# Protein language model embedded geometric graphs power inter-protein contact prediction

Yunda Si, Chengfei Yan*

School of Physics, Huazhong University of Science and Technology, China

Correspondence: chengfeiyan@hust.edu.cn

# Abstract

Accurate prediction of contacting residue pairs between interacting proteins is very useful for structural characterization of protein-protein interactions (PPIs). Although significant improvement has been made in inter-protein contact prediction recently, there is still large room for improving the prediction accuracy. Here we present a new deep learning method referred to as PLMGraph-Inter for inter-protein contact prediction. Specifically, we employ rotationally and translationally invariant geometric graphs obtained from structures of interacting proteins to integrate multiple protein language models, which are successively transformed by graph encoders formed by geometric vector perceptrons and residual networks formed by dimensional hybrid residual blocks to predict inter-protein contacts. Extensive evaluation on multiple test sets illustrates that PLMGraph-Inter outperforms three top inter-protein contact prediction methods, including DeepHomo, GLINTER and DRN-1D2D_Inter by large margins. In addition, we also show that the prediction of PLMGraph-Inter can well complement the result of AlphaFold-Multimer. Finally, we show leveraging the contacts predicted by PLMGraph-Inter as constraints for protein-protein docking can dramatically improve its performance for protein complex structure prediction.

# Introduction

Protein-protein interactions(PPIs) are essential activities of most cellular processes[1,2]. Structure characterization of PPIs is important for mechanistic investigation of these cellular processes and therapeutic development[3]. However, currently experimental structures of many important PPIs are still missing as experimental methods to resolve complex structures such as X-ray crystallography, nuclear magnetic resonance, cryo-electron microscopy are costly and time-consuming[4]. Therefore, it is necessary to develop computational methods to predict protein complex structures[5]. Predicting contacting residue pairs between interacting proteins can be considered as an intermediate step for protein complex structure prediction[6,7], as the predicted contacts can be integrated into protein-protein docking algorithms to assist protein complex structure prediction[8–10]. Besides, the predicted contacts can also be very useful to guide protein interfacial design[11] to and the inter-protein contact prediction methods can be further extended to predict novel PPIs [12,13].

Based on the fact that contacting residue pairs often vary co-operatively during evolution, coevolutionary analysis methods[14] have been used in previous studies to predict inter-protein contacts[6,7]. However, coevolutionary analysis methods do have certain limitations. For examples, effective

coevolutionary analysis requires a large number of interolog sequences, which are often difficult to obtain, especially for heteromeric PPIs[15]; and it is difficult to distinguish inter-protein and intra-protein coevolutionary signals for homomeric PPIs[16]. Inspired by its great success in intra-protein contact prediction[17–21], deep learning has also been applied to predict inter-protein contacts[22–26]. ComplexContact[24] , to the best of our knowledge, the first deep learning method for inter-protein contact prediction, has significantly improved the prediction accuracy over coevolutionary analysis methods. However, its performance on eukaryotic PPIs is still quite limited, partly due to the difficulty to accurately infer interologs for eukaryotic PPIs. In a later study, coming from the same group as ComplexContact, Xie et al. developed GLINTER[25], another deep learning method for inter-protein contact prediction. Comparing with ComplexContact, GLINTER leverages structures of interacting monomers, from which their graph representations are used as additional input features. GLINTER significantly outperforms ComplexContact in the prediction accuracy, although there is still large room for improvement, especially for heteromeric PPIs. Apart from these methods developed to predict inter-protein contacts for both homomeric and heteromeric PPIs, inter-protein contact prediction methods specifically for homomeric PPIs were also developed[23,26,27], as predicting the inter-protein

contacts for homomeric PPIs is generally much easier due to the symmetric restriction, relatively larger interfaces and the trivialness of interologs identification. For example, Yan et al. developed DeepHomo[26], a deep learning method specifically to predict inter-protein contacts of homomeric PPIs, which also significantly outperforms coevolutionary analysis-based methods. However, DeepHomo requires docking maps calculated from structures of interacting monomers, which is computationally expensive. Besides, the docking maps is also sensitive to the quality of monomeric structures. In our recent study, we proved that sequence embeddings from protein language models[28,29] (PLMs) are effective features for inter-protein contact prediction, with which we developed DRN-1D2D_Inter[30]. Extensive benchmark results show that DRN-1D2D_Inter significantly outperforms DeepHomo and GLINTER in inter-protein contact prediction, although DRN-1D2D_Inter makes the prediction purely from sequences.

Recently, AlphaFold2[31] developed by DeepMind has achieved unprecedented breakthrough in protein structure prediction, with which structures of many proteins with unknown experimental structures can be predicted with accuracies close to experimental results. Besides, DeepMind also released over 200 million protein structures predicted by AlphaFold2[32]. Therefore, for most proteins under study, we can always find either their

experimental structures or high-quality predicted structures. Leveraging the structural information effectively may further improve the performance of inter-protein contact prediction. In this study, we developed a structure-informed method to predict inter-protein contacts. Given structures of two interacting proteins, we first build rotationally and translationally (SE(3)) invariant geometric graphs from the two monomeric structures, with which we integrate PLMs of single sequences, multiple sequence alignments (MSAs) and structures to form the PLM embedded geometric graphs, which are then transformed by graph encoders formed by geometric vector perceptrons to generate graph embeddings for interacting monomers. The graph embeddings are further combined with inter-protein pairwise features and transformed by residual networks formed by dimensional hybrid residual blocks to predict inter-protein contacts. The developed method referred to as PLMGraph-Inter was extensive benchmarked on multiple tests with application of either experimental or predicted structures of interacting monomers. The result shows that in both cases, PLMGraph-Inter outperforms other top prediction methods including DeepHomo, GLINTER and DRN-1D2D_Inter by large margins. In addition, we also compared the prediction results of PLMGraph-Inter with the protein complex structures generated by AlphaFold-Multimer[33], the extension of AlphaFold2 for protein

complex structure prediction. The result shows that for many targets which AlphaFold-Multimer made poor predictions, PLMGraph-Inter yielded much better results. Finally, we show leveraging the contacts predicted by PLMGraph-Inter as constraints for protein-protein docking can dramatically improve its performance for protein complex structure prediction.

# Results

## Overview of PLMGraph-Inter

The method of PLMGraph-Inter is summarized in Figure 1. PLMGraph-Inter consists of three modules: the graph representation module, the graph encoder module and the residual network module. Each interacting monomer is first transformed into a PLM-embedded graph by the graph representation module, then the graph is passed through the graph encoder module to obtain a 1D representation of each protein. The two protein representations are transformed into 2D pairwise features through outer concatenation and further concatenated with other 2D pairwise features, which are then transformed by the residual network module to obtain the predicted inter-protein contact map.

**The graph representation module.** The first step of the graph representation module is to represent the protein 3D structure as a geometric

graph, where each residue is represented as a node, and an edge is defined if the $C_\alpha$ atom distance between two residues is less than 18Å. For each node and edge, we use scalars and vectors extracted from the 3D structures as their geometric features. To make the geometric graph SE(3) invariant, we use a set of local coordinate systems to extract the geometric vectors. The SE(3) invariance of representation of each interacting monomer is important, as in principal, the inter-protein contact prediction result should not depend on the initial positions and orientations of protein structures. A detailed description can be found in the Methods section. The second step is to integrate the single sequence embedding from ESM-1b[29], the MSA embedding from ESM-MSA-1b[28], the Position-Specific Scoring Matrix (PSSM) calculated from the MSA and structure embedding from ESM-IF[34] for each interacting monomer using its corresponding geometric graph. Where ESM-1b and ESM-MSA-1b are pretrained PLMs learned from large datasets of sequences and MSAs respectively without label supervision, and ESM-IF is a supervised PLM trained from 12 million protein structures predicted by AlphaFold2 for fixed backbone design. The embeddings from these models contain high dimensional representations of each residue in the protein, which are concatenated and further combined with the PSSM to form additional features of each node in the geometric graph. Since the

sequence embeddings, the MSA embeddings, the PSSM and the structure embeddings are all SE(3) invariant, the PLM-embedded geometric graph of each protein is also SE(3) invariant.

**The graph encoder module.** The graph encoder module is formed by geometric vector perceptron (GVP) and GVP convolutional layer (GVPConv)[35,36]. Where GVP is a graph neural network module consisting of a scalar track and a vector track, which can perform SE(3) invariant transformations on scalar features and SE(3) equivariant on vector features of nodes and edges; GVPConv follows the message passing paradigm of graph neural network and mainly consists of GVP, which updates the embedding of each node by passing information from its neighboring nodes and edges. A detailed description of GVP and GVPConv can be found in the Methods section and also in the work of GVP[35,36]. For each protein graph, we first use a GVP module to reduce the dimension of the scalar features of each node from 2586 to 256, which is then transformed successively by three GVPConv layers. Finally, we stitch the scalar features and the vector features of each node to form the 1D representation of the protein. Since the input protein graph is SE(3) invariant and the GVP and GVPConv transformations are SE(3) equivariant, the 1D representation of each interacting monomer is also SE(3) invariant.

**The residual network module.** The residual network module block is mainly formed by nine dimensional hybrid residual blocks. Our previous study illustrated the effective receptive field can be enlarged with the application of the dimensional hybrid residual block, thus helps improve the model performance[19]. Specifically, we first use a convolution layer with kernel size of 1*1 to reduce the number of channels of the input 2D maps from 1044 to 96, which are then transformed successively by nine dimensional hybrid residual blocks and another convolution layer with kernel size of 1*1 for the channel reduction (from 96 to 1). Finally, we use the sigmoid function to transform the feature map to obtain the predicted inter-protein contact map.

**Evaluation of PLMGraph-Inter on HomoPDB and HeteroPDB test sets**

We first evaluated PLMGraph-Inter on two self-built non-redundant test sets: HomoPDB and HeteroPDB. Where HomoPDB is the test set for homomeric PPIs containing 400 homodimers and HeteroPDB is the test set for heteromeric PPIs containing 200 heterodimers. For comparison, we also evaluated DeepHomo, GLINTER, and DRN-1D2D_Inter on the same datasets. Since DeepHomo was developed to predict inter-protein contacts only for homomeric PPIs, its evaluation was only performed on HomoPDB. In all the evaluations, the structural related features were drawn from

experimental bound structures of interacting monomers separated from complex structures of PPIs (i.e. native structures) after randomizing their initial positions and orientations (DRN-1D2D_Inter does not use structural information). It should be noted that since HomoPDB and HeteroPDB are not de-redundant with the training sets of DeepHomo and GLINTER, the performances of the two methods may be overestimated.

Table 1 shows the mean precision of each method on the HomoPDB and HeteroPDB when top (5, 10, 50, L/10, L/5) predicted inter-protein contacts are considered, where L denotes the sequence length of the shorter protein in the PPI and (Note: GLINTER encountered errors in 81 targets in HomoPDB and 15 targets in HeteroPDB at run time and did not produce predictions, thus we removed these targets in the evaluation of the performance of GLINTER. The performances of other methods on HomoPDB and HeteroPDB after the removal of these targets are shown in Table S1). As can be seen from the table, the mean precision of PLMGraph-Inter far exceeds those of other algorithms in each metric for both datasets. In particular, the mean precision of PLMGraph-Inter is substantially improved in each metric on each dataset compared to our previous method DRN-1D2D_Inter which used most features of PLMGraph-Inter except these drawn from structures of the interacting monomers, illustrating the

importance of the inclusion of structural information. Besides, GLINTER, which also uses structural information and PLMs, has much lower performance than PLMGraph-Inter, illustrating the efficacy of our deep learning framework.

It can also be seen from the result that all the methods tend to have better performances on HomoPDB than those on HeteroPDB. This is reasonable as complex structures of homodimers are generally C2 symmetric, which largely restricts the configurational spaces of PPIs, making the inter-protein contact prediction a much easier task (e.g., the inter-protein contact maps for homodimers are also symmetric). Besides, comparing with heteromeric PPIs, we may more likely to successfully infer the inter-protein coevolutionary information for homomeric PPIs to assist the contact prediction for two reasons: First, it is straightforward to pair sequences in the MSAs for homomeric PPIs, thus the paired MSA for homomeric PPIs may have higher qualities; second, homomeric PPIs may undergo stronger evolutionary constraints, as homomeric PPIs are generally permanent interactions, but many heteromeric PPIs are transient interactions.

In addition to using the mean precision on each test set to evaluate the performance of each method, the performance comparisons between PLMGraph-Inter and other models on the top 50 predicted contacts for each

individual target in HomoPDB and HeteroPDB are shown in Figure 2a-b. Specifically, PLMGraph-Inter achieved the best performance for 81.5% of the targets in HomoPDB and 67.5% of the targets in HeteroPDB. We further group targets in each dataset according to their inter-protein contact densities defined as $\frac{\#\ of\ contacts}{L_A * L_B}$, where $L_A$ and $L_B$ are the protein lengths. In Figure 2c-d, we show the mean precisions of the top 50 predicted contacts on targets within different ranges of contact densities. As can be seen from the figure that the predicted contacts tend to have lower precision when contact densities of the targets are lower, however, PLMGraph-Inter consistently achieved the best performance.

**Impact of the monomeric structure quality on contact prediction**

In real practice, native structures of interacting monomers generally do not exist, and only unbound or predicted structures can be employed in the prediction, which often contain certain levels of structural variances comparing with native structures. An effective inter-protein contact prediction method should be robust to structural variances of interacting monomers. Recently, AlphaFold2 achieved unprecedented progress in protein structure prediction, which can predict protein structures with the accuracy close to experimental results for many proteins. Therefore, it is

necessary to evaluate the performance of PLMGraph-Inter using the predicted structures generated by AlphaFold2.

We used AlphaFold2 to predict structures of interacting monomers for all PPIs in HomoPDB and HeteroPDB. Considering that interacting monomers in HomoPDB and HeteroPDB are not de-redundant with the training set of AlphaFold2, using default settings of AlphaFold2 may overestimate its performance. To mimic the performance of AlphaFold2 in real practice and produce predicted monomeric structures with more diverse qualities, we only used the MSA searched from Uniref100[37] protein sequence database as the input to the model_5_ptm of AlphaFold2, and did not use other sequence databases and prediction models included in AlphaFold2. Besides, we set to not use templates in the structure prediction. The predicted structures yielded a mean TM-score 0.88, which is close to the performance of AlphaFold2 for CASP14 targets (mean TM-score 0.85)[38]. The predicted structures of interacting monomers were then severed as the input to PLMGraph-Inter and other prediction methods (except for DRN-1D2D_Innter, which does not use structural information) for inter-protein contact prediction.

As shown in Figures 3a-b, when the predicted structures were used by PLMGraph-Inter for inter-protein contact prediction, mean precisions of the

predicted inter-protein contacts in each metric on both HomoPDB and HeteroPDB test sets decrease by about 5%, indicating qualities of the input structures do have certain impact on the prediction performance. However, the performance of PLMGraph-Inter using the predicted monomer structures still dramatically outperforms DRN-1D2D_Inter and other prediction methods (see Table S2).

We further explored the impact of the monomeric structure quality on the inter-protein contact prediction performance of PLMGraph-Inter. Specifically, for a given PPI, the TM-score[39] was used to evaluate the quality of the predicted structure for each interacting monomer, and the TM-score of the predicted structure with lower quality was used to evaluate the overall monomeric structure prediction quality for the PPI, denoted as "DTM-score". In Figure 3c, we show the performance gaps (using the mean precisions of the top 50 predicted contacts as the metric) between applying the predicted structures and applying the experimental structures in the inter-protein contact prediction, in which we grouped targets according to DTM-scores of their monomeric structure prediction, and in Figure 3d, we show the performance comparison for each specific target. From Figure 3c-d, we can clearly see that when the DTM-score is lower, the prediction using the prediction structure tends to have lower accuracy. However, when the DTM-

score is greater than or equal to 0.8, there is almost no difference between the applying the predicted structures and applying the experimental structure, which shows the robustness of PLMGraph-Inter to the structure quality.

**Ablation study**

To explore the contribution of each input component to the performance of PLMGraph-Inter, we conducted ablation study on PLMGraph-Inter. The graph representation from the structure of each interacting proteins is the base feature of PLMGraph-Inter, so we first trained the baseline model using only the geometric graphs as the input feature, denoted as model a. Our previous study in DRN-1D2D_Inter has shown that the single sequence embeddings, the MSA 1D features (including the MSA embeddings and PSSMs) and the 2D pairwise features from the paired MSA play important roles in the model performance. To further explore the importance of these features when integrated with the geometric graphs, we trained model b-d separately (model b: geometric graphs + sequence embeddings, model c: geometric graphs + sequence embeddings + MSA 1D features, model d: geometric graphs + sequence embeddings + MSA 1D features + 2D features). Finally, we included the structure embeddings as additional features to train the model e (model e uses all the input features of PLMGraph-Inter). All the five models were trained using

the same protocol as PLMGraph-Inter on the same training and validation partition without cross validation. We further evaluated performances of five models together with PLMGraph-Inter (i.e., model f: model e + cross validation) on HomoPDB and HeteroPDB using native structures of interacting monomers respective.

In Figure 4a, we show the mean precisions of the top 50 predicted contacts by model a-f on HomoPDB and HeteroPDB respectively. It can be seen from Figure 4a that including the sequence embeddings in the geometric graphs has a very good boost to the model performance (model b versus model a), while the additional introduction of MSA 1D features and 2D pairwise features can further improve the model performance (model d versus model c versus model b). DRN-1D2D_Inter also uses the same set of sequence embeddings, MSA 1D features and 2D pairwise features as the input, and our model d shows a significant performance improvement over DRN-1D2D_Inter (single model) (the model trained on the same training and validation partition without cross validation) on both HomoPDB and HeteroPDB (the mean precision improvement: HomoPDB:14%, HeteroPDB:5.6%), indicating that the introduced graph representation is important for the model performance. The head-to-head comparison of model d and DRN-1D2D_Inter (single model) on each specific target in

Figure 4b further demonstrates the value of the graph representation. Besides, the additional introduced structure embeddings from ESM-IF can further improve the mean precisions of the predicted contacts by 3~4% on both HomoPDB and HeteroPDB (model e versus model d) and the application of the cross validation can also improve the precisions by 1.0 % on HomoPDB and 2.7% on HeteroPDB (model f versus model d) (see Table S3).

To demonstrate the efficacy of our proposed graph representation of protein structures, we also trained a model using the structural representation proposed in the work of GVP[35,36] (denoted as "GVP Graph"), as a control. Our structural representation differs significantly from GVP Graph. For example, we extracted inter-residue distances and orientations between five atoms (C,O,Cα,N and a virtual Cβ) from the structure as the geometric scalar and vector features, in which the vector features are calculated in a local coordinate system. However, GVP Graph only uses the distances and orientations between Cα atoms as the geometric scalar and vector features and the vector features are calculated in a global coordinate system. In addition, after the geometric graph is transformed by the graph encoder module, GVP Graph only uses the scalar features of each node as the node representation, while we concatenate the scalar and vector features of the node as the node representation. In Figure 4c, we show the performance

comparison between this model and our base model (model a). From Figure 4c, we can clear see that our base model significantly outperforms the GVP Graph-based model on both HomoPDB and HeteroPDB, illustrating the high efficacy of our proposed graph representation.

**Evaluation of PLMGraph-Inter on DHTest and DB5.5 test sets**

We further evaluated PLMGraph-Inter on DHTest and DB5.5. The DHTest test set was formed by removing PPIs redundant to our training set from the original test set of DeepHomo, which contains 130 homomeric PPIs. The DB5.5 test set was formed by removing PPIs redundant to our training dataset from the heterodimers in Protein-protein Docking Benchmark 5.5, which contains 59 heteromeric PPIs. Still, both the native structures and the predicted structures (generated using the same protocol as in HomoPDB and HeteroPDB) of the interacting monomers were used respectively in the inter-protein contact prediction.

As shown in Figure 5a-b, when using the native structures in the prediction, the mean precisions of the top 50 contacts predicted by PLMGraph-Inter are 71.9% on DHTest and 29.5% on DB5.5 (also see Table S4), which are dramatically higher than DeepHomo, GLINTER and DRN-1D2D_Inter (Note: GLINTER encountered errors for 47 targets in DHTest and 3 targets in DB5.5 at run time and did not produce predictions, thus we

removed these targets in the evaluation of the performance of GLINTER. The performances of other methods on DHTest and DB5.5 after the removal of these targets are shown in Table S5). The distributions of the precisions of top 50 predicted contacts by different methods on DHTest and DB5.5 are shown in Figure 5d-e, from which we can clearly see that PLMGraph-Inter can make high-quality predictions for more targets on both DHTest and DB5.5.

When using the predicted structures in the prediction, the mean precisions of top 50 contacts predicted by PLMGraph-Inter show reasonable decrease to 61.1% on DHTest and 23.8% on DB5.5 respectively (see Figure 5a-b and Table S6). We also analyzed the impact of the monomeric structure quality on the inter-protein contact prediction performance of PLMGraph-Inter. As shown in Figure 5c, when the DTM-score is greater than or equal to 0.8, there is almost no difference between applying the predicted structures and the native structures, which is consistent with our analysis in HomoPDB and HeteroPDB.

We noticed that the performance of PLMGraph-Inter on the DB5.5 is significantly lower than that on HeteroPDB, and so are the performances of other methods. That the targets in DB5.5 have relatively lower mean contact densities (1.01% versus 1.29%) may partly explain this phenomenon. In

Figure 5f, we show the variations of the precisions of predicted contacts with the variation of contact density. As can be seen from Figure 5f, as the contact density increases, precisions of predicted contacts tend to increase regardless of whether the native structures or predicted structures are used in the prediction. 37.29% targets in DB5.5 are with inter-protein contact densities lower than 0.5%, for which precisions of predicted contacts are generally very low, making the overall inter-protein contact prediction performance on DB5.5 relatively low.

**Comparison of PLMGraph-Inter with AlphaFold-Multimer**

After the development of AlphaFold2, DeepMind also released AlphaFold-Multimer, as an extension of AlphaFold2 for protein complex structure prediction. The inter-protein contacts can also be extracted from the complex structures generated by AlphaFold-Multimer. It is worth to making a comparison between the performances of AlphaFold-Multimer and PLMGraph-Inter on inter-protein contact prediction. Therefore, we also employed AlphaFold-Multimer to generate complex structures for all the PPIs in the four datasets which we used to evaluate PLMGraph-Inter. We then selected the 50 inter-protein residue pairs with the shortest heavy atom distances in each generated protein complex structures as the predicted inter-protein contacts. It should be noted that AlphaFold-Multimer used all protein

complex structures in Protein Data Bank deposited before 2018-04-30 for the model development, thus these PPIs may have a large overlap with the training set of AlphaFold-Multimer. Therefore, there is no doubt that the performance of AlphaFold-Multimer would be overestimated here. Besides, AlphaFold-Multimer employed structural templates in the prediction, thus some successes may be achieved by trivially recognizing the templates.

In Figure 6a, we show the relationship between the quality of the generated protein complex structure (evaluated with DockQ) and the precision of the top 50 inter-protein contacts extracted from the protein complex structure for each PPI in the homomeric PPI (DHTest + HomoPDB) and heteromeric PPI (DB5.5 + HeteroPDB) datasets. As it can be seen from the figure that the precision of the predicted contacts is highly correlated with the quality of the generated structure. Especially when the precision of the contacts is higher than 50%, most of the generated complex structures have at least acceptable qualities (DockQ≥0.23), in contrast, almost all the generated complex structures are incorrect (DockQ<0.23) when the precision of the contacts is below 50%. Therefore, 50% can be considered as a critical precision threshold for inter-protein contact prediction (the top 50 contacts).

In Figure 6b-c, we show the comparison of the precisions of top 50 contacts predicted by AlphaFold-Multimer and PLMGraph-Inter for each target when using the native monomeric structures and the AlphaFold2 predicted structures as the input for PLMGraph-Inter respectively. It can be seen from the figure that although for most of the targets, AlphaFold-Multimer yielded better results, but for a significant number of the targets that AlphaFold-Multimer made poor predictions (precision<50%), the results of PLMGraph-Inter can have certain improvement over the AlphaFold-Multimer predictions (see Table S7). Therefore, PLMGraph-Inter can provide a complement to AlphaFold-Multimer for structural characterization of PPIs.

**PLMGraph-Inter can significantly improve protein-protein docking performance**

Prior to AlphaFold-Multimer, protein-protein docking is generally used for protein complex structure prediction. HADDOCK[40,41] is a widely used information-driven protein-protein docking approach to model complex structures of PPIs, which allows us to encode predicted inter-protein contacts as constraints to drive the docking. In this study, we used HADDOCK (version 2.4) to explore the contribution of PLMGraph-Inter to protein complex structure prediction.

We prepared the test set of homomeric PPIs by merging HomoPDB and DHTest and the test set of heteromeric PPIs by merging HeteroPDB and DB5.5, where the monomeric structures generated previously by AlphaFold2 were used as the input to HADDOCK for protein-protein docking, in which the top 50 contacts predicted by PLMGraph-Inter with the application of the predicted monomeric structures were used as the constraints. Since HADDOCK generally cannot model large conformational changes in protein-protein docking, we filtered PPIs in which either of the AlphaFold2 generated interacting monomeric structure has a TM-score lower than 0.8. Finally, the homomeric PPI test set contains 462 targets, denoted as Homodimer, and the heteromeric PPI test set contains 174 targets, denoted as Heterodimer.

For each PPI, we used the top 50 contacts predicted by PLMGraph-Inter and other methods as ambiguous distance restraints between the alpha carbons (CAs) of residues (distance=8Å, lower bound correction=8Å, upper-bound correction=4Å) to drive the protein-protein docking. All other parameters of HADDOCK were set as the default parameters. In each protein-protein docking, HADDOCK output 200 predicted complex structures ranked by the HADDOCK scores. As a control, we also performed protein-protein docking with HADDOCK in ab initio docking

mode (center of mass restraints and random ambiguous interaction restraints definition). Besides, for homomeric PPIs, we additionally added the C2 symmetry constraint in both cases.

As shown in Figure 7a-c, the success rate of docking on Homodimer and Heterodimer test sets can be significantly improved when using the PLMGraph-Inter predicted inter-protein contacts as restraints. Where on the Homodimer test set, the success rate ($\mathrm{DockQ} \geq 0.23$) of the top 1 (top 10) prediction of HADDOCK in ab initio docking mode are 15.37% (39.18%), and when predicted by HADDOCK with PLMGraph-Inter predicted contacts, the success rate of the top 1 (top 10) prediction 57.58% (61.04%). On the Heterodimer test set, the success rate of top 1 (top 10) predictions of HADDOCK in ab initio docking mode is only 1.72% (6.32%), and when predicted by HADDOCK with PLMGraph-Inter predicted contacts, the success rate of top 1 (top 10) prediction is 29.89% (37.93%). From Figure 7a-7b we can also see that integrating PLMGraph-Inter predicted contacts with HADDOCK not only allows for a higher success rate, but also more high-quality models in the docking results. The performances of docking with contacts predicted by GLINTER, DeepHomo, DRN-1D2D_Inter were shown in Table S8 and Table S9.

We further explored the relationship between the precision of top 50 contacts predicted by PLMGraph-Inter and the success rate of the top prediction of HADDOCK with PLMGraph-Inter predicted contacts. It can be clearly seen from Figure 7d that the success rate of protein-protein docking increases with the precision of contact prediction. Especially, when the precision of the predicted contacts reaches 50%, the docking success rate of both homologous and heterologous complexes can reach 80%, which is consistent with our finding in AlphaFold-Multimer. Therefore, we think this threshold can be used as a critical criterion for inter-protein contact prediction. It is important to emphasize that for some targets, although precisions of predicted contacts are very high, HADDOCK still failed to produce acceptable models. We manually checked these targets and found many of these targets have at least one chain totally entangled by another chain (e.g., PDB 3DFU in Figure S1). We doubt large structural rearrangements may exist in forming the complex structures, which is difficult to model by traditional protein-protein docking approach.

Finally, we compared the qualities of the complex structures predicted by HADDOCK (with PLMGraph-Inter predicted contacts) and AlphaFold-Multimer. Although for some targets (e.g., PDB 5HPS in Figure S2a), the qualities of the structures predicted by HADDOCK were higher than those

by AlphaFold-Multimer, for most targets, AlphaFold-Multimer generated higher quality structures (Figure S2b). Several reasons can account for the performance gap. First, precisions of the PLMGraph-Inter predicted contacts are still not enough, especially for heteromeric PPIs; second, HADDOCK cannot model large structural rearrangements in protein-protein docking, as we can see that for some targets, HADDOCK made poor predictions with high precisions of contact constraints (Figure S2c); third, it is difficult to provide an objective evaluation of the true performance of AlphaFold-Multimer, since many targets may have already been included in the training set of AlphaFold-Multimer, and AlphaFold-Multimer employed structural templates in the prediction.

## Discussion

In this study, we proposed a new method to predict inter-protein contacts, denoted as PLMGraph-Inter. PLMGraph-Inter is based on the SE(3) invariant geometric graphs obtained from structures of interacting proteins which are embedded with multiple PLMs. The predicted inter-protein contacts are obtained by successively transforming the PLM embedded geometric graphs with graph encoders and residual networks. Benchmarking results on four test datasets show that PLMGraph-Inter outperforms three

state-of-the-art inter-protein contact prediction methods including GLINTER, DeepHomo and DRN-1D2D_Inter by large margins, regardless of whether the native or predicted monomeric structures are used in building the geometric graphs. The ablation study further shows that the integration of the PLMs with the protein geometric graphs can dramatically improve the model performance, illustrating the efficacy of the PLM embedded geometric graphs in protein representations. The protein representation framework proposed in this work can also be used to develop models for other tasks like protein function prediction, PPI prediction, etc. We further show PLMGraph-Inter can well complement the result of AlphaFold-Multimer and leveraging the inter-protein contacts predicted by PLMGraph-Inter as constraints in protein-protein docking implemented with HADDOCK can dramatically improve its performance for protein complex structure prediction.

We noticed that although PLMGraph-Inter has achieved remarkable progress in inter-protein contact prediction, there is still room for further improvement, especially for heteromeric PPIs. Using more advanced PLMs, larger training datasets and explicitly integrating physicochemical features of interacting proteins are directions worthy of exploration. Besides, since protein-protein docking approach generally have difficulties in modelling

large conformational changes in PPIs, developing new approaches to integrating the predicted inter-protein contacts in the more advanced folding-and-docking framework like AlphaFold-Multimer can also be a future research direction.

# Methods

## Training and test datasets

We used the training set and test sets prepared in our previous work DRN-1D2D_Inter[30] to train and evaluate PLMGraph-Inter. More details for the dataset generation can be found in the previous work. Specifically, we first prepared a non-redundant PPI dataset containing 4828 homomeric PPIs and 3134 heteromeric PPIs (with sequence identity 40% as the threshold), and after randomly selecting 400 homomeric PPIs (denoted as HomoPDB) and 200 heteromeric PPIs (denoted as HeteroPDB) as independent test sets, the remaining 7362 homomeric and heteromeric PPIs were used for training and validation.

DHTest and DB5.5 were also prepared in the work of DRN-1D2D_Inter by removing PPIs which are redundant (with sequence identity 40% as the threshold) to our training and validation set from the test set of DeepHomo and Docking Benchmark 5.5. DHTest contains 130 homomeric PPIs, and

DB5.5 contains 59 heteromeric PPIs. Therefore, all the test sets used in this study are non-redundant (with sequence identity 40% as the threshold) to the dataset for the model development.

**Inter-protein contact definition**

For a given PPI, two residues from the two interacting proteins are defined to be in contact if the distance of any two heavy atoms belonging the two residues is smaller than 8 Å.

**Preparing the Input features**

*Geometric graphs from structures of interacting monomers*

We first represent the protein as a graph, where each residue is represented as a node, and an edge is defined if the heavy atom distance between two residues is less than 18Å. For each node and edge, we use scalars and vectors extracted from the 3D structures as their geometric features.

For each residue, we use its C,O,Cα,N and a virtual Cβ atom coordinates to extract information, the virtual Cβ coordinates are calculated using the following formula[42]: b = Cα - N, c = C - Cα, a = cross(b, c), Cβ = -0.58273431*a + 0.56802827*b - 0.54067466*c + Cα.

To achieve a SE(3) invariant graph representation, as shown in Figure S3b, we define a local coordinate system on each residue[31,43]. Specifically, for each residue, the unit vector in the $\mathrm{C}\alpha - \mathrm{C}$ direction is set as the $\vec{x}$ axis, the unit vector in the Cα-C-N plane and perpendicular $\vec{x}$ to is used as $\vec{y}$, and the z-direction is obtained through the cross product of $\vec{x}$ and $\vec{y}$.

For the ith node, we use the three dihedral angles (□, ψ, ω) of the corresponding residue as the scalar features of the node (Figure S3a), and the unit vectors between the $C_i$, $N_i$,$O_i$, $C\alpha_i$, and $\mathrm{C}\beta_i$ atoms of the corresponding residue and the $C_{i-1}$, $N_{i-1}$,$O_{i-1}$, $C\alpha_{i-1}$, and $\mathrm{C}\beta_{i-1}$ atoms of the forward residue and the $C_{i+1}$, $N_{i+1}$,$O_{i+1}$, $C\alpha_{i+1}$, and $\mathrm{C}\beta_{i+1}$ atoms of backward residue as the vector features of the node. In total, for each node, the dimension of the scalar features is 6 (each dihedral angle is encoded with its sine and cosine) and the dimension of the vector features is 50*3.

For the edge between ith node and jth node, we use the distances and directions between the atoms of the two residues as the scalar features and vector features(See Figure S3b). The distances between the $C_i$, $N_i$,$O_i$, $C\alpha_i$, and $\mathrm{C}\beta_i$ atoms of ith residue and the $C_j$, $N_j$,$O_j$, $C\alpha_j$, and $\mathrm{C}\beta_j$ atoms of the jth residue are used as scalar features after encoded with the 16 Gaussian

radial basis functions[36]. The position difference between i and j (j-i) is also used as a scalar feature after sinusoidal encoding[44]. The unit vectors between the $C_i$, $N_i$,$O_i$, $C\alpha_i$, and $C\beta_i$ atoms of ith residue and the $C_j$, $N_j$,$O_j$, $C\alpha_j$, and $C\beta_j$ atoms of the jth residue are used as vector features. In total, for each edge, the dimension of the scalar features is 432 and the dimension of the vector features is 25*3.

*Embeddings of single sequence, MSA and structure*

The single sequence embedding is obtained by feeding the sequence into ESM-1b, and the structure embedding is obtained by feeding the structure into ESM-IF. To obtain the MSA embedding, we first search the Uniref100 protein sequence database for the sequence using JACKHMMER[45] with default parameters to obtain the MSA, which is then inputted to hhmake[46] to get the HMM file, and to the LoadHMM.py script from RaptorX_Contact[17] to obtain the PSSM. The number of sequences of MSA is limited to 256 by hhfilter[46] and then input to ESM-MSA-1b to get the MSA embedding. The dimensions of the sequence embedding, PSSM, MSA embedding and structural embeddings are 1280, 20, 768 and 512 respectively. After adding embeddings to the scalar features of the nodes, the dimension of the scalar features of each node is 2586.

*2D feature from paired MSA*

For homomeric PPIs, the paired MSA is formed by concatenating two copies of the MSA. For heteromeric PPIs, the paired MSA is formed by pairing the MSAs through the phylogeny-based approach described in (https://github.com/ChengfeiYan/PPI_MSA-taxonomy_rank)[47]. We input the paired MSA into CCMpred[48] to get the evolutionary coupling matrix, and into alnstats[49] to get mutual information matrix, APC-corrected mutual information matrix and contact potential matrix. The number of sequences of paired MSA is limited to 256 by hhfilter[46] and then input to ESM-MSA-1b to get the attention maps. In total, the channel of 2D features is 148.

**GVP and GVPConv**

GVP is a two-track neural network module consisting of a scalar track and a vector track, which can perform SE(3) invariant transformations on scalar features and SE(3) equivariant transformations on vector features. A detailed description can be found in the work of GVP[35,36].

GVPConv is a message passing based graph neural network, which mainly consists of a message function and a feedforward function. Where the message function contains a sequence of three GVP modules and the feedforward function contains a sequence of two GVP modules. GVPConv is used to transform the node features. Specifically, the input node features

are first processed by the message function. We denote the features of node i by $\boldsymbol{h}^i$ , the feature of edge (j→i) by $\boldsymbol{h}^{j\to i}$, the set of nodes connected to node i by $\boldsymbol{\varepsilon}_i$, and the three GVP modules of the message function by *gm*, then the node features processed by the message function can be represented as:

$$\boldsymbol{h}_m^i = \frac{1}{len(\boldsymbol{\varepsilon}_i)}\left(\sum_{j\in\boldsymbol{\varepsilon}_i} gm(concat(\boldsymbol{h}^i,\ \boldsymbol{h}^{j\to i}, \boldsymbol{h}^j))\right) \quad (1)$$

Where $len(\boldsymbol{\varepsilon}_i)$ denotes the number of nodes connected to node i. After sequential normalization (Equation 2) and feedforward function (Equation 3), the features of node i updated by GVPConv Layer are obtained:

$$\boldsymbol{h}_m^i = \mathrm{LayerNorm}(\boldsymbol{h}_m^i + \mathrm{Dropout}(\boldsymbol{h}_m^i)) \quad (2)$$

$$\boldsymbol{h}_{out}^i = \mathrm{LayerNorm}(\boldsymbol{h}_m^i + \mathrm{Dropout}(\mathrm{gs}(\boldsymbol{h}_m^i))) \quad (3)$$

Where *gs* denotes the two GVP modules of the feedforward function, $\boldsymbol{h}_{out}^i$ denotes the outputs of GVPConv Layer.

**Training protocol**

Our training set contains 7362 PPIs, and we used seven-fold cross-validation to train PLMGraph-Inter. Specifically, we randomly divided the training set into seven subsets, and each time, we selected six subsets as the training set and the remaining subset as the validation set. Seven models

were trained in total, and the final prediction was the average of the predictions from the seven models. Each model was trained using AdamW optimizer with 0.001 as the initial learning rate, in which the singularity enhanced loss function proposed by in our previous study[19] was used calculate the training and validation loss. During training, if the validation loss did not decrease within 2 epochs, we would decay the learning rate by 0.1. The training stopped after the learning rate decayed twice and the model with the highest top-50 mean precision on the validation dataset was saved as the prediction model.

PLMGraph-Inter was implemented with pytorch (v.1.11) and trained on one NVIDIA TESLA A100 GPU with batch size equaling to 1. Due to memory limitation of GPU, the length of each protein sequence was limited to 400. When a sequence was longer than 400, a fragment with sequence length equaling to 400 was randomly selected in the model training.

**Quality assessment of the predicted protein complex structures**

We evaluated the models generated by AlphaFold-Multimer and HADDOCK using DockQ[50], a score ranging between 0 and 1. Specifically, a model with DockQ<0.23 means that the prediction is incorrect; 0.23≤ DockQ<0.49 means the model is an acceptable prediction; 0.49≤DockQ<0.8

corresponds to a medium quality prediction; and 0.8≤DockQ corresponds to a high quality prediction.

## Data Availability

The PDB accession codes for the all the training and test sets are provided in https://github.com/ChengfeiYan/PLMGraph-Inter/tree/main/data. Other data for supporting the finds of this study are available from the corresponding author upon request.

## Code Availability

The code for implementing PLMGraph-Inter is provided in https://github.com/ChengfeiYan/PLMGraph-Inter. A demo of reproducible run of PLMGraph-Inter can be found in https://codeocean.com/capsule/4368186/tree/v1.

## Acknowledgements

The work was supported by the National Natural Science Foundation of China (32101001) and new faculty startup grant (3004012167) of Huazhong University of Science and Technology. The computation is completed in the HPC Platform of Huazhong University of Science and Technology.

# Author Contributions

Y.S. and C.Y. designed and performed the experiments. Y.S. and C.Y. wrote the manuscript. C.Y. supervised the work.

# Ethics declarations

## Competing interests

The authors declare no competing interests.

# References


1. Alberts, B. The Cell as a Collection of Protein Machines: Preparing the Next Generation of Molecular Biologists. *Cell* **92**, 291–294 (1998).
2. Spirin, V. & Mirny, L. A. Protein complexes and functional modules in molecular networks. *Proc. Natl. Acad. Sci.* **100**, 12123–12128 (2003).
3. Goodsell, D. S. & Olson, A. J. Structural Symmetry and Protein Function. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 105–153 (2000).
4. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
5. Bonvin, A. M. Flexible protein–protein docking. *Curr. Opin. Struct. Biol.* **16**, 194–200 (2006).
6. Ovchinnikov, S., Kamisetty, H. & Baker, D. Robust and accurate prediction of residue–residue interactions across protein interfaces using evolutionary information. *eLife* **3**, e02030 (2014).
7. Hopf, T. A. *et al.* Sequence co-evolution gives 3D contacts and structures of protein complexes. *eLife* **3**, (2014).

8. Li, H. & Huang, S.-Y. Protein–protein docking with interface residue restraints\ast. *Chin. Phys. B* **30**, 018703 (2021).

9. Dominguez, C., Boelens, R. & Bonvin, A. M. J. J. HADDOCK:? A Protein–Protein Docking Approach Based on Biochemical or Biophysical Information. *J. Am. Chem. Soc.* **125**, 1731–1737 (2003).

10. Sun, D., Liu, S. & Gong, X. Review of multimer protein–protein interaction complex topology and structure prediction\ast. *Chin. Phys. B* **29**, 108707 (2020).

11. Martino, E., Chiarugi, S., Margheriti, F. & Garau, G. Mapping, Structure and Modulation of PPI. *Front. Chem.* **9**, 718405 (2021).

12. Cong, Q., Anishchenko, I., Ovchinnikov, S. & Baker, D. Protein interaction networks revealed by proteome coevolution. *Science* **365**, 185–189 (2019).

13. Green, A. G. *et al.* Large-scale discovery of protein interactions at residue resolution using co-evolution calculated from genomic sequences. *Nat. Commun.* **12**, 1396 (2021).

14. Weigt, M., White, R. A., Szurmant, H., Hoch, J. A. & Hwa, T. Identification of direct residue contacts in protein–protein interaction by message passing. *Proc. Natl. Acad. Sci.* **106**, 67–72 (2009).

15. Rao, R. M. *et al.* MSA Transformer. in *Proceedings of the 38th International Conference on Machine Learning* 8844–8856 (PMLR, 2021).

16. Uguzzoni, G. *et al.* Large-scale identification of coevolution signals across homo-oligomeric protein interfaces by direct coupling analysis. *Proc. Natl. Acad. Sci.* **114**, E2662–E2671 (2017).

17. Wang, S., Sun, S., Li, Z., Zhang, R. & Xu, J. Accurate De Novo Prediction of Protein Contact Map by Ultra-Deep Learning Model. *PLOS Comput. Biol.* **13**, (2017).

18. Li, Y., Hu, J., Zhang, C. & Yu, D. J. ResPRE: High-accuracy protein contact prediction by coupling precision matrix with deep residual neural networks. *Bioinformatics* **35**, 4647–4655 (2019).

19. Si, Y. & Yan, C. Improved protein contact prediction using dimensional hybrid residual networks and singularity enhanced loss function. *Brief. Bioinform.* **22**, bbab341 (2021).

20. Ju, F. *et al.* CopulaNet: Learning residue co-evolution directly from multiple sequence alignment for protein structure prediction. *Nat. Commun.* **12**, (2021).

21. Hanson, J., Paliwal, K., Litfin, T., Yang, Y. & Zhou, Y. Accurate prediction of protein contact maps by coupling residual two-dimensional bidirectional long short-term memory with convolutional neural networks. *Bioinformatics* **34**, 4039–4045 (2018).

22. Guo, Z., Liu, J., Skolnick, J. & Cheng, J. Prediction of inter-chain distance maps of protein complexes with 2D attention-based deep neural networks. *Nat. Commun.* **13**, 6963 (2022).

23. Roy, R. S., Quadir, F., Soltanikazemi, E. & Cheng, J. A deep dilated convolutional residual network for predicting interchain contacts of protein homodimers. *Bioinformatics* **38**, 1904–1910 (2022).

24. Zeng, H. *et al.* ComplexContact: a web server for inter-protein contact prediction using deep learning. *Nucleic Acids Res.* **46**, W432–W437 (2018).

25. Xie, Z. & Xu, J. Deep graph learning of inter-protein contacts. *Bioinformatics* **38**, 947–953 (2022).

26. Yan, Y. & Huang, S. Y. Accurate prediction of inter-protein residue-residue contacts for homo-oligomeric protein complexes. *Brief. Bioinform.* **22**, 1–13 (2021).

27. Quadir, F., Roy, R. S., Halfmann, R. & Cheng, J. DNCON2_Inter: predicting interchain contacts for homodimeric and homomultimeric protein complexes using multiple sequence alignments of monomers and deep learning. *Sci. Rep.* **11**, 1–10 (2021).

28. Rao, R. *et al.* MSA Transformer. *bioRxiv* 2021.02.12.430858 (2021).

29. Rives, A. *et al.* Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proc. Natl. Acad. Sci. U. S. A.* **118**, 1–46 (2021).

30. Si, Y. & Yan, C. Improved inter-protein contact prediction using dimensional hybrid residual networks and protein language models. *Brief. Bioinform.* bbad039 (2023) doi:10.1093/bib/bbad039.

31. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

32. Tunyasuvunakool, K. *et al.* Highly accurate protein structure prediction for the human proteome. *Nature* **596**, 590–596 (2021).

33. Evans, R. *et al.* Protein complex prediction with AlphaFold-Multimer. 2021.10.04.463034 Preprint at https://doi.org/10.1101/2021.10.04.463034 (2022).

34. Hsu, C. *et al.* Learning inverse folding from millions of predicted structures. 2022.04.10.487779 Preprint at https://doi.org/10.1101/2022.04.10.487779 (2022).

35. Jing, B., Eismann, S., Soni, P. N. & Dror, R. O. Equivariant Graph Neural Networks for 3D Macromolecular Structure. Preprint at http://arxiv.org/abs/2106.03843 (2021).

36. Jing, B., Eismann, S., Suriana, P., Townshend, R. J. L. & Dror, R. O. LEARNING FROM PROTEIN STRUCTURE WITH GEOMETRIC VECTOR PERCEPTRONS. 18 (2021).

37. Suzek, B. E. *et al.* UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* **31**, 926–932 (2015).

38. Lin, Z. *et al.* Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **379**, 1123–1130 (2023).

39. Zhang, Y. & Skolnick, J. Scoring function for automated assessment of protein structure template quality. *Proteins Struct. Funct. Bioinforma.* **57**, 702–710 (2004).

40. Honorato, R. V. *et al.* Structural Biology in the Clouds: The WeNMR-EOSC Ecosystem. *Front. Mol. Biosci.* **8**, (2021).

41. van Zundert, G. C. P. *et al.* The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* **428**, 720–725 (2016).

42. Dauparas, J. *et al.* Robust deep learning–based protein sequence design using ProteinMPNN. *Science* **0**, eadd2187 (2022).

43. Pagès, G., Charmettant, B. & Grudinin, S. Protein model quality assessment using 3D oriented convolutional neural networks. *Bioinformatics* **35**, 3313–3319 (2019).

44. Vaswani, A. *et al.* Attention is All you Need. in *Advances in Neural Information Processing Systems* vol. 30 (Curran Associates, Inc., 2017).

45. Potter, S. C. *et al.* HMMER web server: 2018 update. *Nucleic Acids Res.* **46**, W200–W204 (2018).

46. Steinegger, M. *et al.* HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics* **20**, 1–15 (2019).

47. Si, Y. & Yan, C. Protein complex structure prediction powered by multiple sequence alignments of interologs from multiple taxonomic ranks and AlphaFold2. *Brief. Bioinform.* **23**, bbac208 (2022).

48. Seemayer, S., Gruber, M. & Söding, J. CCMpred--fast and precise prediction of protein residue-residue contacts from correlated mutations. *Bioinforma. Oxf. Engl.* **30**, 3128–3130 (2014).

49. Jones, D. T., Singh, T., Kosciolek, T. & Tetchner, S. MetaPSICOV: combining coevolution methods for accurate prediction of contacts and long range hydrogen bonding in proteins. *Bioinforma. Oxf. Engl.* **31**, 999–1006 (2015).

50. Basu, S. & Wallner, B. DockQ: A Quality Measure for Protein-Protein Docking Models. *PLOS ONE* **11**, e0161879 (2016).

# Figures and Tables

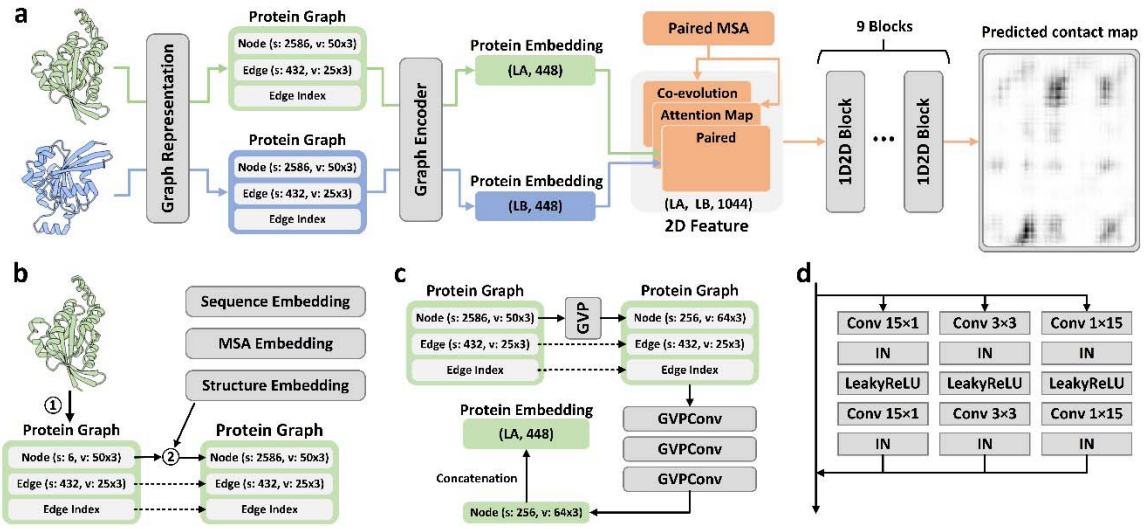



**Figure 1.** Overview of PLMGraph-Inter. (a) The network architecture of PLMGraph-Inter. (b) The graph representation module. (c) The graph encoder module, s denotes scalar features, v denotes vector features. (d) The dimensional hybrid residual block ("IN" denotes Instance Normalization).

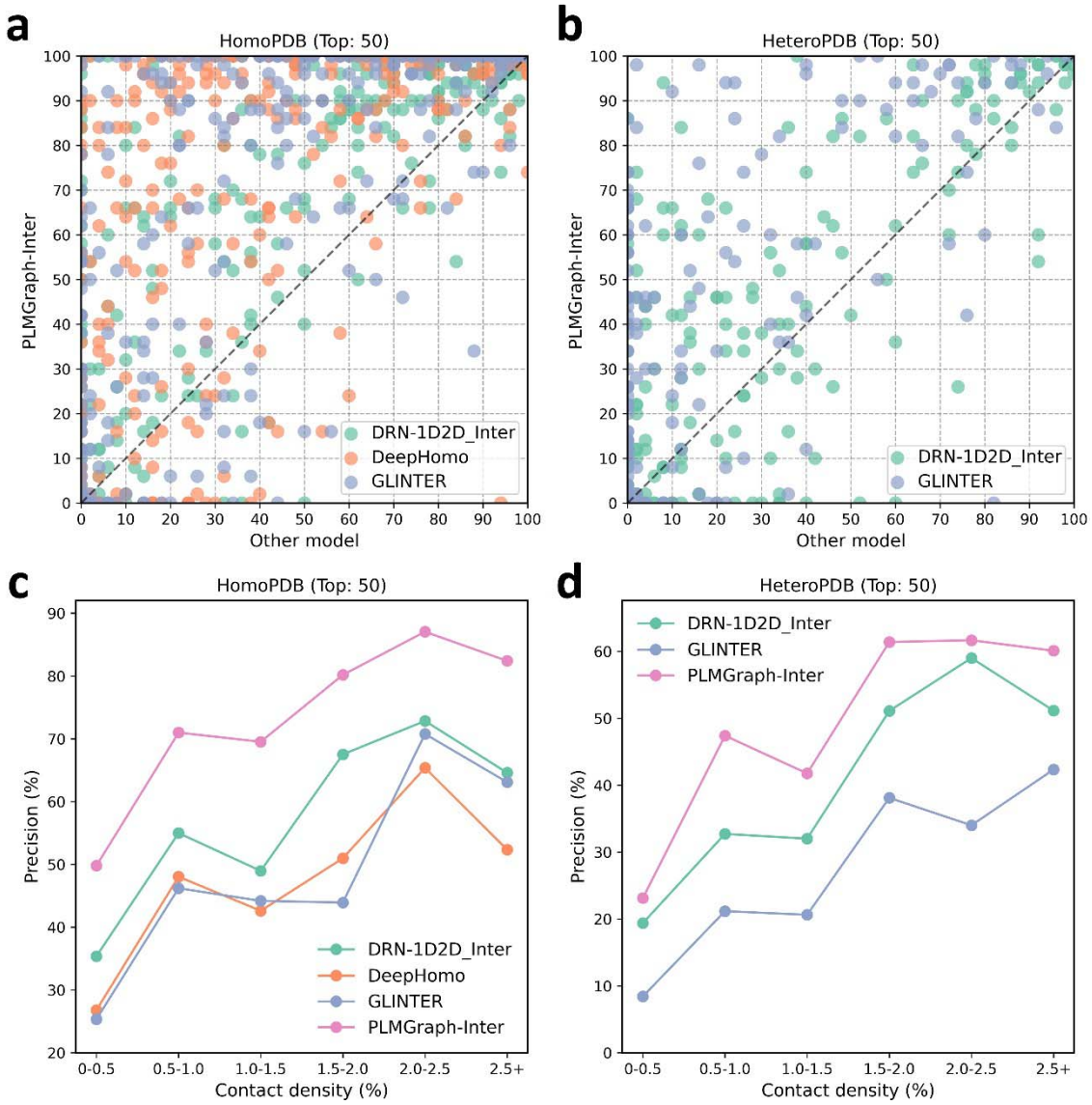



**Figure 2.** The performances of PLMGraph-Inter and other methods on the HomoPDB and HeteroPDB test sets. (a)~(b): The head-to-head comparison of the precisions (%) of the top 50 contacts predicted by PLMGraph-Inter and other methods for each target in (a) HomoPDB and (b) HeteroPDB. (c)~(d): The mean precisions of the top 50 contacts predicted by each method on PPIs within different intervals of contact densities in (c) HomoPDB and (d) HeteroPDB.

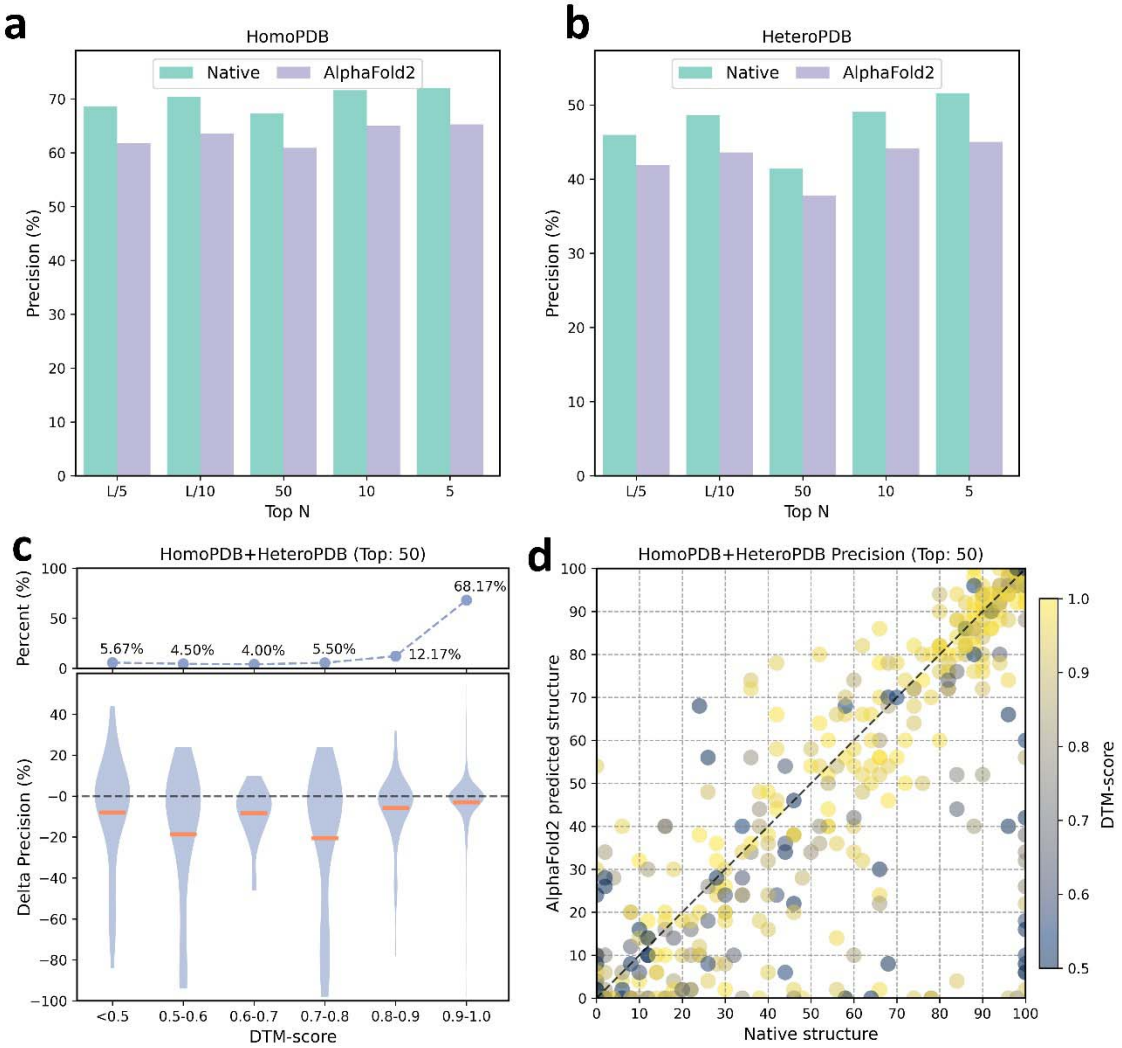



**Figure 3.** The performances of PLMGraph-Inter when using native and AlphaFold2 predicted structures as the input. (a)~(b): The performance comparison of PLMGraph-Inter when using native structures and AlphaFold2 predicted structures as the input on (a) HomoPDB and (b) HeteroPDB. (c) The performance gaps (measured as the difference of the mean precision of the top 50 predicted contacts) of PLMGraph-Inter with the application of AlphaFold2 predicted structures and native structures as the input when the PPIs are within different intervals of DTM-score. (d) The comparison of the precision of top 50 contacts predicted by PLMGraph-Inter for each target when using native structures and AlphaFold2 predicted structures as the input.

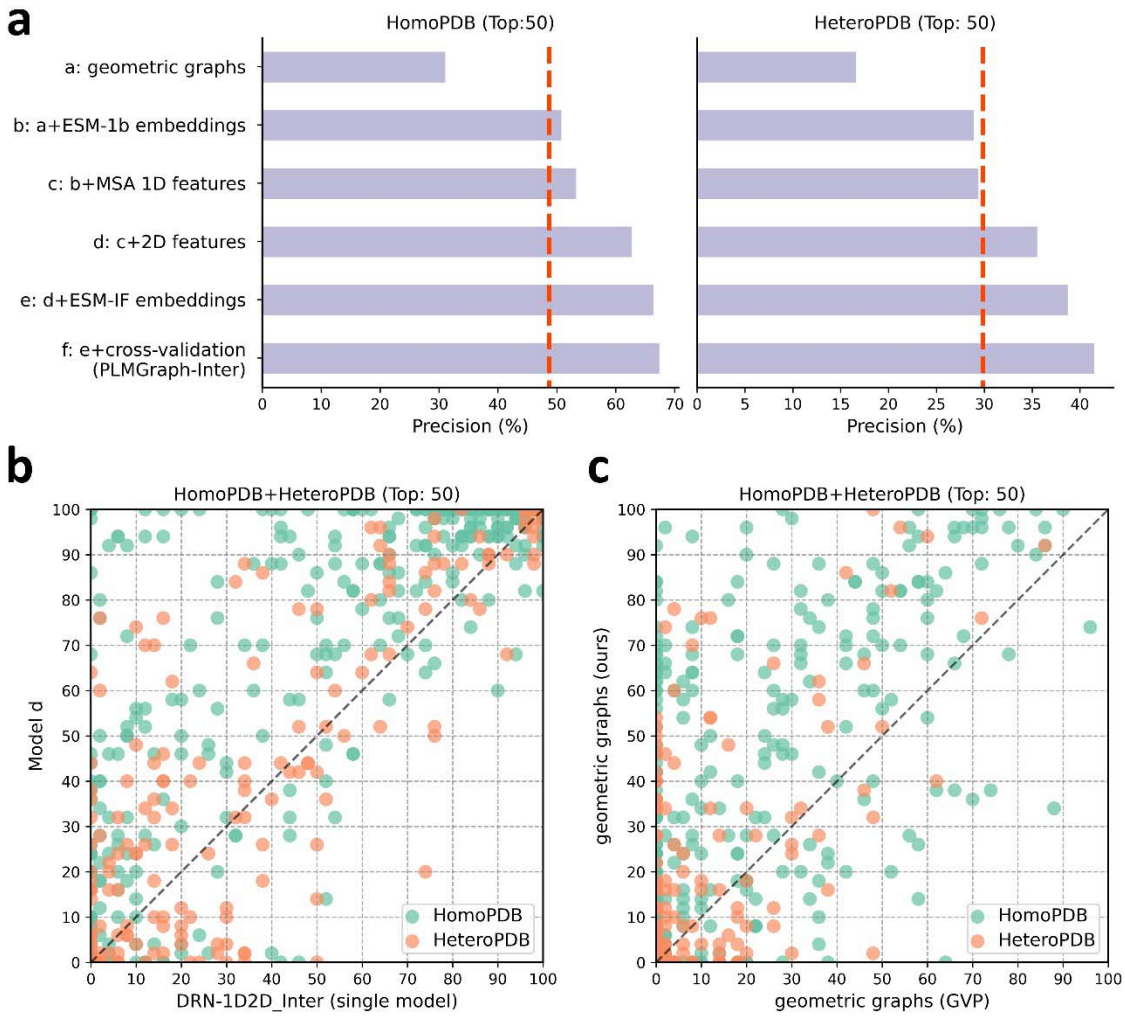



**Figure 4.** The ablation study of PLMGraph-Inter on the HomoPDB and HeteroPDB test sets. (a) The mean precisions of the top 50 contacts predicted by different ablation models on the HomoPDB and HeteroPDB test sets. (b) The head-to-head comparisons of mean precisions of the top 50 contacts predicted by model d and DRN-1D2D_Inter (single model) for each target in HomoPDB and HeteroPDB. (c) The head-to-head comparison of mean precisions of the top 50 contacts predicted by the model using our geometric graphs and the GVP geometric graphs.

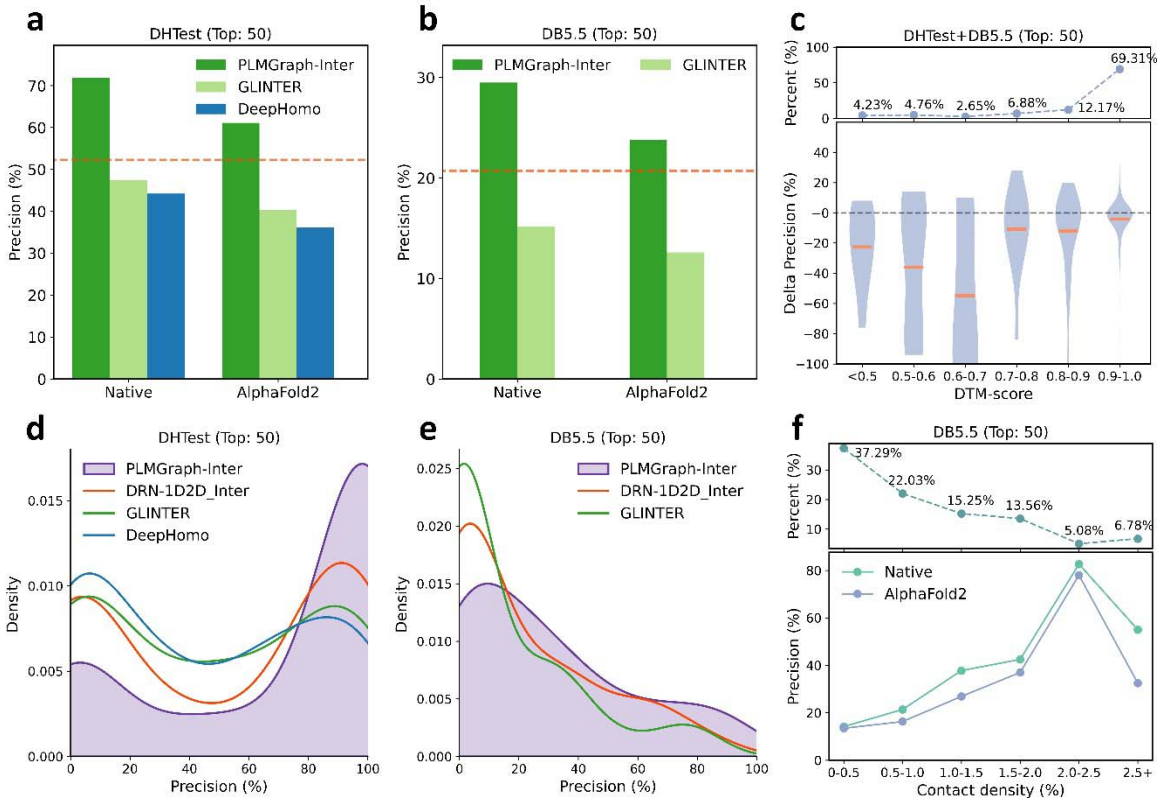



**Figure 5.** The performances of PLMGraph-Inter and other methods on the DHTest and DB5.5 test sets. (a)~(b): The mean precisions of top 50 contacts predicted by PLMGraph-Inter, GLINTER, and DeepHomo on (a) DHTest and (b) DB5.5 when using native structures and AlphaFold2 predicted structures as the input, where the orange lines indicate the performance of DRN-1D2D_Inter. (c) The performance gaps (measured as the difference of the mean precision of the top 50 predicted contacts) of PLMGraph-Inter with the application of AlphaFold2 predicted structures and native structures as the input when the PPIs are within different intervals of DTM-score. (d)~(e): The distributions of precisions of the top 50 contacts predicted by PLMGraph-Inter and other methods for PPIs in (d) DHTest and (e) DB5.5. (f) The mean precisions of the top 50 contacts predicted by PLMGraph-Inter on PPIs within different intervals of contact densities in DB5.5 .

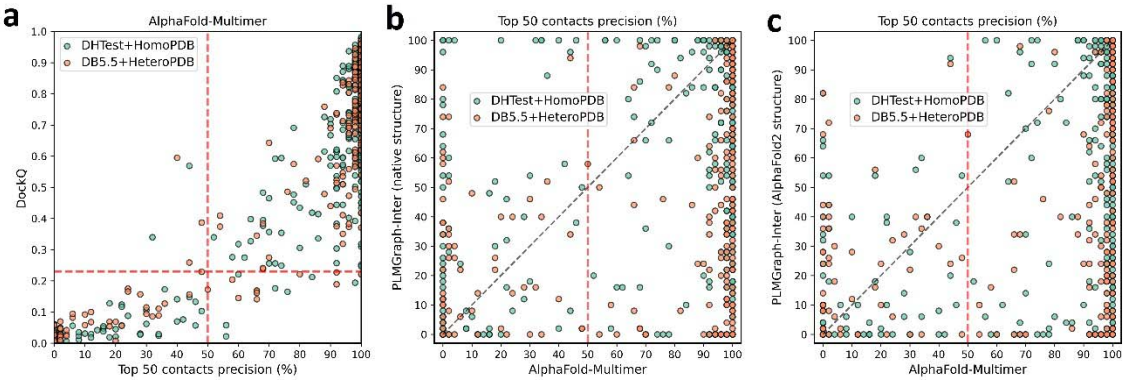



**Figure 6. The comparison of PLMGraph-Inter with AlphaFold-Multimer.** (a) The head-to-head comparison between the qualities of the protein complex structures generated by AlphaFold-Multimer (evaluated with DockQ) and the precision of the top 50 inter-protein contacts extracted from the generated protein complex structures. The red horizontal lines represent the threshold (DockQ=0.23) to determine whether the complex structure prediction is successful or not. (b)~(c) The head-to-head comparisons of precisions of the top 50 inter-protein contacts predicted by PLMGraph-Inter ((b) using AlphaFold2 predicted monomeric structures, (c) using native monomeric structures) and AlphaFold-Multimer for each target in the homomeric PPI and heteromeric PPI datasets.

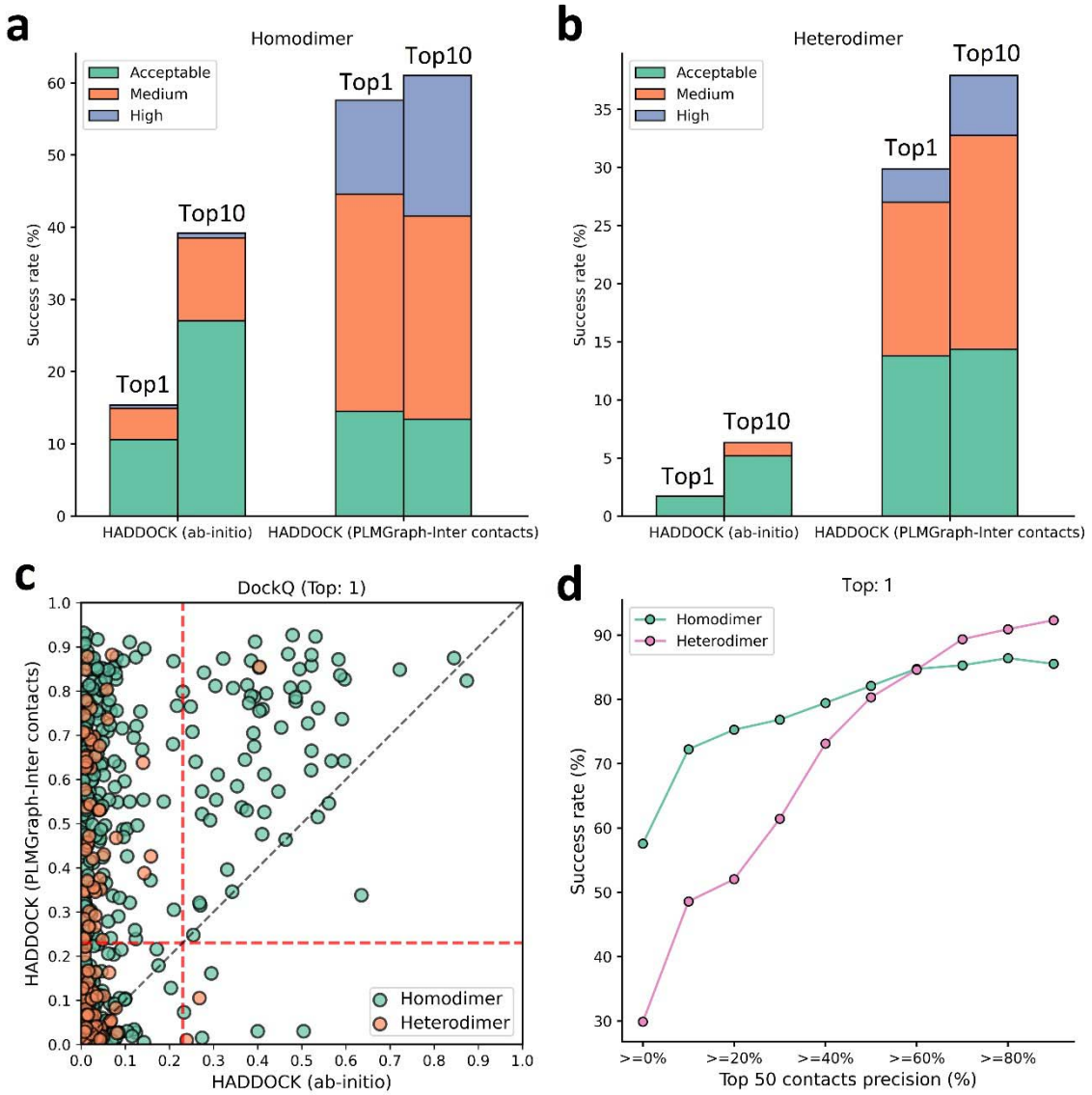



**Figure 7.** Protein-protein docking performances on the Homodimer and Heterodimer test sets. (a)~(b) The protein-protein docking performance comparison between HADDOCK with and without (ab-initio) using PLMGraph-Inter predicted contacts as restraints on (a) Homodimer and (b) Heterodimer. The left side of each column shows the performance when the top 1 predicted model for each PPI is considered, and the right side shows the performance when the top 10 predicted models for each PPI are considered. (c) The head-to-head comparison of qualities of the top 1 model predicted by HADDOCK with and without using PLMGraph-Inter predicted contacts as restraints for each target PPI. The red lines represent the threshold (DockQ=0.23) to determine whether the complex

structure prediction is successful or not. (d) The success rates (the top 1 model) for protein complex structure prediction when only including targets for which precisions of the predicted contacts are higher than certain thresholds.

**Table 1.** The performances of DeepHomo, GLINTER, DRN-1D2D_Inter and PLMGraph-Inter on the HomoPDB and HeteroPDB test sets using native structures.

| Methods | HomoPDB | | | | | HeteroPDB | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | L/5 | L/10 | 50 | 10 | 5 | L/5 | L/10 | 50 | 10 | 5 |
| DeepHomo | 43.2 | 46.7 | 42.4 | 48.5 | 49.9 | | | | | |
| GLINTER | 42.9 | 45.0 | 42.2 | 46.4 | 48.5 | 23.9 | 24.7 | 20.9 | 25.5 | 26.7 |
| DRN-1D2D_Inter | 52.5 | 55.2 | 51.3 | 56.6 | 57.6 | 34.9 | 37.1 | 32.6 | 38.1 | 38.5 |
| PLMGraph-Inter | **68.6** | **70.4** | **67.3** | **71.6** | **72.1** | **45.9** | **48.6** | **41.4** | **49.1** | **51.6** |

Note: The highest mean precision (%) in each column is highlighted in bold.

# Supplementary

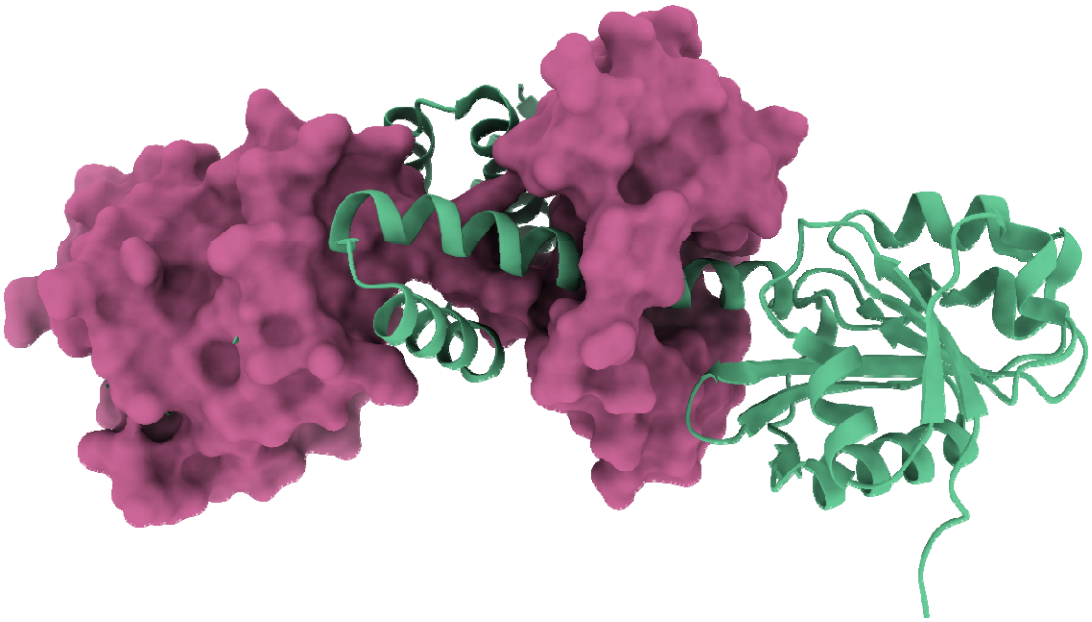


**Figure S1.** 3D structure of the homodimer (PDB: 3DFU).

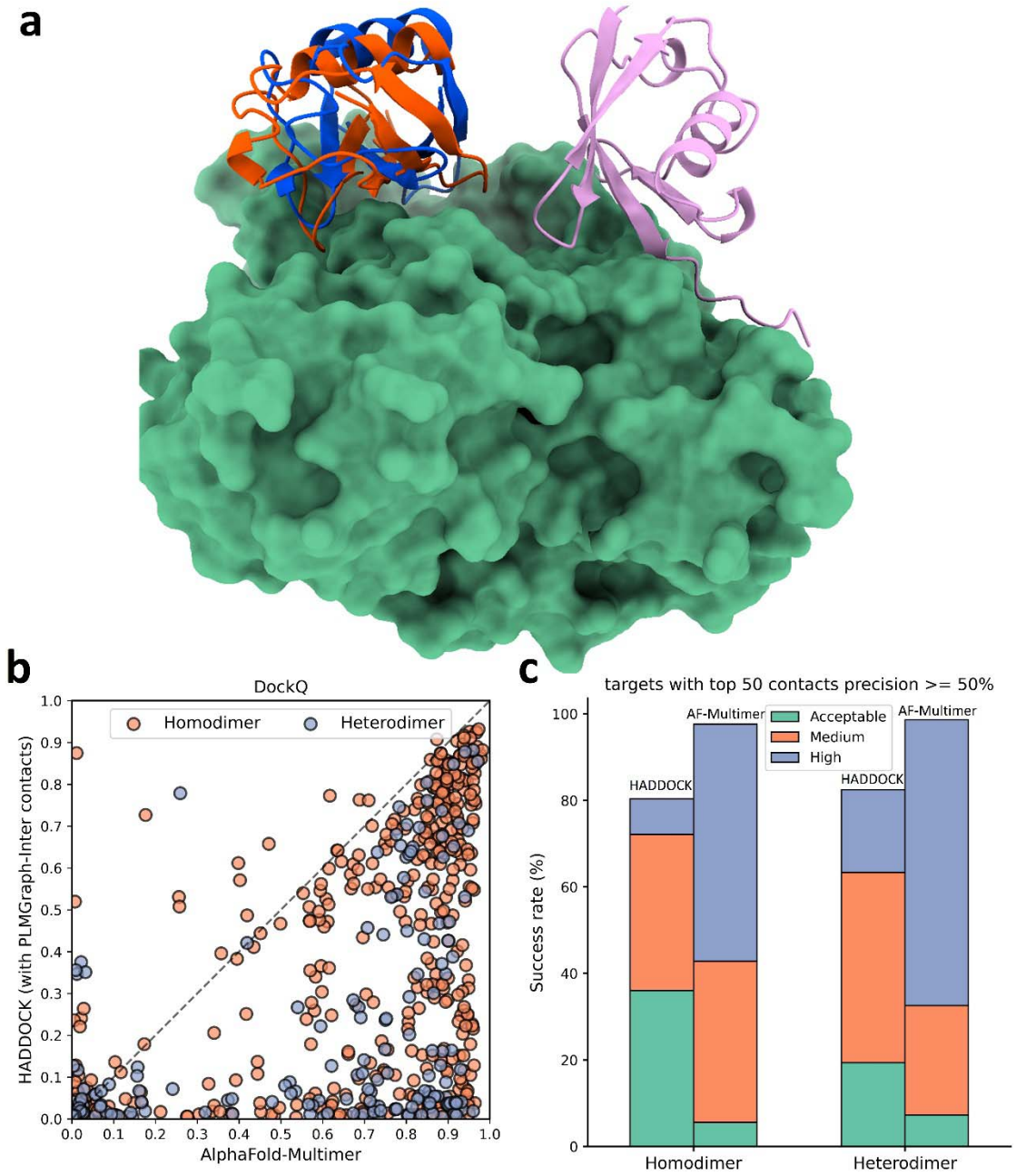



**Figure S2. The comparison of HADDOCK (with PLMGraph-Inter contact constraints) with AlphaFold-Multimer in protein complex structure prediction.** (a) The binding configurations predicted by HADDOCK (colored orange), predicted by AlphaFold-Multimer (colored pink) and the native binding configuration (colored blue) for the chain B of the protein complex structure in PDB 5HPS. The chain A is shown in

the protein surface mode (colored green) (b) The head-to-head comparison of the quality of the top 1 model predicted by HADDOCK(with PLMGraph-Inter contacts) and AlphaFold-Multimer for each target PPI. (c) The comparison of the performances of HADDOCK (with PLMGraph-Inter contact constraints) and AlphaFold-Multimer when the precision of the top 50 inter-protein contacts is higher than or equal to 50%.

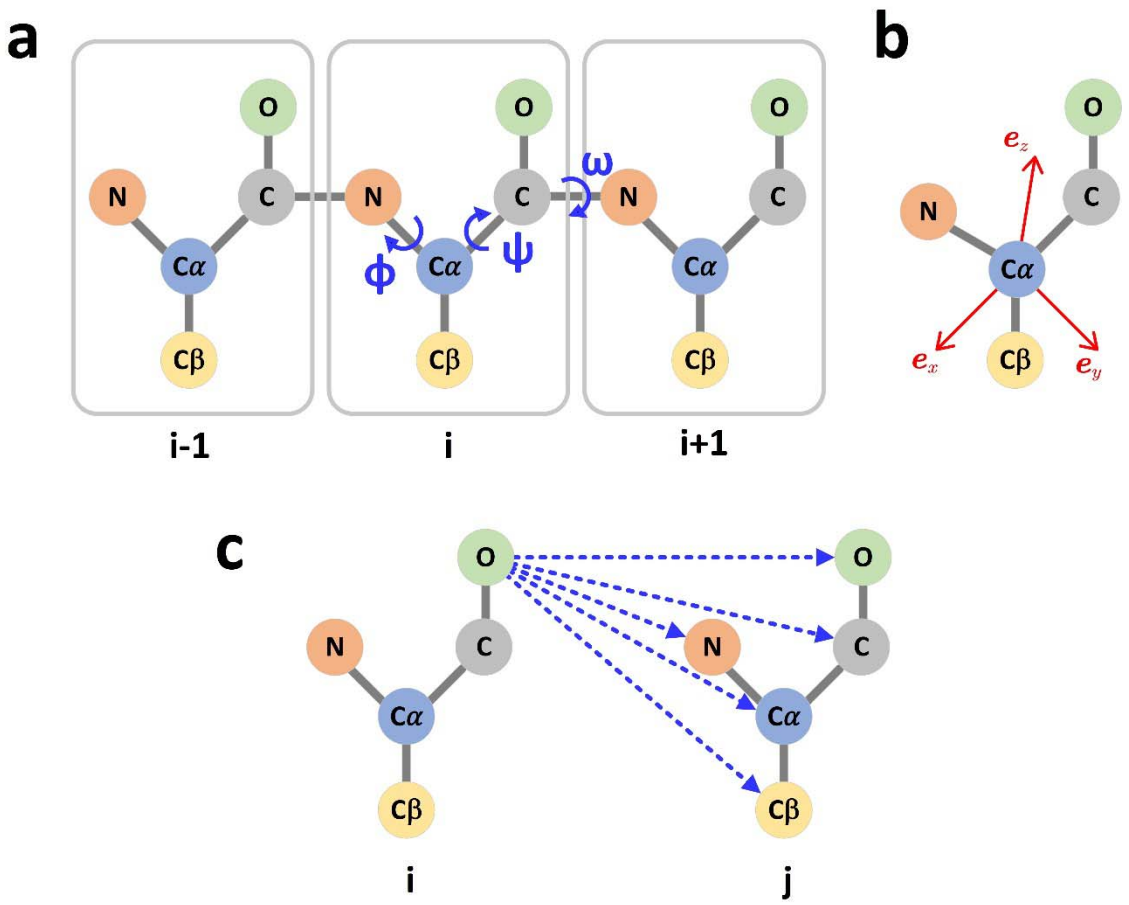



**Figure S3.** The graph representation of protein structures. (a) Dihedral angles of the protein backbone. (b) The local coordinate system of each amino acid. (c) The scalar (distances) and vector (directions) of the edge i->j.

**Table S1. The performances of DeepHomo, GLINTER, DRN-1D2D_Inter and PLMGraph-Inter on HomoPDB and HeteroPDB after the removal of targets which GLINTER failed to make the prediction**

| Methods | HomoPDB | | | | | HeteroPDB | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | L/5 | L/10 | 50 | 10 | 5 | L/5 | L/10 | 50 | 10 | 5 |
| DeepHomo | 41.1 | 44.3 | 40.6 | 46.0 | 46.9 | | | | | |
| GLINTER | 42.9 | 45.0 | 42.2 | 46.4 | 48.5 | 23.9 | 24.7 | 20.9 | 25.5 | 26.7 |
| DRN-1D2D_Inter | 50.6 | 53.0 | 49.5 | 53.9 | 55.2 | 33.6 | 35.6 | 31.5 | 36.2 | 37.2 |
| PLMGraph-Inter | **66.5** | **68.5** | **65.3** | **69.9** | **70.2** | **46.1** | **49.0** | **41.4** | **49.5** | **52.2** |

Note: The highest mean precision (%) in each column is highlighted in bold.

**Table S2. The performances of DeepHomo, GLINTER, DRN-1D2D_Inter and PLMGraph-Inter on HomoPDB and HeteroPDB using AlphaFold2 predicted structures**

| Methods | HomoPDB | | | | | HeteroPDB | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | L/5 | L/10 | 50 | 10 | 5 | L/5 | L/10 | 50 | 10 | 5 |
| DeepHomo | 39.3 | 42.7 | 38.8 | 44.8 | 46.2 | | | | | |
| GLINTER | 47.3 | 50.1 | 52.1 | 51.9 | 53.6 | 25.1 | 27.0 | 21.9 | 25.8 | 26.2 |
| DRN-1D2D_Inter | 52.5 | 55.2 | 51.3 | 56.6 | 57.6 | 34.9 | 37.1 | 32.6 | 38.1 | 38.5 |
| PLMGraph-Inter | **61.8** | **63.6** | **60.9** | **65.0** | **65.25** | **41.9** | **43.6** | **37.8** | **44.1** | **45.0** |

Note: The highest mean precision (%) in each column is highlighted in bold.

**Table S3. The performances of different ablation study models on the HomoPDB and HeteroPDB test sets**

| Methods | HomoPDB | | | | | HeteroPDB | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | L/5 | L/10 | 50 | 10 | 5 | L/5 | L/10 | 50 | 10 | 5 |
| Model a | 31.7 | 34.2 | 30.9 | 35.8 | 37.4 | 18.2 | 19.1 | 16.5 | 19.3 | 19.7 |
| Model b | 52.1 | 54.9 | 50.7 | 57.3 | 58.3 | 32.7 | 32.8 | 28.8 | 32.9 | 34.3 |
| Model c | 54.5 | 57.4 | 52.8 | 58.9 | 60.6 | 32.9 | 34.1 | 29.3 | 34.0 | 34.3 |
| Model d | 64.0 | 65.7 | 62.6 | 67.3 | 68.1 | 37.8 | 39.7 | 35.4 | 40.1 | 41.0 |
| Model e | 67.1 | 69.2 | 66.3 | 70.5 | 71.3 | 42.2 | 43.3 | 38.7 | 43.9 | 44.6 |
| Model f | 68.6 | 70.4 | 67.3 | 71.6 | 72.1 | 45.9 | 48.6 | 41.4 | 49.1 | 51.6 |

**Table S4. The performances of DeepHomo, GLINTER, DRN-1D2D_Inter and PLMGraph-Inter on the DHTest and DB5.5 test sets using native structures**

| Methods | DHTest | | | | | DB5.5 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | L/5 | L/10 | 50 | 10 | 5 | L/5 | L/10 | 50 | 10 | 5 |
| DeepHomo | 45.0 | 47.4 | 44.2 | 50.2 | 50.9 | | | | | |
| GLINTER | 48.5 | 50.8 | 47.4 | 51.2 | 52.3 | 18.8 | 20.5 | 15.2 | 21.9 | 20.7 |
| DRN-1D2D_Inter | 52.1 | 53.3 | 52.3 | 54.9 | 56.0 | 24.4 | 26.9 | 20.7 | 25.4 | 27.5 |
| PLMGraph-Inter | **72.0** | **73.3** | **71.9** | **74.7** | **75.4** | **33.6** | **36.4** | **29.5** | **36.9** | **40.0** |

Note: The highest mean precision (%) in each column is highlighted in bold.

**Table S5. The performances of DeepHomo, GLINTER, DRN-1D2D_Inter and PLMGraph-Inter on DHTest and DB5.5 after the removal of targets which GLINTER failed to make the prediction**

| Methods | DHTest | | | | | DB5.5 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | L/5 | L/10 | 50 | 10 | 5 | L/5 | L/10 | 50 | 10 | 5 |
| DeepHomo | 41.6 | 43.9 | 41.3 | 47.6 | 48.6 | | | | | |
| GLINTER | 48.5 | 50.8 | 47.4 | 51.2 | 52.3 | 18.8 | 20.5 | 15.2 | 21.9 | 20.7 |
| DRN-1D2D_Inter | 45.5 | 46.9 | 45.5 | 48.9 | 49.9 | 23.4 | 25.3 | 19.9 | 24.6 | 25.7 |
| PLMGraph-Inter | **64.7** | **66.2** | **65.3** | **68.3** | **69.2** | **32.4** | **34.4** | **28.5** | **35.5** | **38.6** |

Note: The highest mean precision (%) in each column is highlighted in bold.

**Table S6. The performances of DeepHomo, GLINTER, DRN-1D2D_Inter and PLMGraph-Inter on the DHTest and DB5.5 test sets using AlphaFold2 predicted structures**

| Methods | DHTest | | | | | DB5.5 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | L/5 | L/10 | 50 | 10 | 5 | L/5 | L/10 | 50 | 10 | 5 |
| DeepHomo | 36.5 | 38.2 | 36.1 | 40.4 | 42.9 | | | | | |
| GLINTER | 41.6 | 43.4 | 40.3 | 46.1 | 46.2 | 15.5 | 14.0 | 12.6 | 15.4 | 16.4 |
| DRN-1D2D_Inter | 52.1 | 53.3 | 52.3 | 54.9 | 56.0 | 24.4 | 26.9 | 20.7 | 25.4 | 27.5 |
| PLMGraph-Inter | **61.7** | **62.6** | **61.1** | **63.4** | **63.5** | **28.6** | **31.4** | **23.8** | **30.3** | **31.5** |

Note: The highest mean precision (%) in each column is highlighted in bold.

**Table S7. The percent of PPIs for which the precision of top 50 contacts predicted by PLMGraph-Inter is higher than the precision of top 50 contacts predicted by AlphaFold-Multimer on the PPIs which the precision of top 50 contacts predicted by AlphaFold-Multimer is lower than 50%**

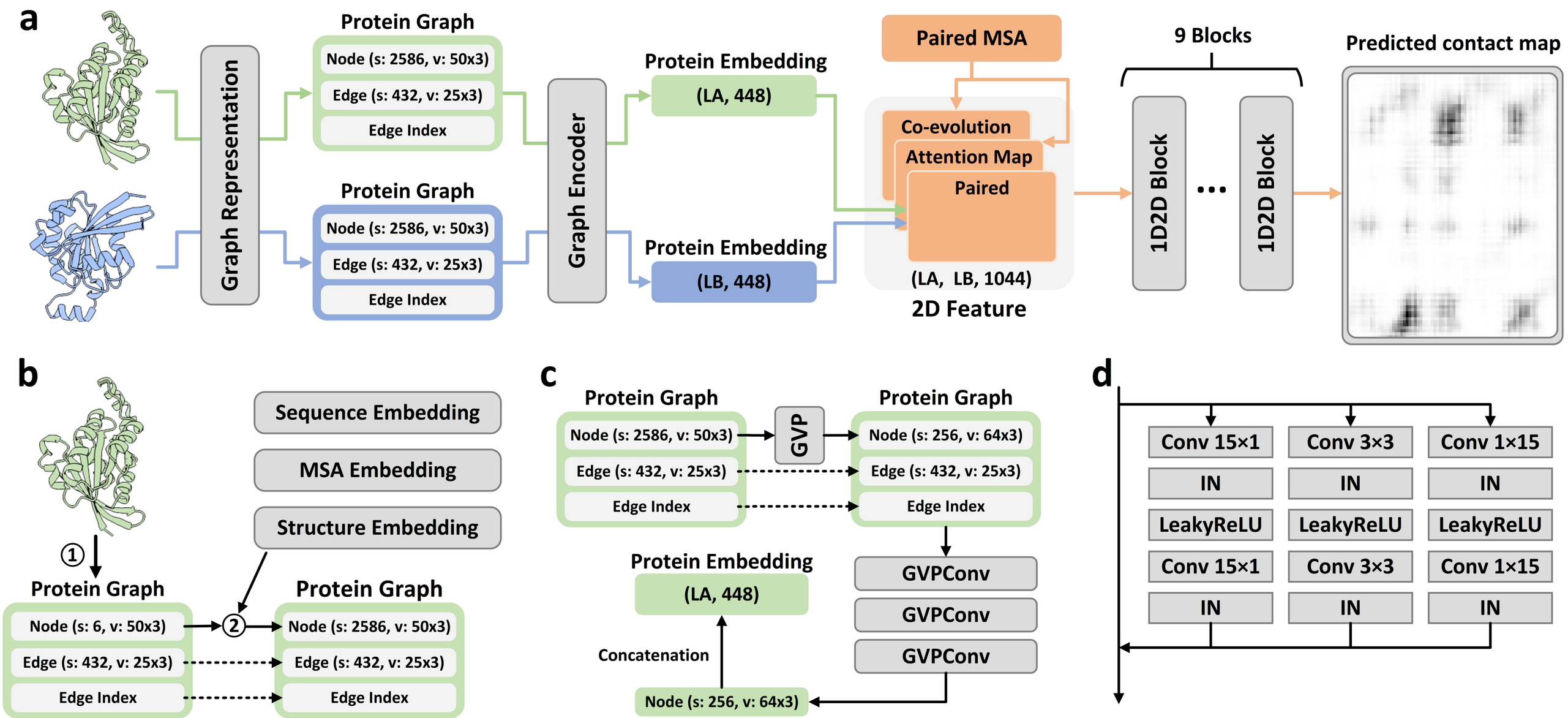
| | Using native monomeric structure | Using AlphaFold2 predicted structure |
|---|---|---|
| DHTest + HomoPDB | 48.3% | 30.8% |
| DB5.5 + HeteroPDB | 48.1% | 51.9% |

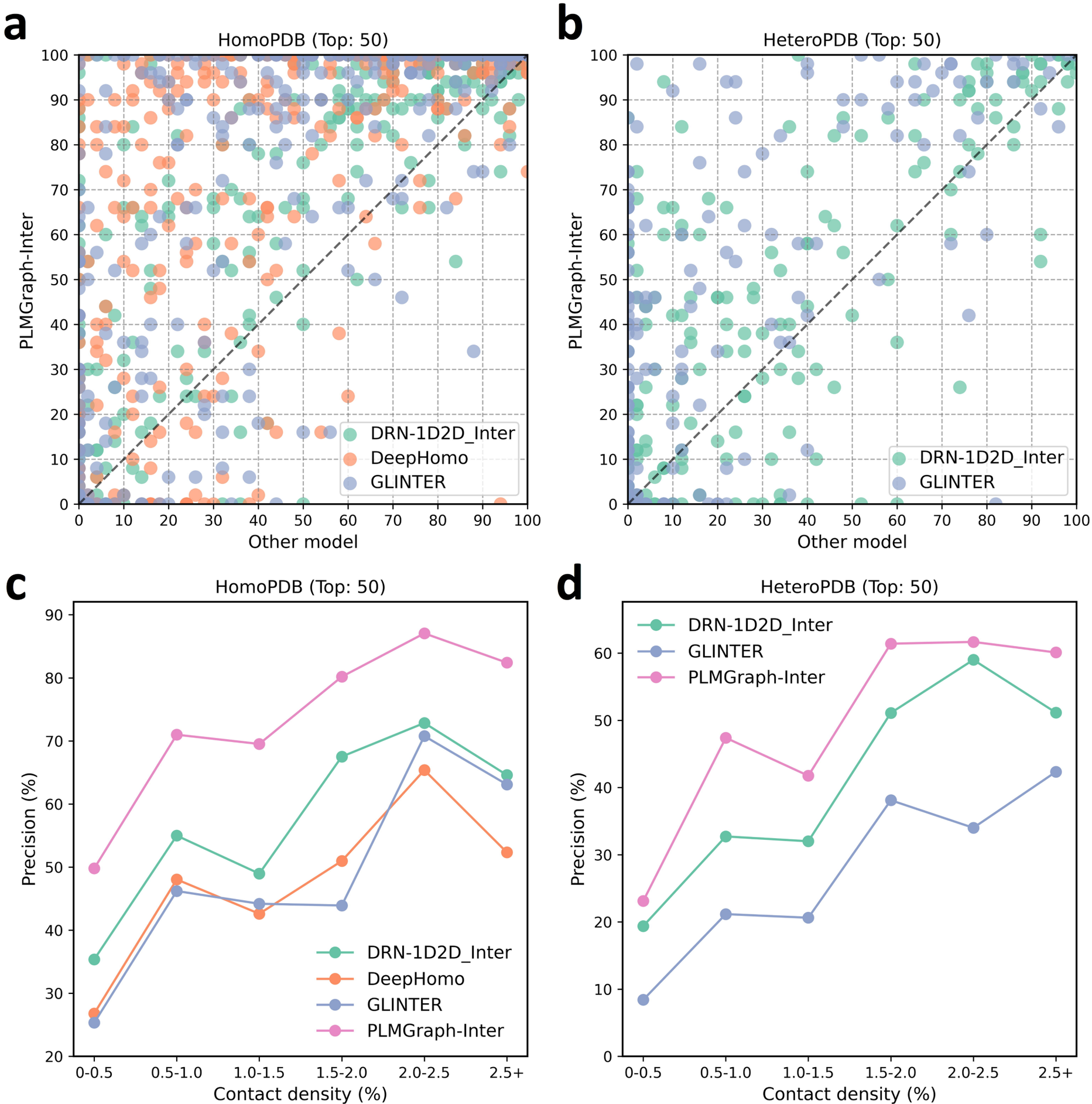
**Table S8. The performances of protein-protein docking by HADDOCK with application of contact constraints predicted by different methods when only the top 1 model generated by HADDOCK for each PPI is considered**

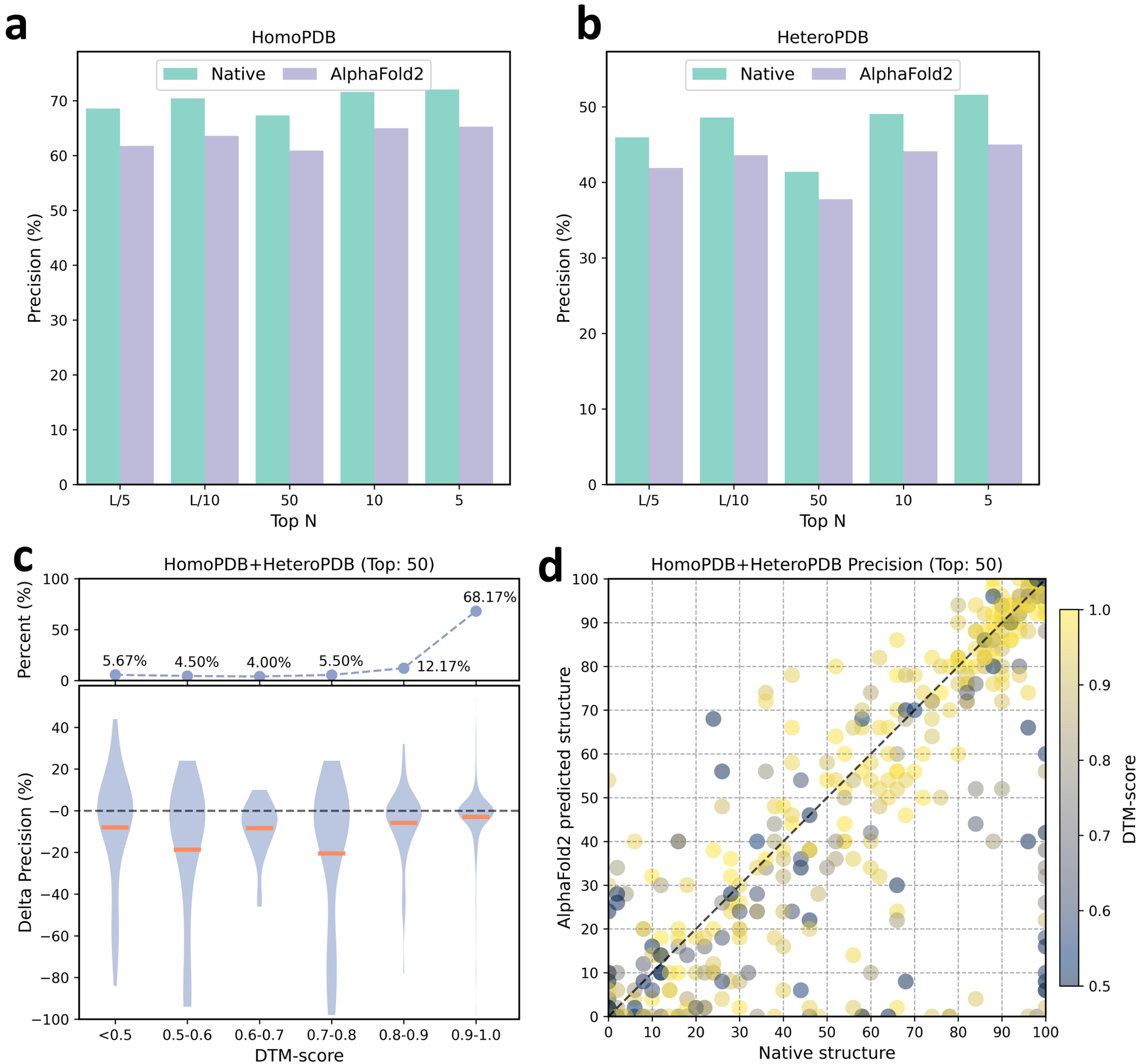
| Dataset | DockQ | PLMGraph-Inter | GLINTER | DeepHomo | DRN-1D2D_Inter |
|---|---|---|---|---|---|
| Homodimer | Acceptable | 14.50 | 11.57 | 7.58 | 11.69 |
| | Medium | 30.09 | 24.89 | 24.03 | 24.89 |
| | High | 12.99 | 11.35 | 8.44 | 13.42 |
| | All | 57.58 | 47.82 | 40.04 | 50.00 |
| | | | | | |
| Heterodimer | Acceptable | 13.79 | 6.94 | | 7.47 |
| | Medium | 13.22 | 7.51 | | 13.22 |
| | High | 2.87 | 1.16 | | 2.30 |
| | All | 29.89 | 15.61 | | 22.99 |

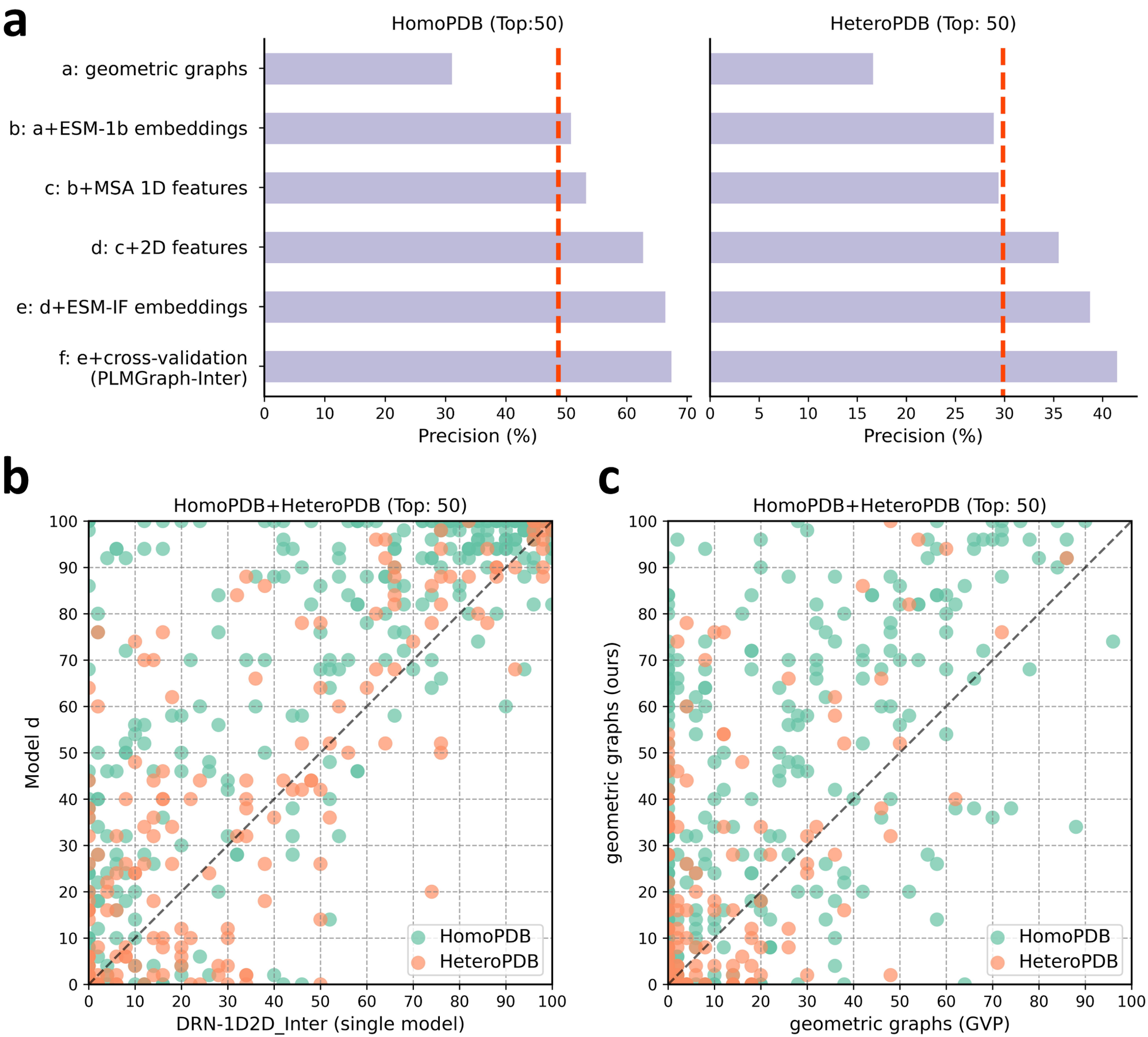
**Table S9. The performances of protein-protein docking by HADDOCK with application of contact constraints predicted by different methods when only the top 10 models generated by HADDOCK for each PPI are considered**

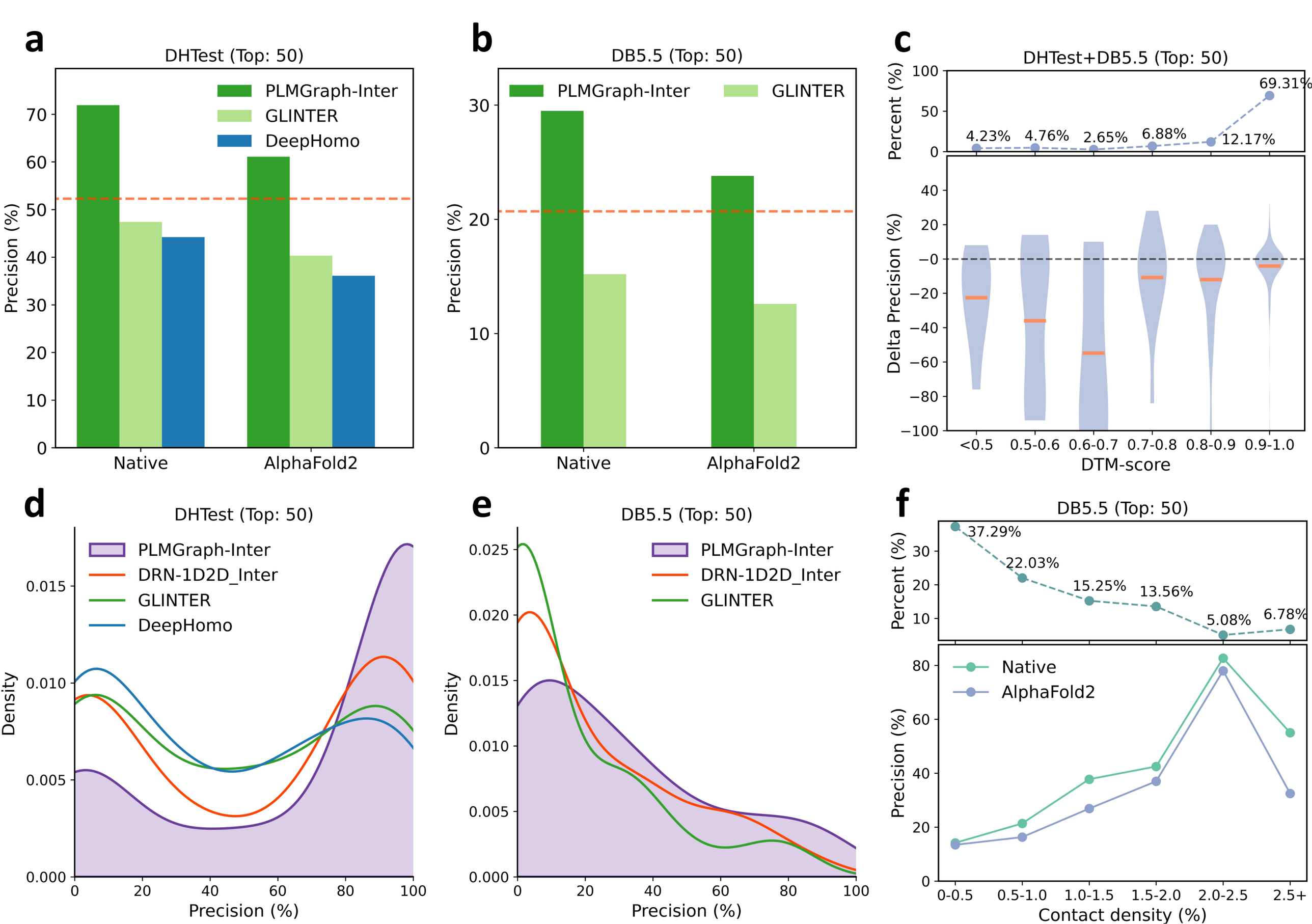
| Dataset | DockQ | PLMGraph-Inter | GLINTER | DeepHomo | DRN-1D2D_Inter |
|---|---|---|---|---|---|
| Homodimer | Acceptable | 13.42 | 11.35 | 8.44 | 11.69 |
| | Medium | 28.14 | 23.80 | 22.94 | 24.68 |
| | High | 19.48 | 15.50 | 12.99 | 17.97 |
| | All | 61.04 | 50.66 | 44.37 | 54.33 |
| | | | | | |
| Heterodimer | Acceptable | 14.37 | 7.51 | | 9.77 |
| | Medium | 18.39 | 8.09 | | 12.07 |
| | High | 5.17 | 1.73 | | 4.60 |
| | All | 37.93 | 17.34 | | 26.44 |

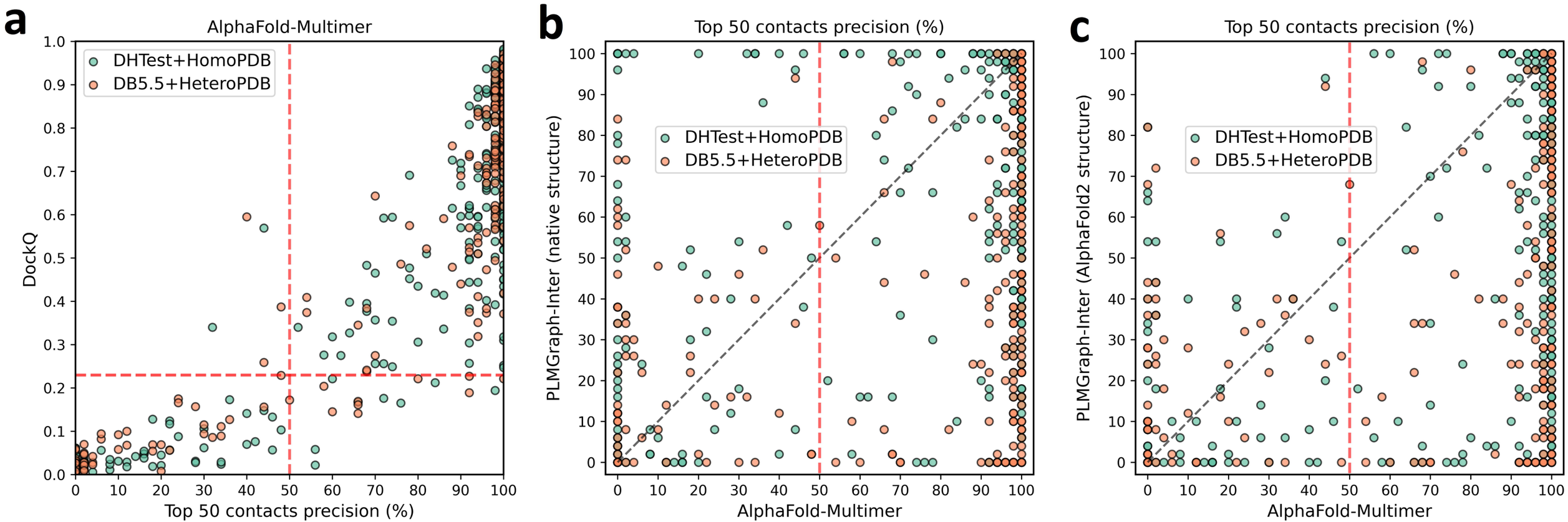

a
Graph Representation
Protein Graph
Node (s: 2586, v: 50x3)
Edge (s: 432, v: 25x3)
Edge Index
Protein Graph
Node (s: 2586, v: 50x3)
Edge (s: 432, v: 25x3)
Edge Index
Graph Encoder
Protein Embedding
(LA, 448)
Protein Embedding
(LB, 448)
Paired MSA
Co-evolution
Attention Map
Paired
(LA, LB, 1044)
2D Feature
9 Blocks
1D2D Block
1D2D Block
Predicted contact map
b
Sequence Embedding
MSA Embedding
Structure Embedding
1
Protein Graph
Node (s: 6, v: 50x3)
Edge (s: 432, v: 25x3)
Edge Index
2
Protein Graph
Node (s: 2586, v: 50x3)
Edge (s: 432, v: 25x3)
Edge Index
c
Protein Graph
Node (s: 2586, v: 50x3)
Edge (s: 432, v: 25x3)
Edge Index
GVP
Protein Graph
Node (s: 256, v: 64x3)
Edge (s: 432, v: 25x3)
Edge Index
GVPConv
GVPConv
GVPConv
Protein Embedding
(LA, 448)
Concatenation
Node (s: 256, v: 64x3)
d
Conv 15×1
IN
LeakyReLU
Conv 15×1
IN
Conv 3×3
IN
LeakyReLU
Conv 3×3
IN
Conv 1×15
IN
LeakyReLU
Conv 1×15
IN

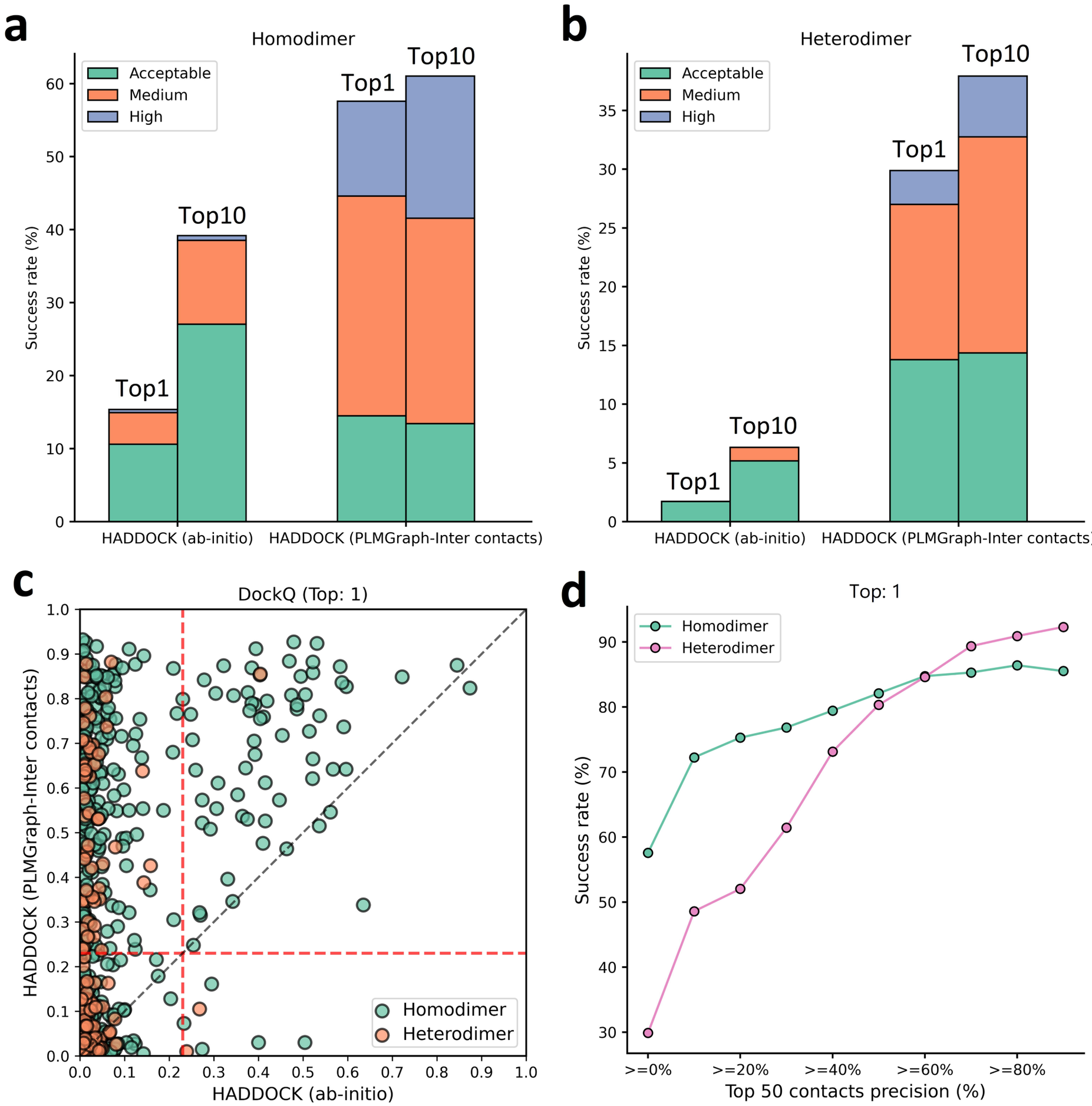

a
HomoPDB (Top: 50)
PLMGraph-Inter
Other model
DRN-1D2D_Inter
DeepHomo
GLINTER
b
HeteroPDB (Top: 50)
PLMGraph-Inter
Other model
DRN-1D2D_Inter
GLINTER
c
HomoPDB (Top: 50)
Precision (%)
Contact density (%)
0-0.5
0.5-1.0
1.0-1.5
1.5-2.0
2.0-2.5
2.5+
DRN-1D2D_Inter
DeepHomo
GLINTER
PLMGraph-Inter
d
HeteroPDB (Top: 50)
Precision (%)
Contact density (%)
DRN-1D2D_Inter
GLINTER
PLMGraph-Inter


a
HomoPDB
Native
AlphaFold2
Precision (%)
Top N
L/5
L/10
50
10
5
b
HeteroPDB
Native
AlphaFold2
Precision (%)
Top N
c
HomoPDB+HeteroPDB (Top: 50)
Percent (%)
5.67%
4.50%
4.00%
5.50%
12.17%
68.17%
Delta Precision (%)
<0.5
0.5-0.6
0.6-0.7
0.7-0.8
0.8-0.9
0.9-1.0
DTM-score
d
HomoPDB+HeteroPDB Precision (Top: 50)
AlphaFold2 predicted structure
Native structure
DTM-score


a
HomoPDB (Top:50)
HeteroPDB (Top: 50)
a: geometric graphs
b: a+ESM-1b embeddings
c: b+MSA 1D features
d: c+2D features
e: d+ESM-IF embeddings
f: e+cross-validation (PLMGraph-Inter)
Precision (%)
Precision (%)
b
HomoPDB+HeteroPDB (Top: 50)
Model d
DRN-1D2D_Inter (single model)
HomoPDB
HeteroPDB
c
HomoPDB+HeteroPDB (Top: 50)
geometric graphs (ours)
geometric graphs (GVP)
HomoPDB
HeteroPDB


a
DHTest (Top: 50)
PLMGraph-Inter
GLINTER
DeepHomo
Precision (%)
Native
AlphaFold2
b
DB5.5 (Top: 50)
PLMGraph-Inter
GLINTER
Precision (%)
Native
AlphaFold2
c
DHTest+DB5.5 (Top: 50)
Percent (%)
4.23%
4.76%
2.65%
6.88%
12.17%
69.31%
Delta Precision (%)
<0.5
0.5-0.6
0.6-0.7
0.7-0.8
0.8-0.9
0.9-1.0
DTM-score
d
DHTest (Top: 50)
PLMGraph-Inter
DRN-1D2D_Inter
GLINTER
DeepHomo
Density
Precision (%)
e
DB5.5 (Top: 50)
PLMGraph-Inter
DRN-1D2D_Inter
GLINTER
Density
Precision (%)
f
DB5.5 (Top: 50)
Percent (%)
37.29%
22.03%
15.25%
13.56%
5.08%
6.78%
Native
AlphaFold2
Precision (%)
0-0.5
0.5-1.0
1.0-1.5
1.5-2.0
2.0-2.5
2.5+
Contact density (%)


a
AlphaFold-Multimer
DHTest+HomoPDB
DB5.5+HeteroPDB
DockQ
Top 50 contacts precision (%)
b
Top 50 contacts precision (%)
DHTest+HomoPDB
DB5.5+HeteroPDB
PLMGraph-Inter (native structure)
AlphaFold-Multimer
c
Top 50 contacts precision (%)
DHTest+HomoPDB
DB5.5+HeteroPDB
PLMGraph-Inter (AlphaFold2 structure)
AlphaFold-Multimer


a
Homodimer
Acceptable
Medium
High
Top1
Top10
Top1
Top10
Success rate (%)
HADDOCK (ab-initio)
HADDOCK (PLMGraph-Inter contacts)
b
Heterodimer
Acceptable
Medium
High
Top1
Top10
Top1
Top10
Success rate (%)
HADDOCK (ab-initio)
HADDOCK (PLMGraph-Inter contacts)
c
DockQ (Top: 1)
HADDOCK (PLMGraph-Inter contacts)
HADDOCK (ab-initio)
Homodimer
Heterodimer
d
Top: 1
Homodimer
Heterodimer
Success rate (%)
>=0%
>=20%
>=40%
>=60%
>=80%
Top 50 contacts precision (%)