nature machine intelligence

Article

https://doi.org/10.1038/s42256-023-00741-2

Protein–protein contact prediction by geometric triangle-aware protein language models

Received: 19 January 2023

Accepted: 19 September 2023

Published online: 19 October 2023

Check for updates

Peicong Lin, Huanyu Tao, Hao Li & Sheng-You Huang

Information regarding the residue–residue distance between interacting proteins is important for modelling the structures of protein complexes, as well as being valuable for understanding the molecular mechanism of protein–protein interactions. With the advent of deep learning, many methods have been developed to accurately predict the intra-protein residue–residue contacts of monomers. However, it is still challenging to accurately predict inter-protein residue–residue contacts for protein complexes, especially hetero-protein complexes. Here we develop a protein language model-based deep learning method to predict the inter-protein residue–residue contacts of protein complexes—named DeepInter—by introducing a triangle-aware mechanism of triangle update and triangle self-attention into the deep neural network. We extensively validate DeepInter on diverse test sets of 300 homodimeric, 28 CASP-CAPRI homodimeric and 99 heterodimeric complexes and compare it with state-of-the-art methods including CDPred, DeepHomo2.0, GLINTER and DeepHomo. The results demonstrate the accuracy and robustness of DeepInter.

Proteins conduct their functions by interacting with other molecules or assembling to form symmetric homo-oligomers[1]. Elucidation of the interface of complex structures is a fundamental step towards understanding their biological function[2]. However, the prediction of quaternary protein structures has remained a long-standing challenge. Given the low throughput and high cost of experimental structure determination, computational methods have become a valuable way to predict the protein structures of monomers and complexes. Recently, AlphaFold2[3] was proposed as an end-to-end deep learning architecture to directly predict the structures of monomers. It achieved a remarkable performance in the 14th Critical Assessment of Protein Structure Prediction (CASP14) experiment[4]. With the great progress provided by AlphaFold2, some studies have adopted a polyglycine to link the two chains of a complex[5] or modify the relative position of the two chains to predict the structures of a complex[6,7]. Furthermore, DeepMind extended AlphaFold2 to AlphaFold-Multimer by training with multi-chain proteins[8]. Although the accuracy of structures predicted by AlphaFold2 is competitive with that of experiments in many monomer cases, it is far from achieving comparable accuracy in the structure prediction of protein complexes. Accordingly, there remains a great need to develop further computational methods for predicting protein–protein interactions.

Inter-protein residue–residue contact prediction plays an important role in protein complex structure prediction that requires critical interfacial information. The DeepHomo method converts predicted inter-protein contacts as the restraints of an energy function to filter putative binding modes[9–12]. DeepComplex, meanwhile, applies a gradient descent optimization algorithm to model quarternary structures by utilizing the predicted inter-protein contacts as distance restraints[13]. Given the importance of interfacial interactions, various deep learning-based methods have been developed to predict inter-protein contacts[14–20] and protein–protein interactions[21–24]. Compared with the intra-protein

School of Physics and Key Laboratory of Molecular Biophysics of MOE, Huazhong University of Science and Technology, Wuhan, Hubei, China.
e-mail: huangsy@hust.edu.cn

https://doi.org/10.1038/s42256-023-00741-2

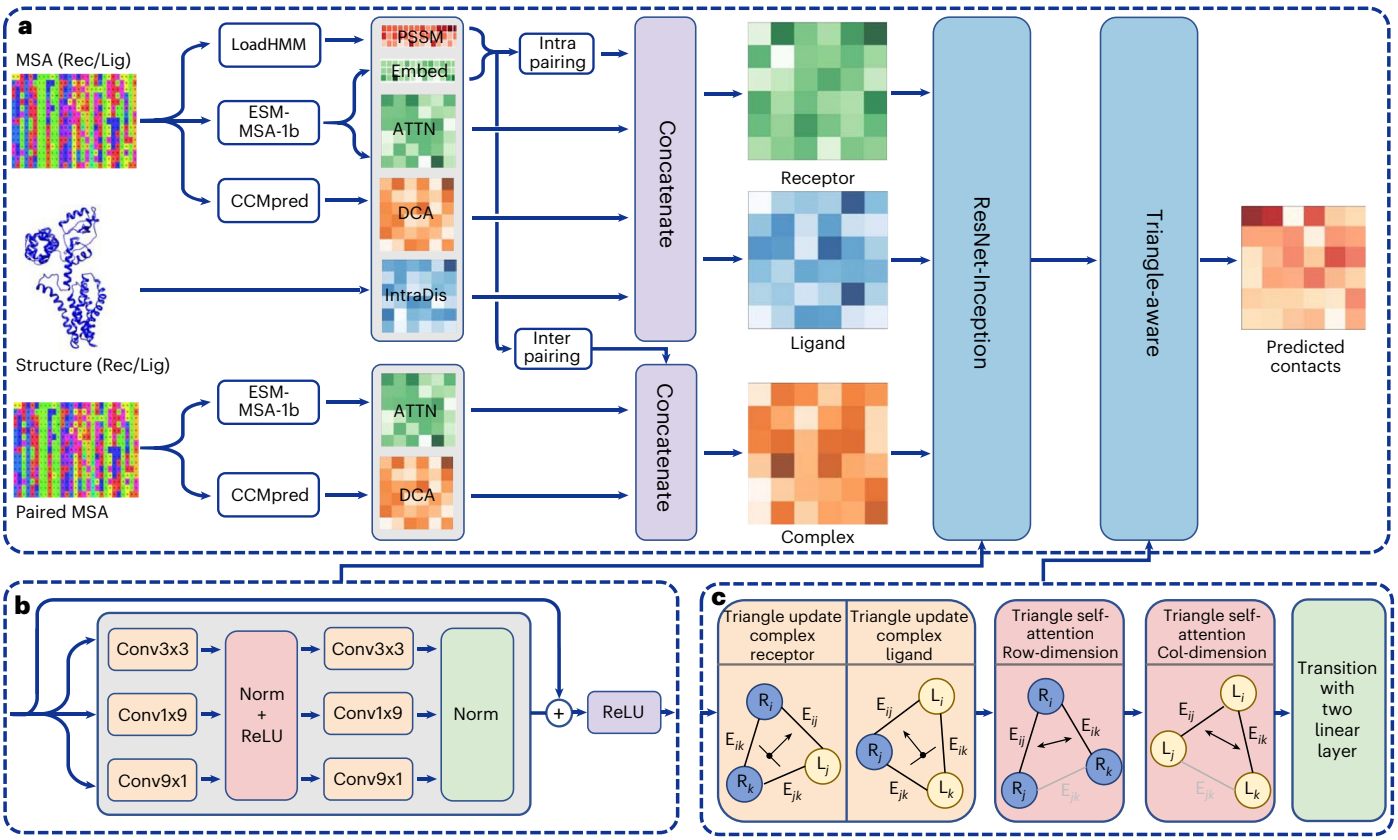

Fig. 1 | The framework of DeepInter. a, Overview of DeepInter. The inputs mainly contain the receptor, ligand and complex features. The features of the receptor and ligand include the PSSM, ESM-MSA-1b features, DCA features and intra-protein distances. The features of the complex exclude the intra-protein distance. **b**,**c**, DeepInter includes two main modules: the ResNet-Inception module (**b**) and the triangle-aware module (**c**). Here, the R, L and E stand for the receptor, ligand and the edge between their residues, respectively.

contact predictions utilizing co-evolution data based on direct coupling analysis (DCA) from multiple sequence alignment (MSA)[25–33], advanced inter-protein contact predictions make use of the advantages of protein language models to capture the inter-protein interaction and improve the performance. For example, DeepHomo2.0[34], GLINTER[35] and CDPred[36] apply distributed vector representations of the sequence and the multi-head attention matrix from the unsupervised learning-based protein language model, and improve the ability to predict inter-protein contacts. However, the performances of these methods depend on the homologies available from the MSA. For inter-protein contact prediction, it is necessary to conduct paired MSA (interlogs)[37–39]—this is a major bottleneck and challenge. Although there are several approaches for pairing the MSAs from different monomers of a complex, including genome-based[40], phylogeny-based[41] and block-diagonal-based methods[6], it is still difficult to accurately predict interfacial contacts.

Current methods for inter-protein contact prediction have several limitations. First, the existing mainstream frameworks of inter-protein contact predictions are mainly based on a residual convolutional network, which can only capture local features. Second, the predicted contact maps of those methods have a large geometric inconsistency that violates the triangle inequality. Third, some methods directly utilize the attention mechanism on the two-dimensional (2D) matrix of features, which only considers the interaction for each pair of residues. To overcome these limitations, we have developed a deep learning-based method for inter-protein contact predictions of protein complexes, named DeepInter, by applying the hidden features generated from a pre-trained protein language model and utilizing a triangle-aware module.

Compared with existing methods like DeepHomo2.0, there are three further main contributions from DeepInter. First, DeepInter introduces a ResNet-Inception module in the network to process the intra-protein features; this can efficiently capture the long-range interaction between pairs of residues by increasing the effective receptive field. Second, DeepInter applies a geometric triangle-aware module to update the inter-protein features by utilizing an attention mechanism on the pair representations of three residues that satisfy geometric consistency; this is able to consider the many-body effect in residue–residue interactions. Third, DeepInter designs a new network architecture to simultaneously consider intra-protein and inter-protein features through an attention mechanism[42], which can predict the inter-chain contacts for both homodimers and heterodimers. As such, DeepInter provides a substantial performance improvement over existing approaches on diverse test sets of both homodimers and heterodimers.

Results

Overview of DeepInter

Figure 1 presents an overview of DeepInter. It has three main components: preprocessing of the input features, the ResNet-Inception module and the triangle-aware module. The inputs of DeepInter are the two monomer structures and the corresponding MSAs (Fig. 1a). Coevolutionary conservation (position-specific scoring matrices, PSSM) and co-evolution information from DCA are generated from the MSA of the monomer sequence. The paired MSAs are generated from the MSAs of two monomers using the organism information. The sequence representation and the multi-head attention matrix are then

https://doi.org/10.1038/s42256-023-00741-2

Table 1 | Comparison of the precisions (in %) of DeepInter, DeepHomo2.0, CDPred, GLINTER and DeepHomo on the Homodimer300 test set with the input of experimental (main entries) and AlphaFold2-predicted (in parentheses) monomer structures when the top 1, 10, 25, 50, *L*/10, *L*/5 and *L* predicted contacts are considered

Method	Top 1	Top 10	Top 25	Top 50	Top *L*/10	Top *L*/5	Top *L*
DeepInter	**80.0 (69.7)**	**78.3 (67.1)**	**77.5 (66.9)**	**76.6 (65.8)**	**77.6 (66.7)**	**76.7 (66.0)**	**71.0 (59.0)**
CDPred	74.0 (68.3)	71.8 (67.1)	69.8 (66.2)	67.9 (64.2)	69.6 (66.4)	67.8 (64.4)	58.3 (54.4)
DeepHomo2.0	73.7 (62.7)	71.4 (62.2)	69.3 (60.5)	66.8 (58.2)	69.4 (60.8)	67.0 (58.3)	54.7 (47.0)
GLINTER	68.3 (60.7)	63.8 (56.7)	59.8 (54.2)	56.1 (51.1)	60.7 (54.6)	56.9 (51.6)	43.0 (39.1)
DeepHomo	60.7 (55.0)	56.8 (49.2)	53.4 (46.1)	49.8 (43.8)	54.2 (47.3)	51.3 (44.3)	38.6 (33.4)

Numbers in bold indicate the best performances for the corresponding metrics.

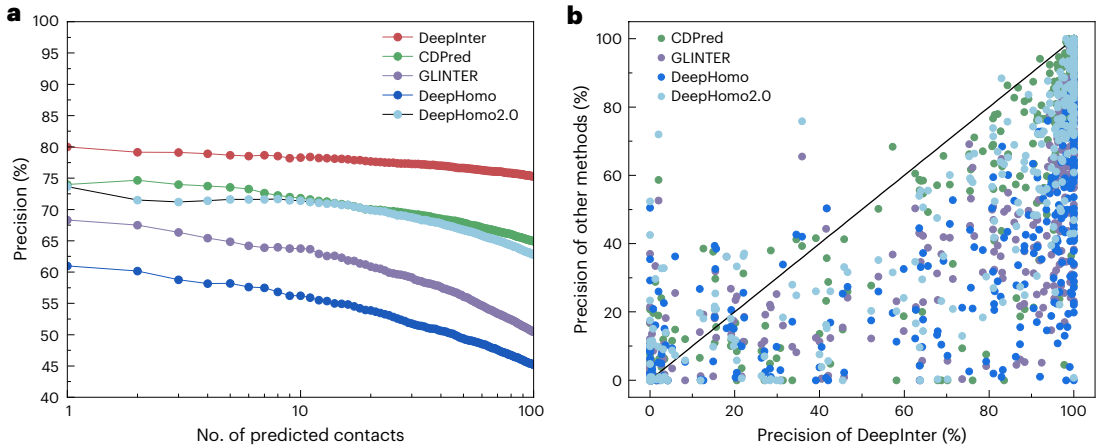

Fig. 2 | Comparison of DeepInter with other methods on the Homodimer300 test set with experimental structures as input. a, Average precisions of DeepInter, DeepHomo2.0, CDPred, GLINTER and DeepHomo as a function of the number of top predicted contacts. **b**, Comparison of the top *L* precisions of DeepInter and the other four methods.

generated from the three MSAs by a protein language model. Meanwhile, the intra-protein distances of the monomers are extracted and combined with the sequence features. After the preparation of input features, the integrated receptor, ligand and complex features are fed into the ResNet-Inception module (Fig. 1b) to extract the high-order intra-protein (receptor and ligand) and inter-protein interactions (complex). The inter-protein interactions are updated by utilizing a triangle-aware module (Fig. 1c) on three pairwise features. Finally, the inter-protein contacts are predicted from the hidden pairwise representations of the complex.

Evaluation of DeepInter on homodimeric complexes

We first evaluated our DeepInter model on the Homodimer300 test set with experimental and AlphaFold2-predicted monomer structures, comparing it with CDPred, DeepHomo2.0, GLINTER and DeepHomo. Table 1 lists the results for the five methods for the top 1, 10, 25, 50, *L*/10, *L*/5 and *L* predicted inter-protein contacts (Supplementary Data 1), where *L* is the length of the monomer sequence of the homodimers. The precision is defined as the percentage of true contacts among a certain number of top predicted contacts. It can be seen from the table that, among the five methods, DeepInter achieves the best performance for all criteria. For example, DeepInter obtains precisions of 78.3%, 77.5%, 76.7% and 71.0% with the inputs of experimental structures for the top 10, 50, *L*/5 and *L* predictions, better than the corresponding precisions of CDPred, DeepHomo2.0, GLINTER and DeepHomo. With the use of Evolutionary Scale Modeling-Multiple Sequence Alignment-1b (ESM-MSA-1b) features and the triangle-aware module, DeepInter achieves an improvement of 32.4% for the top *L* precision on experimental structures compared with DeepHomo. Compared with CDPred, which applies the ESM-MSA-1b features and a 2D-attention mechanism, DeepInter still yields an improvement of 12.7% for the top *L* precision with the input of experimental structures. These results suggest that both the ESM-MSA-1b features and the triangle-aware module can effectively improve the ability of DeepInter to capture correct inter-protein residue–residue contacts.

Figure 2a shows the average precisions of DeepInter, CDPred, DeepHomo2.0, GLINTER and DeepHomo as a function of the number of top predicted contacts on the Homodimer300 test set with experimental monomer structures as input. It can be seen from the figure that the gaps in the precisions between DeepInter and the other methods gradually widen. Figure 2b shows the precision of DeepInter versus those of the other methods for the top *L* predicted contacts on the test set of 300 homodimers with experimental monomer structures. It can be seen from the figure that most of the points are distributed in the lower triangle, where DeepInter has a higher precision than the other methods. For 230 of the 300 homodimers where DeepInter has a higher precision than CDPred, DeepInter achieved an average top *L* precision of 82.5%, compared with 63.9% for CDPred. Meanwhile, DeepInter obtained a higher precision for 251, 254 and 254 dimers than DeepHomo2.0, GLINTER and DeepHomo, respectively. These results suggest the superiority of DeepInter over the other methods.

In real applications, the experimental structures of monomers are unavailable because of the experimental difficulties in obtaining them. To further investigate the robustness of DeepInter, we thus tested DeepInter on the Homodimer300 test set with the structures predicted by AlphaFold2. The corresponding precisions of the five methods are listed in Table 1 in parentheses. Although the AlphaFold2-predicted structures are similar to the experimental structures, there are still

https://doi.org/10.1038/s42256-023-00741-2

Table 2 | Comparison of the precisions (in %) of DeepInter, DeepHomo2.0, CDPred, GLINTER and DeepHomo on the CASP-CAPRI test set of 28 realistic targets with the input of experimental (predicted structures by AlphaFold2) monomer structures when the top 1, 10, 25, 50, *L*/10, *L*/5 and *L* predicted contacts are considered

Method	Top 1	Top 10	Top 25	Top 50	Top *L*/10	Top *L*/5	Top *L*
DeepInter	**75.0 (67.9)**	**72.1 (64.6)**	**72.4 (64.4)**	**70.7 (63.9)**	**72.0 (65.6)**	**70.4 (64.6)**	**62.4 (56.6)**
CDPred	64.3	65.4 (65.7)	61.9	60.8	58.5 (63.1)	60.7 (60.9)	49.8 (49.0)
DeepHomo2.0	71.4 (71.4)	64.6 (63.6)	62.6 (59.6)	59.9 (56.5)	61.7 (58.1)	58.4 (55.4)	45.2 (41.8)
GLINTER	59.3 (63.0)	57.0 (62.2)	53.9 (58.7)	49.9 (54.1)	52.3 (58.5)	49.3 (53.4)	35.5 (35.6)
DeepHomo	53.6 (57.1)	48.9 (51.8)	44.7 (45.7)	42.0 (43.0)	42.9 (44.5)	41.5 (42.2)	29.7 (28.5)

Numbers in bold indicate the best performances for the corresponding metrics.

Table 3 | Comparison of the precisions (in %) of DeepInter, CDPred and GLINTER on the Heterodimer99 test set with the input of experimental (predicted structures by AlphaFold2) monomer structures when the top 1, 10, 25, 50, *L*/10, *L*/5 and *L* predicted contacts are considered

Method	Top 1	Top 10	Top 25	Top 50	Top *L*/10	Top *L*/5	Top *L*
DeepInter	**45.5 (42.3)**	**46.1 (37.4)**	**44.7 (35.5)**	**43.7 (33.7)**	**44.9 (36.4)**	**44.4 (35.5)**	**40.0 (29.4)**
CDPred	39.8 (35.1)	35.4 (32.1)	33.1 (29.7)	30.4 (26.9)	35.2 (32.3)	33.3 (30.3)	26.5 (22.8)
GLINTER	37.4 (36.1)	33.0 (27.5)	28.9 (25.3)	26.1 (22.9)	32.3 (27.1)	29.0 (24.8)	22.2 (19.2)

Numbers in bold indicate the best performances for the corresponding metrics.

some deviations in performance. Nevertheless, DeepInter is more accurate than the other methods in the top *L* precision. Specifically, DeepInter obtains a top *L* precision of 59.0%, compared with 54.4% for CDPred, 47.0% for DeepHomo2.0, 39.1% for GLINTER and 33.4% for DeepHomo (Table 1).

Application to realistic CASP-CAPRI targets

To test DeepInter in real applications, we evaluated DeepInter on the 28 targets obtained from the CASP_CAPRI test set[43,44]. Table 2 lists the precisions of DeepInter, DeepHomo2.0, CDPred, GLINTER and DeepHomo in the inter-protein contact predictions with experimental and predicted monomer structures for top 1, 10, 25, 50, *L*/10, *L*/5 and *L* predicted contacts (Supplementary Data 2). Because GLINTER can only handle homodimers with a sequence length of less than 500, only 27 homodimers in the CASP_CAPRI test set were evaluated for GLINTER. In addition, the precisions of CDPred with predicted structures are directly taken from the corresponding literature[36]. It can be seen from the table that DeepInter achieves the best performance for almost all metrics. Specifically, DeepInter obtained precisions of 75.0% (67.9%), 72.1% (64.6%), 72.4% (64.4%), 70.7% (63.9%), 72.0% (65.6%), 70.4% (64.6%) and 62.4% (56.6%) for the top 1, 10, 25, 50, *L*/10, *L*/5 and *L* predicted contacts with experimental (AlphaFold2-predicted) structures, respectively, outperforming the other methods. The comparison on real targets from the CASP_CAPRI test set indicates the robustness and accuracy of DeepInter in inter-protein contact predictions.

Performance of DeepInter on heterodimeric complexes

To investigate the performance of DeepInter on heterodimers, we also evaluated DeepInter on the Heterodimer99 test set, comparing it with CDPred and GLINTER. Although we use an Expectation-value (*E*-value) of 0.1 to remove the redundant heterodimers between the training and validation sets, there may still be some redundancy between the HeteroDimer99 test set and the training sets of the other methods, which may lead to an over-estimated performance for the other methods. In addition, CDPred only yields predictions on 98 heterodimers because of a failure to conduct MSA on one complex. Table 3 lists the precisions of DeepInter for the top 1, 10, 25, 50, *L*/10, *L*/5 and *L* predicted contacts on the HeteroDimer99 test set with experimental and predicted structures (Supplementary Data 3), where *L* is the minimum length of two monomer sequences for the heterodimer.

It can be seen from Table 3 that DeepInter achieves the best performance in all metrics among the three methods. For example, DeepInter obtains a precision of 46.1% for the top ten predicted contacts, which is substantially higher than the 35.4% achieved by CDPred and 33.0% by GLINTER. This indicates that DeepInter can also capture the important inter-protein interactions on heterodimers and obtain a better performance than the other methods. Figure 3a shows the average precisions of DeepInter, CDPred and GLINTER as a function of the number of top predicted contacts on the Heterodimer99 test set. On the whole, DeepInter obtains a higher precision with a gap of at least 5% (8%) compared with CDpred (GLINTER). CDPred shows a performance similar to that of GLINTER when the number of predicted contacts ranges from 3 to 8. Although CDPred applies a 2D-attention mechanism to enhance its ability to predict inter-protein contacts, CDPred only shows a small improvement over GLINTER. By contrast, the precisions of DeepInter are substantially higher than those of GLINTER. This may suggest the superiority of the triangle-aware module over the 2D-attention mechanism.

Figure 3b compares the top *L* precisions of DeepInter with the other methods for each heterodimer in the test set. It can be seen from the figure that most of the points are located in the lower triangle where DeepInter has a better performance. Specifically, DeepInter obtains a higher top *L* precision than CDPred and GLINTER for 52 and 63 heterodimers, respectively. For structures predicted by AlphaFold2, there were two failed targets. Accordingly, only 97 heterodimers were evaluated by the three methods and the results are listed in Table 3 in parentheses. It can be seen from the table that DeepInter achieves the best performance in all metrics, even with the predicted structures. In summary, DeepInter exhibits a greater ability to predict the correct inter-protein contacts on heterodimers.

Ablation study

To investigate the roles of the different modules and features, we retrained three ablation models by removing the PSSM feature (No_PSSM), the triangle-aware module (No_T) and both the triangle-aware module and ESM-MSA-1b features (No_T_E), respectively. DeepInter trained on the homodimeric training dataset was set as the

https://doi.org/10.1038/s42256-023-00741-2

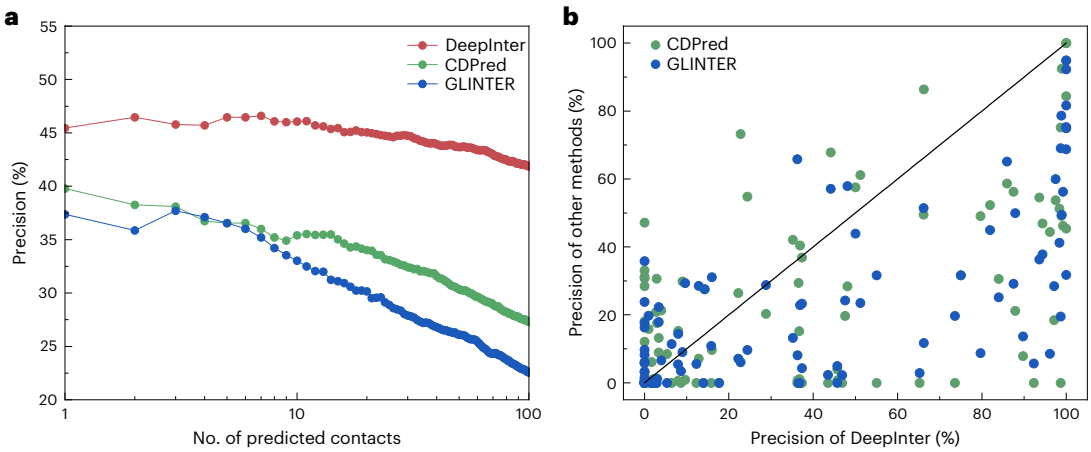

Fig. 3 | Comparison of DeepInter with other methods on the Heterodimer99 test set with the input of experimental structures. a, The precisions of DeepInter, CDPred and GLINTER as a function of the number of top predicted contacts. **b**, Comparison of the top *L* precisions between DeepInter and the other four methods.

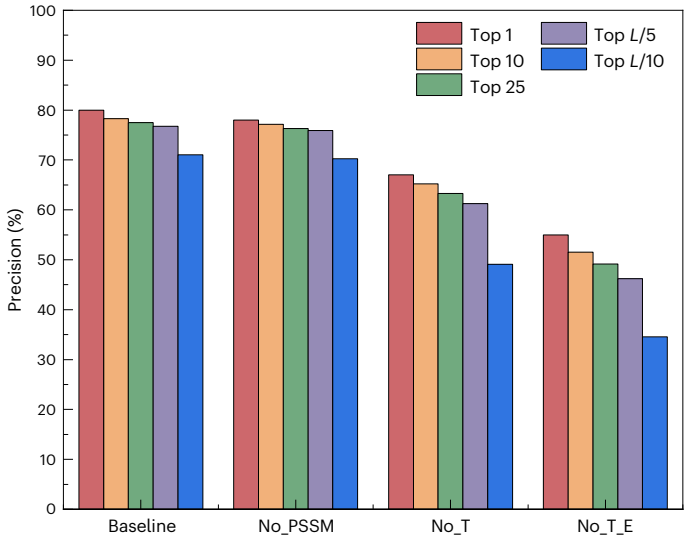

Fig. 4 | Ablation experiments with DeepInter on the Homodimer300 test set with the input of experimental structures. Three ablation models were retrained: No_PSSM model excluding the PSSM features, No_T model excluding the triangle-aware module, and No_T_E excluding the triangle-aware module and ESM-MSA-1b features.

baseline model. To ensure consistency, all four models were trained using the same hyperparameters. Figure 4 shows the precisions of the four models for the top 1, 10, 25, *L*/5 and *L* predicted contacts on the Homodimer300 test set, with the input of experimental structures.

First, the baseline model shows a slightly better performance than the No_PSSM model, which indicates that the PSSM features only add a small improvement for DeepInter (Fig. 4). Second, the No_T model yields a precision that is improved by ~15% over the No_T_E model for the top *L* predicted contacts (Fig. 4). This suggests that the row attention features generated by ESM-MSA-1b include important information about the inter-protein interaction and thus can help the deep learning-based architecture correctly predict the inter-protein contacts[45]. Third, the baseline model achieves an increase of nearly 21% precision for the top *L* predicted contacts compared with the No_T model (Fig. 4). Because the No_T model is to a certain extent similar to DeepHomo2.0, utilization of the triangle-aware module will be the main reason why DeepInter achieves a better performance. The larger increase in precision obtained by applying the triangle-aware module also demonstrates its larger contribution to the DeepInter model. In summary, the ESM-MSA-1b features contain the important residue–residue interaction information, and the triangle-aware module can effectively reduce the geometric inconsistency. Both factors are important for the DeepInter model and can effectively improve its capability to predict complex interfacial residue–residue interactions.

Impact factors on DeepInter

In addition to establishing the importance of the network features, we also extensively investigated the impact of other factors on DeepInter, including the direct use of intra-protein distance instead of the combination of a ResNet-Inception module and other intra-protein features, MSA depth, contact density, sequence length cutoff, conformational changes upon binding, intrinsically disordered proteins[46] and structure similarity[47] (Supplementary Tables 1–4 and Supplementary Fig. 1). It was found that the direct use of intra-protein distance gives a slightly lower performance than the baseline DeepInter model. All the tested methods tend to have a lower precision with decreasing MSA depth or contact density, although DeepInter always performs best within the same category. The performance decrease for targets with longer sequences compared with shorter sequences is attributed to the smaller contact density rather than the sequence length cutoff. DeepInter is also robust against the conformational changes occurring upon binding, intrinsically disordered proteins, and structure similarity, although it may have some degree of performance decrease with larger conformational changes, more intrinsically disordered proteins, and low structurally similar targets. Nevertheless, DeepInter still outperforms the other methods within the same category. These results suggest the accuracy and robustness of DeepInter in different situations.

Case study of inter-protein contact prediction

From Fig. 2, it can be seen that the precision of DeepInter decreases more slowly than the other methods as the number of considered contacts increases. As we only evaluate the average precisions of DeepInter and the other methods on the specified numbers of contacts, it is valuable to show the whole predicted contact map. Figure 5 shows the predicted inter-protein contact maps of an example dimer (PDB 2ZY9) from the Homodimer300 test set. Because the GLINTER and No_T models both utilize the ESM-MSA-1b features and exclude the attention mechanism, and the figure also shows the corresponding predicted contact maps, we can investigate the influence of the triangle-aware module compared with the DeepInter model. Similarly, because the DeepHomo and No_T_E models both use the DCA-DI and DCA-APC features and exclude the ESM-MSA-1b features, the predicted

https://doi.org/10.1038/s42256-023-00741-2

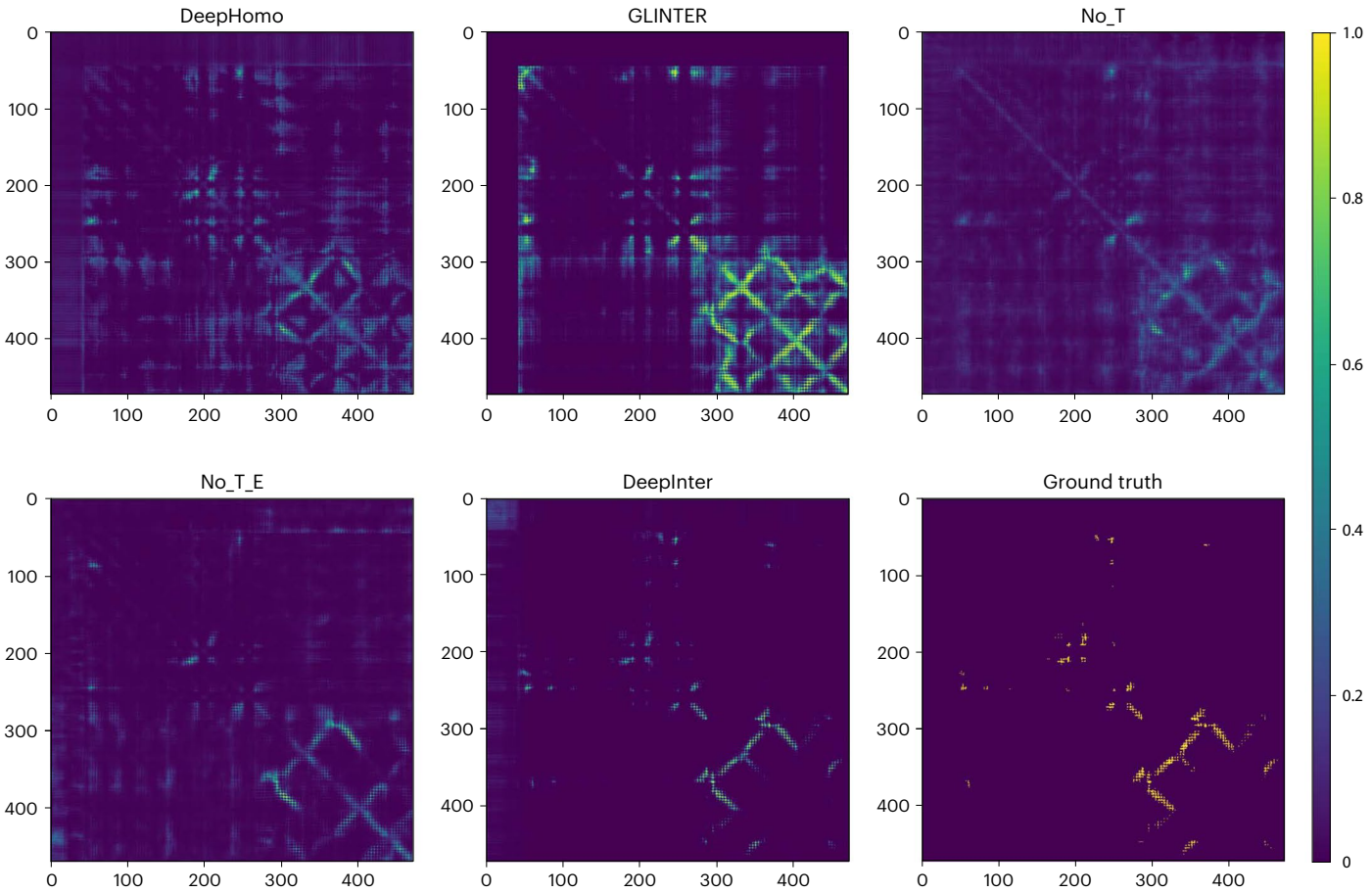

Fig. 5 | Examples of predicted contact maps. Whole predicted inter-protein contact maps of five models (DeepHomo, GLINTER, No_T, No_T_E and DeepInter) on the 2ZY9 dimeric complex. The inter-protein contact map of the native complex is shown as the ground truth. The numbers on the axes indicate the corresponding residue number in the monomer sequence, and the value on the colour bar indicates the predicted contact probability of inter-protein residue pairs.

contact maps of the two models are also shown in Fig. 5 to analyse the effect of the ESM-MSA-1b features.

It can be seen from Fig. 5 that the predicted inter-protein contact maps of DeepHomo, GLINTER, No_T_E and No_T tend to have more noise than that of DeepInter. The lower performance of the four models compared with DeepInter will be due to the defects of traditional convolutional networks. Although a convolutional network can predict correct inter-protein contacts in local regions, it cannot guarantee the geometric triangle inequality and effectively capture long-range interactions. In contrast, DeepInter can predict the corresponding region of ground truth with little noise, as well as the long-range contacts, via the attention mechanism in the triangle-aware module. On the contact map for 2ZY9, DeepInter obtains a precision of 98.5% for the top L predicted contacts, which is substantially higher than the 82.7% achieved by DeepHomo2.0, 75.3% by CDPred, 71.0% by DeepHomo and 43.8% by GLINTER. Through observation of the whole predicted inter-protein contact map and the comparison of the top L precisions, it may be concluded that DeepInter can effectively capture the correct inter-protein interactions by reducing the violation of triangle inequality.

Discussion

We have presented a deep learning-based method for the inter-protein contact prediction of protein complexes by integrating ESM-MSA-1b features, a ResNet-Inception module and triangle-aware attention modules. When compared with other advanced methods, DeepInter correctly predicted inter-protein contacts and achieved the best performance for almost all metrics on two diverse homodimeric test sets and one heterodimeric test set. Through an ablation experiment, it is revealed that the improvements are mainly contributed by the ESM-MSA-1b features and the triangle-aware module. The former, which is trained on large sequence data, interleaves row and column attention to extract the residue–residue interaction. The latter applies an attention mechanism on pair representations to consider the many-body effect by satisfying the geometric triangle inequality. Further analysis regarding MSA depth and the contact density of interfaces indicates that DeepInter can obtain a higher precision than the other methods on hard dimers that have a small interactions interface and few homologies.

In further developments, the predicted inter-protein contacts may be used for the structure-modelling of complexes, for example, in protein–protein docking and gradient descent optimization. The architecture of DeepInter can be further improved for use in relation to the inter-protein contacts of higher-order oligomeric complexes. In addition, the present architecture may also be adapted to predicting residue-residue distance maps, working with AlphaFold2 or AlphaFold-Multimer to improve complex structure prediction. A future direction might also be the prediction of the oligomeric state of a protein, a critical problem that remains open in the field.

Despite the present success of DeepInter, there are still some limitations that may be improved in future work. First, because we have used the MSA representation and the attention matrix generated from the ESM-MSA-1b model, the maximum sequence length of the dimers is limited to 1,024. This means that DeepInter cannot predict the inter-protein contacts of extremely large hetero-complexes. Second,

some large protein complexes with a small interface are hard to correctly predict with DeepInter and existing predictors. Third, the quality of the predicted structures will affect the precision of structure-based methods. Finally, large conformational changes in the protein will have a substantial impact on performance. In future work, we may use the structure representation of a large encoder model to enhance the robustness of predictors and further improve our network to predict the multiple-chain contacts of hetero-complexes.

Methods

Overall architecture of DeepInter

Figure 1 shows the overall architecture of DeepInter, which is composed of three main elements: (1) the encoding features of the protein sequence and pairwise interaction; (2) implementation of the ResNet-Inception module; (3) the triangle-aware module, including the triangle update, triangle self-attention and the transition layer.

Representations of sequence and structure. DeepInter takes the representations of sequences and structures to predict inter-protein contacts. All the sequence features are generated from the MSA, so we first generate the MSA for each monomer against the UniRef30_2020_03 database[48] by using the hhblits tool[49], with similar hyperparameters as used in our previous DeepHomo study[9]. We then build the paired MSA for the heterodimer according to the organism information. Specifically, the sequences in the MSA are first separated according to the organism information. The sequences from the same organism are sorted by sequence identity with the corresponding monomer sequence of heterodimers. The top two sequences for the same organism from two different MSAs are paired. For homodimers, we directly use the MSA generated by the sequence of the monomer structure as the paired MSA.

After preparation of the MSA, we calculate the sequence features for the triangle-aware module, which requires three kinds of input feature, for the receptor, ligand and complex, respectively. For convenience, we use $L1$ and $L2$ to represent the sequence lengths of the receptor and ligand of a dimer. As shown in Fig. 1a, PSSMs are generated from the MSAs of the receptor and ligand using the LoadHMM script in the RaptorX-Contact package[31]. Two kinds of DCA, including raw scores and average product correction of the DCA (DCA-DI and DCA-APC) for the receptor and ligand, are then produced by CCMPred[50]. For the internal features of the protein language model (ESM-MSA-1b), we first select the top 512 diverse sequences in each MSA using the hhfilter tool[51] and extract the ESM-MSA-1b vector embedding and row attentions for the receptor and ligand, respectively.

In addition to the sequence features, the monomer structure information of the dimeric complex is also critical for inter-protein contact predictions. We therefore calculate the intra-protein residue–residue distance map of the receptor and ligand, where the distance between two residues is represented by their minimal heavy-atom distance. We then use the radial basis function to convert the intra-protein distance as follows[52]:

$$f(d) = \mathrm{e}^{-\left(\frac{d-d^k}{\sigma}\right)^2} \tag{1}$$

where d is the intra-protein distance, and d^k is a hyperparameter, which ranges from 2 to 22 Å and is divided into 64 bins with variance $\sigma = 0.3125$ Å.

Before the ResNet-Inception module, we need to integrate the preprocessed features. For the input features of the receptor and ligand from the dimeric complex, we use the 'intra-pairing' option to convert the concatenated 1D features of the PSSM and ESM-MSA-1b vector into embedded 2D features, whose dimensions for the receptor and ligand are represented as $L1 \times L1 \times 1{,}576$ and $L2 \times L2 \times 1{,}576$, where $1{,}576 = (20 + 768) \times 2$. We then concatenate the 2D features with the DCA, ESM-MSA-1b row attention and intra-protein distance map to create the input features of the ResNet-Inceptor module for the receptor and ligand. For the complex, we use the 'inter-pairing' option to convert the concatenated 1D features of the receptor and ligand to 2D features with a dimension of $L1 \times L2 \times 1{,}576$. We then crop the inter-protein region on the DCA-DI, DCA-APC and ESM-MSA-1b row attentions generated by the paired MSA. Finally, we concatenate the inter-pairing 2D features with the DCA and ESM-MSA-1b row attention features as the input features of the complex for the ResNet-Inception module.

Implementation of the ResNet-Inception module. To make up for the shortcomings of traditional ResNet, recent methods combine the residual connection and Inception module to increase the effective receptive field of the network with fewer parameters[53,54]. The latter modules use a multi-branch convolutional architecture to obtain multiscale features. Here we use three parallel branches of convolutional architectures to create the basic block (Fig. 1b), termed ResNet-Inception. The detailed formulae of the ResNet-Inception module are as follows:

$$h_{ij}^{1(k)} = \mathrm{Conv}_{1\times 9}(\mathrm{ReLU}(\mathrm{Norm}(\mathrm{Conv}_{1\times 9}(h_{ij})))) \tag{2}$$

$$h_{ij}^{2(k)} = \mathrm{Conv}_{3\times 3}(\mathrm{ReLU}(\mathrm{Norm}(\mathrm{Conv}_{3\times 3}(h_{ij}))))) \tag{3}$$

$$h_{ij}^{3(k)} = \mathrm{Conv}_{9\times 1}(\mathrm{ReLU}(\mathrm{Norm}(\mathrm{Conv}_{9\times 1}(h_{ij})))) \tag{4}$$

$$\widetilde{h}_{ij}^{(k)} = \mathrm{ReLU}\left(\mathrm{Norm}\left(h_{ij}^{1(k)}\right) + \mathrm{Norm}\left(h_{ij}^{2(k)}\right) + \mathrm{Norm}\left(h_{ij}^{3(k)}\right)\right) \tag{5}$$

where h_{ij} represents the input features, including the pair representations of the receptor, ligand and complex, Conv is the 2D convolutional kernel, Norm is the InstanceNorm, ReLU is the activation function, and k is the number of blocks in ResNet-Inception. We use four continuous ResNet-Inception modules to preliminarily capture the intra-protein and inter-protein interactions.

Implementation of the triangle-aware module. Due to the distance constraint, residue C, which is far away from residue A in the receptor, cannot contact residue B when residues A and B form the interfacial contact. Therefore, we use an inter-protein triangle multiplicative mechanism to update the inter-protein residue–residue interaction. We combine the triangle update with triangle self-attention and the transition layer to build the triangle-aware module, shown in Fig. 1c. Specifically, the triangle update utilizes an attention mechanism on the pair representations of three residues to ensure that the pair representation satisfies the geometric triangle inequality. The detailed formulae for the triangle update are as follows:

$$z_{ij}^{\prime(k)} = \mathrm{Linear}\left(\phi\left(z_{ij}^{(k)}\right)\right); z_{ij}^{\prime\prime(k)} = \mathrm{Linear}\left(\phi\left(z_{ij}^{(k)}\right)\right) \tag{6}$$

$$r'_{ij} = \mathrm{Linear}\left(\phi(r_{ij})\right); l'_{ij} = \mathrm{Linear}\left(\phi(l_{ij})\right) \tag{7}$$

$$\tilde{z}_{ij}^{(k)} = z_{ij}^{(k)} + \varphi\left(\sum_{m=1}^{L} r'_{im} z_{mj}^{\prime(k)} + \sum_{n=1}^{L} z_{in}^{\prime\prime(k)} l'_{nj}\right) \odot \phi\left(z_{ij}^{(k)}\right) \tag{8}$$

where r_{ij}, l_{ij} and z_{ij} are the output features from the last ResNet-Inception module, k is the number of blocks in the triangle update ($z_{ij}^{(1)} = z_{ij}$), the function ϕ is a sigmoid function followed by linear transformation, and the function φ is a LayerNorm function followed by a linear transformation.

We then adopt triangle self-attention to calculate the relative strengths of the interaction between pairwise residues in two different directions. First, the multi-head attention on the row dimension is applied on all residue pairs between specific receptor residues and all ligand residues. The detailed formulae are

$$q_{ij}, k_{ij}, v_{ij} = \varphi_q\left(\tilde{z}_{ij}^{(k)}\right), \varphi_k\left(\tilde{z}_{ij}^{(k)}\right), \varphi_v\left(\tilde{z}_{ij}^{(k)}\right) \quad (9)$$

$$w_{ijm} = \mathrm{softmax}\left(q_{ij}^T k_{im}\right) \odot g(d) \quad (10)$$

$$g(d) = \exp\left(-\frac{d^2}{2\sigma^2}\right) \quad (11)$$

$$\hat{z}_{ij}^{(k)} = \tilde{z}_{ij}^{(k)} + \varphi\left(\sum_{m=1}^{L}(w_{ijm} v_{im}) \odot \phi\left(\tilde{z}_{ij}^{(k)}\right)\right) \quad (12)$$

where q_{ij}, k_{ij}, v_{ij} is the linear transformations of the output of the triangle update, d is the intra-protein distance, and σ is set to 8. Second, we also update the pair representation on the column dimension. In addition, we use a mask mechanism like a classical transformer for the surface residues defined by the NACCESS program[55] in the triangle-aware module. Finally, the transition layer, including two linear transformations, is used to update the pair representation $\hat{z}_{ij}^{(k)}$.

Datasets

For the homo-protein model, we use the same datasets as those for our previous DeepHomo study for training and testing[9]. Specifically, all the homodimers are downloaded from the Protein Data Bank (PDB)[56] and clustered into a non-redundant dataset using MMSeq2[57,58]. The non-redundant dimers are divided into a training set with 3,504 homodimers, a validation set with 296 homodimers, and a test set with 300 homodimers (Homodimer300 test set) (Supplementary Data 10). In addition, the 28 CASP_CAPRI targets prepared by DeepHomo are also used as another test set (CASP_CAPRI test set) to evaluate the performance in real experiments (Supplementary Data 10). To verify the robustness of DeepInter, the AlphaFold2-predicted structures of the two test sets are also used as input to evaluate performance.

For the hetero-protein model, we collect all the heterodimers from the PDB and filter them with criteria similar to those for the homodimers. Specifically, the heterodimers are distilled with the following criteria: (1) the oligomeric state should be heterodimer; (2) the biological assemblies only contain two single chains; (3) the resolution of the X-ray structure is better than 3.0 Å; and (4) the length of the single chain is larger than 50 and the total length of the complex is less than 1,023. We then use a sequence identity cutoff of 40% to cluster the heterodimers and choose the heterodimer with the best resolution in each cluster. The resulting 2,174 heterodimeric complexes are randomly divided into a training set of 1,974 heterodimers, a validation set of 100 heterodimers, and a test set of 100 heterodimers. In addition, we also filter the test set of 100 heterodimers with a sequence identity cutoff of 70% according to the two chains of a complex, which results in 99 heterodimers (Heterodimer99; Supplementary Data 10). We use an E-value of 0.1 to remove the redundant heterodimeric complex from the training and validation set with the heterodimeric test set, which results in 1,881 and 96 heterodimers (Supplementary Data 10). To increase the data abundance, we combine the training sets (validation set) of homodimers and heterodimers for training our hetero-protein model.

Implementation and training

The pseudo-code for the training implementation of DeepInter is shown in Supplementary Algorithm 1. Unlike the intra-protein contact predictions, the interfacial contacts between proteins are more difficult to predict due to the extreme imbalance between the non-contact and contact residue pairs. DeepInter adopts a Focal Loss[59] loss function with the same parameters as those of DeepHomo. DeepInter is trained on one A100 graphics processing unit (GPU) with PyTorch (v1.8.0) and the optimizer Adam[60]. In addition, the training process uses a learning rate of 0.001 without any regularization and a dropout of 0.1 for the triangle-aware module. Because of the large use of GPU memory by the triangle self-attention layer, we apply a mini-batch size of 1 and set a maximum length of 256 for each monomer sequence in the training process. Specifically, for the monomer proteins of the dimer with a sequence length of more than 256, we use a window of size 256 and a stride of 1 to scan the sequence. Fragments with the maximum inter-protein contacts are kept, and the final cropped sequences are randomly selected from those fragments to represent the protein sequence.

During training, the ground-truth inter-protein contacts are defined as those inter-protein residue–residue pairs with a distance of <8 Å. Here, the distance between two residues is represented by the minimal distance between the heavy atoms of the two residues.

Data availability

The data that support the findings of this study are available from the corresponding author upon request. A full list of the protein complexes used in this study is also provided in Supplementary Data 10. The protein structures used in this study are all available in the PDB, and the sequence database of Uniref30_2020_03 used in this study is available at https://www.uniprot.org/help/uniref/. Source data are provided with this paper.

Code availability

The DeepInter package is freely available at http://huanglab.phys.hust.edu.cn/DeepInter/ and https://doi.org/10.5281/zenodo.8304327 ref. 61.

References

1. Yadid, I. & Tawfik, D. S. Reconstruction of functional beta-propeller lectins via homo-oligomeric assembly of shorter fragments. *J. Mol. Biol.* **365**, 10–17 (2007).
2. Goodsell, D. S. & Olson, A. J. Structural symmetry and protein function. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 105–153 (2000).
3. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
4. Kryshtafovych, A., Schwede, T., Topf, M., Fidelis, K. & Moult, J. Critical assessment of methods of protein structure prediction (CASP)-Round XIV. *Proteins* **89**, 1607–1617 (2021).
5. Ko, J. & Lee, J. Can AlphaFold2 predict protein-peptide complex structures accurately? Preprint at https://www.biorxiv.org/content/10.1101/2021.07.27.453972v2 (2021).
6. Bryant, P., Pozzati, G. & Elofsson, A. Improved prediction of protein–protein interactions using AlphaFold2. *Nat. Commun.* **13**, 1265 (2022).
7. Mirdita, M. et al. ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
8. Evans, R. et al. Protein complex prediction with AlphaFold-Multimer. Preprint at https://www.biorxiv.org/content/10.1101/2021.10.04.463034v2 (2021).
9. Yan, Y. & Huang, S. Y. Accurate prediction of inter-protein residue-residue contacts for homo-oligomeric protein complexes. *Brief. Bioinformatics.* **22**, bbab038 (2021).
10. Yan, Y., Tao, H., He, J. & Huang, S. Y. The HDOCK server for integrated protein–protein docking. *Nat. Protoc.* **15**, 1829–1852 (2020).
11. Du, Z. et al. The trRosetta server for fast and accurate protein structure prediction. *Nat. Protoc.* **16**, 5634–5651 (2021).
12. Yan, Y., Tao, H. & Huang, S. Y. HSYMDOCK: a docking web server for predicting the structure of protein homo-oligomers with Cn or Dn symmetry. *Nucleic Acids Res.* **46**, W423–W431 (2018).

13. Soltanikazemi, E., Quadir, F., Roy, R. S., Guo, Z. & Cheng, J. Distance-based reconstruction of protein quaternary structures from inter-chain contacts. *Proteins* **90**, 720–731 (2022).
14. Hopf, T. A. et al. Sequence co-evolution gives 3D contacts and structures of protein complexes. *eLife* **3**, e03430 (2014).
15. 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).
16. 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**, 12295 (2021).
17. Sanchez-Garcia, R., Sorzano, C. O. S., Carazo, J. M. & Segura, J. BIPSPI: a method for the prediction of partner-specific protein-protein interfaces. *Bioinformatics* **35**, 470–477 (2019).
18. Sanchez-Garcia, R., Macias, J. R., Sorzano, C. O. S., Carazo, J. M. & Segura, J. BIPSPI+: mining type-specific datasets of protein complexes to improve protein binding site prediction. *J. Mol. Biol.* **434**, 167556 (2022).
19. Zhao, Z. & Gong, X. Protein-protein interaction interface residue pair prediction based on deep learning architecture. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **16**, 1753–1759 (2019).
20. Liu, J. & Gong, X. Attention mechanism enhanced LSTM with residual architecture and its application for protein-protein interaction residue pairs prediction. *BMC Bioinformatics* **20**, 609 (2019).
21. Soleymani, F., Paquet, E., Viktor, H., Michalowski, W. & Spinello, D. Protein-protein interaction prediction with deep learning: a comprehensive review. *Comput. Struct. Biotechnol. J.* **20**, 5316–5341 (2022).
22. Baranwal, M. et al. Struct2Graph: a graph attention network for structure based predictions of protein-protein interactions. *BMC Bioinformatics* **23**, 370 (2022).
23. Hu, X., Feng, C., Zhou, Y., Harrison, A. & Chen, M. DeepTrio: a ternary prediction system for protein-protein interaction using mask multiple parallel convolutional neural networks. *Bioinformatics* **38**, 694–702 (2022).
24. Soleymani, F., Paquet, E., Viktor, H. L., Michalowski, W. & Spinello, D. ProtInteract: a deep learning framework for predicting protein-protein interactions. *Comput. Struct. Biotechnol. J.* **21**, 1324–1348 (2023).
25. Morcos, F. et al. Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proc. Natl Acad. Sci. USA* **108**, E1293–E1301 (2011).
26. Jones, D. T., Buchan, D. W., Cozzetto, D. & Pontil, M. PSICOV: precise structural contact prediction using sparse inverse covariance estimation on large multiple sequence alignments. *Bioinformatics* **28**, 184–190 (2012).
27. Ekeberg, M., Lövkvist, C., Lan, Y., Weigt, M. & Aurell, E. Improved contact prediction in proteins: using pseudolikelihoods to infer Potts models. *Phys. Rev. E* **87**, 012707 (2013).
28. Yang, J. et al. Improved protein structure prediction using predicted interresidue orientations. *Proc. Natl Acad. Sci. USA* **117**, 1496–1503 (2020).
29. Li, Y. et al. Deducing high-accuracy protein contact-maps from a triplet of coevolutionary matrices through deep residual convolutional networks. *PLoS Comput. Biol.* **17**, e1008865 (2021).
30. Adhikari, B., Hou, J. & Cheng, J. DNCON2: improved protein contact prediction using two-level deep convolutional neural networks. *Bioinformatics* **34**, 1466–1472 (2018).
31. 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**, e1005324 (2017).
32. Senior, A. W. et al. Improved protein structure prediction using potentials from deep learning. *Nature* **577**, 706–710 (2020).
33. Wang, W., Peng, Z. & Yang, J. Single-sequence protein structure prediction using supervised transformer protein language models. *Nat. Comput. Sci.* **2**, 804–814 (2022).
34. Lin, P., Yan, Y. & Huang, S. Y. DeepHomo2.0: improved protein-protein contact prediction of homodimers by transformer-enhanced deep learning. *Brief. Bioinformatics* **24**, bbac499 (2023).
35. Xie, Z. & Xu, J. Deep graph learning of inter-protein contacts. *Bioinformatics* **38**, 947–953 (2022).
36. 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).
37. Bitbol, A. F., Dwyer, R. S., Colwell, L. J. & Wingreen, N. S. Inferring interaction partners from protein sequences. *Proc. Natl Acad. Sci. USA* **113**, 12180–12185 (2016).
38. Szurmant, H. & Weigt, M. Inter-residue, inter-protein and inter-family coevolution: bridging the scales. *Curr. Opin. Struct. Biol.* **50**, 26–32 (2018).
39. Gueudr'e, T., Baldassi, C., Zamparo, M., Weigt, M. & Pagnani, A. Simultaneous identification of specifically interacting paralogs and interprotein contacts by direct coupling analysis. *Proc. Natl Acad. Sci. USA* **113**, 12186–12191 (2016).
40. Ovchinnikov, S., Kamisetty, H. & Baker, D. Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. *eLife* **3**, e02030 (2014).
41. Zeng, H. et al. ComplexContact: a web server for inter-protein contact prediction using deep learning. *Nucleic Acids Res.* **46**, W432–W437 (2018).
42. Vaswani, A. et al. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **30**, 5998–6008 (2017).
43. Lensink, M. F. et al. The challenge of modeling protein assemblies: the CASP12-CAPRI experiment. *Proteins* **86**, 257–273 (2018).
44. Lensink, M. F. et al. Blind prediction of homo- and hetero-protein complexes: the CASP13-CAPRI experiment. *Proteins* **87**, 1200–1221 (2019).
45. Rao, R. et al. MSA transformer. *Proc. 38th International Conference on Machine Learning* **139**, 8844–8856 (PMLR, 2021).
46. Hatos, A. et al. DisProt: intrinsic protein disorder annotation in 2020. *Nucleic Acids Res.* **48**, D269–D276 (2020).
47. Xu, J. & Zhang, Y. How significant is a protein structure similarity with TM-score=0.5? *Bioinformatics* **26**, 889–895 (2010).
48. Mirdita, M. et al. Uniclust databases of clustered and deeply annotated protein sequences and alignments. *Nucleic Acids Res.* **45**, D170–D176 (2017).
49. Remmert, M., Biegert, A., Hauser, A. & Söding, J. HHblits: lightning-fast iterative protein sequence searching by HMM–HMM alignment. *Nat. Methods* **9**, 173–175 (2011).
50. Seemayer, S., Gruber, M. & Söding, J. CCMpred—fast and precise prediction of protein residue-residue contacts from correlated mutations. *Bioinformatics* **30**, 3128–3130 (2014).
51. Steinegger, M. et al. HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics* **20**, 473 (2019).
52. Baek, M. et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).
53. Si, Y. & Yan, C. Improved protein contact prediction using dimensional hybrid residual networks and singularity enhanced loss function. *Brief. Bioinformatics* **22**, bbab341 (2021).
54. Su, H. et al. Improved protein structure prediction using a new multi-scale network and homologous templates. *Adv. Sci.* **8**, e2102592 (2021).

55. Hubbard, S. J. & Thornton, J. M. *NACCESS: computer program* (Department of Biochemistry and Molecular Biology, University College London, 1993).
56. Berman, H. M. et al. The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
57. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive datasets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).
58. Steinegger, M. & Söding, J. Clustering huge protein sequence sets in linear time. *Nat. Commun.* **9**, 2542 (2018).
59. Lin, T.-Y., Goyal, P., Girshick, R., He, K. & Dollar, P. Focal loss for dense object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **42**, 318–327 (2020).
60. Kinga, D. & Adam, J. B. A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)* (2015).
61. Lin, P., Tao, H., Li, H. & Huang, S.-Y. Protein-protein contact prediction by geometric triangle-aware protein language models. *Zenodo* (2023); https://doi.org/10.5281/zenodo.8304327

Acknowledgements

This work was supported by the National Natural Science Foundation of China (grants nos. 32161133002 and 62072199) and a startup grant of Huazhong University of Science and Technology.

Author contributions

S.-Y.H. conceived and supervised the project. P.L. and S.-Y.H. designed and performed the experiments. P.L., H.T. and H.L. analysed the data. P.L., H.T., H.L. and S.-Y.H. wrote the paper. All authors reviewed and approved the final version of the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42256-023-00741-2.

Correspondence and requests for materials should be addressed to Sheng-You Huang.

Peer review information *Nature Machine Intelligence* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editor: Jacob Huth, in collaboration with the *Nature Machine Intelligence* team.