Protein-Protein Docking with Iterative Transformer

Lee-Shin Chu
Johns Hopkins University
Baltimore, MD 21218
lchu11@jhu.edu

Jeffrey A. Ruffolo
Johns Hopkins University
Baltimore, MD 21218
jruffolo@jhu.edu

Jeffrey J. Gray
Johns Hopkins University
Baltimore, MD 21218
jgray@jhu.edu

Abstract

Conventional protein-protein docking algorithms usually rely on heavy candidate sampling and re-ranking, but these steps are time-consuming and hinder applications that require high-throughput complex structure prediction, e.g., structure-based virtual screening. Existing deep learning methods for protein-protein docking, despite being much faster, suffer from low docking success rates. In addition, they simplify the problem to assume no conformational changes within any protein upon binding (rigid docking). This assumption precludes applications when binding-induced conformational changes play a role, such as allosteric inhibition or docking from uncertain unbound model structures. To address the limitations, we designed a novel iterative transformer network that predicts the 3D transformation from a randomized initial docking pose to a refined docked pose. Our method, GeoDock, is flexible at the protein residue level, allowing the prediction of rigid-body movement as well as conformational changes upon binding. For two benchmark sets of rigid docking targets, GeoDock successfully docks 32% and 20% of the protein pairs, outperforming the baseline deep learning method EquiDock [1] (8% and 0% success rates). Additionally, GeoDock achieves comparable docking success rates to the conventional docking algorithms while being 80-500 times faster. Although binding-induced conformational changes are still a challenge owing to limited training and evaluation data, our architecture sets up the foundation to capture flexibility going ahead.

1 Introduction

Protein-protein interactions are involved in nearly all cellular functions in living organisms, from signaling and regulation to recognition. These cellular functions are crucially dependent on the precise assembly of proteins to become functional multi-protein complexes. Understanding the 3D structures of protein-protein complexes at the atomic level can give insight into the mechanisms that underlie these functions. While experimental approaches can determine protein structures, they are expensive, low-throughput, and not applicable to all proteins. Recent breakthroughs of AlphaFold2 [2] and other follow-up works [3–6] have demonstrated the prediction of 3D models of protein structures comparable to experimental accuracy. Along with genome-wide protein sequencing, the AlphaFold database [7] provides open access to 214M protein structure predictions from the sequences deposited in UniProt [8]. Because of the expanding number of known protein monomer structures, fast and reliable computational approaches for modeling protein-protein interactions are critical. Protein-protein docking methods provide computational tools for fundamental studies of protein interactions by predicting the favorable protein binding sites and possibly binding induced conformational changes.

Protein-protein docking methods predict a protein complex structure given the structures of its unbound monomeric partners. Classical protein docking approaches generally comprise a sampling algorithm that generates ensembles of candidate docked structures, and a scoring function that evalu-

Machine Learning for Structural Biology Workshop, NeurIPS 2022.

ates the candidates generated from the sampling stage. The sampling strategies can be classified into exhaustive global search, local shape feature matching, and randomized search [9]. The exhaustive global search methods [10–17] mostly use fast Fourier transforms (FFTs) [18] to extensively search over the complete 6D (3D translational plus 3D rotational) space, assuming no conformational changes of the docking partners. The local shape matching methods [19–24] typically represent a protein by the shape of its molecular surface and find matches of high shape complementarity between two proteins. The randomized search methods [25–37] use stochastic algorithms such as Monte Carlo methods to search the docking poses through the free energy landscape starting from a randomized initial pose, with the protein represented as all-atom or coarse grain models. Depending on the sampling algorithms used, the scoring can take place after or can be coupled during the sampling process. In contrast to sampling and scoring scheme, template-based docking methods [38] use information such as sequence similarity, evolutionary conservation, and interface complementarity to search for complexes that are homologous to the proteins to be docked, and then use the complexes as docking templates. Due to the massive candidate searching and evaluation, these docking methods are usually time-consuming, hindering applications that require high-throughput complex structure prediction.

Recent breakthroughs in machine learning, especially in deep learning techniques, have been applied to several protein docking related tasks. MaSIF [39] and ScanNet [40] use geometric deep learning methods to predict the binding sites of a protein based on the structural and chemical features of its molecular surface, independent of the binding partners. For partner specific protein interface prediction, several methods [41–48] predict protein inter-chain contacts (residues within some cut-off distance) using Graph Neural Networks (GNNs) that input structural and possibly evolutionary features of the proteins. The predicted binding sites or inter-chain contact maps from these deep learning methods can guide docking and generate protein complex structures [49, 50], but the resulting complex structures have not achieved the accuracy comparable to the conventional docking approaches. End-to-end deep learning methods that predict protein complex structures from protein sequences and multiple sequences alignments (MSAs) have also been developed. These methods [51–53] extend the ability of AlphaFold2 [2] that was originally designed for protein folding to protein multimer structure prediction. Remarkably, AlphaFold-Multimer (AF-M) [54] includes multiple chains during training and outperforms other AlphaFold2-based approaches. However, searching for MSAs slows the inference process, hampering its applications to fast protein complex structure predictions. Furthermore, MSAs do not apply to the docking of important protein families like antibodies or T-cell receptors that evolve at different timescales than their binding partners [55].

Several deep learning methods have been developed for fast rigid docking. Ganea et al. [1] developed EquiDock, an equivariant graph matching neural network that predicts the rotation and translation to place one of the proteins at the correct docked position relative to the second protein. Sverrisson et al. [49] developed a generative model based on the interface features derived from dMaSIF [56] for generating ensembles of docking candidates that are then scored by a trained discriminative model. However, due to the rigid-body assumption, these methods are not capable of predicting the conformational changes upon protein binding. In addition, the docking success rates of these methods are lower than the conventional docking methods. In this work, we develop a iterative transformer network for fast and flexible protein-protein docking. Our method, GeoDock, is to our knowledge the first deep learning method allowing docking with backbone flexibility. We demonstrate superior performance compared to EquiDock and several conventional docking methods.

2 Methods

Datasets. For training and evaluating our models, we use (1) Database of Interacting Protein Structures (DIPS) [42] and (2) Docking Benchmark 5.5 (DB5.5) [57, 58]. DIPS comprises 42,826 non-redundant experimentally resolved binary protein complexes, after excluding the complexes that have any individual protein with over 30% sequence identity to any protein in the DB5.5. Following EquiDock, We partitioned the DIPS dataset into train/validation/test splits of size 37,402/983/941. DB5.5 is the gold standard dataset for evaluating the performance of a docking algorithm; it contains 271 complex structures with both bound and unbound conformations. We curated the 25 complexes that were used as the test set in EquiDock and use the remaining 246 complexes for fine-tuning the model pre-trained on the DIPS dataset. For the final evaluation, we use the same test sets as used in the EquiDock work for both DIPS and DB5.5.

Random Initialization We generate the initial docking pose by first separating the centers of mass of the two docking partners by 50.0 Å, and randomizing the orientations of each of them by unbiased sampling of unit quaternions. The two partners are then perturbed by a random direction and magnitude specified by a Gaussian distribution around 3.0 Å. The random initialization process ensures fair comparisons between our method and other global docking algorithms.

Interface Biased Cropping. The interface region between docking partners is smaller than the non-interface region, since only some of the residues are in contact between the partners. To achieve a better balance between the interface and non-interface regions, we used a cropping procedure that maximizes the coverage of the interface regions while maintaining the continuity in sequence space. Specifically, for each docking partner, we find the first and the last contact residues (within 12 Å between the docking partners) in sequence space and crop the region between the two residues. After cropping the interface regions, we randomly crop a continuous block of residues on each partner up to 384 residues in total (summing over the docking partners) to fit the memory size of the GPU (NVIDIA A100 40GB) we used. Cropping is used only when training the models.

Protein Docking Graph Representation. We represent the initial docking pose as a fully connected graph. Each residue is represented as a node with an amino-acid identity. Each pair of nodes is connected by a unique edge capturing relations between the nodes. For node embeddings, we use the ESM-2 (650M) pre-trained language model [5], which inputs a protein sequence and outputs per-residue embeddings. The embeddings from the language model have been shown to improve downstream structure prediction tasks such as protein folding [59, 5]. For edge embeddings, we adopt the relative spatial encodings from Ingraham et al. [60] constructed by SE(3)-invariant distance, direction, and orientation between pairwise residues, and the relative positional encodings from Evans et al. [54] representing distances between residues in the sequence and identifying if residues are on the same docking partner or not. The graph representation contains rich structural information about the docking pose without any task-specific geometric or hand-crafted chemical features.

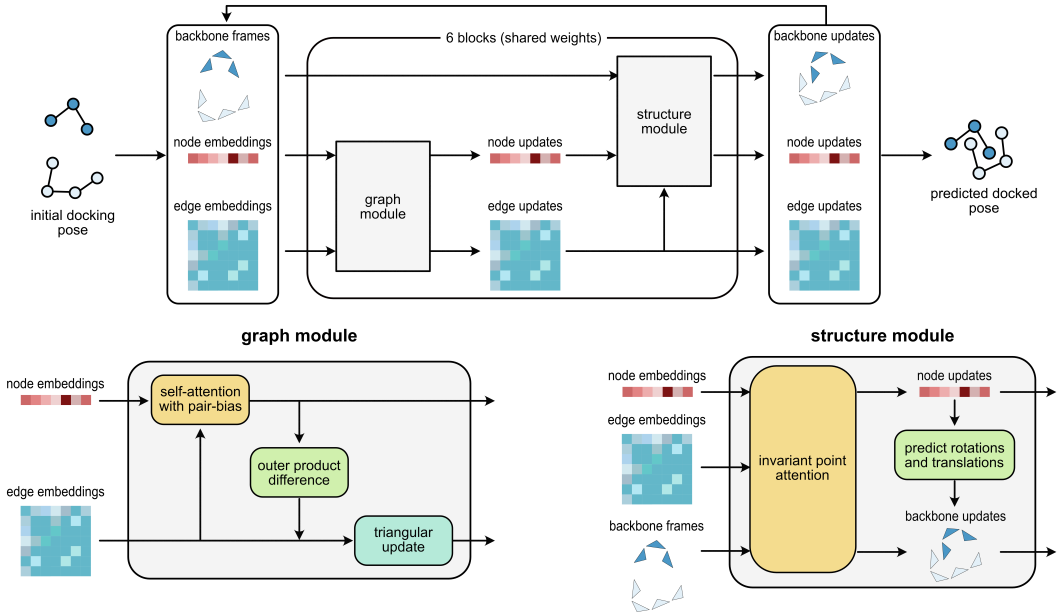

Figure 1: Diagram of the iterative transformer network for predicting a docked pose from a randomized initial docking pose.

Protein Backbone Frames. Along with the graph representation, we represent the 3D atom coordinates of a docking pose as backbone frames. The representation is adopted from the Structure Module in AlphaFold2 [2]. Each residue is represented as a backbone frame encoding the Euclidean transform from the origin frame. The Euclidean transform comprises a 3D rotation matrix and a 3D translation vector. The rotation matrix is computed using the Gram-Schmidt process with the backbone atom (N, C_α, C) coordinates. The translation vector is defined as the atom coordinate of C_α. The backbone frames map each residue from the origin frame to the current locations and are updated through each iteration of the transformer network.

Iterative Transformer Network. We consider docking as an iterative update process from a randomized initial docking pose to a refined docked pose. We update the pose via an iterative transformer network as shown in Figure 1. Inspired by the Evoformer and the Structure Module in AlphaFold2 [2], the network comprises a graph module followed by a structure module. The graph module updates the node and edge embeddings using self-attention with pair-bias and a triangular update module. The structure module updates the node embedding by an invariant point attention (IPA) layer and predicts the rotations and translations for updating the backbone frames. The network inputs the graph representation (node and edge embeddings) along with the backbone frames from the previous iteration, and it outputs their updates. We iterate the network six times with shared weights. The number of trainable parameters is around 2M. The last iteration outputs a predicted docked pose and is compared against the ground truth docked pose by the masked FAPE loss described below.

Masked FAPE Loss. We adopt the FAPE loss used in AlphaFold2 [2]. FAPE loss calculates the deviation of predicted coordinates from the ground truth coordinates when aligning each one of the predicted backbone frames to the corresponding ground truth backbone frame. In the original paper, FAPE loss is applied to all pairs of residues. Here we modify to only consider the loss for residue pairs with C_α within 12 Å. This choice emphasizes the accurate positioning of residue pairs within the binding interfaces and de-emphasizes the non-interacting residue pairs.

3 Results

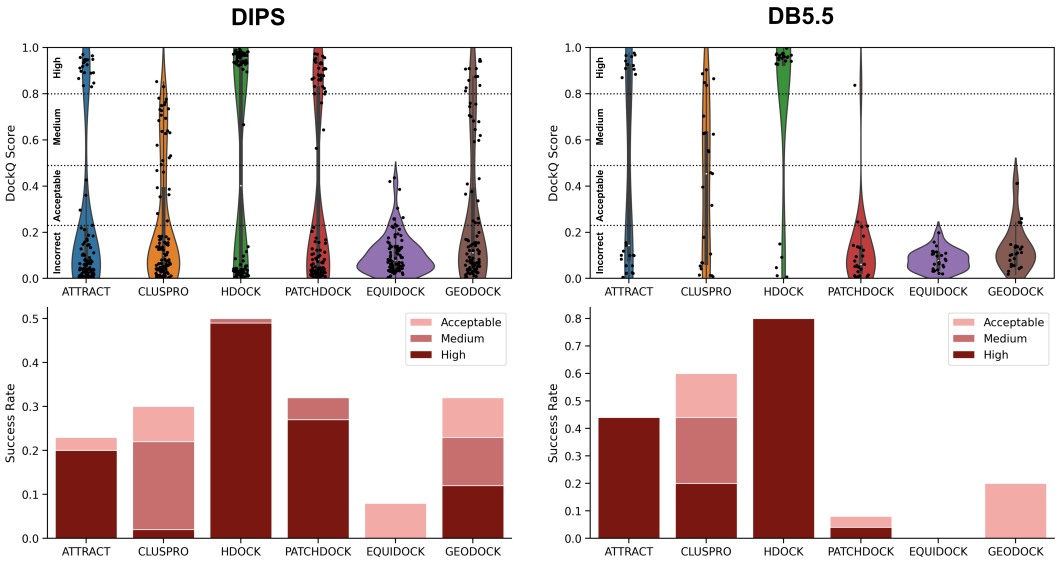

Figure 2: DIPS and DB5.5 (bound) test sets results of DockQ score distributions and docking success rates with CAPRI acceptable, medium, or high quality ranking. The success rates are calculated as the fraction of cases within a specific range of DockQ scores. Scores above 0.23 are considered acceptable, scores between 0.49 and 0.80 are considered medium quality, and scores above 0.80 are considered high quality [61].

Training Details. We train our models on the DIPS training dataset first, using Adam with learning rate 10^{-4} and weight decay 10^{-6}. The best DIPS validated model was then tested on the DIPS test set. For DB5.5, we fine-tuned the model on the DB5.5 training set using the same optimization settings. The best DB5.5 validated model was finally tested on the DB5.5 test set.

Baselines and Evaluation. We compare GeoDock with the existing methods compared in the EquiDock paper. Each method belongs to one of the categories mentioned in the introduction. ClusPro [16] is an exhaustive global search method using FFT; PatchDock [21] is a local shape feature matching approach; ATTRACT [31] is a randomized search docking algorithm; HDOCK [62] is a template-based docking method; EquiDock [1] is a deep learning rigid-body docking approach. Following EquiDock, we only used the bound conformation for both DIPS and DB5.5 test sets. The predicted protein complex structures of these methods are taken from the EquiDock GitHub repository. We evaluate models using DockQ score [61], which evaluates the quality of the interface

and yields a score between [0, 1]. To be consistent with the EquiDock paper, we only use Cα coordinates to calculate the DockQ scores. Following the practice in the CAPRI blind prediction challenge [63], we use a DockQ score threshold of 0.23 to count as a successful dock, corresponding to an "acceptable" CAPRI ranking.

Docking Results. Results on both the DIPS and DB5.5 (bound) sets are shown in Figure 2. We note that in the EquiDock paper, they used average interface root mean square deviation (IRMSD) as an evaluation metric, which is only one component of the DockQ scores and CAPRI assessments. Although EquiDock achieved a better average IRMSD compared to the other methods, the DockQ success rate of EquiDock is the lowest. From the distribution of DockQ score, ATTRACT, HDOCK and PatchDock show bimodal distributions, clustering around low (failure) and high (success) DockQ scores. ClusPro and GeoDock have more continuous distributions, with tails extending from medium to high DockQ scores. EquiDock has the largest cluster at low DockQ scores, which indicates the predictions are generally classified as incorrect. HDOCK scores on DB5.5 are high (80% success); Since HDOCK is a template based docking mehtod, we suspect this accuracy results from the DB5.5 targets being in the HDOCK's template set. With the same training dataset as EquiDock, our method outperforms EquiDock (Appendix Figure 4) and achieves comparable docking success rates to the conventional docking methods on both DIPS and DB5.5 test set.

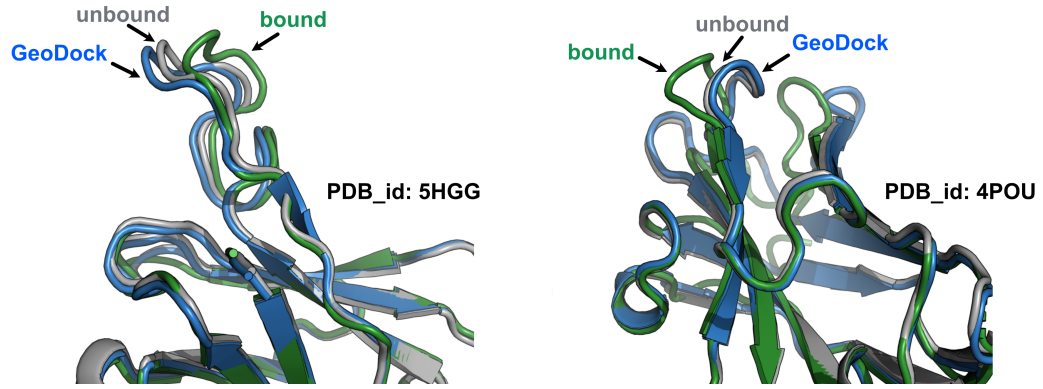

Figure 3: Comparison of the unbound structure (grey) superimposed over the bound (green), and the GeoDock predicted structure (blue).

Backbone Flexibility. We show in Figure 3 that GeoDock is able to move the backbone at the flexible regions. The current GeoDock model pushes the backbone away from the bound state similar to previous conformer selection methods [64]. We will next fine-tuning our model with an ensemble of sampled backbone conformations (e.g. from ReplicaDock2 [29]) to test whether the model can better capture the unbound to bound conformational changes.

Inference Speed. The inference times of GeoDock are 7 ± 7 seconds for the DIPS test set (residue sizes 100-1,600) and 8 ± 15 seconds for the DB5.5 test set (residue sizes 200-2,500) on a 48-core CPU, which is close to EquiDock (5 ± 5 and 5 ± 10 seconds) and is between 80-500 times faster than the other methods [1]. Inference time is important for applications that require high throughput protein complex structure prediction, such as structure-based virtual screening.

4 Conclusion

We have presented GeoDock, a fast, end-to-end protein docking approach with an iterative transformer network. Our method outperforms EquiDock and achieves comparable success rates to the conventional docking methods. Despite the model being currently only trained on the bound protein complexes, it moves the backbone and can be extended to flexible docking with adequate fine-tuning on datasets with conformational changes upon binding. This is particularly important for cases such as allosteric inhibition or docking from uncertain unbound model structures, where the flexible regions on the proteins hinder the success of rigid docking. For target-specific protein binder design, e.g., screening antibodies for a specific antigen, a high-throughput docking algorithm is usually required to scan over a vast search space of potential hits. With further development, GeoDock can serve as a fast and flexible protein-protein docking tool and facilitate the design of protein binders and drugs for a wide variety of targets.

References

[1] Octavian-Eugen Ganea, Xinyuan Huang, Charlotte Bunne, Yatao Bian, Regina Barzilay, Tommi Jaakkola, and Andreas Krause. Independent se (3)-equivariant models for end-to-end rigid protein docking. *arXiv preprint arXiv:2111.07786*, 2021.

[2] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, 2021.

[3] Minkyung Baek, Frank DiMaio, Ivan Anishchenko, Justas Dauparas, Sergey Ovchinnikov, Gyu Rie Lee, Jue Wang, Qian Cong, Lisa N Kinch, R Dustin Schaeffer, et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, 373(6557):871–876, 2021.

[4] Ruidong Wu, Fan Ding, Rui Wang, Rui Shen, Xiwen Zhang, Shitong Luo, Chenpeng Su, Zuofan Wu, Qi Xie, Bonnie Berger, et al. High-resolution de novo structure prediction from primary sequence. *bioRxiv*, 2022.

[5] Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Allan dos Santos Costa, Maryam Fazel-Zarandi, Tom Sercu, Sal Candido, et al. Language models of protein sequences at the scale of evolution enable accurate structure prediction. *bioRxiv*, 2022.

[6] Guoxia Wang, Xiaomin Fang, Zhihua Wu, Yiqun Liu, Yang Xue, Yingfei Xiang, Dianhai Yu, Fan Wang, and Yanjun Ma. Helixfold: An efficient implementation of alphafold2 using paddlepaddle. *arXiv preprint arXiv:2207.05477*, 2022.

[7] Mihaly Varadi, Stephen Anyango, Mandar Deshpande, Sreenath Nair, Cindy Natassia, Galabina Yordanova, David Yuan, Oana Stroe, Gemma Wood, Agata Laydon, et al. Alphafold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic acids research*, 50(D1):D439–D444, 2022.

[8] Rolf Apweiler, Amos Bairoch, Cathy H Wu, Winona C Barker, Brigitte Boeckmann, Serenella Ferro, Elisabeth Gasteiger, Hongzhan Huang, Rodrigo Lopez, Michele Magrane, et al. Uniprot: the universal protein knowledgebase. *Nucleic acids research*, 32(suppl_1):D115–D119, 2004.

[9] Sheng-You Huang. Search strategies and evaluation in protein–protein docking: principles, advances and challenges. *Drug discovery today*, 19(8):1081–1096, 2014.

[10] Henry A Gabb, Richard M Jackson, and Michael JE Sternberg. Modelling protein docking using shape complementarity, electrostatics and biochemical information. *Journal of molecular biology*, 272(1):106–120, 1997.

[11] Ilya A Vakser. Evaluation of gramm low-resolution docking methodology on the hemagglutinin-antibody complex. *Proteins: Structure, Function, and Bioinformatics*, 29(S1):226–230, 1997.

[12] Jeffrey G Mandell, Victoria A Roberts, Michael E Pique, Vladimir Kotlovyi, Julie C Mitchell, Erik Nelson, Igor Tsigelny, and Lynn F Ten Eyck. Protein docking using continuum electrostatics and geometric fit. *Protein engineering*, 14(2):105–113, 2001.

[13] Rong Chen and Zhiping Weng. Docking unbound proteins using shape complementarity, desolvation, and electrostatics. *Proteins: Structure, Function, and Bioinformatics*, 47(3): 281–294, 2002.

[14] Rong Chen, Li Li, and Zhiping Weng. Zdock: an initial-stage protein-docking algorithm. *Proteins: Structure, Function, and Bioinformatics*, 52(1):80–87, 2003.

[15] Alexander Heifetz, Ephraim Katchalski-Katzir, and Miriam Eisenstein. Electrostatics in protein–protein docking. *Protein Science*, 11(3):571–587, 2002.

[16] Stephen R Comeau, David W Gatchell, Sandor Vajda, and Carlos J Camacho. Cluspro: a fully automated algorithm for protein–protein docking. *Nucleic acids research*, 32(suppl_2): W96–W99, 2004.

[17] Dima Kozakov, Ryan Brenke, Stephen R Comeau, and Sandor Vajda. Piper: an fft-based protein docking program with pairwise potentials. *Proteins: Structure, Function, and Bioinformatics*, 65(2):392–406, 2006.

[18] Ephraim Katchalski-Katzir, Isaac Shariv, Miriam Eisenstein, Asher A Friesem, Claude Aflalo, and Ilya A Vakser. Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. *Proceedings of the National Academy of Sciences*, 89(6):2195–2199, 1992.

[19] Irwin D Kuntz, Jeffrey M Blaney, Stuart J Oatley, Robert Langridge, and Thomas E Ferrin. A geometric approach to macromolecule-ligand interactions. *Journal of molecular biology*, 161 (2):269–288, 1982.

[20] Dina Duhovny, Ruth Nussinov, and Haim J Wolfson. Efficient unbound docking of rigid molecules. In *International workshop on algorithms in bioinformatics*, pages 185–200. Springer, 2002.

[21] Dina Schneidman-Duhovny, Yuval Inbar, Ruth Nussinov, and Haim J Wolfson. Patchdock and symmdock: servers for rigid and symmetric docking. *Nucleic acids research*, 33(suppl_2): W363–W367, 2005.

[22] Vishwesh Venkatraman, Yifeng D Yang, Lee Sael, and Daisuke Kihara. Protein-protein docking using region-based 3d zernike descriptors. *BMC bioinformatics*, 10(1):1–21, 2009.

[23] Juan Esquivel-Rodríguez, Yifeng David Yang, and Daisuke Kihara. Multi-lzerd: multiple protein docking for asymmetric complexes. *Proteins: Structure, Function, and Bioinformatics*, 80(7):1818–1833, 2012.

[24] Eleanor J Gardiner, Peter Willett, and Peter J Artymiuk. Protein docking using a genetic algorithm. *Proteins: Structure, Function, and Bioinformatics*, 44(1):44–56, 2001.

[25] Jeffrey J Gray, Stewart Moughon, Chu Wang, Ora Schueler-Furman, Brian Kuhlman, Carol A Rohl, and David Baker. Protein–protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *Journal of molecular biology*, 331(1):281–299, 2003.

[26] Chu Wang, Philip Bradley, and David Baker. Protein–protein docking with backbone flexibility. *Journal of molecular biology*, 373(2):503–519, 2007.

[27] Sidhartha Chaudhury and Jeffrey J Gray. Conformer selection and induced fit in flexible backbone protein–protein docking using computational and nmr ensembles. *Journal of molecular biology*, 381(4):1068–1087, 2008.

[28] Nicholas A Marze, Shourya S Roy Burman, William Sheffler, and Jeffrey J Gray. Efficient flexible backbone protein–protein docking for challenging targets. *Bioinformatics*, 34(20): 3461–3469, 2018.

[29] Ameya Harmalkar, Sai Pooja Mahajan, and Jeffrey J Gray. Induced fit with replica exchange improves protein complex structure prediction. *PLOS Computational Biology*, 18(6):e1010124, 2022.

[30] Martin Zacharias. Protein–protein docking with a reduced protein model accounting for side-chain flexibility. *Protein Science*, 12(6):1271–1282, 2003.

[31] Martin Zacharias. Attract: protein–protein docking in capri using a reduced protein model. *Proteins: Structure, Function, and Bioinformatics*, 60(2):252–256, 2005.

[32] Karine Bastard, Chantal Prévost, and Martin Zacharias. Accounting for loop flexibility during protein–protein docking. *Proteins: Structure, Function, and Bioinformatics*, 62(4):956–969, 2006.

[33] Cyril Dominguez, Rolf Boelens, and Alexandre MJJ Bonvin. Haddock: a protein- protein docking approach based on biochemical or biophysical information. *Journal of the American Chemical Society*, 125(7):1731–1737, 2003.

[34] Sjoerd J De Vries, Aalt DJ Van Dijk, Mickaël Krzeminski, Mark van Dijk, Aurelien Thureau, Victor Hsu, Tsjerk Wassenaar, and Alexandre MJJ Bonvin. Haddock versus haddock: new features and performance of haddock2. 0 on the capri targets. *Proteins: structure, function, and bioinformatics*, 69(4):726–733, 2007.

[35] Ruben Abagyan, Maxim Totrov, and Dmitry Kuznetsov. Icm—a new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. *Journal of computational chemistry*, 15(5):488–506, 1994.

[36] Juan Fernández-Recio, Maxim Totrov, and Ruben Abagyan. Icm-disco docking by global energy optimization with fully flexible side-chains. *Proteins: Structure, Function, and Bioinformatics*, 52(1):113–117, 2003.

[37] Iain H Moal and Paul A Bates. Swarmdock and the use of normal modes in protein-protein docking. *International journal of molecular sciences*, 11(10):3623–3648, 2010.

[38] Andras Szilagyi and Yang Zhang. Template-based structure modeling of protein–protein interactions. *Current opinion in structural biology*, 24:10–23, 2014.

[39] Pablo Gainza, Freyr Sverrisson, Frederico Monti, Emanuele Rodola, D Boscaini, MM Bronstein, and BE Correia. Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nature Methods*, 17(2):184–192, 2020.

[40] Jérôme Tubiana, Dina Schneidman-Duhovny, and Haim J Wolfson. Scannet: An interpretable geometric deep learning model for structure-based protein binding site prediction. *Nature Methods*, pages 1–10, 2022.

[41] Alex Fout, Jonathon Byrd, Basir Shariat, and Asa Ben-Hur. Protein interface prediction using graph convolutional networks. *Advances in neural information processing systems*, 30, 2017.

[42] Raphael Townshend, Rishi Bedi, Patricia Suriana, and Ron Dror. End-to-end learning on 3d protein structure for interface prediction. *Advances in Neural Information Processing Systems*, 32, 2019.

[43] Yi Liu, Hao Yuan, Lei Cai, and Shuiwang Ji. Deep learning of high-order interactions for protein interface prediction. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 679–687, 2020.

[44] Ziwei Xie and Jinbo Xu. Deep graph learning of inter-protein contacts. *Bioinformatics*, 38(4): 947–953, 2022.

[45] Bowen Dai and Chris Bailey-Kellogg. Protein interaction interface region prediction by geometric deep learning. *Bioinformatics*, 37(17):2580–2588, 2021.

[46] Farhan Quadir, Raj S Roy, Randal Halfmann, and Jianlin Cheng. Dncon2_inter: predicting interchain contacts for homodimeric and homomultimeric protein complexes using multiple sequence alignments of monomers and deep learning. *Scientific reports*, 11(1):1–10, 2021.

[47] Yumeng Yan and Sheng-You Huang. Accurate prediction of inter-protein residue–residue contacts for homo-oligomeric protein complexes. *Briefings in bioinformatics*, 22(5):bbab038, 2021.

[48] Raj S Roy, Farhan Quadir, Elham Soltanikazemi, and Jianlin Cheng. A deep dilated convolutional residual network for predicting interchain contacts of protein homodimers. *Bioinformatics*, 38(7):1904–1910, 2022.

[49] Freyr Sverrisson, Jean Feydy, Joshua Southern, Michael M Bronstein, and Bruno Correia. Physics-informed deep neural network for rigid-body protein docking. In *ICLR2022 Machine Learning for Drug Discovery*, 2022.

[50] Elham Soltanikazemi, Raj S Roy, Farhan Quadir, Nabin Giri, Alex Morehead, and Jianlin Cheng. Drlcomplex: Reconstruction of protein quaternary structures using deep reinforcement learning. *arXiv preprint arXiv:2205.13594*, 2022.

[51] Usman Ghani, Israel Desta, Akhil Jindal, Omeir Khan, George Jones, Nasser Hashemi, Sergey Kotelnikov, Dzmitry Padhorny, Sandor Vajda, and Dima Kozakov. Improved docking of protein models by a combination of alphafold2 and cluspro. *BioRxiv*, pages 2021–09, 2022.

[52] Milot Mirdita, Konstantin Schütze, Yoshitaka Moriwaki, Lim Heo, Sergey Ovchinnikov, and Martin Steinegger. Colabfold: making protein folding accessible to all. *Nature Methods*, pages 1–4, 2022.

[53] Mu Gao, Davi Nakajima An, Jerry M Parks, and Jeffrey Skolnick. Af2complex predicts direct physical interactions in multimeric proteins with deep learning. *Nature communications*, 13(1): 1–13, 2022.

[54] Richard Evans, Michael O'Neill, Alexander Pritzel, Natasha Antropova, Andrew W Senior, Timothy Green, Augustin Žídek, Russell Bates, Sam Blackwell, Jason Yim, et al. Protein complex prediction with alphafold-multimer. *BioRxiv*, 2021.

[55] Rui Yin, Brandon Y Feng, Amitabh Varshney, and Brian G Pierce. Benchmarking alphafold for protein complex modeling reveals accuracy determinants. *Protein Science*, 31(8):e4379, 2022.

[56] Freyr Sverrisson, Jean Feydy, Bruno E Correia, and Michael M Bronstein. Fast end-to-end learning on protein surfaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15272–15281, 2021.

[57] Thom Vreven, Iain H Moal, Anna Vangone, Brian G Pierce, Panagiotis L Kastritis, Mieczyslaw Torchala, Raphael Chaleil, Brian Jiménez-García, Paul A Bates, Juan Fernandez-Recio, et al. Updates to the integrated protein–protein interaction benchmarks: docking benchmark version 5 and affinity benchmark version 2. *Journal of molecular biology*, 427(19):3031–3041, 2015.

[58] Johnathan D Guest, Thom Vreven, Jing Zhou, Iain Moal, Jeliazko R Jeliazkov, Jeffrey J Gray, Zhiping Weng, and Brian G Pierce. An expanded benchmark for antibody-antigen docking and affinity prediction reveals insights into antibody recognition determinants. *Structure*, 29(6): 606–621, 2021.

[59] Alexander Rives, Joshua Meier, Tom Sercu, Siddharth Goyal, Zeming Lin, Jason Liu, Demi Guo, Myle Ott, C Lawrence Zitnick, Jerry Ma, et al. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proceedings of the National Academy of Sciences*, 118(15):e2016239118, 2021.

[60] John Ingraham, Vikas Garg, Regina Barzilay, and Tommi Jaakkola. Generative models for graph-based protein design. *Advances in neural information processing systems*, 32, 2019.

[61] Sankar Basu and Björn Wallner. Dockq: a quality measure for protein-protein docking models. *PloS one*, 11(8):e0161879, 2016.

[62] Yumeng Yan, Di Zhang, Pei Zhou, Botong Li, and Sheng-You Huang. Hdock: a web server for protein–protein and protein–dna/rna docking based on a hybrid strategy. *Nucleic acids research*, 45(W1):W365–W373, 2017.

[63] Joël Janin, Kim Henrick, John Moult, Lynn Ten Eyck, Michael JE Sternberg, Sandor Vajda, Ilya Vakser, and Shoshana J Wodak. Capri: a critical assessment of predicted interactions. *Proteins: Structure, Function, and Bioinformatics*, 52(1):2–9, 2003.

[64] Daisuke Kuroda and Jeffrey J Gray. Pushing the backbone in protein-protein docking. *Structure*, 24(10):1821–1829, 2016.

5 Appendix

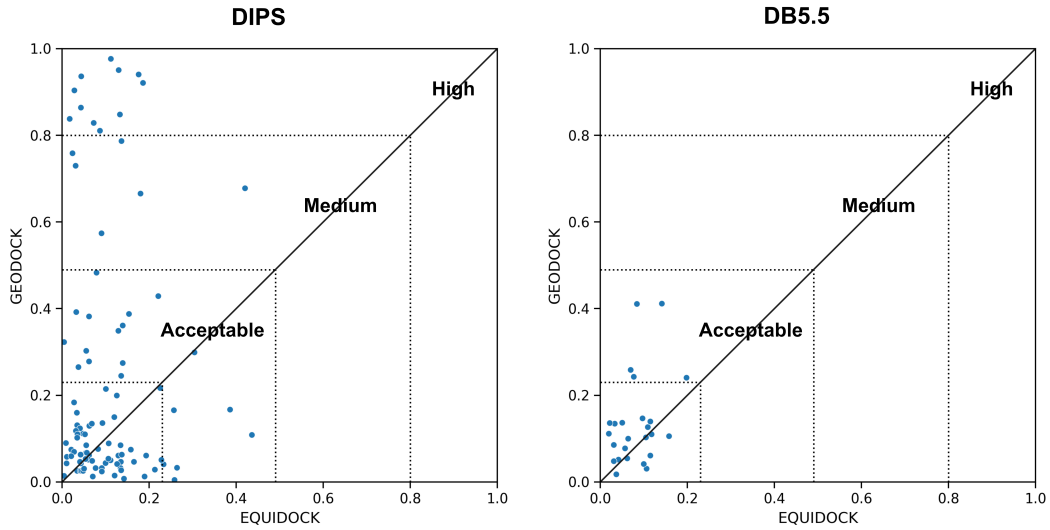

Figure 4: Comparison of DockQ scores of GeoDock versus EquiDock on the DIPS and DB5.5 test sets.

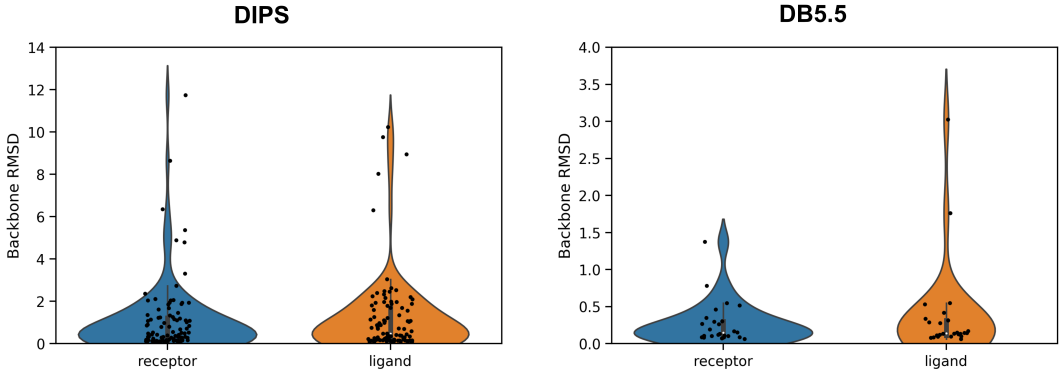

Figure 5: Receptor and ligand backbone RMSD (input versus prediction) distributions of GeoDock on the DIPS and DB5.5 (unbound) test sets.

To demonstrate our method is capable of predicting binding-induced conformational changes, we show in Figure 5 the receptor and ligand backbone RMSD distributions respectively on both DIPS and DB5.5 (unbound) test sets. The backbone RMSD measures the conformational difference between the input and predicted structures of the individual proteins (receptor and ligand) from a docking pair. For the rigid docking methods, the backbone RMSD is always zero.