

Flexible Protein-Protein Docking with a Multi-Track Iterative Transformer

Lee-Shin Chu¹ Jeffrey A. Ruffolo² Ameya Harmalkar¹ Jeffrey J. Gray^{1,2}

¹Department of Chemical and Biomolecular Engineering

²Program in Molecular Biophysics

Johns Hopkins University, Baltimore, MD 21218, USA.

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 these limitations, we present GeoDock, a multi-track iterative transformer network to predict a docked structure from separate docking partners. Unlike deep learning models for protein structure prediction that input multiple sequence alignments (MSAs), GeoDock inputs just the sequences and structures of the docking partners, which suits the tasks when the individual structures are given. GeoDock is flexible at the protein residue level, allowing the prediction of conformational changes upon binding. For a benchmark set of rigid targets, GeoDock obtains a 41% success rate, outperforming all the other tested methods. For a more challenging benchmark set of flexible targets, GeoDock achieves a similar number of top-model successes as the traditional method ClusPro [1], but fewer than ReplicaDock2 [2]. GeoDock attains an average inference speed of under one second on a single GPU, enabling its application in large-scale structure screening. Although binding-induced conformational changes are still a challenge owing to limited training and evaluation data, our architecture sets up the foundation to capture this backbone flexibility. Code and a demonstration Jupyter notebook are available at <https://github.com/Graylab/GeoDock>.

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 [3] and other follow-up works [4–7] have demonstrated the prediction of 3D models of protein structures comparable to experimental accuracy. Along with genome-wide protein sequencing, the AlphaFold database [8] provides open access to 214M protein structure predictions from the sequences deposited in UniProt [9]. To provide proper biological context to the expanding number of predicted 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 (1) a sampling algorithm that generates ensembles of candidate docked structures and (2) a scoring function that evaluates the candidates generated from the sampling stage. The sampling strategies can be classified into exhaustive global search, local shape feature matching, or randomized search [10]. The exhaustive global search methods [11–16, 1, 17] mostly use fast Fourier transforms (FFTs) [18] to cover 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 [25–28, 2, 29–36] use stochastic algorithms such as Monte Carlo methods to explore docking poses through the free energy landscape starting from a randomized initial pose, with the protein represented as an all-atom or coarse-grain model. 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 [37] 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 [38] and ScanNet [39] 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 [40–47] 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 [48, 49], but the resulting complex structures have not yet achieved 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 [50–52] extend the ability of AlphaFold2 [3] that was originally designed for protein folding to protein multimer structure prediction. Remarkably, AlphaFold-Multimer (AF-M) [53] 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 [54].

Several deep learning methods have been developed for fast rigid docking. Ganea *et al* [55] 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* [48] developed a generative model based on the interface features derived from dMaSIF [56] to produce ensembles of docking candidates that are then scored by a trained discriminative model. 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. More recently, McPartlon & Xu [57] proposed DockGPT for flexible and site-specific protein docking and design, demonstrating high docking success rates when partial binding site information is provided.

In this work, we develop GeoDock, a multi-track iterative transformer network for fast and flexible protein-protein docking. We use the ESM language model [6] to embed the sequences, thereby avoiding MSAs during training. Our approach involves a multi-track architecture inspired by AlphaFold2 [3], and it incorporates the input structure information and the invariant point attention (IPA) module to enable backbone motion. We demonstrate a superior performance of the docking success rate on a benchmark dataset consisting of bound docking tasks. We measure the extent of backbone motion the network can capture, and we highlight limited success on a benchmark set of unbound docking tasks.

2 Methods

Datasets. For training, fine-tuning, and evaluating our models, we used datasets from (1) Database of Interacting Protein Structures (DIPS) [41] and (2) Docking Benchmark 5.5 (DB5.5) [58, 59]. DIPS comprises 42,826 non-redundant experimentally resolved binary protein complexes (< 30% sequence identity to any protein in the DB5.5). We filtered out the complexes with less than 50 residues in a single chain, resulting in train/validation/test splits of size 36318/927/100. DB5.5 contains 271 protein complexes with both bound and unbound conformations. We curated two subsets from DB5.5 for evaluation. The first subset (DB5.5 bound) consists of 24 complexes that were used as the test set in the EquiDock [55] report. The second subset (DB5.5 unbound) comprises 69 complexes that are classified as medium difficulty or difficult targets. We used the remaining 178 complexes for fine-tuning the model pre-trained on DIPS.

Continuous Cropping for Model Training. We randomly cropped a continuous block of residues on each chain up to 500 residues in total (summing over the chains) to fit the memory size of the GPU (NVIDIA A100 40GB) we used. The cropping strategy ensures that the cropped structure still has two continuous chains, and the length of the chains always sums to 500. Cropping is used for training the models only but not for inference.

Input Protein Graphs. We represent the input docking partners as two separate graphs. Each graph is fully-connected, with nodes representing the amino-acid identity and edges representing the geometric relations between the intra-chain residues. For node embeddings, we used the ESM-2 (650M) pre-trained language model [6], 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 [60, 6]. For edge embeddings, we used the inter-residue distance and orientation distributions adopted from trRosetta [61] and the relative positional encodings from Evans *et al* [53]. The two separate graphs from the two docking partners are concatenated before input to the neural network, with the inter-chain information masked. The graph representation contains rich geometric information without any task-specific or hand-crafted features.

Inter-chain Contact Information for Model Training. Inspired by the work from McPartlon & Xu [57], we provided a small fraction of inter-chain contact (distance between C_β atoms under 10 Å) information as a binary embedding $C_{ij} \in \{0, 1\}$ concatenated to the edge embeddings during training. Specifically, for each training example, we generated a random integer $i \in [0, 3]$ and randomly sampled i inter-chain contacts. Thus, during training, the model sees 0 to 3 contacts with equal probability. During inference for the blind docking tests, no inter-chain contact information is provided ($C_{ij} = 0$).

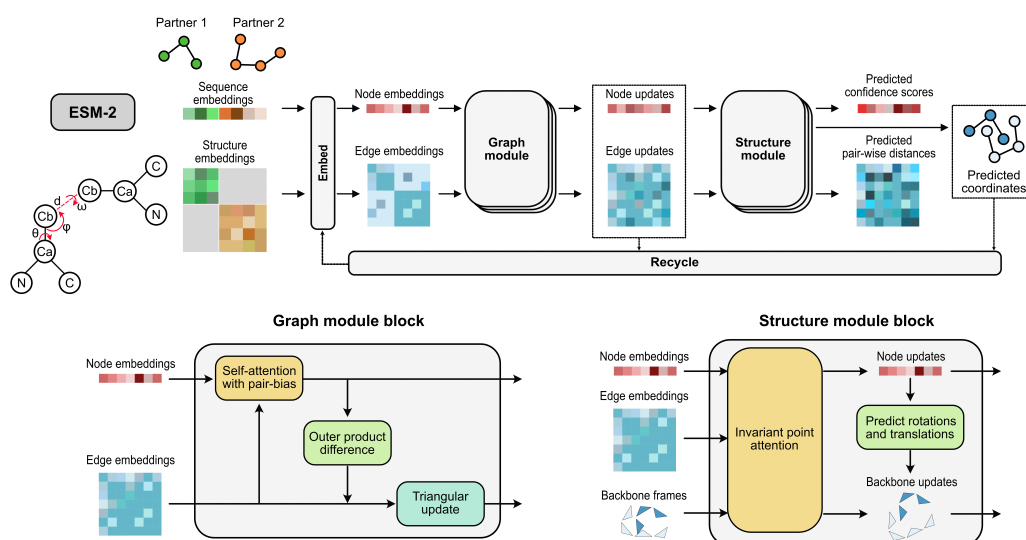


Figure 1: Diagram of the multi-track iterative transformer network for predicting a docked protein complex structure from input docking partners.

Multi-Track Iterative Transformer Network. We predict a docked protein complex structure from the input docking partners end-to-end via a multi-track iterative transformer network as shown in Figure 1. Inspired by the model architecture of AlphaFold2 [3], the network comprises a graph module followed by a structure module. The graph module communicates 1D (node embeddings) and 2D (edge embeddings) information and outputs the updated embeddings. The structure module inputs the updated embeddings from the graph module along with the backbone frames and outputs the predicted rotations and translations for updating the backbone frames. The main difference between our network and AlphaFold2 is that we input sequences and structures instead of MSAs. In addition, the network is much shallower compared to AlphaFold2, with only 3 layers for both the graph and structure modules. During training, we iterate the network 4 times with no gradient except for the last iteration. During each iteration, we recycle the outputs from both the graph and structure modules. Our implementation of recycling differs from AlphaFold2, as we additionally recycle the inter-residue orientation distribution from the output backbone frames of the structure module. The number of trainable parameters is around 4.3M. The last iteration outputs the predicted backbone frames, pair-wise distance bins, and per-residue confidence scores.

Loss Functions. To train our neural network, we used intra-chain and inter-chain frame-alignment point error (FAPE), pairwise distance loss, the C_α local distance difference test (IDDT- C_α) [62] loss, and structure violation loss. The loss functions are largely adopted from AlphaFold-Multimer [53], but we omit the cross-entropy loss for sequences, as we do not input MSAs. We also exclude the within-residue violation term in the structure violation loss because our network does not predict side-chain atoms.

Training Stages. We found that including the inter-chain FAPE at the beginning hinders the convergence of the intra-chain FAPE, as the gradients from the two loss functions compete. Therefore, we trained our network in three stages with gradually increasing loss terms. In the first stage, we only include the intra-chain FAPE, pair-wise distance loss, and IDDT- C_α loss. In the second stage, we add the inter-chain FAPE. At the last stage, we add the structure violation loss. With this training procedure, the neural network initially learns how to position the intra-chain residues, and subsequently acquires the ability to assemble the docking partners. Our model was first trained on the DIPS training set for 50 epochs, using Adam with a learning rate 5×10^{-5} . The best DIPS-validated model was then fine-tuned on the DB5.5 training set alone with the same optimization settings.

3 Results

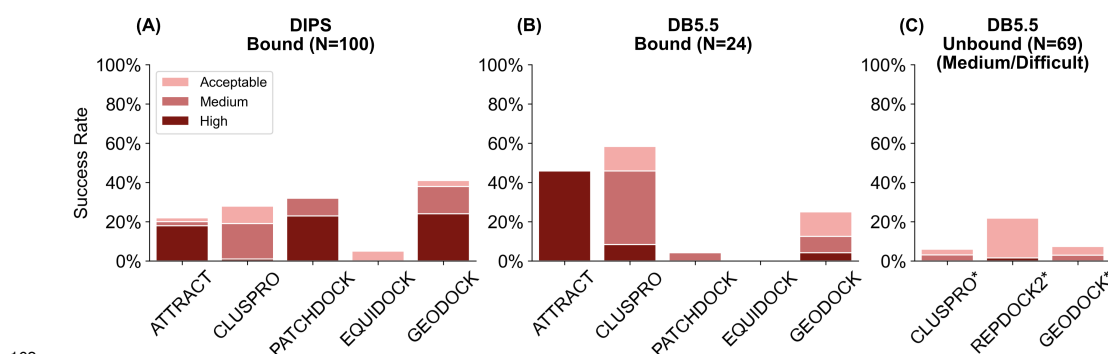


Figure 2: (A) DIPS, (B) DB5.5 bound, and (C) DB5.5 unbound test sets results of docking success rates with CAPRI acceptable, medium, or high quality ranking. The success rates are calculated as the fraction of cases within a specific range of DockQ scores. Scores above 0.23 are considered acceptable, scores between 0.49 and 0.80 are considered medium quality, and scores above 0.80 are considered high quality [63]. * indicates inference starting from the unbound conformation.

Evaluation. We evaluate the predicted structures using the DockQ score [63], which assesses the structural quality of the interface and provides a score ranging from 0 to 1. Following the practice in the CAPRI blind prediction challenge [64], we use a DockQ score threshold of 0.23 to count as a successful dock, corresponding to an "acceptable" CAPRI ranking.

Bound-conformation Docking Results. For the bound docking task, we compared GeoDock with the existing methods mentioned in EquiDock [55]. Each method belongs to one of the categories specified in the introduction. ClusPro [1] is an exhaustive global search method using FFT; PatchDock [21] is a local shape feature matching approach; ATTRACT [30] is energy minimization with a reduced protein model; EquiDock [55] is a deep learning rigid-body docking approach. The predicted protein complex structures of these methods are taken from the EquiDock GitHub repository https://github.com/octavian-ganea/equidock_public. We excluded HDOCK [65] in our comparison, as it is a template-based docking method and the test targets are in HDOCK's template set. We cannot compare GeoDock to the dMaSIF rigid docking method [48] since they have not reported the per-target docking success rate, and their code is not accessible. We also do not compare with AlphaFold-Multimer [53] due to the presence of test targets within their training sets. We employed the best DIPS-validated model and the DB5.5 fine-tuned model to evaluate DIPS and DB5.5 held-out test sets, respectively. Our method GeoDock successfully docks 41% of the targets in the DIPS test set, outperforming all the other tested methods (Fig 2A). On the DB5.5 bound set of targets tested in EquiDock [55] paper, GeoDock obtains a 25% success rate, outperforming EquiDock and PatchDock but underperforming other traditional methods (Fig 2B). The lower success rate on DB5.5 relative to DIPS suggests that the DIPS training set, which is predominantly homo-multimers, may not adequately cover the range of interface diversity in the DB5.5 set.

Unbound-conformation Docking Results. For the unbound docking task, we compared GeoDock with the conventional methods ClusPro [1] (rigid docking) and ReplicaDock2 [2] (flexible docking). For ClusPro, we used the single top model success rate reported in their paper [66]. For ReplicaDock2, we selected the single top model (ranked by Rosetta interface score) obtained from blind docking runs. We evaluate our method on the DB5.5 unbound set (Fig 2C) of flexible targets (classified as medium or difficult). The success rate of GeoDock performs similarly to the traditional rigid docking method ClusPro (7% vs 6%), demonstrating GeoDock's capability of unbound docking. However, GeoDock underperforms the physics-based flexible docking method ReplicaDock2 (7% vs 22%). Similar to GeoDock, DockGPT [57] reported a blind unbound docking success rate (7%) on a DB5.5 set of 42 targets (classified as rigid, medium, or difficult). The generally lower docking success rates for unbound docking reflect the current difficulty of deep-learning docking approaches to capture backbone conformational changes.

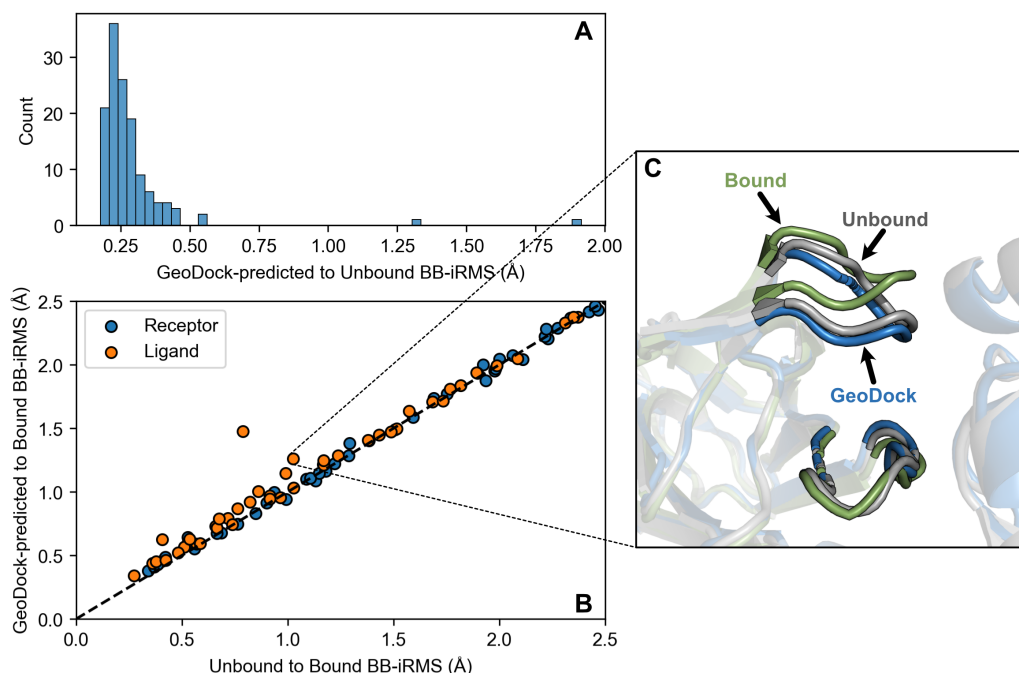


Figure 3: Analysis of backbone flexibility measured with backbone interface RMSD (BB-iRMS) for DB5.5 flexible targets. (A) BB-iRMS distributions of GeoDock-predicted structures to unbound structures. (B) BB-iRMS of GeoDock-predicted structures to bound structures versus unbound to bound. (C) Comparison of the unbound structure (grey) superimposed over the bound (green), and

the GeoDock-predicted structure (blue) for the DsbC-DsbDalpha complex (PDB accession code: 1JZD_AB:C (bound); 1JZO_AB and 1JPE_A (unbound)).

Backbone Flexibility. As shown in Figure 3, GeoDock is able to move the backbone at the flexible regions. The current GeoDock model typically pushes the backbone away from the bound state (Fig 3B) similar to previous conformer selection methods [67]. The observed movement is minimal ($< 1\text{\AA}$ Fig 3A) and the predicted structures closely resemble the input unbound structures (Fig 3C). The current training examples had no conformational change, so a next step might be to fine-tune our model with an ensemble of sampled backbone conformations (e.g. from ReplicaDock2 [2]) to better capture conformational changes from the unbound to bound states.

Inference Speed. The inference times of GeoDock are 0.76 ± 0.62 seconds for the DIPS test set (residue sizes 100-800) and 0.80 ± 0.52 seconds for the DB5.5 test set (residue sizes 200-700) on a single NVIDIA V100 GPU, which is faster than the reported runtimes for EquiDock (5 ± 5 and 5 ± 10 seconds) and is at least 10^3 times faster than the other methods [55]. Inference time is important for applications that require high throughput protein complex structure prediction, such as structure-based virtual screening.

4 Discussion and Conclusion

GeoDock is a fast, end-to-end approach for flexible protein-protein docking with a multi-track iterative transformer network. Unlike previous deep learning methods for protein complex structure prediction that require MSAs as input [3, 4], GeoDock inputs structure and sequence information of individual proteins and predicts a docked protein complex structure, which suits the task when the unbound structures are given. Furthermore, the model allows for backbone movements, which is in contrast to previous deep learning methods that assume no conformational changes [55, 48]. With the same training dataset and faster inference speed, our method GeoDock outperforms EquiDock [55] on both DIPS and DB5.5 test sets. The observed improvement can be attributed to the iterative updating of coordinates, which contrasts with the EquiDock paper's approach of predicting a single transformation for one of the docking partners.

While this method significantly accelerates the inference process, in terms of docking success rates, it does not surpass conventional methods that use sampling and re-ranking strategies. This may be attributed to the training objective of the model not aligning with the evaluation of docking success, which is determined by the percentage of predictions with a DockQ score above a certain threshold. The objective function of GeoDock is to minimize the expected error between the predicted and ground truth structures. Trained with this objective, the model learns to predict structures with low average error, although each individual structure may not necessarily be considered a success case if the error is not sufficiently low. A more viable objective is to identify a low-error docking pose distribution and sample from it, which can be framed as a generative model similar to a prior study on the small molecule docking using diffusion model [68]. This suggests a future research direction of deep generative models in sampling protein docking poses.

We have demonstrated that GeoDock can move the backbone despite being currently trained only on the bound protein complexes. However, the movement is minimal, and the predicted structures closely resemble the input unbound structures. We leave for future work the fine-tuning of the model with sampled backbone conformations to test whether the model can better capture the conformational change 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.

Acknowledgements

This work was supported by National Institutes of Health grants R35-GM141881. Computational resources were provided by the Advanced Research Computing at Hopkins (ARCH).

Code Availability

Code, pre-trained models, and a demonstration Jupyter notebook for GeoDock are available at <https://github.com/Graylab/GeoDock>.

References

- [1] 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.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] Sheng-You Huang. Search strategies and evaluation in protein–protein docking: principles, advances and challenges. *Drug Discovery Today*, 19(8):1081–1096, 2014.
- [11] 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.
- [12] 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.
- [13] Jeffrey G Mandell, Victoria A Roberts, Michael E Pique, Vladimir Kotlovyyi, 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.
- [14] Rong Chen and Zhiping Weng. Docking unbound proteins using shape complementarity, desolvation, and electrostatics. *Proteins: Structure, Function, and Bioinformatics*, 47(3): 281–294, 2002.

- 304 [15] Rong Chen, Li Li, and Zhiping Weng. Zdock: an initial-stage protein-docking algorithm.
305 *Proteins: Structure, Function, and Bioinformatics*, 52(1):80–87, 2003.
- 306 [16] Alexander Heifetz, Ephraim Katchalski-Katzir, and Miriam Eisenstein. Electrostatics in protein–
307 protein docking. *Protein Science*, 11(3):571–587, 2002.
- 308 [17] Dima Kozakov, Ryan Brenke, Stephen R Comeau, and Sandor Vajda. Piper: an fft-based protein
309 docking program with pairwise potentials. *Proteins: Structure, Function, and Bioinformatics*,
310 65(2):392–406, 2006.
- 311 [18] Ephraim Katchalski-Katzir, Isaac Shariv, Miriam Eisenstein, Asher A Friesem, Claude Aflalo,
312 and Ilya A Vakser. Molecular surface recognition: determination of geometric fit between
313 proteins and their ligands by correlation techniques. *Proceedings of the National Academy of
314 Sciences*, 89(6):2195–2199, 1992.
- 315 [19] Irwin D Kuntz, Jeffrey M Blaney, Stuart J Oatley, Robert Langridge, and Thomas E Ferrin. A
316 geometric approach to macromolecule-ligand interactions. *Journal of Molecular Biology*, 161
317 (2):269–288, 1982.
- 318 [20] Dina Duhovny, Ruth Nussinov, and Haim J Wolfson. Efficient unbound docking of rigid
319 molecules. In *International workshop on algorithms in bioinformatics*, pages 185–200. Springer,
320 2002.
- 321 [21] Dina Schneidman-Duhovny, Yuval Inbar, Ruth Nussinov, and Haim J Wolfson. Patchdock and
322 symmdock: servers for rigid and symmetric docking. *Nucleic Acids Research*, 33(suppl_2):
323 W363–W367, 2005.
- 324 [22] Vishwesh Venkatraman, Yifeng D Yang, Lee Sael, and Daisuke Kihara. Protein-protein docking
325 using region-based 3d zernike descriptors. *BMC Bioinformatics*, 10(1):1–21, 2009.
- 326 [23] Juan Esquivel-Rodríguez, Yifeng David Yang, and Daisuke Kihara. Multi-lzard: multiple
327 protein docking for asymmetric complexes. *Proteins: Structure, Function, and Bioinformatics*,
328 80(7):1818–1833, 2012.
- 329 [24] Eleanor J Gardiner, Peter Willett, and Peter J Artymiuk. Protein docking using a genetic
330 algorithm. *Proteins: Structure, Function, and Bioinformatics*, 44(1):44–56, 2001.
- 331 [25] Jeffrey J Gray, Stewart Moughon, Chu Wang, Ora Schueler-Furman, Brian Kuhlman, Carol A
332 Rohl, and David Baker. Protein–protein docking with simultaneous optimization of rigid-body
333 displacement and side-chain conformations. *Journal of Molecular Biology*, 331(1):281–299,
334 2003.
- 335 [26] Chu Wang, Philip Bradley, and David Baker. Protein–protein docking with backbone flexibility.
336 *Journal of Molecular Biology*, 373(2):503–519, 2007.
- 337 [27] Sidhartha Chaudhury and Jeffrey J Gray. Conformer selection and induced fit in flexible
338 backbone protein–protein docking using computational and NMR ensembles. *Journal of
339 Molecular Biology*, 381(4):1068–1087, 2008.
- 340 [28] Nicholas A Marze, Shourya S Roy Burman, William Sheffler, and Jeffrey J Gray. Efficient
341 flexible backbone protein–protein docking for challenging targets. *Bioinformatics*, 34(20):
342 3461–3469, 2018.
- 343 [29] Martin Zacharias. Protein–protein docking with a reduced protein model accounting for side-
344 chain flexibility. *Protein Science*, 12(6):1271–1282, 2003.
- 345 [30] Martin Zacharias. Attract: protein–protein docking in capri using a reduced protein model.
346 *Proteins: Structure, Function, and Bioinformatics*, 60(2):252–256, 2005.
- 347 [31] Karine Bastard, Chantal Prévost, and Martin Zacharias. Accounting for loop flexibility during
348 protein–protein docking. *Proteins: Structure, Function, and Bioinformatics*, 62(4):956–969,
349 2006.

- 350 [32] Cyril Dominguez, Rolf Boelens, and Alexandre MJJ Bonvin. HADDOCK: a protein- protein
351 docking approach based on biochemical or biophysical information. *Journal of the American*
352 *Chemical Society*, 125(7):1731–1737, 2003.
- 353 [33] Sjoerd J De Vries, Aalt DJ Van Dijk, Mickaël Krzeminski, Mark van Dijk, Aurelien Thureau,
354 Victor Hsu, Tsjerk Wassenaar, and Alexandre MJJ Bonvin. HADDOCK versus HADDOCK:
355 new features and performance of HADDOCK2.0 on the capri targets. *Proteins: Structure,*
356 *Function, and Bioinformatics*, 69(4):726–733, 2007.
- 357 [34] Ruben Abagyan, Maxim Totrov, and Dmitry Kuznetsov. Icm—a new method for protein
358 modeling and design: Applications to docking and structure prediction from the distorted native
359 conformation. *Journal of Computational Chemistry*, 15(5):488–506, 1994.
- 360 [35] Juan Fernández-Recio, Maxim Totrov, and Ruben Abagyan. Icm-disco docking by global energy
361 optimization with fully flexible side-chains. *Proteins: Structure, Function, and Bioinformatics*,
362 52(1):113–117, 2003.
- 363 [36] Iain H Moal and Paul A Bates. Swarmdock and the use of normal modes in protein-protein
364 docking. *International Journal of Molecular Sciences*, 11(10):3623–3648, 2010.
- 365 [37] Andras Szilagyi and Yang Zhang. Template-based structure modeling of protein–protein
366 interactions. *Current Opinion in Structural Biology*, 24:10–23, 2014.
- 367 [38] Pablo Gainza, Freyr Sverrisson, Frederico Monti, Emanuele Rodola, D Boscaini, MM Bronstein,
368 and BE Correia. Deciphering interaction fingerprints from protein molecular surfaces using
369 geometric deep learning. *Nature Methods*, 17(2):184–192, 2020.
- 370 [39] Jérôme Tubiana, Dina Schneidman-Duhovny, and Haim J Wolfson. Scannet: An interpretable
371 geometric deep learning model for structure-based protein binding site prediction. *Nature*
372 *Methods*, pages 1–10, 2022.
- 373 [40] Alex Fout, Jonathon Byrd, Basir Shariat, and Asa Ben-Hur. Protein interface prediction using
374 graph convolutional networks. *Advances in Neural Information Processing Systems*, 30, 2017.
- 375 [41] Raphael Townshend, Rishi Bedi, Patricia Suriana, and Ron Dror. End-to-end learning on 3d
376 protein structure for interface prediction. *Advances in Neural Information Processing Systems*,
377 32, 2019.
- 378 [42] Yi Liu, Hao Yuan, Lei Cai, and Shuiwang Ji. Deep learning of high-order interactions for
379 protein interface prediction. In *Proceedings of the 26th ACM SIGKDD international conference*
380 *on knowledge discovery & data mining*, pages 679–687, 2020.
- 381 [43] Ziwei Xie and Jinbo Xu. Deep graph learning of inter-protein contacts. *Bioinformatics*, 38(4):
382 947–953, 2022.
- 383 [44] Bowen Dai and Chris Bailey-Kellogg. Protein interaction interface region prediction by geo-
384 metric deep learning. *Bioinformatics*, 37(17):2580–2588, 2021.
- 385 [45] Farhan Quadir, Raj S Roy, Randal Halfmann, and Jianlin Cheng. Dncon2_inter: predicting
386 interchain contacts for homodimeric and homomultimeric protein complexes using multiple
387 sequence alignments of monomers and deep learning. *Scientific Reports*, 11(1):1–10, 2021.
- 388 [46] Yumeng Yan and Sheng-You Huang. Accurate prediction of inter-protein residue–residue
389 contacts for homo-oligomeric protein complexes. *Briefings in Bioinformatics*, 22(5):bbab038,
390 2021.
- 391 [47] Raj S Roy, Farhan Quadir, Elham Soltanikazemi, and Jianlin Cheng. A deep dilated convolu-
392 tional residual network for predicting interchain contacts of protein homodimers. *Bioinformatics*,
393 38(7):1904–1910, 2022.
- 394 [48] Freyr Sverrisson, Jean Feydy, Joshua Southern, Michael M Bronstein, and Bruno Correia.
395 Physics-informed deep neural network for rigid-body protein docking. In *ICLR2022 Machine*
396 *Learning for Drug Discovery*, 2022.

- 397 [49] Elham Soltanikazemi, Raj S Roy, Farhan Quadir, Nabin Giri, Alex Morehead, and Jianlin Cheng.
398 Drlcomplex: Reconstruction of protein quaternary structures using deep reinforcement learning.
399 *arXiv preprint arXiv:2205.13594*, 2022.
- 400 [50] Usman Ghani, Israel Desta, Akhil Jindal, Omeir Khan, George Jones, Nasser Hashemi, Sergey
401 Kotelnikov, Dzmitry Padhorny, Sandor Vajda, and Dima Kozakov. Improved docking of protein
402 models by a combination of alphafold2 and cluspro. *BioRxiv*, pages 2021–09, 2022.
- 403 [51] Milot Mirdita, Konstantin Schütze, Yoshitaka Moriwaki, Lim Heo, Sergey Ovchinnikov, and
404 Martin Steinegger. Colabfold: making protein folding accessible to all. *Nature Methods*, pages
405 1–4, 2022.
- 406 [52] Mu Gao, Davi Nakajima An, Jerry M Parks, and Jeffrey Skolnick. Af2complex predicts direct
407 physical interactions in multimeric proteins with deep learning. *Nature Communications*, 13(1):
408 1–13, 2022.
- 409 [53] Richard Evans, Michael O’Neill, Alexander Pritzel, Natasha Antropova, Andrew W Senior,
410 Timothy Green, Augustin Židek, Russell Bates, Sam Blackwell, Jason Yim, et al. Protein
411 complex prediction with AlphaFold-Multimer. *BioRxiv*, 2021.
- 412 [54] Rui Yin, Brandon Y Feng, Amitabh Varshney, and Brian G Pierce. Benchmarking AlphaFold
413 for protein complex modeling reveals accuracy determinants. *Protein Science*, 31(8):e4379,
414 2022.
- 415 [55] Octavian-Eugen Ganea, Xinyuan Huang, Charlotte Bunne, Yatao Bian, Regina Barzilay, Tommi
416 Jaakkola, and Andreas Krause. Independent se (3)-equivariant models for end-to-end rigid
417 protein docking. *arXiv preprint arXiv:2111.07786*, 2021.
- 418 [56] Freyr Sverrisson, Jean Feydy, Bruno E Correia, and Michael M Bronstein. Fast end-to-end
419 learning on protein surfaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision
420 and Pattern Recognition*, pages 15272–15281, 2021.
- 421 [57] Matthew McPartlon and Jinbo Xu. Deep learning for flexible and site-specific protein docking
422 and design. *BioRxiv*, 2023.
- 423 [58] Thom Vreven, Iain H Moal, Anna Vangone, Brian G Pierce, Panagiotis L Kastritis, Mieczyslaw
424 Torchala, Raphael Chaleil, Brian Jiménez-García, Paul A Bates, Juan Fernandez-Recio, et al.
425 Updates to the integrated protein–protein interaction benchmarks: docking benchmark version
426 5 and affinity benchmark version 2. *Journal of Molecular Biology*, 427(19):3031–3041, 2015.
- 427 [59] Johnathan D Guest, Thom Vreven, Jing Zhou, Iain Moal, Jeliasko R Jeliaskov, Jeffrey J Gray,
428 Zhiping Weng, and Brian G Pierce. An expanded benchmark for antibody-antigen docking and
429 affinity prediction reveals insights into antibody recognition determinants. *Structure*, 29(6):
430 606–621, 2021.
- 431 [60] Alexander Rives, Joshua Meier, Tom Sercu, Siddharth Goyal, Zeming Lin, Jason Liu, Demi Guo,
432 Myle Ott, C Lawrence Zitnick, Jerry Ma, et al. Biological structure and function emerge from
433 scaling unsupervised learning to 250 million protein sequences. *Proceedings of the National
434 Academy of Sciences*, 118(15):e2016239118, 2021.
- 435 [61] Jianyi Yang, Ivan Anishchenko, Hahnbeom Park, Zhenling Peng, Sergey Ovchinnikov, and
436 David Baker. Improved protein structure prediction using predicted interresidue orientations.
437 *Proceedings of the National Academy of Sciences*, 117(3):1496–1503, 2020.
- 438 [62] Valerio Mariani, Marco Biasini, Alessandro Barbato, and Torsten Schwede. lddt: a local
439 superposition-free score for comparing protein structures and models using distance difference
440 tests. *Bioinformatics*, 29(21):2722–2728, 2013.
- 441 [63] Sankar Basu and Björn Wallner. Dockq: a quality measure for protein-protein docking models.
442 *PloS One*, 11(8):e0161879, 2016.
- 443 [64] Joël Janin, Kim Henrick, John Moult, Lynn Ten Eyck, Michael JE Sternberg, Sandor Vajda, Ilya
444 Vakser, and Shoshana J Wodak. Capri: a critical assessment of predicted interactions. *Proteins:
445 Structure, Function, and Bioinformatics*, 52(1):2–9, 2003.

- 446 [65] Yumeng Yan, Di Zhang, Pei Zhou, Botong Li, and Sheng-You Huang. Hdock: a web server
447 for protein–protein and protein–DNA/RNA docking based on a hybrid strategy. *Nucleic Acids*
448 *Research*, 45(W1):W365–W373, 2017.
- 449 [66] Israel T Desta, Kathryn A Porter, Bing Xia, Dima Kozakov, and Sandor Vajda. Performance
450 and its limits in rigid body protein-protein docking. *Structure*, 28(9):1071–1081, 2020.
- 451 [67] Daisuke Kuroda and Jeffrey J Gray. Pushing the backbone in protein-protein docking. *Structure*,
452 24(10):1821–1829, 2016.
- 453 [68] Gabriele Corso, Hannes Stärk, Bowen Jing, Regina Barzilay, and Tommi Jaakkola. DiffDock:
454 Diffusion steps, twists, and turns for molecular docking. *arXiv preprint arXiv:2210.01776*,
455 2022.