness and prediction performance versus baselines. FOTF-CPI's modeling of fragment interactions likely improves the prediction of a given protein's interaction with a drug.

**Table 3. The results of the Davis dataset**

Model	AUC (SD)	PRC (SD)	Sensitivity (SD)	Specificity (SD)	F1 (SD)	Cost (h)
DeepDTA	0.880 (0.007)	0.602 (0.024)	0.785 (0.031)	0.825 (0.025)	0.633 (0.018)	2.07
TransformerCPI	0.875 (0.006)	0.632 (0.010)	0.661 (0.013)	0.922 (0.015)	0.606 (0.006)	1.44
Moltrans	0.897 (0.002)	0.708 (0.013)	0.818 (0.016)	0.842 (0.022)	0.665 (0.008)	0.39
ML-DTI	0.828 (0.005)	0.563 (0.007)	0.692 (0.042)	<b>0.965</b> (0.030)	0.622 (0.012)	0.58
IIFDTI	0.899 (0.005)	0.582 (0.006)	0.779 (0.021)	0.884 (0.022)	0.616 (0.003)	3.67
FOTF-CPI	<b>0.902</b> (0.003)	<b>0.718</b> (0.016)	<b>0.844</b> (0.019)	0.843 (0.011)	<b>0.687</b> (0.009)	0.34

The best results are shown in bold font.

**Table 4. The results of the Biosnap dataset**

Model	AUC (SD)	PRC (SD)	Sensitivity (SD)	Specificity (SD)	F1 (SD)	Cost (h)
DeepDTA	0.876 (0.002)	0.880 (0.006)	0.781 (0.015)	0.824 (0.012)	0.804 (0.011)	11.06
TransformerCPI	0.878 (0.003)	0.869 (0.004)	0.821 (0.011)	0.804 (0.012)	0.814 (0.006)	3.93
Moltrans	0.893 (0.004)	0.891 (0.002)	0.786 (0.024)	0.848 (0.009)	0.815 (0.004)	1.08
ML-DTI	0.900 (0.003)	0.897 (0.004)	<b>0.836</b> (0.032)	0.819 (0.014)	0.817 (0.004)	2.22
IIFDTI	0.894 (0.003)	0.891 (0.003)	0.734 (0.016)	0.856 (0.013)	0.810 (0.005)	17.6
FOTF-CPI	<b>0.907</b> (0.002)	<b>0.895</b> (0.004)	0.786 (0.012)	<b>0.869</b> (0.014)	<b>0.821</b> (0.006)	0.93

The best results are shown in bold font.

### Effect of different slicing methods

The BindingDB dataset was used to evaluate how different string slicing methods affect outcomes. For comparison, character-based slicing (CS), BRICS, BPE, and FCS were tested as baseline methods.<sup>28,35–37</sup> Table 5 shows slicing method impact on CPI results. The FSOT method proposed by FOTF-CPI achieved the highest AUC, PRC, specificity, and F1 scores compared to other methods. Although BPE had higher sensitivity than FOTF-CPI, it significantly reduced specificity. Aside from BPE, FOTF-CPI had the best sensitivity, demonstrating it provides superior model robustness and prediction. FOTF-CPI's rational fragmentation enables fully utilizing molecular fragment interactions to map to compound-protein affinity.

### Ablation study

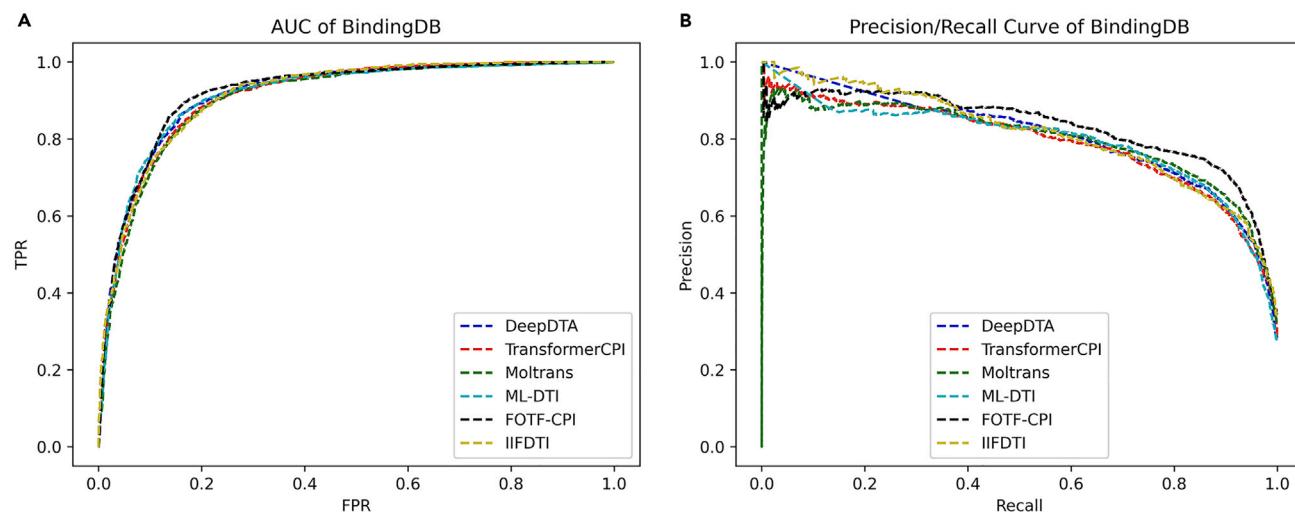
Ablation studies were conducted with the following settings.

- (1) W/O Fusion: The fusion module is removed, and protein and compound features are fed directly into the fully connected layer.
- (2) W/O FSOT: FSOT is removed and BPE is used directly for fragmentation.
- (3) W/O Both: The fusion module and FSOT are removed, using a basic transformer network.

Table 6 shows W/O Fragments achieved the highest Sensitivity but lower Specificity. Thus, appropriate feature fusion and fragmentation both contribute to overall model performance. Specifically, the attention-guided fusion network effectively analyzes subunit affinity relationships from BPE fragmentation.

### Cross-domain experiments

In addition, cross-domain experiments assessed model adaptability on unseen data. The training set was BindingDB, while Biosnap and Davis were test sets. All datasets contain small molecule drug-protein affinity results. As Table 7 shows, FOTF-CPI significantly outperformed baselines, achieving higher AUC, sensitivity, specificity, and F1 than the state-of-the-art model. Although IIFDTI had slightly higher PRC, FOTF-CPI overall provided more accurate predictions.



**Figure 2. AUC and PRC curves for the BindingDB dataset**

(A) AUC curves for the BindingDB dataset.

(B) PRC curves for the BindingDB dataset.

**Table 5. Effect of different slicing methods**

Model	AUC (SD)	PRC (SD)	Sensitivity (SD)	Specificity (SD)	F1 (SD)
CS	0.864 (0.017)	0.773 (0.032)	0.781 (0.022)	0.854 (0.019)	0.736 (0.011)
BRICS	0.882 (0.008)	0.792 (0.012)	0.783 (0.016)	0.842 (0.013)	0.747 (0.006)
BPE	0.899 (0.005)	0.786 (0.004)	<b>0.832</b> (0.013)	0.806 (0.009)	0.755 (0.004)
FCS	0.902 (0.004)	0.801 (0.007)	0.759 (0.005)	0.878 (0.007)	0.769 (0.003)
FSOT	<b>0.929</b> (0.003)	<b>0.834</b> (0.002)	0.822 (0.006)	<b>0.924</b> (0.012)	<b>0.789</b> (0.005)

The best results are shown in bold font.

### 3D dataset experiments

To demonstrate FOTF-CPI's performance in CPI experiments using only sequence information, we compare it to commonly used methods on the 3D DUD-E dataset. The methods we compare include the traditional machine learning methods NNScore and RFScore, the virtual docking method Autodock Vina, methods using 3D information such as 3D-CNN, AtomNet, PocketGCN, and Deffini, and methods using 2D graph information such as DrugVQA and AttentionSiteDTI.<sup>18,22,38–43</sup> In this experiment, FOTF-CPI uses the same training and test sets as the other methods, and its AUC and RE are calculated at 0.5%, 1.0%, 2.0%, and 5.0%. The specific experimental results are shown in the following table.

As shown in Table 8, FOTF-CPI achieves a level of accuracy useful for drug discovery that is similar to methods using 3D dataset information. In Table 8, the best results are shown in bold font, and "/" indicates that this result was missing from the original literature report. Note that the results are calculated on a per-target basis, and then we report the average results across 102 targets. On the full DUD-E dataset, although FOTF-CPI falls short of the highest AUC scores, our RE scores demonstrate high statistical performance at different thresholds. This shows that FOTF-CPI can achieve relatively similar results to methods using 3D information, even though it only analyzes relatively simple sequence data. FOTF-CPI provides new neural network insights for performing affinity analysis with limited computational resources. Although the 3D-CNN is trained on pure 3D structures, it has the lowest performance among several deep learning methods. This may be due to sparse data in 3D space, with pure CNN neural networks not adept at mining affinity information from sparse data. In contrast, FOTF-CPI based on sequence slicing provides good feature analysis. We also show the performance of cross-validation and how the FOTF-CPI model differs in its performance on the DUD-E benchmark compared to the Vina scoring function and 3D-CNN.

### Case studies and visualization

Previous experimental results have confirmed FOTF-CPI's outstanding affinity prediction performance. To further demonstrate FOTF-CPI's practical use, we conduct experiments on the Biosnap dataset. As shown in Table 9, FOTF-CPI predicts the affinity for compounds and proteins randomly chosen from the test set. The first eight predicted CPI combinations have either been supported by prior research or are included in DrugBank (the Biosnap gold standard). The last two predictions differ from Biosnap's gold standard, implying FOTF-CPI identifies possible interactions between compound-protein pairs not yet discovered.

In this chapter, Autodock Vina and PyMOL are used respectively for molecular docking and visualization analysis. In Figure 3A, compounds GLU-87, PRO-89, ARG-112, and ARG-114 form hydrogen bonds with compound C-135398510. In Figure 3B, ALA-289, GLU-293, and VAL-312 form hydrogen bonds with compound C-156422. In both cases, Autodock Vina successfully docks the relevant compounds to the proteins and identifies the hydrogen bonds formed, demonstrating FOTF-CPI's capacity to find potential affinity relations.

For the molecular docking examples, we further conduct visualization experiments of fragment affinity, as seen in Figure 4, to confirm the impact of fragments on overall affinity. For visualization, we only intercept part of the amino acid residue chain containing hydrogen bonds, as the full protein sequence was too long. These findings verify whether fragment interactions influence affinity prediction. Figure 4A shows the affinity of fragment-VR (ARG-112) with fragment- = O is greater than surrounding fragments. Similarly, fragment-NRG (ARG-114) affinity with fragment- = O is higher than surrounding fragments. As shown in Figure 4B, the affinity of fragment-ADS (ALA-289) with fragment-NC(=O)N, and fragment-YE (GLU-293) with fragment-c1ccc (OCCN), is higher than surrounding fragments. However, the affinity between fragment-VC (VAL-312) and fragment-NC(=O)N was indistinguishable from other nearby affinities, possibly because

**Table 6. Experimental results of ablation study**

Model	AUC (SD)	PRC (SD)	Sensitivity (SD)	Specificity (SD)	F1 (SD)
W/O Two	0.903 (0.004)	0.806 (0.004)	0.762 (0.003)	0.908 (0.004)	0.752 (0.002)
W/O FSOT	0.924 (0.003)	0.835 (0.006)	<b>0.831</b> (0.003)	0.898 (0.005)	0.780 (0.004)
W/O Fusion	0.908 (0.001)	0.810 (0.007)	0.811 (0.004)	0.879 (0.006)	0.771 (0.005)
FOTF-CPI	<b>0.929</b> (0.003)	<b>0.834</b> (0.002)	0.822 (0.006)	<b>0.924</b> (0.012)	<b>0.789</b> (0.005)

The best results are shown in bold font.

**Table 7. Experimental results of independent scene experiment**

Model	AUC	PRC	Sensitivity	Specificity	F1
DeepDTA	0.854	0.724	0.933	0.775	0.807
TransformerCPI	0.917	0.852	0.833	0.831	0.782
Moltrans	0.883	0.735	0.900	0.866	0.832
ML-DTI	0.880	0.760	0.936	0.823	0.839
IIFDTI	0.929	<b>0.874</b>	0.842	0.852	0.801
FOTF-CPI	<b>0.936</b>	0.867	<b>0.965</b>	<b>0.907</b>	<b>0.915</b>

The best results are shown in bold font.

this affinity link has no clear bearing on the final affinity outcome. Thus, FOTF-CPI can analyze fragment affinity and uncover potential CPAs.

## DISCUSSION

In this article, a transformer based on optimal transmission fragment fusion (FOTF-CPI) is proposed and used for CPI. As CPA is influenced by the affinity between its substructures, hidden substructures, i.e., fragments, can be effectively sliced out using the molecular structure and a rational feature cutting method. The transformer's encoder is employed as the feature extraction module of fragments. It is a neural network architecture that performs well in NLP tasks. Applying transformer to feature the extraction of CPI can capture the affinity between fragments at a greater distance than the simple one-hot, word2vec method.

The current mainstream CPI methods simply splice or multiply compound and protein features. These methods ignore the interaction between compound fragments and protein fragments, reducing the overall affinity detection performance. FOTF-CPI, which can capture the affinity of different compound fragments and different protein fragments, is a feature fusion method based on the influence of both substructures on each other. The affinity prediction of compounds and proteins can be enhanced by a reasonable fragment-slicing method with fragment-based feature information fusion. The feature fusion approach, in particular, significantly influences the model's interpretability. Fragment's slicing method also significantly influences the model's interpretability.

Prior research just simply combined compound and protein features such as concatenation or multiplication. The FOTF-CPI is distinct from past research in three main aspects. First, we suggest a CPI method FOTF-CPI that can accelerate the drug discovery process without requiring 3D structural information. Secondly, to enhance the quality of the input data, we also provide a dependable fragment-slicing approach. Finally, an improved feature fusion method is applied to completely account for the influence of the interaction forces between fragments.

## Limitations of the study

- (1) Due to computational cost constraints, we did not consider using the protein 3D structures generated by AlphaFlod2. In future work, we will try to further develop models for affinity prediction between compounds and proteins based on AlphaFlod2.
- (2) In this article we did not measure the actual experimental affinities to further validate the accuracy of the predicted affinities, FOTF-CPI may lack validation in practical affinity application scenarios.

**Table 8. 3D dataset experiments**

Model	AUC	0.5% RE	1.0% RE	2.0% RE	5.0% RE
NNscore	0.584	4.166	2.980	2.460	1.891
RFscore	0.622	5.628	4.274	3.499	2.678
Autodock Vina	0.716	9.139	7.321	5.811	4.444
3D-CNN	0.868	42.559	26.655	19.363	10.710
AtomNet	0.895	/	/	/	/
PocketGCN	0.886	44.406	29.748	19.408	10.735
Deffini	0.921	/	21.597	/	11.861
DrugVQA	<b>0.972</b>	88.17	58.71	35.06	<b>17.39</b>
AttentionSiteDTI	0.971	101.74	59.92	<b>35.07</b>	16.74
FOTF-CPI	0.942	<b>102.911</b>	<b>61.945</b>	34.178	17.368

The best results are shown in bold font.

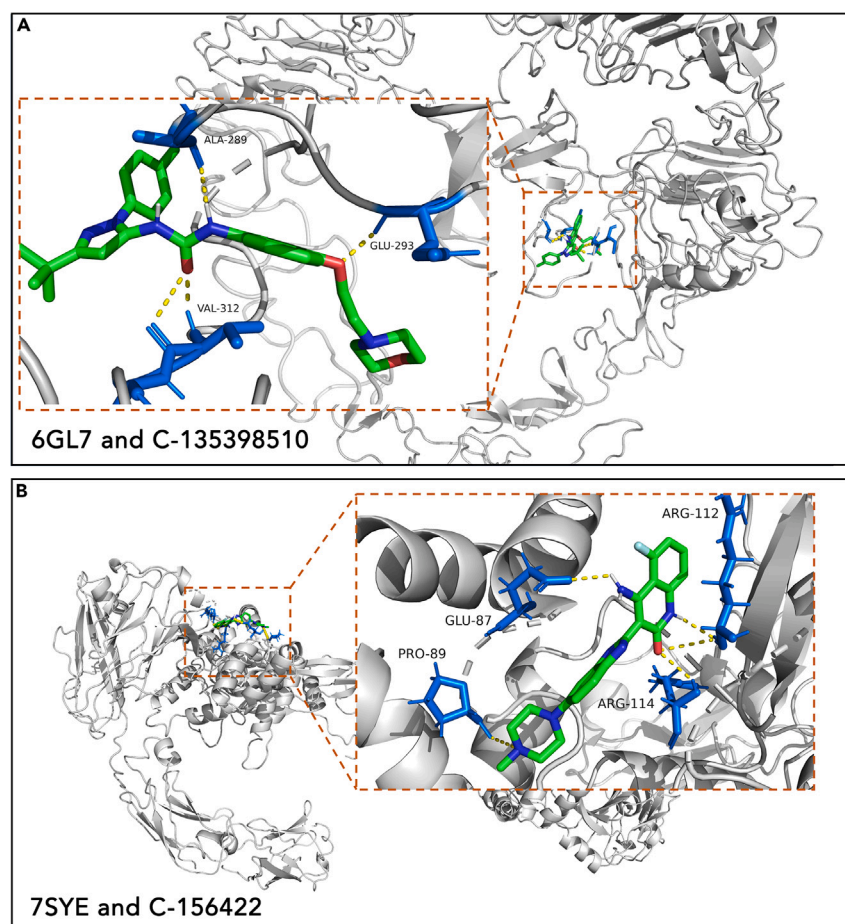
**Table 9. Case studies on Biosnap**

Protein	Compound (CID)	Label	Prediction	Evidence
7E2Y	C-44447073	1	1	Biosnap record
2R4R	C-54756928	1	1	Biosnap record
7SY1	C-5287969	1	1	Biosnap record
5F1A	C-126842800	1	1	Biosnap record
5HLN	C-9926791	1	1	Biosnap record
2EC8	C-11213558	1	1	Biosnap record
4LP5	C-694593	1	1	Biosnap record
4ZK9	C-10008367	1	1	Biosnap record
6GL7	C-135398510	0	1	Autodock vina
7SYE	C-156422	0	1	Autodock vina

**STAR★METHODS**

Detailed methods are provided in the online version of this paper and include the following:

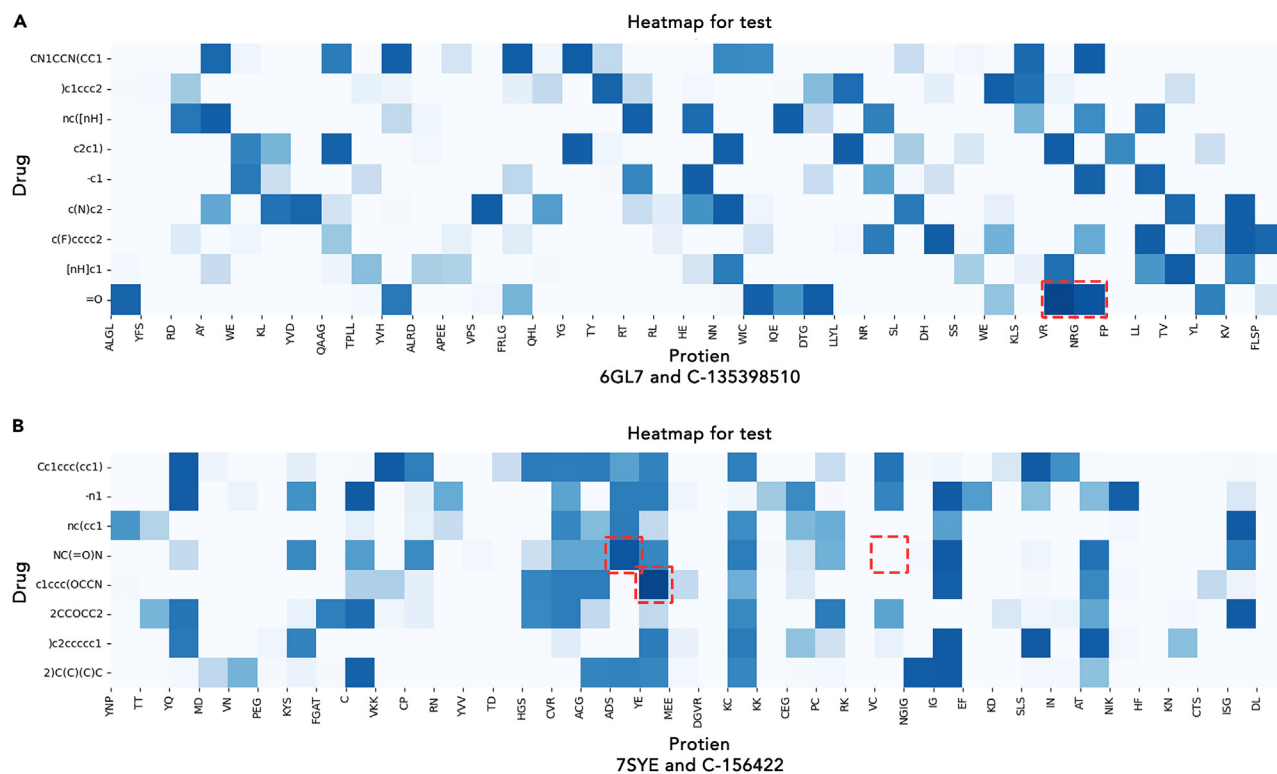
- [KEY RESOURCES TABLE](#)

**Figure 3. Molecular docking the visualization of Autodock vina results**

(A) Docking result of protein 6GL7 and compound C-135398510.

(B) Docking result of protein 7SYE and compound C-156422.





**Figure 4. Heat map of affinity between compound fragment and partial protein fragment**

(A) Heatmap of affinity between C-135398510 compound fragment and partial 6GL7 protein fragment.

(B) Heatmap of affinity between C-156422 compound fragment and partial 7YSE protein fragment.

- **RESOURCE AVAILABILITY**

- Materials availability
- Lead contact
- Data and code availability

- **METHOD DETAILS**

- FSOT
- Feature extraction
- Feature fusion of proteins and compounds
- Loss function
- AutoDock vina
- Datasets
- BindingDB database
- Davis database
- Biosnap database
- DUD-E database
- Compound-protein interaction prediction
- Baseline models
- Evaluation metrics

- **QUANTIFICATION AND STATISTICAL ANALYSIS**

## ACKNOWLEDGMENTS

The authors acknowledge Dr. Qineng Gong of Fudan University for his support on molecular docking technology. This work was supported by National Natural Science Foundation of China (No.81873915) and Foreign Youth Talent Program of the Ministry of Science and Technology, China (No. QN2022014011L).

## AUTHOR CONTRIBUTIONS

Zeyu Yin: conceptualization, methodology, software, validation, and writing - original draft. Yu Chen: survey, resources, data organization, writing, and re-sourcing - review and editing. Yajie Hao: survey, resource, and writing - review and editing. Jingsong Shao: survey, writing, and drawing - comments and editing. Sanjeevi Pandiyan: writing - commenting and editing. Li Wang: conceptualization, methodology, formal analysis, survey, resources, writing - review and editing, visualization, and project management.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 25, 2023

Revised: November 5, 2023

Accepted: December 13, 2023

Published: December 15, 2023

## REFERENCES

- Paul, S.M., Mytelka, D.S., Dunwiddie, C.T., Persinger, C.C., Munos, B.H., Lindborg, S.R., and Schacht, A.L. (2010). How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat. Rev. Drug Discov.* 9, 203–214.
- Martinez-Mayorga, K., Madariaga-Mazon, A., Medina-Franco, J., and Maggiora, G. (2020). The impact of chemoinformatics on drug discovery in the pharmaceutical industry. *Expet Opin. Drug Discov.* 15, 293–306.
- Bleakley, K., and Yamanishi, Y. (2009). Supervised prediction of drug–target interactions using bipartite local models. *Bioinformatics* 25, 2397–2403.
- Ma, W., Yang, L., and He, L. (2018). Overview of the detection methods for equilibrium dissociation constant KD of drug-receptor interaction. *J. Pharm. Anal.* 8, 147–152.
- Wang, P., Zhuo, X., Chu, W., and Tang, X. (2017). Exenatide-loaded microsphere/thermosensitive hydrogel long-acting delivery system with high drug bioactivity. *Int. J. Pharm.* 528, 62–75.
- Zhou, H., Gao, M., and Skolnick, J. (2015). Comprehensive prediction of drug-protein interactions and side effects for the human proteome. *Sci. Rep.* 5, 11090–11113.
- Liu, T., Lin, Y., Wen, X., Jorissen, R.N., and Gilson, M.K. (2007). BindingDB: a web-accessible database of experimentally determined protein–ligand binding affinities. *Nucleic Acids Res.* 35, 198–201.
- Lavecchia, A., and Di Giovanni, C. (2013). Virtual screening strategies in drug discovery: a critical review. *Curr. Med. Chem.* 20, 2839–2860.
- Atiya, A., Alhumaydhi, F.A., Sharaf, S.E., Al Abdulmonem, W., Elsbali, A.M., Al Enazi, M.M., Shamsi, A., Jawaid, T., Alghamdi, B.S., Hashem, A.M., et al. (2022). Identification of 11-Hydroxytyphrosin and Torosflavone A as Potential Inhibitors of 3-Phosphoinositide-Dependent Protein Kinase 1 (PDK1): Toward Anticancer Drug Discovery. *Biology* 11, 1230.
- Gupta, R., Srivastava, D., Sahu, M., Tiwari, S., Ambasta, R.K., and Kumar, P. (2021). Artificial intelligence to deep learning: machine intelligence approach for drug discovery. *Mol. Divers.* 25, 1315–1360.
- Dhakal, A., McKay, C., Tanner, J.J., and Cheng, J. (2022). Artificial intelligence in the prediction of protein–ligand interactions: recent advances and future directions. *Briefings Bioinf.* 23, 476.
- Lo, Y.C., Rensi, S.E., Torng, W., and Altman, R.B. (2018). Machine learning in chemoinformatics and drug discovery. *Drug Discov. Today* 23, 1538–1546.
- Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., Li, B., Madabhushi, A., Shah, P., Spitzer, M., and Zhao, S. (2019). Applications of machine learning in drug discovery and development. *Nat. Rev. Drug Discov.* 18, 463–477.
- Zhang, L., Tan, J., Han, D., and Zhu, H. (2017). From machine learning to deep learning: progress in machine intelligence for rational drug discovery. *Drug Discov. Today* 22, 1680–1685.
- Costello, J.C., Heiser, L.M., Georgii, E., Gönen, M., Menden, M.P., Wang, N.J., Bansal, M., Ammad-ud-din, M., Hintsanen, P., Khan, S.A., et al. (2014). A community effort to assess and improve drug sensitivity prediction algorithms. *Nat. Biotechnol.* 32, 1202–1212.
- Lapins, M., Avidsson, S., Lampa, S., Berg, A., Schaal, W., Alvarsson, J., and Spjuth, O. (2018). A confidence predictor for logD using conformal regression and a support-vector machine. *J. Cheminf.* 10, 17.
- Shi, H., Liu, S., Chen, J., Li, X., Ma, Q., and Yu, B. (2019). Predicting drug-target interactions using Lasso with random forest based on evolutionary information and chemical structure. *Genomics* 111, 1839–1852.
- Wallach, I., Dzamba, M., and Heifets, A. (2015). AtomNet: a deep convolutional neural network for bioactivity prediction in structure-aware drug discovery. Preprint at arXiv.
- Gomes, J., Ramsundar, B., Feinberg, E.N., and Pande, V.S. (2017). Atomic convolutional networks for predicting protein–ligand binding affinity. Preprint at arXiv.
- Jiménez, J., Skalić, M., Martínez-Rosell, G., and De Fabritiis, G. (2018). Kdeep: protein–ligand absolute binding affinity prediction via 3d-convolutional neural networks. *J. Chem. Inf. Model.* 58, 287–296.
- Li, F., Zhang, Z., Guan, J., and Zhou, S. (2022). Effective drug–target interaction prediction with mutual interaction neural network. *Bioinformatics* 38, 3582–3589.
- Zheng, S., Li, Y., Chen, S., Xu, J., and Yang, Y. (2020). Predicting drug–protein interaction using quasi-visual question answering system. *Nat. Mach. Intell.* 2, 134–140.
- Li, S., Wan, F., Shu, H., Jiang, T., Zhao, D., and Zeng, J. (2020). MONN: a multi-objective neural network for predicting compound–protein interactions and affinities. *Cell Systems* 10, 308–322.e11.
- Lu, W., Wu, Q., Zhang, J., Rao, J., Li, C., and Zheng, S. (2022). Tankbind: Trigonometry-aware neural networks for drug-protein binding structure prediction. *Adv. Neural Inf. Process. Syst.* 35, 7236–7249.
- Li, M., Lu, Z., Wu, Y., and Li, Y. (2022). BACPI: a bi-directional attention neural network for compound–protein interaction and binding affinity prediction. *Bioinformatics* 38, 1995–2002.
- Watanabe, N., Ohnuki, Y., and Sakakibara, Y. (2021). Deep learning integration of molecular and interactome data for protein–compound interaction prediction. *J. Cheminf.* 13, 36.
- Karimi, M., Wu, D., Wang, Z., and Shen, Y. (2019). DeepAffinity: interpretable deep learning of compound–protein affinity through unified recurrent and convolutional neural networks. *Bioinformatics* 35, 3329–3338.
- Huang, K., Xiao, C., Glass, L.M., and Sun, J. (2021). MolTrans: Molecular Interaction Transformer for drug–target interaction prediction. *Bioinformatics* 37, 830–836.
- Öztürk, H., Özgür, A., and Ozkirimli, E. (2018). DeepDTA: deep drug–target binding affinity prediction. *Bioinformatics* 34, 821–829.
- Magnan, C.N., and Baldi, P. (2014). SSpro/ACCpro 5: almost perfect prediction of protein secondary structure and relative solvent accessibility using profiles, machine learning and structural similarity. *Bioinformatics* 30, 2592–2597.
- Cheng, Z., Zhao, Q., Li, Y., and Wang, J. (2022). IIFDTI: predicting drug–target interactions through interactive and independent features based on attention mechanism. *Bioinformatics* 38, 4153–4161.
- Song, T., Zhang, X., Ding, M., Rodriguez-Paton, A., Wang, S., and Wang, G. (2022). DeepFusion: A deep learning based multi-scale feature fusion method for predicting drug–target interactions. *Methods* 204, 269–277.
- Agyemang, B., Wu, W.P., Kpiebaareh, M.Y., Lei, Z., Nanor, E., and Chen, L. (2020). Multi-view self-attention for interpretable

- drug–target interaction prediction. *J. Biomed. Inf.* 110, 103547.
34. Chen, L., Tan, X., Wang, D., Zhong, F., Liu, X., Yang, T., Luo, X., Chen, K., Jiang, H., and Zheng, M. (2020). TransformerCPI: improving compound–protein interaction prediction by sequence-based deep learning with self-attention mechanism and label reversal experiments. *Bioinformatics* 36, 4406–4414.
  35. Sennrich, R., Haddow, B., and Birch, A. (2015). Neural machine translation of rare words with subword units. Preprint at arXiv.
  36. Xu, J., Zhou, H., Gan, C., Zheng, Z., and Li, L. (2020). Vocabulary learning via optimal transport for neural machine translation. Preprint at arXiv.
  37. Degen, J., Wegscheid-Gerlach, C., Zaliani, A., and Rarey, M. (2008). On the Art of Compiling and Using ‘Drug-Like’ Chemical Fragment Spaces. *ChemMedChem* 3, 1503–1507.
  38. Durrant, J.D., and McCammon, J.A. (2011). NNScore 2.0: a neural-network receptor–ligand scoring function. *J. Chem. Inf. Model.* 51, 2897–2903.
  39. Ballester, P.J., and Mitchell, J.B.O. (2010). A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking. *Bioinformatics* 26, 1169–1175.
  40. Ragoza, M., Hochuli, J., Idrobo, E., Sunseri, J., and Koes, D.R. (2017). Protein–ligand scoring with convolutional neural networks. *J. Chem. Inf. Model.* 57, 942–957.
  41. Torng, W., and Altman, R.B. (2019). Graph convolutional neural networks for predicting drug–target interactions. *J. Chem. Inf. Model.* 59, 4131–4149.
  42. Zhou, D., Liu, F., Zheng, Y., Hu, L., Huang, T., and Huang, Y.S. (2022). Deffini: A family-specific deep neural network model for structure-aware virtual screening. *Comput. Biol. Med.* 151, 106323.
  43. Yazdani-Jahromi, M., Yousefi, N., Tayebi, A., Kolanthai, E., Neal, C.J., Seal, S., and Garibay, O.O. (2022). AttentionSiteDTI: an interpretable graph-based model for drug–target interaction prediction using NLP sentence-level relation classification. *Briefings Bioinf.* 23, bbac272.
  44. Yang, Z., Zhong, W., Zhao, L., and Chen, C.Y. (2021). ML-DTI: Mutual Learning Mechanism for Interpretable Drug–Target Interaction Prediction. *J. Phys. Chem. Lett.* 13, 4247–4261.
  45. Trott, O., and Olson, A.J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31, 455–461.

## STAR★METHODS

## KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<i>Deposited data</i>		
BindingDB	Binding Database	<a href="http://www.bindingdb.org">www.bindingdb.org</a>
Davis	Huang et al. <sup>28</sup>	<a href="http://staff.cs.utu.fi/~aatapa/data/DrugTarget/">http://staff.cs.utu.fi/~aatapa/data/DrugTarget/</a>
Biosnap	Song et al. <sup>32</sup>	<a href="http://snap.stanford.edu/biodata/datasets/10002/10002-ChG-Miner.html">http://snap.stanford.edu/biodata/datasets/10002/10002-ChG-Miner.html</a>
DUD-E	A Database of Useful (Docking) Decoys — Enhanced	<a href="https://dude.docking.org/">https://dude.docking.org/</a>
PROTAC-DB	An online database of PROTACs	<a href="http://cadd.zju.edu.cn/protacdb/about">http://cadd.zju.edu.cn/protacdb/about</a>
<i>Software and algorithms</i>		
FOTF-CPI	This paper	<a href="https://github.com/NTU-MedAI/FOTF-CPI">https://github.com/NTU-MedAI/FOTF-CPI</a>
AlphaFold2	DeepMind	<a href="https://alphafold.com/">https://alphafold.com/</a>
linear and random forest (RF)	Scikit-learn	<a href="https://scikit-learn.org/stable/">https://scikit-learn.org/stable/</a>
DeepDTA	Öztürk et al. <sup>29</sup>	<a href="https://github.com/hkmztrk/DeepDTA">https://github.com/hkmztrk/DeepDTA</a>
TransformerCPI	Chen et al. <sup>34</sup>	<a href="https://github.com/lifanchen-simm/transformerCPI">https://github.com/lifanchen-simm/transformerCPI</a>
Moltrans	Huang et al. <sup>28</sup>	<a href="https://github.com/kexinhuang12345/moltrans">https://github.com/kexinhuang12345/moltrans</a>
ML-DTI	Yang et al. <sup>44</sup>	<a href="https://github.com/guaguabujianle/ML-DTI.git">https://github.com/guaguabujianle/ML-DTI.git</a>
IIFDTI	Cheng et al. <sup>31</sup>	<a href="https://github.com/czjczj/IIFDTI">https://github.com/czjczj/IIFDTI</a>
character-based slicing (CS)	Xu et al. <sup>36</sup>	<a href="https://github.com/Jingjing-NLP/VOLT">https://github.com/Jingjing-NLP/VOLT</a>
BRICS	Degen et al. <sup>37</sup>	<a href="https://www.zbh.uni-hamburg.de/en/forschung/amd/datasets/brics.html">https://www.zbh.uni-hamburg.de/en/forschung/amd/datasets/brics.html</a>
BPE	Sennrich et al. <sup>35</sup>	<a href="https://github.com/rsennrich/subword-nmt">https://github.com/rsennrich/subword-nmt</a>
FCS	Huang et al. <sup>28</sup>	<a href="https://github.com/kexinhuang12345/MolTrans">https://github.com/kexinhuang12345/MolTrans</a>
NNscore	Durrant et al. <sup>38</sup>	<a href="https://github.com/durrantlab/nnscore2">https://github.com/durrantlab/nnscore2</a>
RFscore	Ballester et al. <sup>39</sup>	<a href="https://doi.org/10.1093/bioinformatics/btq112">https://doi.org/10.1093/bioinformatics/btq112</a>
Autodock Vina	Trott et al. <sup>45</sup>	<a href="https://github.com/ccsb-scripps/AutoDock-Vina">https://github.com/ccsb-scripps/AutoDock-Vina</a>
3D-CNN	Ragoza et al. <sup>40</sup>	<a href="https://github.com/gnina/gnina">https://github.com/gnina/gnina</a>
AtomNet	Wallach et al. <sup>18</sup>	<a href="https://doi.org/10.48550/arXiv.1510.02855">https://doi.org/10.48550/arXiv.1510.02855</a>
PocketGCN	Tornig et al. <sup>41</sup>	<a href="https://pubs.acs.org/doi/abs/10.1021/acs.jcim.9b00628">https://pubs.acs.org/doi/abs/10.1021/acs.jcim.9b00628</a>
Deffini	Zhou et al. <sup>42</sup>	<a href="https://github.com/joewood/Deffini">https://github.com/joewood/Deffini</a>
DrugVQA	Zheng et al. <sup>22</sup>	<a href="https://github.com/prokia/drugVQA">https://github.com/prokia/drugVQA</a>
AttentionSiteDTI	Yazdani-Jahromi et al. <sup>43</sup>	<a href="https://github.com/yazdanimehdi/AttentionSiteDTI">https://github.com/yazdanimehdi/AttentionSiteDTI</a>
PyMOL	PyMOL by Schrödinger	<a href="https://pymol.org/">https://pymol.org/</a>
FOTF-CPI	This paper	<a href="https://github.com/NTU-MedAI/FOTF-CPI">https://github.com/NTU-MedAI/FOTF-CPI</a>

## RESOURCE AVAILABILITY

## Materials availability

This study did not generate any new unique reagents.

## Lead contact

Further information should be directed to Li Wang ([wangli@ntu.edu.cn](mailto:wangli@ntu.edu.cn)).

## Data and code availability

- This study analyzes existing, publicly available data. The sources for the datasets are listed in the [key resources table](#).

- All original code has been deposited at Github and is publicly available as of the date of publication. DOI is listed in the [key resources table](#).
- any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

## METHOD DETAILS

### FSOT

FSOT (Fragment Slicing based on Optimal Transport) is a fragments slicing method for proteins and compounds based on optimal transport algorithm. Similar to BPE merging rules, the original compound and protein sequences are split into separate markers. For simplicity, BPE generates the fragment candidate lists, using ZINC Clean Lead datasets for compound token generation and PDBbind datasets for protein token generation. All tokens are initialized with probabilities for the optimal transport algorithm. At each time step, the vocabulary with maximum entropy is obtained based on the transport matrix. Due to relaxed restrictions, cases of illegal marginal utility occur, so tokens with frequencies below 0.001 are removed.

Formally, Marginal Utility of Vocabulary (MUV) represents the negative derivation of entropy to size. For simplification, we leverage a smaller vocabulary to estimate MUV in implementation. Specially, MUV is calculated as

$$M_{v(k+m)} = \frac{-(H_{v(k+m)} - H_{v(k)})}{m} \quad (\text{Equation 1})$$

where  $v(k)$ ,  $v(k+m)$  are two vocabularies with  $k$  and  $k+m$  tokens, respectively.  $H_v$  represents the corpus entropy with the vocabulary  $v$ , which is defined by the sum of token entropy. To avoid the effects of token length, here we normalize entropy with the average length of tokens and the final entropy is defined as:

$$H_v = -\frac{1}{l_v} \sum_{i \in v} P(i) \log P(i) \quad (\text{Equation 2})$$

where  $P(i)$  is the relative frequency of token  $i$  from the training corpus and  $l_v$  is the average length of tokens in vocabulary  $v$

$S = \{i, 2 \cdot i, \dots, (t-1) \cdot i, \dots\}$  where each timestep  $t$  represents a set of vocabularies with the number up to  $S[t]$ . For any vocabulary, its MUV score can be calculated based on a vocabulary from its previous timestep. With sequence  $S$ , the target to find the optimal vocabulary  $v(t)$  with the highest MUV can be formulated as:

$$\operatorname{argmax}_{v(t-1) \in V_{S[t-1]}, v(t) \in V_{S[t]}} M_{v(t)} = \operatorname{argmax}_{v(t-1) \in V_{S[t-1]}, v(t) \in V_{S[t]}} -\frac{1}{i} [H_{v(t)} - H_{v(t-1)}] \quad (\text{Equation 3})$$

where  $V_{S[t-1]}$  and  $V_{S[t]}$  are two sets containing all vocabularies with upper bound of size  $S[t-1]$  and  $S[t]$ . Due to exponential search space, we propose to optimize its lower bound:

$$\operatorname{argmax}_t \frac{1}{t} \left[ \max_{v(t) \in V_{S[t]}} H_{v(t)} - \max_{v(t-1) \in V_{S[t-1]}} H_{v(t-1)} \right] \quad (\text{Equation 4})$$

where  $i$  means the size difference between  $t-1$  vocabulary and  $t$  vocabulary. MUV requires the size difference as a denominator. Based on this equation, the whole solution is split into two steps: 1) searching for the optimal vocabulary with the highest entropy at each time step  $t$ ; 2) enumerating all timesteps and outputting the vocabulary corresponding to the time step satisfying [Equation 4](#).

The first step of our approach is to search for the vocabulary with the highest entropy from  $V_{S[t]}$ . Formally, the goal is to find a vocabulary  $v(t)$  such that entropy is maximized,

$$\operatorname{argmax}_{v(t) \in V_{S[t]}} -\frac{1}{l_v} \sum_{i \in v} P(i) \log P(i) \quad (\text{Equation 5})$$

Given a set of vocabularies  $V_{S[t]}$ , we want to find the vocabulary with the highest entropy. Consequently, the objective function in [Equation 5](#) becomes:

$$\min_{v \in V_{S[t]}} \frac{1}{l_v} \sum_{i \in v} P(i) \log P(i), \text{ s.t. } P(i) = \frac{\text{Token}(i)}{\sum_{i \in v} \text{Token}(i)}, l_v = \frac{\sum_{i \in v} \text{len}(i)}{|v|} \quad (\text{Equation 6})$$

After the vocabulary is generated, VOLT uses the greedy strategy to encode the text like a BPE. To encode a compound, it first breaks the compound down into character-level markers. If the combined token is in the vocabulary, the two consecutively tokens are merged into one token. This process runs until there are no more tokens to merge. The out-of-vocabulary tags are split into smaller tags.

### Feature extraction

The sequences of proteins and compounds cannot be directly applied to neural networks. Therefore, the sequences need to be segmented, and then extract features. The residue sequences of proteins are cut into protein fragments  $s^p = \{s_1^p, s_2^p, \dots, s_n^p\}$  according to the FSOT. Resolution and cleavage of SMILES sequences of compounds into compound fragments  $s^d = \{s_1^d, s_2^d, \dots, s_n^d\}$ . The pre-processed compound sequence  $s^d$  and protein sequence  $s^p$  are randomly initialized with fragment as the base entity to obtain compound embeddings  $e^d = \{e_1^d, e_2^d, \dots, e_n^d\}$  and protein embeddings respectively  $e^p = \{e_1^p, e_2^p, \dots, e_n^p\}$ . Sequences  $s^d$  and  $s^p$  are both regarded as sentences in that each fragment is a word. After encoder, the embedding  $e^{ds} = \{e_1^{ds}, e_2^{ds}, \dots, e_n^{ds}\}$  and  $e^{ps} = \{e_1^{ps}, e_2^{ps}, \dots, e_n^{ps}\}$  are obtained separately from the compound embedding  $e^d$  and protein embedding  $e^p$ .

$$e^d = \text{Encoder}(e^d) \quad (\text{Equation 7})$$

$$e^p = \text{Encoder}(e^p) \quad (\text{Equation 8})$$

### Feature fusion of proteins and compounds

Feature fusion is based on the mutual influence between two substructures to capture the local affinity relationships between different fragments using the attention mechanism. When the affinity information of a certain fragment combination has a positive or negative impact on the overall affinity of the complete sequence, the feature fusion network will record this positive or negative interaction information and train the neural network weights accordingly. Ultimately, FOTF-CPI will learn the affinity information of the entire protein and compound.

To obtain the affinity relationship between fragments,  $e^d$  is multiplied with  $e^p$  to acquire the local affinity matrix  $W_{SA}$ . The normalization technique was used to generate the normalized  $W'_{SA}$  in order to avoid the high affinity score of  $W_{SA}$ .

$$W'_{SA} = \text{Norm}(\text{Softmax}(W_{SA})) \quad (\text{Equation 9})$$

$W'_{SA}$  is used to obtain the affinity relationship matrix  $W'_{DA}$  of each drug fragment with different protein fragments by *Softmax*. Meanwhile,  $W'_{SA}$  is transposed to obtain the affinity relationship  $W'_{PA}$  of each protein fragment with different drug fragments by *Softmax*.  $W'_{DA}$  is multiplied with  $e^d$  to obtain the compound embedding  $e^{ds'} = \{e_1^{ds'}, e_2^{ds'}, \dots, e_n^{ds'}\}$  under the local fragment correction. Similarly,  $W'_{PA}$  is multiplied by  $e^p$  to obtain the protein representation  $e^{ps'} = \{e_1^{ps'}, e_2^{ps'}, \dots, e_n^{ps'}\}$ .

$$W'_{DA} = \text{Softmax}(W'_{SA}) \quad (\text{Equation 10})$$

$$W'_{PA} = \text{Softmax}(W'_{SA}^T) \quad (\text{Equation 11})$$

$$e^{ds'} = W'_{DA} \times e^d \quad (\text{Equation 12})$$

$$e^{ps'} = W'_{PA} \times e^p \quad (\text{Equation 13})$$

The  $e^{ds'}$  is concatenated with  $e^d$  under the local fragment affinity correction in vector dimensions to obtain the hybrid compound representation  $h^d = \{h_1^d, h_2^d, \dots, h_n^d\}$ . Similarly, the  $e^{ps'}$  and  $e^p$  are concatenated to obtain a hybrid protein representation  $h^p = \{h_1^p, h_2^p, \dots, h_n^p\}$ . The hybrid representations  $h^d$  and  $h^p$  obtained by concatenating are pooled by global adaptive pooling. The compound representation  $h^d$  and the protein representation  $h^p$  are obtained after fusion of global features with local features. The fused  $h^d$  and  $h^p$  are concatenated in the fragment dimension to obtain the affinity representation  $h^a$ .

$$h_i^d = [e_i^{ds'} \parallel e_i^d] \quad (\text{Equation 14})$$

$$h_i^p = [e_i^{ps'} \parallel e_i^p] \quad (\text{Equation 15})$$

$$h^d = \text{AdatiPool}(h^d) \quad (\text{Equation 16})$$

$$h^p = \text{AdatiPool}(h^p) \quad (\text{Equation 17})$$

$$h^a = [h^d \parallel h^p] \quad (\text{Equation 18})$$

### Loss function

We used the binary cross-entropy loss function in the CPI task. The obtained affinity feature  $h^a$  is sequentially passed through the pooling layer with the activation function to obtain the prediction results  $y_s$  for a pair of compounds and proteins. Based on the predicted results

$y = \{y_1, y_2, \dots, y_s, \dots, y_N\}$  and the true labels  $l = \{l_1, l_2, \dots, l_s, \dots, l_N\}$ , the entire network is continuously optimized by combining the binary cross-entropy loss function.

$$y_s = \text{ReLu}\left(\text{AdatiPool}\left(h^d\right)\right) \quad (\text{Equation 19})$$

$$\text{loss} = -\frac{1}{N} \sum_{i=1}^N \left( l_i \times \log(y_i) + (1 - l_i) \times \log(1 - y_i) \right) \quad (\text{Equation 20})$$

### AutoDock vina

AutoDock Vina is a next-generation molecular docking software developed by the Scripps Research Institute, succeeding AutoDock.<sup>45</sup> Molecular docking plays a crucial role in computer-aided drug research, and AutoDock Vina is one of the most widely used programs.

To validate the feasibility of FOTF-CPI for affinity prediction, we employ 3D molecular structure information for molecular docking using AutoDock Vina. Moreover, PyMOL is employed to visualize the docking results and highlight the hydrogen bonds formed between amino acids and compounds. In this section, we provide a brief overview of our molecular docking and visualization process.

- (1) Receptor preparation: Retrieve the relevant protein 3D structure files from the PDB Bank and eliminate ligands, water molecules, and other unwanted components.
- (2) Ligand Preparation: Acquire the 2D structure files of the corresponding compounds from PubChem and utilize Open Babel software to convert them into the requisite 3D structure files of the ligand molecules.
- (3) Grid Box Setup: Employ AutoDockTools to manually define the docking search space for the protein and compound.
- (4) Autodock vina: Execute Autodock vina using the prepared Grid Box files, receptor 3D structure files, and ligand 3D structure files to obtain the docking result files.
- (5) Result visualization: Utilize PyMOL to visualize and analyze the docking result files.

### Datasets

Four benchmark datasets were used for affinity analysis experiments, [Table 1](#) shows dataset statistics.

#### BindingDB database

BindingDB is a widely used database for biomolecular interactions, providing extensive binding data between proteins and small molecule compounds. The database comprises various bioactive molecules, including drugs, compounds, and natural products. The interactions between these molecules and proteins are experimentally determined and recorded in the database. The binding affinity data in the dataset are represented using different measurement units such as Kd, Ki, IC50, among others. Since IC50 values can be influenced by experimental conditions and may contain significant noise, we utilize the negative logarithm of Kd values as the target for binding affinity prediction. We obtained and filtered the dataset to include the mentioned 10,665 drug Kd values and 1,413 protein Kd values.

#### Davis database

Davis dataset is a publicly available dataset widely used for drug discovery and chemical biology research, containing binding affinity data between drugs and proteins. The primary purpose of this dataset is to predict the binding affinity between compounds and proteins. It collects experimental data on drug-protein binding from sources such as BindingDB, ChEMBL, and other public resources. The dataset undergoes rigorous screening and quality control measures to ensure the reliability and consistency of the data. For this translation, we have selected 68 drug compounds and 379 protein targets from the Davis dataset, considering those drug-protein pairs with a binding affinity of less than 30 units as positive samples.

#### Biosnap database

Biosnap dataset is a publicly available dataset used for research in bioinformatics and chemical biology, primarily focusing on predicting the interactions between compounds and proteins. The dataset is constructed based on the DrugBank database, which contains a vast collection of compound-protein interaction pairs. Biosnap dataset consists of 13,741 CPI pairs, involving 4,510 distinct drugs and 2,181 protein targets. These CPI pairs represent known interactions between drugs and their corresponding protein targets. It is important to note that the Biosnap dataset only includes positive samples of CPI pairs. To maintain a balanced distribution of positive and negative samples, negative pairs are sampled from unseen CPI pairs.

#### DUD-E database

DUD-E<sup>30</sup> database is a highly regarded benchmark used for structure-based virtual screening. It serves as a valuable resource for evaluating the performance of various methods in this field. The 3D dataset DUD-E stands out by differentiating FOTF-CPI from other 3D methods. The DUD-E database consists of 22,886 active ligands and their corresponding affinities against 102 targets, which are further divided into seven subsets. On average, each target is associated with 224 active ligands, while each active ligand is accompanied by approximately 50 decoy ligands.

### Compound-protein interaction prediction

We utilized the FOTF-CPI to convert the survival prediction problem into a classification problem. The FOTF-CPI algorithm takes a table with four columns as input: compound-protein pair IDs, compound sequences, protein sequences, and affinity label. It outputs affinity result associated with each compound-protein pair ID.

### Baseline models

Advanced deep learning techniques are used as baselines to demonstrate the effectiveness of FOTF-CPI. The performance of these deep learning models is superior to that of shallow models.

**Moltrans<sup>28</sup>:** MolTrans is a Transformer-based architecture that focuses on molecular interactions. It utilizes a BPE-based tokenization method to first decompose the input drug and target protein sequences. The FCS module then performs substructure concatenation and fusion based on the corresponding vocabulary's word frequency. Improved Transformers are employed to obtain vector representations for both drug and target protein substructures. This architecture is used for predicting the interaction between drugs and target proteins.

**DeepDTA<sup>29</sup>:** DeepDTA model applies CNN to the original SMILES strings and protein sequences to extract local residue patterns, with the task of predicting binding affinity values. Finally, a Sigmoid activation function is added to convert it into a binary classification problem, and hyperparameter search is performed to ensure fairness.

**IIFDTI<sup>31</sup>:** IIFDTI model aims to capture the relationship between drugs and targets by introducing both interactive and independent features. It allows the model to adaptively weight the representations of drugs and targets based on the importance of different features. By combining interactive and independent features, the model can more accurately capture the interaction patterns between drugs and targets.

**TransformerCPI<sup>34</sup>:** TransformerCPI model utilizes the sequence representations of compounds and proteins as input to capture their relationship through self-attention mechanism and multi-head attention mechanism. Firstly, the compound and protein sequences are encoded into vector representations separately. Then, the model performs information interaction and extraction using multiple layers of self-attention and feedforward neural network layers. Finally, the model utilizes an output layer to make predictions on the interactions.

**ML-DTI<sup>44</sup>:** ML-DTI has developed a cross-dependent network architecture that enables the collaborative work of protein and drug encoders while utilizing CNN encoding, thus capturing the interaction between drug molecules and proteins during the encoding phase.

The performance of different models is evaluated using commonly used evaluation metrics such as AUC, PRC, sensitivity, specificity, and F1 score.

### Evaluation metrics

We used different metrics to evaluate the relevance of binding affinity prediction, including AUC (Area Under the ROC Curve), PRC (Area Under the Precision-Recall Curve), Sensitivity, and Specificity. These metrics are defined as follows:

**AUC (Area Under the ROC Curve):** AUC measures the model's ability to distinguish between positive and negative samples. It plots the true positive rate (sensitivity) against the false positive rate (1-specificity) at various threshold settings. A higher AUC indicates better discrimination performance.

**PRC (Area Under the Precision-Recall Curve):** PRC measures the trade-off between precision and recall in the classification model. It plots the precision (positive predictive value) against the recall (sensitivity) at different classification thresholds. A higher PRC value indicates better precision-recall balance.

**Sensitivity:** Sensitivity, also known as the true positive rate or recall, measures the proportion of true positive predictions out of all actual positive samples. It indicates the model's ability to correctly identify positive instances.

**Specificity:** Specificity measures the proportion of true negative predictions out of all actual negative samples. It represents the model's ability to correctly identify negative instances.

These metrics provide comprehensive evaluations of the binding affinity prediction performance and help assess the model's ability to accurately classify compounds based on their affinity.

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (\text{Equation 21})$$

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (\text{Equation 22})$$

$$\text{Recall} = \frac{TP}{TP+FP} \quad (\text{Equation 23})$$

$$\text{Precision} = \frac{TP}{TP+FN} \quad (\text{Equation 24})$$

$$\text{FPR} = \frac{FP}{TN+FP} \quad (\text{Equation 25})$$



Where TP stands for True Positive, FP stands for False Positive, TN stands for True Negative, and FN stands for False Negative.

F1 and x% RE (ROC enrichment metric) are among the metrics used to evaluate the performance of binding affinity prediction models.

F1 is a comprehensive metric that assesses the balance between precision and recall of a model. It is the harmonic mean of precision and recall and measures the model's ability to achieve a balance between predicting positives and negatives. A higher F1 score indicates a better balance between precision and recall.

x% RE (ROC enrichment metric) is a metric used to evaluate the predictive performance of a model at a specific recall rate. It measures the model's screening ability by calculating the proportion of true active compounds identified at a specific recall rate. A higher x% RE value indicates that the model can better distinguish true active compounds at the given recall rate.

$$F1 = \frac{2 * Recall * Precision}{Recall + Precision} \quad (\text{Equation 26})$$

$$x\% RE = \frac{Recall}{FPR} \text{ at a give threshold} \quad (\text{Equation 27})$$

These metrics provide a comprehensive evaluation of the performance of binding affinity prediction models and help determine their accuracy, balance, and screening ability.

### QUANTIFICATION AND STATISTICAL ANALYSIS

The results of multiple experiments in each table are supplemented with the corresponding mean squared deviation (SD), by which in this paper we measure the stability of the model. In addition, all the software involved is available in [key resources table](#).