Synergistic application of molecular docking and machine learning for improved binding pose

Yaqi Li1,2,#, Hongrui Lin2,#, He Yang2, Yannan Yuan2, Rongfeng Zou2, Gengmo Zhou2,3, Linfeng Zhang2,4,* & Hang Zheng2,*

1College of Life Sciences, Hunan Normal University, Changsha 410006, China;
2DP Technology, Beijing 100089, China;
3Gaoling School of Artificial Intelligence, Renmin University of China, Beijing 100872, China;
4AI for Science Institute, Beijing 100080, China

#Contributed equally to this work.
*Corresponding authors (emails: zhanglf@dp.tech (Linfeng Zhang); zhengh@dp.tech (Hang Zheng))

Received 10 September 2023; Revised 1 March 2024; Accepted 5 March 2024; Published online 8 March 2024

Abstract: Accurate prediction of protein-ligand complex structures is a crucial step in structure-based drug design. Traditional molecular docking methods exhibit limitations in terms of accuracy and sampling space, while relying on machine-learning approaches may lead to invalid conformations. In this study, we propose a novel strategy that combines molecular docking and machine learning methods. Firstly, the protein-ligand binding poses are predicted using a deep learning model. Subsequently, position-restricted docking on predicted binding poses is performed using Uni-Dock, generating physically constrained and valid binding poses. Finally, the binding poses are re-scored and ranked using machine learning scoring functions. This strategy harnesses the predictive power of machine learning and the physical constraints advantage of molecular docking. Evaluation experiments on multiple datasets demonstrate that, compared to using molecular docking or machine learning methods alone, our proposed strategy can significantly improve the success rate and accuracy of protein-ligand complex structure predictions.

Keywords: binding pose, molecular docking, machine learning, machine learning scoring function

INTRODUCTION

Protein-ligand complex structure prediction is one of the essential aspects of drug design. Accurately predicting the protein-ligand complex structure can provide a basis for structure-based drug design, thereby facilitating the design and selection of potential drug molecules. Furthermore, reasonable complex structures can help medicinal chemists understand the binding mechanism of small molecules with target proteins, laying the foundation for structure-activity relationship analysis and rational drug design [1, 2]. Consequently, developing accurate protein-ligand complex structure prediction methods is of great importance for structure-based drug design.

Existing molecular docking schemes, such as AutoDock Vina [3], Uni-Dock [4], and LeDock (Zhao H. LeDock. http://www.lephar.com/software.htm), typically rely on conformational sampling algorithms and empirical scoring functions to search for protein and ligand binding poses and predict ligand conformations.
at the target protein binding site based on factors such as ligand internal energy and protein-ligand interaction energy [5]. However, these methods struggle to accurately describe various interaction forms between proteins and ligands, mainly due to simplified scoring functions for ensuring computational speed. Moreover, the search complexity of conformational sampling algorithms limits their coverage of chemical space [6]. These factors result in limited capabilities of traditional molecular docking software in protein-ligand complex structure prediction.

In recent years, machine learning-based scoring functions, such as GNINA [7] and RTMScore [8], have gained wide attention. These methods establish more refined and accurate scoring functions by learning from protein-ligand complex structures and affinity data. Studies have shown that compared to traditional empirical scoring functions, machine learning models can improve docking prediction success rates [9]. However, machine learning model inference speeds are relatively slow and are generally only used for re-scoring and ranking several protein-ligand binding poses obtained through molecular docking to select the optimal structure [10]. If the preceding molecular docking step fails to sample conformations close to the crystal structure, machine learning scoring functions may become ineffective.

On the other hand, some deep learning models, such as DeepDock [11], KarmaDock [12] and Uni-Mol Docking [13], attempt to directly predict protein-ligand complex structures end-to-end without explicit conformational search, achieving promising performance. These methods can avoid the limited sampling space size of traditional molecular docking algorithms and exhibit higher prediction success rates. However, the lack of physical constraints on chemical structures in deep learning models may result in predicted conformations that do not adhere to basic physical laws, such as invalid bond lengths and angles, and protein collisions [14].

To fully exploit the advantages of traditional molecular docking methods and machine learning approaches while avoiding their respective shortcomings, we propose a novel strategy that combines molecular docking and machine learning approaches for more accurate and valid protein-ligand complex structure prediction. Firstly, we use the deep learning model to predict the binding poses of protein-ligand complexes. Subsequently, based on the predicted binding poses, we perform position-restricted Docking (PRDock) using the Uni-Dock molecular docking software, generating a series of physically constrained docked binding poses. Finally, we employ scoring functions such as GNINA, RTMScore, and Vina to re-score and rank the docked binding poses, yielding the optimal protein-ligand complex conformation. This approach combines traditional molecular docking methods with machine learning techniques, providing an efficient means for structure-based drug design.

METHODS

The proposed workflow that combines machine learning methods and molecular docking is shown in Figure 1, consisting of three steps.

Predicting

Utilize the deep learning models like DeepDock, KarmaDock and Uni-Mol Docking to predict protein-ligand binding poses. During prediction, we extract amino acid residues within a 10 Å range of the ligand as the
binding pocket and re-generate the ligand’s 3D conformation using RDKit, which is then input into the trained deep learning models along with the binding pocket to obtain the “predicted binding poses”.

**Docking**

Perform PR Docking using Uni-Dock based on the “predicted binding poses”. Uni-Dock is a high-performance GPU-accelerated molecular docking software [4] that can search a larger conformational space. We use the coordinates of each heavy atom in “predicted binding poses” as position-restricted offsets for PR Docking (specific details are in the Supplementary information S1.4), guiding Uni-Dock to focus on the binding pose region for conformation sampling, and generating “docked binding poses” that comply with the position constraints.

**Rescoring**

Apply a machine learning scoring function to re-score the “docked binding poses”. We use various traditional and machine learning scoring functions, including GNINA [7] and RTMScore [8], to score the “docked binding poses”. The conformations are then ranked, and the highest-ranked binding pose is selected as the “final binding pose”.

This strategy takes full advantage of the machine learning conformation prediction capabilities and the physical constraints of traditional molecular docking, avoiding their respective limitations, and is expected to effectively improve the success rate and accuracy of protein-ligand complex structure prediction.

This strategy is available at https://github.com/dptech-corp/Uni-Dock.

**DATASETS**

To evaluate the performance of the proposed method, we used several commonly used protein-ligand complex datasets, including Astex Diverse set [15], CASF-2016 [16], and PoseBusters [14]. Due to the significant differences in docking sampling space brought by varying numbers of rotatable bonds in ligands, we classified the test sets based on the number of rotatable bonds in ligands into different difficulty levels: ligands with 0–5 rotatable bonds were classified as “easy”, 6–12 as “medium”, and ligands with more than 12 rotatable bonds as “difficult”. Given that several scoring functions were trained using the PDBBind PL dataset, we opted to exclude this dataset from our analysis to prevent data overlap and ensure the integrity of our evaluation.
We performed the following preparation steps for the proteins and ligands in the datasets. After obtaining the protein structures from the RCSB database [17] based on the PDB code, we retained the crystal waters and cofactors that affect the binding mode and completed missing protein side chains and lost hydrogen atoms. For ligands, we searched the RCSB database for the isomer SMILES corresponding to the PDB code and determined the correct protonation state according to the receptor pocket environment. Then, we generated 3D conformations for each ligand. After excluding systems with failed preparation and those with large natural products or polypeptide ligands, 84 systems from the Astex Diverse set, 271 systems from CASF-2016 and 428 systems from PoseBusters were used as test sets.

The key statistical information is summarized in Table 1. These datasets broadly represent protein-small molecule systems of varying difficulty levels and complexities.

All the processed data has been compiled and is available for download at the following link: https://bohrium-api.dp.tech/ds-dl/PRDock-Release-5oyj-v1.zip.

### RESULTS AND DISCUSSION

#### Predicting protein-ligand binding poses using deep learning models

We first evaluated the performance of deep learning models in predicting protein-ligand binding poses on Astex, CASF-2016 and PoseBuster datasets (success was defined as the root-mean-square deviation (RMSD) between the predicted pose and the crystal pose less than 2 Å). As shown in Figure 2, the binding poses with RMSD less than 2 Å predicted by Uni-Mol Docking are around 80% for all the datasets, outperforming DeepDock and KarmaDock. Notably, in the PoseBuster dataset, where none of the protein-ligand complexes were part of the training data (as described in Supplementary information S1.1) for any of the methods, Uni-Mol Docking still demonstrated good generalization capabilities and predictive performance.

Interestingly, we found that Uni-Mol Docking had a very low success rate in predicting detailed structures, specifically in the regions with smaller RMSD values. Therefore, we selected a few representative systems for demonstration in Figure 3. From the overlay of the predicted binding poses and crystal structures, we can see that Uni-Mol Docking can accurately predict the overall trend of the molecules. However, in the prediction of symmetric structures, such as phenyl rings and isopropyl groups, Uni-Mol Docking exhibited non-physical bond lengths and angles.

Therefore, we subsequently employed Uni-Dock, a physics-based molecular docking method, to refine and optimize the predicted binding poses obtained from Uni-Mol Docking.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Total number</th>
<th>Number of easy cases</th>
<th>Number of medium cases</th>
<th>Number of hard cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Astex diverse</td>
<td>84</td>
<td>50</td>
<td>34</td>
<td>0</td>
</tr>
<tr>
<td>CASF-2016</td>
<td>271</td>
<td>182</td>
<td>73</td>
<td>16</td>
</tr>
<tr>
<td>PoseBusters</td>
<td>428</td>
<td>239</td>
<td>169</td>
<td>20</td>
</tr>
</tbody>
</table>
Getting binding poses by PRDock

We utilized the predicted binding poses generated by deep learning models as a basis and transformed them into position-restricted bias potentials (Figure 4) during the docking process. Uni-Dock was then employed for PRDock to generate more reasonable binding poses. During docking processing, when the atoms of the ligand molecule enter the range of the bias potential, the binding pose score receives a reward. Consequently, Uni-Dock makes the final docked binding pose more inclined towards the parts with bias potential, as shown in Figure 4. Since Uni-Dock explicitly avoids physical clash, such as ligand-protein proximity, and generates conformations based on rotatable bonds, this workflow can effectively leverage the binding structure prediction ability of deep learning models while ensuring the physical reliability of binding poses.

We applied this workflow to three test datasets. Following the results in the section of “Predicting protein-ligand binding poses using deep learning models”, we will adopt Uni-Mol Docking as the deep learning model in the workflow to ensure superior outcomes. In addition, we conducted PRDock using crystal structures, which can be considered as an upper bound for this workflow. On the other hand, we conducted unbiased molecular docking as a lower bound. The results are shown in Figure S1.

We observed that the PRDock results based on Uni-Mol Docking prediction consistently improved the success rate of binding conformation prediction compared to Uni-Dock molecular docking. In particular, for systems with a higher number of rotatable bonds in the ligand, this combined method had a more significant improvement in prediction accuracy, indicating that PRDock effectively reduced the complexity of searching in chemical space, helping the molecular docking method to converge rapidly around the true structure position. Compared to Uni-Mol Docking’s results, this combined method significantly increased the success rate for RMSD less than 1 Å, proving that Uni-Dock can effectively correct structures that do not conform to physical constraints and improve the local structure prediction accuracy.

Although the protein-ligand complex prediction success rate of Uni-Mol Docking in some systems with RMSD less than 2 Å is even higher than that of PRDock which uses crystal structures, we found that the accuracy of PRDock with Uni-Mol docking predicted structures did not exceed that of PRDock using crystal structures.
Figure 3  The binding conformation predicted by Uni-Mol Docking (left) and the crystal conformation (right). (A) Comparative overlay of Uni-Mol docking prediction with the crystal structure from the PDBBind refined set (PDB ID: 1BJU) (left: prediction, right: crystal structure). (B) Evaluation of symmetric structure elements in Uni-Mol docking prediction versus the crystal structure from PoseBuster (PDB ID: 7XQZ) (left: prediction, right: crystal structure).

Figure 4  (A) Conversion of Uni-Mol Docking predicted conformations to bias potentials. (B) Results obtained by PRDock on Uni-Mol Docking prediction.

structures, and there was even a significant gap. This indicates that the predicted structures of Uni-Mol Docking cannot yet serve as a perfect solution to guide molecular docking in conformation search.

Re-ranking docking poses by machine learning scoring functions

We further investigated the binding poses obtained by PRDock and molecular docking to assess their actual sampling capabilities. The abilities of these methods to reproduce crystal structures when retaining a certain number of docking poses are shown in Figure 5.

We observed that as the number of considered conformations increases, the probability of finding a conformation with an RMSD less than 2.0 Å from the ligand’s crystal conformation also increases for both PRDock and conventional molecular docking. This suggests that to further improve the accuracy of conformation prediction, we need better methods to compare the sampled conformations and identify those that are close to
Figure 5 Sampling capabilities of Uni-Dock and PRDock. The data for Uni-Dock+PRDock refers to selecting the top $N/2$ results from Uni-Dock and the top $N/2$ results from PRDock, combining them, and then determining the minimum RMSD value. Consequently, there is no corresponding result for Top1.

the crystallographic structure. Additionally, when considering all possible docking poses, the success rates of PRDock and conventional molecular docking are comparable.

Furthermore, we computed the results of considering both the molecular docking and PRDock (Uni-Dock+PRDock) together. Under a substantial number of samples, the integrative approach of selecting top results from both methods yielded a much higher success rate than either method alone. For example, the Top2 result for Uni-Dock + PRDock is to take the best prediction from Uni-Dock with the best prediction from PRDock, then evaluate whether any of these two poses match the crystal structure within an RMSD of $<2\,\text{Å}$. It is observed that the success rate is significantly enhanced, implying that Uni-Dock can succeed in cases where Uni-Mol fails, thus revealing a strong complementary relationship. Nonetheless, the persistent challenge is in selecting the superior docking conformation from the ensemble of predictions.

The information above indicates that molecular docking methods can effectively collect conformations close to the crystal structure of the ligand. The challenge lies in selecting excellent binding poses and placing them in the forefront. In particular, when the top-ranked complex conformation structures given by Uni-Dock and PRDock with Uni-Mol Docking prediction are inconsistent, it is challenging to determine which structure is better. Therefore, we subsequently tested the rescoring and re-ranking performance of machine learning scoring functions GNINA, RTMScore and alongside the physics-based scoring function Vina, by assessing whether rescoring the Top2, Top6, Top20 and Top100 could improve the prediction success rate of binding conformations. The results are shown in Figure 6.

Firstly, we compared the efficacy of rescoring and reranking the predicted binding poses from Uni-Dock or PRDock alone using scoring functions. When reevaluating a limited number of poses, such as the Top2 results, we observed an overall increase in the success rate for selecting the correct binding poses. Moreover, when attempting to identify the more crystal-structure-like pose from the best predictions of Uni-Dock+PRDock (performance on Top2 of Uni-Dock+PRDock), GNINA and RTMScore showcased their adeptness in selection. They achieved higher success rates on all three datasets compared to taking the best predictions from Uni-Dock or PRDock individually. This evidence supports the notion that a synergistic workflow that combines deep learning predictions with molecular docking results, followed by the application of MLSFs to select the final binding pose, can indeed improve the prediction of protein-ligand binding.
Figure 6  Rescoring and re-ranking predicted binding poses using Vina, GNINA and RTMScore.

poses, resulting in more accurate conformations.

However, the success rate of pose selection decreased when a larger number of poses were re-scored, such as the Top20 or Top100. We suspect that this is due to the training datasets for GNINA and RTMScore primarily including crystal binding poses and closely related decoys, which may cause a decline in performance when re-scoring poses that significantly deviate from the crystal structure (e.g., changes in the overall direction of the ligand rather than just shifts in the positions of functional groups). This suggests that future training of MLSFs should focus on generating molecular conformations that cover a larger conformational space. On the other hand, while the success rates of MLSFs significantly dropped with an increased number of poses, traditional scoring functions maintained a consistent level of performance across different numbers of poses, exhibiting greater robustness in assessing less common structures due to the presence of physical
constraints. Finally, we observed that RTMScore performed more consistently on datasets closer to the training data, such as Astex and CASF-2016, compared to PoseBuster, which differs more from the training set. This indicates a potential overfitting issue with deep learning models that employ more complex architectures and a larger number of parameters. Conversely, GNINA, with fewer parameters and a simpler model, still performed well even with less training data.

CONCLUSION

In this paper, we propose a novel method that combines molecular docking and machine learning to enhance the accuracy of protein-ligand binding pose prediction. First, we employ the deep learning model to predict protein-ligand binding poses. Next, we use PRDock on predicted binding poses to perform molecular docking, generating physically constrained binding poses. Finally, we re-score multiple binding poses using MLSFs to identify the best binding pose as the final predicted protein-ligand complex structure.

Evaluation experiments on multiple benchmark datasets demonstrate that, compared to using traditional docking or machine learning methods alone, this combined strategy significantly improves the success rate and accuracy of binding pose prediction, particularly for systems with high ligand flexibility. This shows that machine learning-predicted binding poses can effectively guide molecular docking searches, while the physical constraints provided by molecular docking prevent the generation of non-physically plausible conformations.

However, our work also reveals limitations of the current methods: (1) the binding pose prediction accuracy of the deep learning models in terms of structural plausibility still needs improvement, especially for symmetric structures; (2) the re-scoring process on a large number of binding poses using MLSFs brings significant declines on success rates, suggesting potential issues in the training data of the current MLSFs. Based on these findings, we will attempt to incorporate more physical constraints into the deep learning models for docking process and test various combinations of MLSFs and workflows further to enhance the prediction ability of protein-ligand complex structures.

Funding
This work was supported by the National Key Research and Development Program of China (2022YFA1004302).

Author contributions
Y.L. and H.Y. were responsible for conducting the experiments and writing the manuscript. H.L. designed the methodologies and contributed to the manuscript writing. Y.Y. was in charge of constructing the workflow. R.Z. took care of data collection. G.Z. was responsible for the development and deployment of the machine learning model Uni-Mol Docking. L.Z. proposed and supervised the project. H.Z. designed the methodologies, carried out the experiments, took part in writing the manuscript, and also proposed and supervised the project.

Conflict of interest
The authors declare no conflict of interest.

Supplementary information
The supporting information is available online at https://doi.org/10.1360/nso/20230058. The supporting materials are published
as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References