A deep learning method for predicting interactions for intrinsically disordered regions of proteins

Kartik Majila[1,*], Varun Ullanat[1] and Shruthi Viswanath[1,*]

[1]National Center for Biological Sciences, Tata Institute of Fundamental Research, Bangalore, India 560065.

[*]Corresponding author.
Email: kartikm@ncbs.res.in (KM)
Email: shruthiv@ncbs.res.in (SV)

Keywords: Intrinsically disordered proteins (IDP), intrinsically disordered regions (IDR), protein structure, deep learning (DL), protein language model (pLMs)

Abstract

Intrinsically disordered proteins or regions (IDPs or IDRs) exist as ensembles of conformations in the monomeric state and can adopt diverse binding modes, making their experimental and computational characterization challenging. Here, we developed Disobind, a deep-learning method that predicts inter-protein contact maps and interface residues for an IDR and a partner protein, leveraging sequence embeddings from a protein language model. Several current methods, in contrast, provide partner-independent predictions, require the structure of either protein, and/or are limited by the MSA quality. Disobind performs better than AlphaFold-multimer and AlphaFold3. Combining the Disobind and AlphaFold-multimer predictions further improves the performance. However, Disobind is limited to binary IDP-partner complexes, where the two proteins are known to bind, and the input sequence fragments are less than one hundred residues long. The predictions can be used to localize IDRs in integrative structures of large assemblies, characterize protein-protein interactions involving IDRs, and modulate IDR-mediated interactions.

Introduction

Intrinsically disordered proteins or regions (IDPs or IDRs) lack a well-defined three-dimensional structure in their monomeric state (Oldfield & Dunker, 2014; Wright & Dyson, 1999). They exist as an ensemble of interconverting conformers in equilibrium and hence are structurally heterogeneous (Lindorff-Larsen & Kragelund, 2021; Oldfield & Dunker, 2014; Uversky, 2013). Here, we use the term IDR for both IDPs and IDRs. The heterogeneity of IDRs provides several functional advantages including the ability to overcome steric restrictions, have a large capture radius, undergo functional misfolding, and interact with multiple partners (Fonin et al., 2018). They play critical roles in cellular processes including signaling, intracellular transport, protein folding, and condensate formation (Fonin et al., 2018; Uversky, 2013).

IDRs are known to mediate a large number of protein-protein interactions (Tompa et al., 2014). They exhibit considerable diversity in their binding modes within complexes (Olsen et al., 2017). Upon binding to a partner protein, they may adopt an ordered state (disorder-to-order or DOR) or may remain disordered (disorder-to-disorder or DDR) (Miskei et al., 2017, 2020). Moreover, depending on the binding partner, the same IDR may undergo DOR and DDR transitions (Miskei et al., 2020). IDRs may also form multivalent interactions with their partners, where they may bind to multiple sites on the same partner in different conformations (Olsen et al., 2017; J. Zhang et al., 2024). This heterogeneity in binding makes their structural characterization in complexes challenging.

Here, we develop a deep learning method to determine the binding sites of an IDR and an interacting partner protein given their sequences. Several methods have been developed for predicting interface residues, inter-protein contact maps, and structures of complexes formed by two proteins. Some of them are limited by the requirement of a structure of the input proteins (Christoffer & Kihara, 2020; Dai & Bailey-Kellogg, 2021; Gainza et al., 2020; Guo et al., 2022; Lin et al., 2023; Pittala & Bailey-Kellogg, 2020). Sequence-based methods that require a multiple sequence alignment (MSA) are often limited by the low sequence conservation of IDRs (Abramson et al., 2024; Bret et al., 2024; Jumper et al., 2021; Lee et al., 2024).

Other sequence-based methods for IDRs use physicochemical features or embeddings from protein language models (pLMs) but do not account for the partner protein (Jahn et al., 2024; Mészáros et al., 2009; Singh et al., 2022; F. Zhang et al., 2022). However, the knowledge of the partner protein is crucial in understanding IDR interactions as these are context-dependent.

State-of-the-art methods like AlphaFold2 (AF2) and AlphaFold3 (AF3) tend to provide low-confidence predictions for IDRs (Abramson et al., 2024; Evans et al., 2022; Jumper et al., 2021). Even though the low-confidence regions predicted by AF2 often correspond to the presence of IDRs, the predicted structure cannot be considered a representative IDR structure (Alderson et al., 2023; Ruff & Pappu, 2021). Recent studies showed that AlphaFold-multimer (referred to as AF2 hereon) could be successfully used to predict the structures of complexes where the IDR undergoes a DOR transition, including protein-peptide complexes and domain-motif complexes (Bret et al., 2024; Lee et al., 2024; Omidi et al., 2024). However, the results may be sensitive to the inputs, such as the sequence fragment size, the fragment delimitation, and the alignment mode for the MSA (Bret et al., 2024; Lee et al., 2024). In general, predicting the binding interfaces for IDRs at high resolution remains a challenge for current methods (Verburgt et al., 2022).

Our method, Disobind, predicts the inter-protein contact maps and interface residues for an IDR and its partner from their sequences. Predicting inter-protein contact maps for IDRs presents several challenges. There are fewer structures of IDRs in complexes, the IDRs may retain their heterogeneity in complexes, and the inter-protein contact maps can be sparse. Here, we leverage pLMs to train our model with limited data. Predicting coarse-grained inter-protein contact maps and interface residues reduces problems associated with sparsity and heterogeneity. Disobind performs better than AF2 and AF3. Combining the Disobind and AF2 predictions further improves the performance. This is one of the first methods for sequence-based prediction of inter-protein contact maps and interface residues for an IDR and its partner. We demonstrate the performance of Disobind+AF2 on a prion protein dimer and a heterodimer of Interlukin-8 (IL-8)-chemokine receptor (CXCR1). However, its applicability is currently limited to binary IDR-partner complexes, where the two proteins are known to bind, and the input sequence fragments are less than

one hundred residues long. Predictions from the method can be used to localize IDRs in integrative structures of large assemblies. They can be used to characterize protein-protein interactions involving IDRs, identify novel motifs, and modulate IDR-mediated interactions.

Results

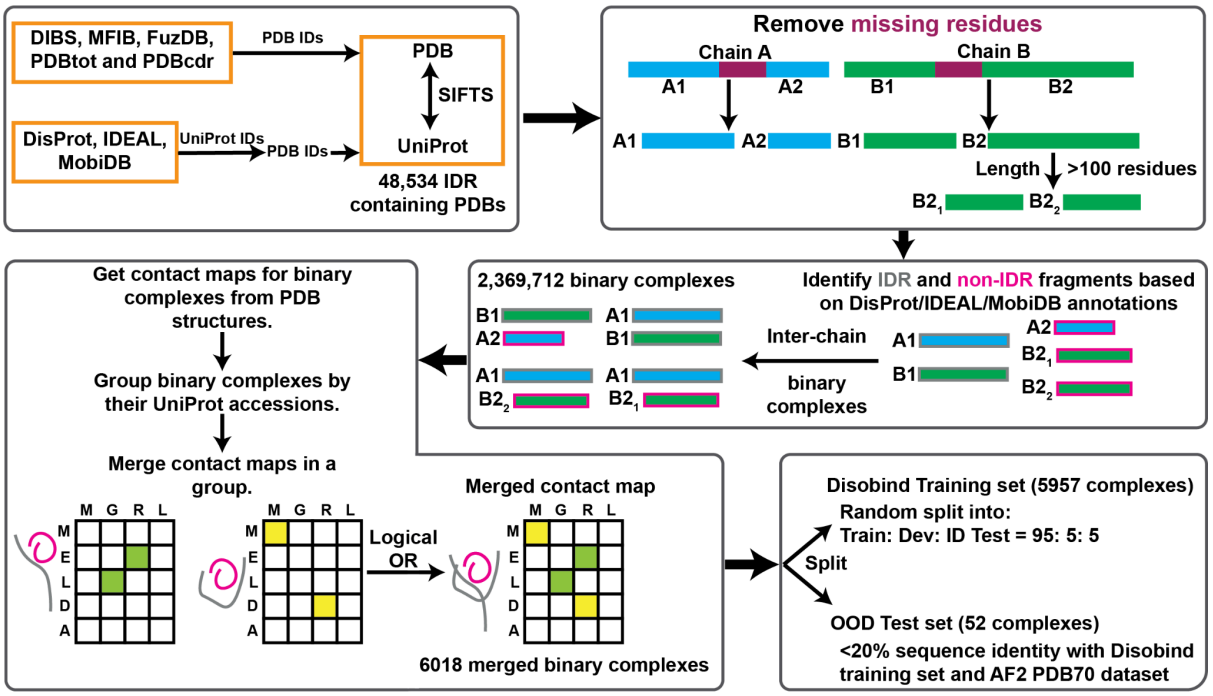

Figure 1: Disobind dataset creation pipeline. We compiled our dataset by gathering structures of IDR-containing complexes from DIBS, MFIB, FuzDB, PDBtot, PDBcdr, DisProt, IDEAL, and MobiDB. For each PDB chain, missing residues in the structures are excluded, resulting in sequence fragments. Further, we restrict the maximum length of a sequence fragment to 100 residues. We use annotations from DisProt, IDEAL, and MobiDB to classify a sequence fragment as an IDR. Next, we obtain inter-chain binary complexes comprising at least one IDR fragment. Contact maps are obtained for all binary complexes which are subsequently grouped based on the UniProt accessions of the sequence fragment pair. The contact maps of all binary complexes in a group are merged, creating merged contact maps. A sequence fragment pair and its corresponding merged contact map comprise a merged binary complex. The merged binary complexes are redundancy reduced at a

20% sequence identity cutoff to create an out-of-distribution (OOD) test set. The remaining complexes comprise the Disobind training set which is split into train, dev, and in-distribution (ID) test sets.

Disobind dataset creation

Given a pair of protein sequences, at least one of which is an IDR, Disobind predicts binary inter-protein contact maps and interface residues for the pair (Fig 1). We compile our dataset by gathering structures of IDR-containing complexes from an array of existing IDR databases and datasets, including DIBS, MFIB, FuzDB, PDBtot, PDBcdr, DisProt, IDEAL, and MobiDB, covering a range of IDR binding modes (Fig 1) (See Methods). Missing residues in the structures are excluded from the input sequences, resulting in sequence fragments for each PDB chain. Further, we restrict the maximum length of a sequence fragment to 100 residues for computational reasons.

We classify a sequence fragment as an IDR based on annotations supported by experimental evidence and sequence-based homology obtained from DisProt, IDEAL, and MobiDB; an IDR fragment has at least 20% residues annotated as disordered (See Methods). Next, we obtain inter-chain binary complexes from each PDB structure, in which at least one of the two fragments is an IDR (Fig 1). We obtain contact maps for all binary complexes. The contact maps are grouped based on the UniProt accessions of the corresponding sequence fragment pair. Next, we merge all contact maps with overlapping sequences in a group using a logical OR operation to obtain a merged contact map, ensuring that the length of each sequence fragment is up to 100 residues. Together, the sequence fragments for the pair and the corresponding merged contact map form a merged binary complex. Disobind is trained on a dataset of merged binary complexes with the sequence fragments as input and the merged contact map as output. The merged contact map encompasses contacts formed across all available structures of the complex, accounting for the multiple conformations of an IDR in a complex. This provides for a less sparse output representation compared to an ensemble of contact maps, while

also mitigating the bias caused by having an unequal number of structures for different sequence fragment pairs.

We then create an out-of-distribution (OOD) test set comprising merged binary complexes where the input sequence fragments share less than 20% identity with the Disobind dataset and with the AlphaFold2 PDB70 dataset. The remaining complexes are further split into a train, validation, and an in-distribution (ID) test set. The ID and OOD test sets comprise 297 and 52 merged binary complexes respectively. Adding more complexes to the OOD test set may be achieved by increasing the sequence identity cutoff but would also increase similarity to the AF2 training dataset, which we attempted to avoid.

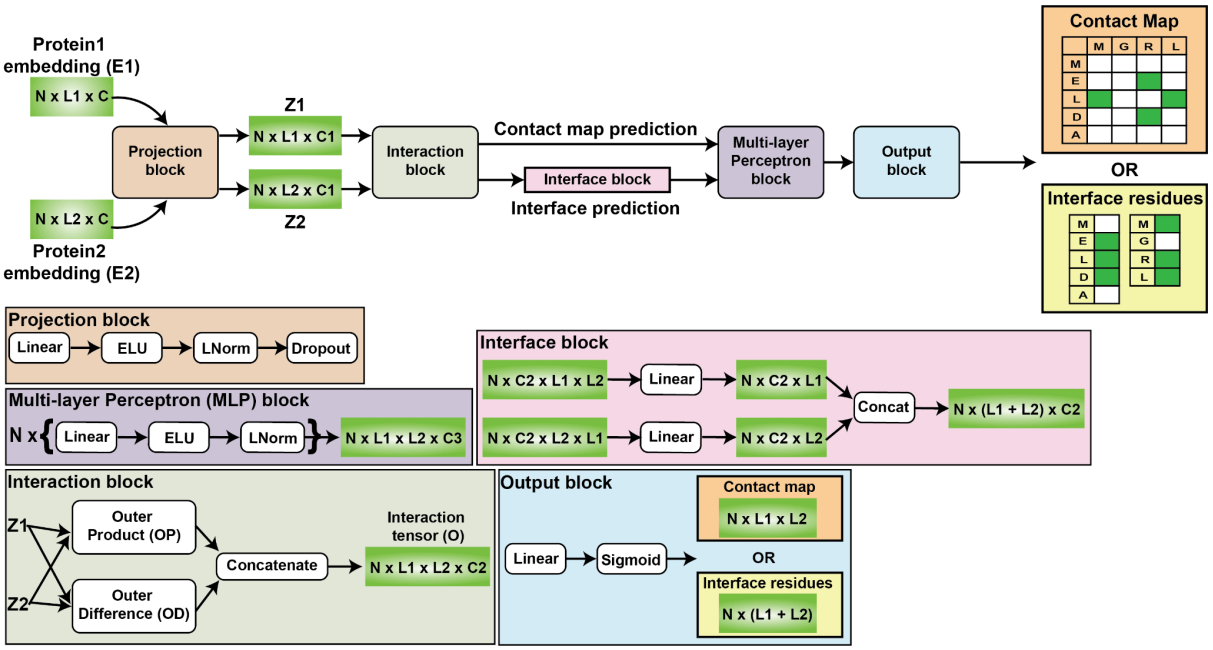

Figure 2: Disobind architecture. The inputs to Disobind are the sequence embeddings for the sequence pair obtained using a pLM. The sequence embeddings are projected to a lower dimension using a projection block. Next, in the interaction block, the projected embeddings are used to compute the outer product and outer difference, which is subsequently concatenated along the feature dimension resulting in an interaction tensor. For contact map prediction, this interaction tensor is processed by a multi-layer perceptron (MLP). For interface residue prediction, the interaction tensor is first reduced along the rows and columns using a linear layer for

both proteins and further processed by an MLP. Next, an output block provides element-wise sigmoid scores, with a score greater than 0.5 representing a contact or interface residue.

Model architecture

Disobind uses a shallow feedforward neural network architecture with the sequence embeddings obtained from a protein language model as input (Fig 2). The input embeddings are projected to a lower dimension using a projection block. Next, in the interaction block, the projected embeddings are used to compute the outer product and outer difference, which are subsequently concatenated along the feature dimension. This allows the model to capture complementarity between the input sequences and yields an interaction tensor. This interaction tensor is processed by a multi-layer perceptron (MLP) for contact map prediction. For interface residue prediction, the interaction tensor is first reduced along the rows and columns using a linear layer for both proteins and further processed by an MLP. Finally, an output block provides element-wise sigmoid scores, with a score greater than 0.5 representing a contact or interface residue. The outputs are binary contact maps or binary interface residue predictions on the fragments. The model is trained with the singularity-enhanced loss (SE loss) (Si & Yan, 2021) using the AdamW optimizer with weight decay (Fig S1). This loss is a modified form of the binary cross-entropy (BCE) loss used for class-imbalanced training. We varied the hyperparameters of the network including the projection dimension, the number of layers in the MLP, and the SE loss parameters (Table S1-S4, Fig S1). We use recall, precision, and F1-score to evaluate model performance.

Evaluating Disobind

Inter-protein contact map prediction

First, we evaluate Disobind on predicting residue-wise inter-protein contact maps on the ID and OOD test sets. Disobind achieves an F1-score of 0.57 and 0.33 on the ID and OOD test sets respectively (Table 1 and Table S5, contact map prediction at

coarse-grained (CG) resolution 1). It performs better than a random baseline (see Methods) on the OOD test set (Table 1). However, predicting residue-wise inter-protein contact maps is challenging due to their sparsity (Table S6). This leads to a class imbalance, biasing the model to predict just the majority class, *i.e.,* 0's or non-contacts. The heterogeneity of IDRs in complexes poses another challenge. For example, the multivalency of IDRs may result in different sets of contacts in different conformations with the same partner (J. Zhang et al., 2024).

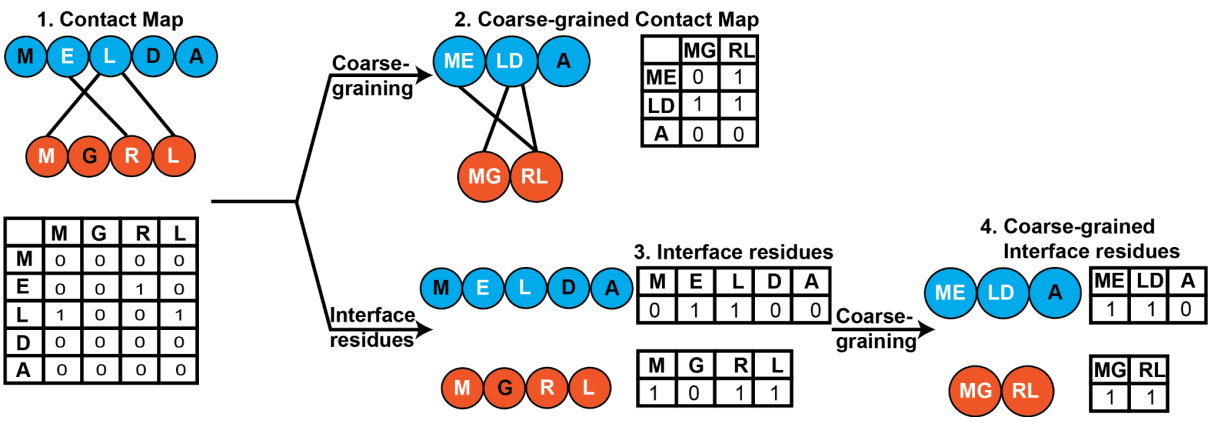

Figure 3: Predicting coarse-grained contact maps and interface residues. As an alternative to predicting inter-protein contact maps, we predict interface residues. Further, we predict coarse-grained contact maps and coarse-grained interface residues.

Interface residue prediction

As an alternative to predicting inter-protein residue-wise contact maps, next, we predict interface residues for the IDR and its partner (Fig 3). For two input sequence fragments of length L_1 and L_2, this reduces the number of predicted elements from $L_1 \times L_2$ in the contact map prediction to $L_1 + L_2$ in the interface residue prediction (Fig S3). This helps mitigate the class imbalance as the interface residue predictions are less sparse (Fig 3, Fig S3, Table S6, interface residue prediction at coarse-grained (CG) resolution 1). Interface residue prediction is also less affected by the heterogeneity of IDRs in complexes. The prediction for each residue is

simplified to determining whether it binds to the partner, rather than identifying the specific residues within the partner it interacts with.

As expected, Disobind performs better in interface residue prediction than in contact map prediction. It achieves an F1 score of 0.71 and 0.50 on the ID and OOD test sets respectively (Fig S2, Table 1 and Table S5, interface residue prediction at coarse-grained (CG) resolution 1). It performs better than a random baseline evaluated on the OOD test set (Table 1).

CG	Model	Recall	Precision	F1-score
Contact map				
1	**Disobind**	0.22	0.63	0.33
	Random baseline	0.51	0.01	0.02
5	**Disobind**	0.38	0.61	0.47
	Random baseline	0.52	0.08	0.13
10	**Disobind**	0.42	0.49	0.45
	Random baseline	0.54	0.17	0.26
Interface residues				
1	**Disobind**	0.50	0.50	0.50
	Random baseline	0.50	0.26	0.34
5	**Disobind**	0.72	0.56	0.63
	Random baseline	0.53	0.45	0.49
10	**Disobind**	0.83	0.61	0.70
	Random baseline	0.48	0.52	0.50

Table 1: Evaluating Disobind performance on the OOD test sets. We compare the performance of Disobind on the OOD test set for the contact map and interface residue prediction tasks for different coarse-grained (CG) resolutions. We compare the model performance to a random baseline generated by randomly sampling contacts or interface residues based on their proportion in the training set.

Coarse-graining improves the performance

To further improve the performance of our model, we predict coarse-grained contact maps and interface residues (Fig 3, Fig S3). Coarse-graining has been widely used for modeling biological systems (Arvindekar et al., 2024; Noid, 2013). Coarse-graining over a set of residue pairs in a contact map or a set of residues in the interfaces further reduces the sparsity and class imbalance, and mitigates the problems posed by the multivalency of IDR interactions (Fig 3, Fig S3, Table S6).

We train Disobind to predict coarse-grained contact maps and coarse-grained interface residues at resolutions of 5 and 10 contiguous residues along the backbone. The coarse-grained embeddings are obtained by averaging the input embeddings (Elnaggar et al., 2022). As expected, coarse-graining further improves the performance of Disobind on the ID and OOD test sets for both the contact map and interface residue prediction (Fig S2, Table 1, Table S5). For all cases, Disobind performs better than a random baseline on the OOD test set.

Comparison to AlphaFold2 and AlphaFold3

We further compare the performance of Disobind with the state-of-the-art methods AF2 and AF3 on the OOD test set. For evaluating AF2 and AF3 outputs, we mask out interactions predicted at low confidence based on the interface-predicted template matching (ipTM) score, predicted local distance difference test (pLDDT), and predicted aligned error (PAE) metrics (see Methods).

Using different ipTM cutoffs for AF2 and AF3

The ipTM score is a metric provided by AF2 and AF3 to assess the confidence in the predicted interfaces in a complex (Abramson et al., 2024; Evans et al., 2022). A prediction with an ipTM score greater than 0.75 is considered highly confident, while an ipTM lower than 0.6 indicates a likely failed prediction (O'Reilly et al., 2023; Yin et al., 2022). We first evaluate the performance of AF2 and AF3 with an ipTM cutoff of 0.75. Disobind performs better than AF2 and AF3 for all tasks, *i.e.,* contact map and interface residue predictions across different coarse-grained resolutions (Table 2-3, ipTM = 0.75). Disordered residues, however, are known to impact the pTM and ipTM

metrics negatively (Abramson et al., 2024; Magana & Kovalevskiy, 2024). Considering this, we relaxed the ipTM cutoff to 0.4 to assess the performance of AF2 and AF3, keeping the per-residue metric, *i.e.* pLDDT and PAE cutoffs as is, guided by recent benchmarks and the AF3 guide (Abramson et al., 2024; Omidi et al., 2024). Here also, Disobind performs better than both AF2 and AF3 for both contact map and interface residue prediction (Table 2-3, ipTM = 0.4). AF3 performs worse than both Disobind and AF2. Furthermore, this trend remains more or less the same even when we do not apply an ipTM cutoff (Table 2-3, ipTM = 0).

Model	Recall	Precision	F1-score
Coarse Graining 1			
Disobind	0.22	0.63	0.33
AF2_0.75	0.12	0.58	0.20
AF3_0.75	0.02	0.41	0.04
AF2_0.4	0.21	0.45	0.29
AF3_0.4	0.10	0.31	0.15
AF2_0	0.22	0.45	0.30
AF3_0	0.12	0.31	0.17
Coarse Graining 5			
Disobind	0.38	0.61	0.47
AF2_0.75	0.15	0.58	0.24
AF3_0.75	0.02	0.34	0.03
AF2_0.4	0.30	0.52	0.38
AF3_0.4	0.17	0.43	0.25
AF2_0	0.31	0.52	0.39
AF3_0	0.20	0.44	0.28
Coarse Graining 10			
Disobind	0.42	0.49	0.45
AF2_0.75	0.16	0.62	0.25
AF3_0.75	0.02	0.4	0.04
AF2_0.4	0.33	0.56	0.41
AF3_0.4	0.2	0.49	0.28

AF2_0	0.34	0.56	0.42
AF3_0	0.23	0.49	0.32

Table 2: Comparing Disobind, AF2, and AF3 performance for contact map predictions. We compare the performances of Disobind, AF2, and AF3 on the OOD set for contact map prediction for different coarse-grained (CG) resolutions. For AF2 and AF3 we only consider the confident predictions (pLDDT >= 70 and PAE <= 5) at multiple ipTM cutoffs (0, 0.4, 0.75) for evaluation. The number after the '_' represents the ipTM cutoff used.

Model	Recall	Precision	F1-score
Coarse Graining 1			
Disobind	0.50	0.50	0.50
AF2_0.75	0.18	0.76	0.29
AF3_0.75	0.02	0.43	0.04
AF2_0.4	0.37	0.69	0.48
AF3_0.4	0.23	0.6	0.33
AF2_0	0.38	0.69	0.49
AF3_0	0.26	0.58	0.36
Coarse Graining 5			
Disobind	0.72	0.56	0.63
AF2_0.75	0.20	0.79	0.32
AF3_0.75	0.02	0.50	0.04
AF2_0.4	0.44	0.74	0.55
AF3_0.4	0.31	0.73	0.43
AF2_0	0.45	0.73	0.56
AF3_0	0.35	0.71	0.47
Coarse Graining 10			
Disobind	0.83	0.61	0.70
AF2_0.75	0.20	0.83	0.32
AF3_0.75	0.03	0.59	0.05

AF2_0.4	0.45	0.80	0.58
AF3_0.4	0.32	0.80	0.46
AF2_0	0.47	0.79	0.59
AF3_0	0.38	0.77	0.51

Table 3: Comparing Disobind, AF2, and AF3 performance for interface residue predictions. We compare the performances of Disobind, AF2, and AF3 on the OOD test set for interface residue prediction for different coarse-grained (CG) resolutions. For AF2 and AF3 we only consider the confident predictions (pLDDT >= 70 and PAE <= 5) across multiple ipTM scores (0, 0.4, 0.75) for evaluation. The number after the '_' represents the ipTM cutoff used.

AF2 performs better than AF3

Similar to Disobind, coarse-graining improves the performance of both AF2 and AF3 for both contact map and interface predictions (Table 2, Table 3). Across all tasks, AF2 performs better than AF3. Moreover, AF3 predictions have lower confidence as measured by their ipTM scores (Fig S3). For 35 of 52 OOD test set entries, the ipTM score of the best AF2 model was greater than that of the best AF3 model. Overall, only 8 of 52 predictions from AF2 and 2 of 52 predictions from AF3 had a high confidence prediction, *i.e.*, ipTM score higher than 0.75.

Disobind performs better in predicting interactions for disordered residues

Next, we assess the Disobind and AF2 predictions specifically for residues annotated as disordered. For this, we mask out interface residues (contacts) not involving disordered residues (disordered residue pairs) (See Methods). For AF2 predictions, we do not apply the ipTM cutoff. Disobind performs better than AF2 in predicting contacts and interface residues for disordered residues across different coarse-grained resolutions (Table S7).

Combining Disobind and AlphaFold2 predictions

Next, we sought to combine the predictions from Disobind and AF2, to see if we could further improve over either method. We combine the predictions from Disobind and AF2 using a logical OR operation and evaluate the performance on the OOD test set. The combined model, “Disobind+AF2”, performs better than either of the methods (Table 4-5). For disordered residues alone, Disobind+AF2 performs comparably to Disobind and is much better than AF2 alone (Table S7).

Model	Recall	Precision	F1-score
Coarse Graining 1			
Disobind	0.22	0.63	0.33
AF2	0.22	0.45	0.30
Disobind+AF2	0.34	0.48	0.40
Coarse Graining 5			
Disobind	0.38	0.61	0.47
AF2	0.31	0.52	0.39
Disobind+AF2	0.51	0.50	0.50
Coarse Graining 10			
Disobind	0.42	0.49	0.45
AF2	0.34	0.56	0.42
Disobind+AF2	0.57	0.46	0.51

Table 4: Evaluating the performance of Disobind+AF2 for contact map prediction on the OOD test set. We combine Disobind and AF2 predictions using a logical OR operation and evaluate the performance on the OOD test set for contact map prediction. For AF2 we consider the confident predictions (pLDDT >= 70 and PAE <= 5) for evaluation. No ipTM cutoff is applied for AF2 and Disobind+AF2 predictions.

Model	Recall	Precision	F1-score
Coarse Graining 1			
Disobind	0.50	0.50	0.50

AF2	0.38	0.69	0.49
Disobind+AF2	0.62	0.49	0.55
Coarse Graining 5			
Disobind	0.72	0.56	0.63
AF2	0.45	0.73	0.56
Disobind+AF2	0.79	0.55	0.65
Coarse Graining 10			
Disobind	0.83	0.61	0.70
AF2	0.47	0.79	0.59
Disobind+AF2	0.88	0.60	0.72

Table 5: Evaluating the performance of Disobind+AF2 for interface residue prediction on the OOD test set. We combine Disobind and AF2 predictions using a logical OR operation and evaluate the performance on the OOD test set for interface residue prediction. For AF2 we consider the confident predictions (pLDDT >= 70 and PAE <= 5) for evaluation. No ipTM cutoff is applied for AF2 and Disobind+AF2 predictions.

Input Embeddings

Next, we investigate the effect of using embeddings from various pLMs and the different embedding types on the performance of Disobind.

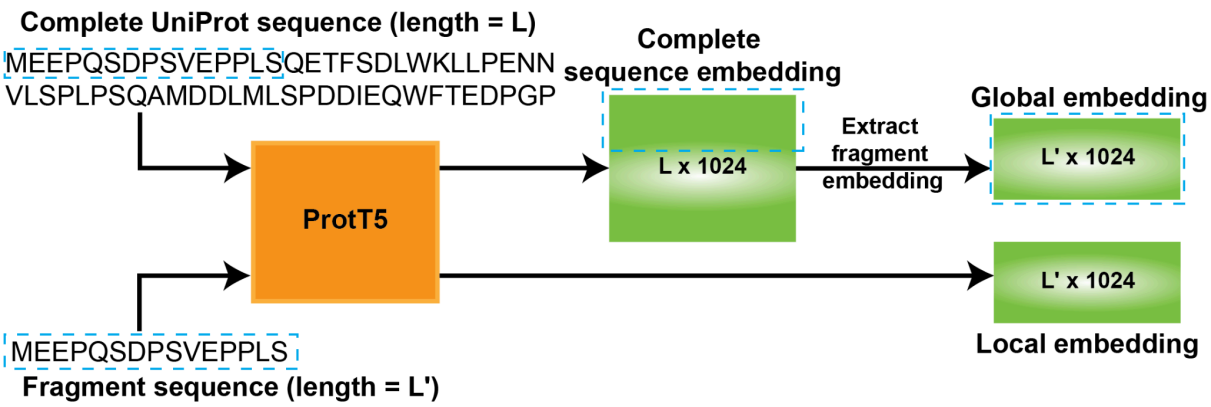

Figure 4: Global and local sequence embeddings. We obtain global and local embeddings for the input protein sequence fragments. For global embedding, the complete UniProt protein sequence corresponding to the sequence fragment (dashed blue box) is used as input to the pLM from which the sequence fragment embedding is extracted. For local embeddings, the sequence fragment (dashed blue box) is used as input to the pLM to obtain the embedding.

Protein language models allow for a shallow architecture

Protein language models (pLMs) provide context-aware representations of protein sequences that facilitate training models for downstream tasks like contact map prediction *via* transfer learning. They allow using a shallow downstream architecture when training with limited data (Bepler & Berger, 2021; Jahn et al., 2024). Several pLMs have been developed including the ESM series of models (Rives et al., 2021), ProstT5 (Heinzinger et al., 2024), ProtT5 (Elnaggar et al., 2022), ProtBERT (Elnaggar et al., 2022), and ProSE (Bepler & Berger, 2019). Disobind was trained using embeddings obtained from ProtT5 which achieves the best performance among all the embeddings tested (Table S8). Due to limitations in memory, we could not generate ESM2-650M-global embeddings for the entire training set. Hence, we do not include ESM in our comparison.

Global embeddings perform better than local embeddings

We then evaluate whether models trained with global sequence embeddings perform better than those with local embeddings. Local embeddings are obtained by providing the sequence of the fragment as input to the pLM. Global embeddings, in contrast, are obtained by providing the entire protein sequence corresponding to the fragment as input to the pLM, followed by extracting the embeddings corresponding to the fragment (Fig 4). Models with global embeddings perform better than those with local embeddings across all the pLMs we tested (Table S8). Using the global embeddings provides the context of the complete protein sequence corresponding to the fragment, plausibly resulting in better performance. This context may contain information relevant to binding: for example, the binding of IDRs is known to be affected by the flanking regions (Olsen et al., 2017).

Case studies

Here, we demonstrate the performance of our method on two biologically relevant complexes. The cellular prion protein (PrP^C) plays an important role in several cellular processes including synapse formation, regulating circadian rhythm, and maintaining ion homeostasis (Kraus et al., 2021; Sigurdson et al., 2019). PrP^C has an N-terminal disordered region and a primarily α-helical C-terminal region. The conversion of the PrP^C to the infectious PrP^{Sc} form involves significant structural changes resulting in a β-sheet-rich protein**.** The PrP^{Sc} form is known to form amyloid fibrils associated with several diseases including Creutzfeldt-Jakob disease, and bovine spongiform encephalopathy (Kraus et al., 2021; Sigurdson et al., 2019). We used Disobind+AF2 to predict interface residues for the homodimer of the PrP^{Sc} form (P04273) residues 95-193 (Fig 5a). The Disobind+AF2 model achieves an F1 score of 0.92 whereas Disobind and AF2 achieve an F1 score of 0.90 and 0.45 respectively.

Chemokines are small secretary molecules that play several important roles including immune cell trafficking, wound healing, and lymphoid cell development (Rossi & Zlotnik, 2000; Sepuru et al., 2020). They bind to chemokine receptors that belong to the G-protein coupled receptor (GPCR) family. Dysregulation of chemokine release is associated with several diseases including chronic pancreatitis, inflammatory bowel disease, and psoriasis (Rossi & Zlotnik, 2000; Sepuru et al., 2020). Chemokine receptors comprise an intrinsically disordered, extracellular N-terminal domain, seven transmembrane helices, and an intracellular C-terminal domain. The chemokine receptor CXCR1 forms a heterodimer with the chemokine Interleukin-8, where the flexible N-terminus of the receptor becomes structured and adopts an extended conformation upon binding (Sepuru et al., 2020). We used Disobind+AF2 to predict interface residues for CXCR1 (P25024) residues 1-29 with Interleukin-8 (P10145) residues 28-93 (Fig 5b). In this case, Disobind+AF2 achieves an F1 score of 0.71 whereas Disobind and AF2 achieve an F1 score of 0.44 and 0.64 respectively.

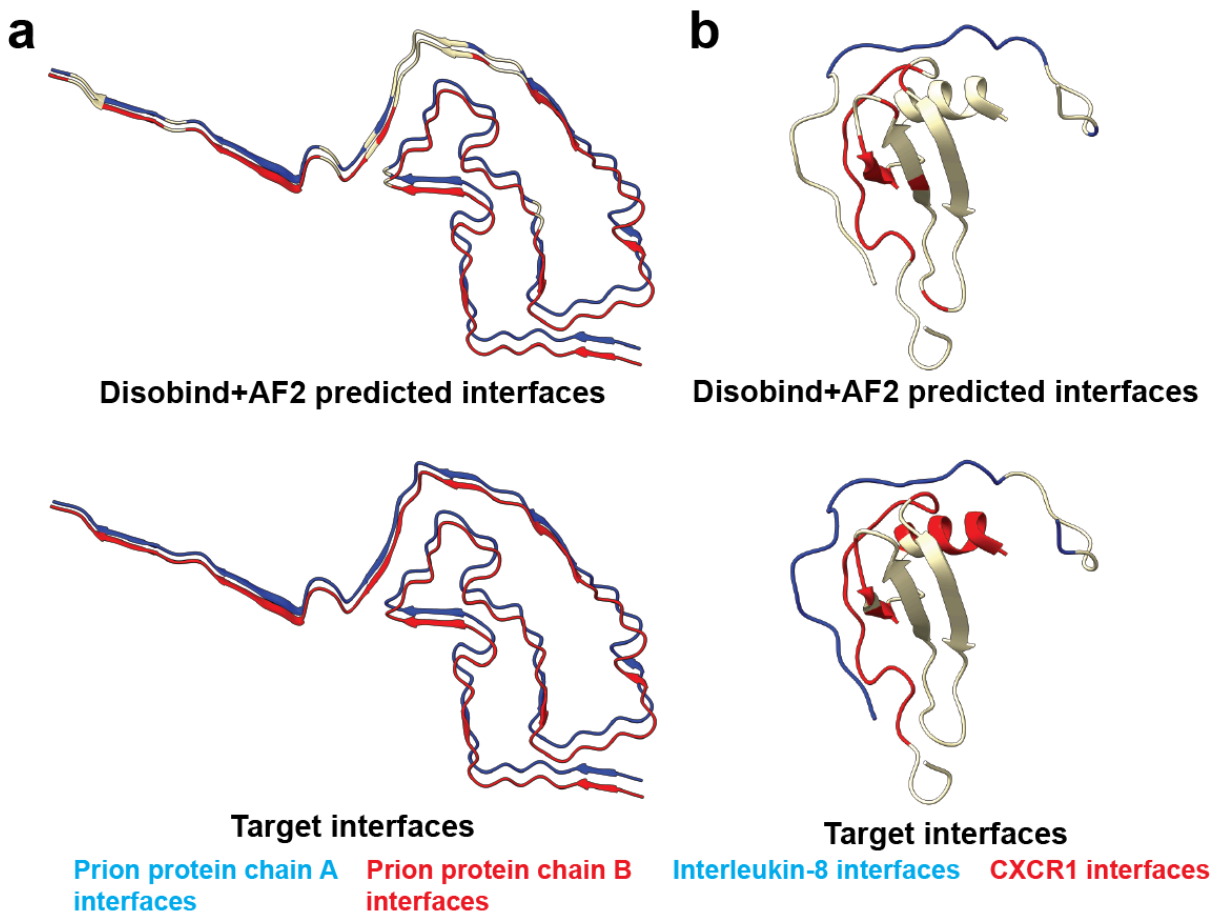

Figure 5: Case studies visualizing Disobind+AF2 predictions. Visualizing the Disobind+AF2 predicted and target interface residues for two OOD test set examples. The interface residues for the two proteins are highlighted in "blue" and "red". **(a)** Interface residues predicted for a homodimer of the prion protein (P04273) residues 95-193. Chains A and B from the PDB entry 7LNA are shown. **(b)** Interface residues predicted for heterodimer of the chemokine receptor CXCR1 (P25024) residues 1-29 with Interlukin-8 (P10145) residues 28-93. Chains A and B for the 1st model from the PDB entry 6XMN are shown.

Discussion

Here, we developed Disobind, a deep-learning method for sequence-based prediction of partner-dependent contact maps and interface residues for an IDR and its binding partner. Leveraging embeddings from pLMs, Disobind predicts contact

maps and interface residues across various coarse-grained resolutions: 1, 5, and 10 residues. We further combine Disobind with AF2 and show that the combined method outperforms Disobind, AF2, and AF3. Ours is one of the first methods to predict inter-protein contact maps and interface residues for the IDR and its partner from their sequences.

Uses and limitations

Predictions from Disobind+AF2 may be used to improve the localization of IDRs in integrative models of large assemblies, our primary motivation for developing the method. Macromolecular assemblies contain significant portions of IDRs, for example, the Fg Nups in the nuclear pore complex, the MBD3-IDR in the nucleosome remodeling and deacetylase complex, and the N-terminus of Plakophilin1 in the desmosome (Akey et al., 2022; Arvindekar et al., 2022; Pasani et al., 2024). These regions typically lack data for structural modeling, resulting in integrative models of poor precision (Arvindekar et al., 2022; Pasani et al., 2024). The predictions from Disobind+AF2 can be used in integrative modeling methods such as IMP, HADDOCK, and Assembline as inter-protein distance restraints (Dominguez et al., 2003; Rantos et al., 2022; Russel et al., 2012). Even the coarse-grained contact map and interface residue predictions would be useful in such cases to improve the precision of these regions in the integrative model. Additionally, our method can be used to characterize interactions involving IDRs across proteomes. This may aid in identifying new binding motifs for IDRs, potentially linked to their sub-cellular localization or function. Finally, predictions from our method may aid in modulating interactions involving IDRs, for example, by suggesting plausible mutations.

However, our method also has several limitations. First, it is limited to binary IDR-partner complexes. Second, it assumes that the IDR and its partner are known to bind. Third, the input sequence fragments must be less than one hundred residues long. Fourth, the accuracy of the predictions depends on the ability of the pLM to provide accurate representations of the IDR and its partner. Finally, our method cannot be used to assess the effects of post-translational modifications as the pLM used does not distinguish post-translationally modified amino acids.

Challenges

Predicting contact maps and interface residues for IDRs in a complex with a partner is challenging. First, a limited number of experimental structures are available for IDRs in complexes (Jahn et al., 2024; J. Zhang et al., 2024). Second, IDRs adopt an ensemble of conformations, and the available structures may only partially capture this conformational diversity. Third, inter-protein contact maps are typically sparse, with only a few residues forming contacts. Although not sufficient, we gather all available structures for IDRs in complexes. With these, we create a dataset of merged binary complexes for training Disobind. Using merged binary complexes helps overcome the issue of sparsity associated with training on an ensemble of contact maps. Predicting interface residues instead of contacts and using coarse-graining helps overcome the challenges associated with the multivalency of IDRs and the sparsity of inter-protein contact maps (Fig S3).

Performance of AF2 and AF3

Several studies indicate that the low-confidence regions in AF2 predictions overlap with the presence of disordered regions, although these low-confidence regions cannot be considered as a representative conformation for IDRs (Alderson et al., 2023; Escobedo et al., 2024; Ruff & Pappu, 2021). On our OOD test set, most of the predictions from AF2 and AF3 had low confidence, with an ipTM score lower than 0.75. Interestingly, compared to AF2, AF3 predictions were less accurate and had lower confidence. It is possible that the confidence cutoffs used for ordered proteins do not apply to IDRs and predictions involving the latter might require different cutoffs. A recent benchmark suggested a lower ipTM score of 0.4 for assessing interfaces involving IDRs (Omidi et al., 2024). The per-residue metrics such as pLDDT and PAE may be more relevant for assessing AF2 and AF3 predictions than the global metrics such as the ipTM and pTM, as is also suggested by a recent benchmark and the AF3 guide (Abramson et al., 2024; Omidi et al., 2024). However, more benchmarking and assessments are required to evaluate AF2 and AF3 predictions for IDRs.

AF2 has been shown to successfully predict the structures of protein-peptide complexes and domain-motif complexes, although the results could be sensitive to the sequence fragment size, fragment delimitation, and the MSA alignment mode (Bret et al., 2024; Lee et al., 2024). Some studies show that MSA subsampling (del Alamo et al., 2022), clustering the sequences in the MSA (Wayment-Steele et al., 2024), and using sliding fragments as input to AF2 (Bret et al., 2024) result in better predictions and can be used to predict multiple conformations. In our comparison, we did not explore these strategies, though they may improve the model predictions.

More generally, AF2 can predict the structures of complexes where the IDR is ordered upon binding (DOR cases), as shown in recent benchmarks (Alderson et al., 2023; Omidi et al., 2024). In contrast, AF2 does not provide accurate predictions in cases where the IDR is flexible in the complex (DDR cases), highlighting an open direction for future research (Omidi et al., 2024).

Future Directions

One of the major roadblocks in training methods such as Disobind is the lack of data. More experimental structures of IDRs in complexes would be valuable ((Jahn et al., 2024; J. Zhang et al., 2024)). Whereas IDR ensembles derived from molecular dynamics can be used, generating MD ensembles for IDR complexes is computationally expensive and challenging, and the existing databases like PED contain very few such ensembles for IDR complexes (Ghafouri et al., 2024; Majila et al., 2024). Alternatively, deep generative models can be used to generate ensembles for IDRs in complexes (Janson & Feig, 2024; Majila et al., 2024; Mansoor et al., 2024). However, the current methods are limited to generating ensembles for monomers.

Disobind and similar methods can be further enhanced by improving the existing pLMs to provide better representations for IDRs plausibly by incorporating physical priors and/or structural information (Majila et al., 2024; Rogers et al., 2023; Wang et al., 2024). Additionally, these methods could be extended to predict protein-protein interactions (PPIs) involving IDRs, and aid in the design of IDR binders. The behaviour of IDRs within cells remains largely unexplored. Methods like Disobind are

expected to facilitate an improved understanding of the interactions, function, and modulation of IDRs.

Methods

Creating the dataset for Disobind

Gathering PDB structures of IDRs in complexes

First, we gathered the available PDB entries for protein complexes containing IDRs from databases on structures of disordered proteins: DIBS (Schad et al., 2018), MFIB (Fichó et al., 2017), and FuzDB (Miskei et al., 2017) (Fig 1). We also included PDB entries from the PDBtot and PDBcdr datasets (Miskei et al., 2020). PDBtot consists of IDRs that either undergo a DOR or DDR transition, whereas PDBcdr consists of IDRs that undergo both DOR and DDR transitions depending upon the partner. We considered all PDB entries associated with each database. This resulted in a set of 4252 PDB entries. PEDS is another source of conformational ensembles of IDRs (Ghafouri et al., 2024). However, PEDS entries were not used: several PED entries were already incorporated into our pipeline and the remaining entries correspond to monomers.

We then supplemented this set by querying the PDB for additional complexes containing IDRs. For this, we first obtained Uniprot identifiers (IDs) of proteins containing IDRs from DisProt (Aspromonte et al., 2024), IDEAL (Fukuchi et al., 2014), and MobiDB (Piovesan et al., 2024) (Fig 1). Specifically, from Disprot, we obtained all UniProt IDs corresponding to each DisProt entry. From IDEAL, we obtained UniProt IDs for entries annotated as "verified ProS". These sequences are experimentally verified to be disordered in isolation and ordered upon binding. From MobiDB, we obtained UniProt IDs for entries annotated "curated-disorder-priority", "homology-disorder-priority", "curated-lip-priority", and "homology-lip-priority". These correspond to sequences with curated experimental evidence for disorder (“curated-disorder”), their homologs (“homology-disorder”), sequences with experimental evidence for disorder with binding motifs such as SLIMs (“curated-lip”), and their homologs (“homology-lip”), respectively. Homology-based annotations were

incorporated due to the lack of sufficient experimentally curated data available for IDRs. Importantly, the annotations were based on the source with the highest confidence ("priority"), and annotations based on indirect sources of evidence or predictions were not considered. Choosing the highest confidence annotation may reduce false positive "disorder" or "lip" annotations that may arise from homology-based evidence.

A total of 357734 unique Uniprot IDs across MobiDB, Disprot, and IDEAL were obtained. The PDB REST API was queried to obtain all the PDB entries associated with a given Uniprot ID, resulting in 48534 PDB entries (Table S1) (Rose et al., 2021). Combining these with the PDB entries obtained in the previous step, we obtained 50294 unique PDB entries. As an aside, we note that MobiDB now provides the PDB IDs for IDR-containing complexes, however, this functionality was not available when this project started.

Defining binary complexes containing IDRs

The above PDB entries were downloaded using the Python requests library and mapped to the sequences in UniProt using the SIFTS (Structure Integration with Function, Taxonomy, and Sequence) tool (Velankar et al., 2013). Entries with an obsolete PDB ID and those lacking a SIFTS mapping were removed, and deprecated PDB IDs were replaced with superseding PDB IDs. Further, entities corresponding to chimeric proteins or non-protein molecules, those having an obsolete UniProt ID, and those containing more than 10000 residues were removed.

Next, for each PDB entry, we obtained sequence fragments by removing missing residues from the sequence of each chain. The length of these fragments was kept to between 20 and 100 residues; fragments longer than 100 residues were further divided. We then identified the disordered residues in these fragments (Fig 1). This was achieved by cross-referencing the UniProt mapping of the fragment with the disorder annotations from DisProt, IDEAL, and MobiDB as obtained above; a residue was considered disordered if it was annotated as such in any of these three databases. Fragments comprising at least 20% disordered residues were considered as IDR fragments whereas the others were considered as non-IDR fragments.

Subsequently, a set of binary complexes was constructed from each PDB entry. Each binary complex consisted of an IDR fragment paired with another IDR or non-IDR fragment. Binary complexes comprising solely of intra-chain fragments or non-IDR fragments were excluded (Fig 1). This resulted in a total of 2369712 binary complexes.

Creating merged binary complexes

We then created contact maps from these binary complexes. A contact map is a binary matrix of zeros (non-contacts) and ones (contacts); two residues are in contact if the distance between their $C\alpha$ atoms is less than 8 Å. We merged the contact maps across different binary complexes of a sequence fragment pair, to create “merged contact maps” (Fig 1). This allows us to account for contacts from all available complexes. Further, a merged contact map is less sparse and therefore easier to learn compared to an ensemble of contact maps. It also mitigates the bias caused by having different numbers of structures for different sequence fragment pairs.

We first grouped the binary complexes across all the PDB entries by the UniProt ID of the constituent proteins in the sequence fragment pair, resulting in 14599 UniProt ID pairs/groups. For computational efficiency, all binary complexes with no contacts were eliminated. The remaining binary complexes in a group were sorted based on the UniProt start residue positions of the protein fragments in the pair and further partitioned into sets. Each set comprised binary complexes containing overlapping sequences (more than one residue overlap) for both fragments such that the sequence length for each fragment, merged across the overlapping sequences, does not exceed the maximum fragment length (100 residues). Contact maps for all binary complexes in a set were aggregated using a logical OR operation to form “merged contact maps”. A sequence fragment pair with the corresponding merged contact map comprises a merged binary complex. We obtained 24442 merged binary complexes. Finally, we filtered merged binary complexes for which the contact density was very high (>5%) or very low (<0.5%), resulting in 6018 merged binary complexes (Fig 1).

Creating the training set and OOD test set

We created an out-of-distribution (OOD) test set that is sequence non-redundant with our training dataset and the PDB70 dataset used for training AlphaFold. We first removed merged binary complexes whose sequence fragments were derived from a PDB chain in PDB70. We then obtained a non-redundant set of sequences by clustering our dataset of sequence fragments and the PDB70 cluster representatives using MMSeqs2 (Steinegger & Söding, 2017). MMSeqs2 was used at a 20% sequence identity threshold and cluster mode 1. Since the PDB70 dataset is already clustered at 40% sequence similarity, we consider only the representative cluster members for sequence identity comparisons.

For the OOD test set, we selected clusters without any chain from PDB70 that are singleton (only one cluster member *i.e.* self) or doublet (two member sequences, both belonging to the same merged binary complex) clusters. This resulted in 53 merged binary complexes for the OOD test set. We removed a OOD tests et entry for which AF2 predictions could not be obtained, resulting in 52 OOD test set complexes. The remaining dataset was split into a training (train), development (dev), and an in-distribution (ID) test set in the ratio 0.9:0.05:0.05, resulting in 5362 train, 298 dev, and 297 ID test set merged complexes.

Disobind Model Architecture and Training

Notations

Here, we define the notation used in this paper.

$\mathbb{R}^{A \times B \times C \times D}$: 4D tensor with the axes i, j, k, l corresponding to the four dimensions of size A, B, C, D respectively

N: batch size used for training.

L_1, L_2: lengths of protein sequence fragments 1 and 2 indicating the number of residues in each protein.

C, C_1, C_2, C_3: the sizes of embedding or feature dimensions.

E_1, E_2: embeddings for protein sequence fragments 1 and 2.

O: interaction tensor obtained post-concatenation.

O_{red}: output from the multi-layer perceptron block, reduced from the interaction tensor.

I_1, I_2 interface tensors for proteins 1 and 2.

I: concatenated interface tensors.

$Linear$: applies a linear transformation, $Linear(X) = XW^T + b$.

ELU: exponential linear unit, $ELU(x) = max(x, \alpha(e^x - 1))$.

$Sigmoid(x)$: $\frac{1}{1+e^{-x}}$.

$Dropout$: regularization technique that randomly sets weights to 0.

$LNorm$: layer normalization.

$Concat([A, B]_i)$: concatenate tensors A, B along the axis i.

$A_{ijkl \to ikjl}$: permute tensor A along the specified axes.

Inputs and outputs for training

We train six separate models, corresponding to six prediction tasks. These correspond to two types of predictions: a) inter-protein contact map and b) interface residue predictions at three coarse-grained (CG) resolutions (1, 5, and 10 sequence-contiguous residues). Coarse-graining at resolution 1 is equivalent to no coarse-graining, *i.e.*, residue-level predictions. The input to the model is the embedding for the two sequence fragments from the ProtT5 pLM $E_1 \in \mathbb{R}^{L_1 \times C}$, $E_2 \in \mathbb{R}^{L_2 \times C}$ (Elnaggar et al., 2022). The outputs for the two types of predictions respectively are a) a binary contact map, $ContactMap \in \mathbb{R}^{L_1/k \times L_2/k}$ and b) a binary interface residue vector, $Interface \in \mathbb{R}^{L_1/k + L_2/k}$ at a particular CG resolution, where $k \in [1, 5, 10]$ represents the CG resolution.

For coarse-grained predictions, the input embeddings were coarse-grained using an AvgPool1d whereas the corresponding contact maps were coarse-grained using a MaxPool2d; the kernel size and stride were set to the CG resolution in both cases.

The target vectors for interface residue predictions were obtained by reducing the rows and columns of the corresponding contact maps by an OR operation. The input embeddings, output contact maps, and interface residue vectors were zero-padded to the maximum length (100 residues).

Model architecture

Disobind comprises a projection block, an interface block, a multi-layer perceptron (MLP) block, and an output block (Fig 2). The models for contact map and interface residue prediction employ a contact map block and interface block respectively.

Projection block

The input embeddings E_1, E_2 are projected to a lower dimension, C_1 using a linear layer.

$$Z_1 = Dropout(\, ELU(\, Linear(\, E_1 \,)\,)\,)\,;\; Z_1 \in \mathbb{R}^{N \times L_1 \times C_1} \text{ – (1)}$$

$$Z_2 = Dropout(\, ELU(\, Linear(\, E_2 \,)\,)\,)\,;\; Z_2 \in \mathbb{R}^{N \times L_2 \times C_1} \text{ – (2)}$$

Interaction block

The model captures complementarity between the projected embeddings Z_1, Z_2 by concatenating the outer product and the absolute value of the outer difference between the projected embeddings along the feature dimension.

$$O = Concat(\, [OuterProduct(\, Z_1, Z_2 \,),\; |OuterDifference(\, Z_1, Z_2 \,)|]_l);\; O \in \mathbb{R}^{N \times L_1 \times L_2 \times C_2} \text{ – (3)}$$

Interface block

To capture the interface residues for both proteins I, we reduce O along its rows and columns using linear layers followed by concatenating along the length.

$$I_1 = Linear(O_{ijkl \to iljk});\; I_1 \in \mathbb{R}^{N \times C_2 \times L_1} \text{ – (4)}$$

$$I_2 = Linear(O_{ijkl \to ilkj});\; I_2 \in \mathbb{R}^{N \times C_2 \times L_2} \text{ – (5)}$$

$I = Concat([I_1, I_2]_k);\ I_2 \in \mathbb{R}^{N \times C_2 \times (L_1 + L_2)}$ – (6)

$I = I_{ijk \to ikj};\ I \in \mathbb{R}^{N \times (L_1 + L_2) \times C_2}$ – (7)

Multi-layer perceptron (MLP) block

The feature dimension of the interaction tensor O is progressively decreased using a multilayer perceptron (MLP) to obtain O_{red}.

$O_{red} = m \times \{ LNorm(ELU(Linear(x))) \};\ O_{red} \in \mathbb{R}^{N \times L_1 \times L_2 \times C_3}$ – (8)

We used a 3-layer ($m = 3$) MLP for contact map prediction whereas it was not used for interface residue prediction ($m = 0$).

Output block

It comprises a $Linear$ layer followed by a $Sigmoid$ activation.

$ContactMap = Sigmoid(Linear(O_{red}));\ ContactMap \in \mathbb{R}^{N \times L_1 \times L_2}$ – (9)

$Interface = Sigmoid(Linear(I));\ Interface \in \mathbb{R}^{N \times (L_1 + L_2)}$ – (10)

Training

The models were trained using PyTorch version 2.0.1 and CUDA version 11.8, using an NVIDIA A6000 GPU. The model was trained using the Singularity Enhanced (SE) loss function (Si & Yan, 2021), a modified form of the binary cross-entropy (BCE) loss used for class-imbalanced training, with $\alpha = 0.9$, $\beta = 3$ and AdamW (amsgrad variant) optimizer with a weight decay of 0.05. We used an exponential learning rate scheduler for slowly decaying the base learning rate.

$SELoss = \sum(\alpha \cdot y\, log(\hat{y}) + (1 - \alpha) \cdot (1 - y)\, log((1 - \hat{y})) \cdot (1 + \hat{y})^{\beta})$ – (10)

A binary mask was used to exclude the padding from the loss calculation. The train and dev set loss were monitored to check for over/under-fitting (Fig S1).

Evaluating predictions

The Disobind predicted score was converted to a binary output using a threshold of 0.5. The model performance was evaluated on the train, dev, and ID test set using recall, precision, and F1-score. All metrics were calculated using the torchmetrics library.

$$Recall = \frac{TP}{TP + FN} \qquad Precision = \frac{TP}{TP + FP} \qquad F1\ score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \qquad - (11)$$

Hyperparameter tuning and ablations

Here, we explore the effect of select hyperparameters on the model performances in the dev and ID test sets. Hyperparameters were tuned for the contact map and interface prediction at CG 1 and minimally modified for the CG 5 and CG 10 models. The final set of hyperparameters is provided (Table S1).

Projection dimension

We tuned the projection dimension C_1 for the projection block (Table S2).

Number of layers in the MLP

$Linear$ layers in the MLP were used to either upsample (upsampling layers, US) or downsample (downsampling layers, DS) the feature dimension by a factor of 2. We tuned the number of US (DS) layers. Additionally, we removed the MLP for interface residue prediction (Table S3).

α, β for SE loss

We tuned the α and β parameters for SE loss; these weigh the contribution of the contact and non-contact elements to the loss (Table S4).

pLM embeddings

We compared the model performance using embeddings from ProtT5 ($C = 1024$), ProstT5 ($C = 1024$), ProSE ($C = 6165$), and ProtBERT ($C = 1024$) (Bepler & Berger, 2021; Elnaggar et al., 2022; Heinzinger et al., 2024).

Further, we compared two types of embeddings: global and local embeddings (Fig 4, Table S8). Local embeddings were obtained using the sequence of the fragment as input to the pLM. In contrast, global embeddings were obtained by using the complete UniProt sequence as input to the pLM and extracting fragment embedding from it.

Random baseline

To compute random baseline predictions, we predicted random contacts and interface residues based on their frequencies in the training set. For the prediction tasks involving coarse-graining, these frequencies were determined from the coarse-grained contact maps and coarse-grained interface residues in the training set.

AlphaFold predictions on the OOD test set

We used a local AF2 installation and the AF3 webserver for obtaining OOD test set predictions (Abramson et al., 2024; Evans et al., 2022, 2022). AF2 was run with the default arguments as specified in their GitHub repository. Due to the failure of AF2 to relax several of the OOD test set entries we switch off the Amber relaxation step and consider the unrelaxed models for further evaluation. Additionally, we removed an OOD test set entry for which we were unable to obtain the AF2 prediction. We considered the best models from AF2 (based on the ipTM+pTM score) and AF3 (based on the ranking score) for comparison with Disobind. The ipTM score was used to assess the overall model confidence and the pLDDT and PAE metrics assessed the per-residue confidence. pLDDT scores higher than 70 and PAE values lower than 5 are considered as confident (Edmunds et al., 2024). An ipTM score higher than 0.8 represents a high-confidence prediction whereas an ipTM score lower than 0.6 represents a likely incorrect prediction; predictions with ipTM scores between 0.6-0.8 require further validation (Abramson et al., 2024; Evans et al., 2022; O'Reilly et al., 2023; Yin et al., 2022). Given that only five AF2 predictions and none

of the AF3 predictions in the OOD test set have an ipTM score greater than 0.8, we initially applied an ipTM cutoff of 0.75 as recommended in benchmarks on AlphaFold2 (O'Reilly et al., 2023; Yin et al., 2022). Later we relaxed the ipTM cutoffs to 0.4 and 0 (*i.e.*, no ipTM cutoff).

To compare the AlphaFold outputs to those from our models, we first derive the binary contact maps from the best-predicted model. We then apply binary masks on this output to ignore low-confidence interactions (pLDDT<70 and PAE>5). Both the pLDDT and PAE masks are applied on the CG 1 contact map obtained from AF2 and AF3 before proceeding further. To apply an ipTM cutoff, we zero the contact maps for AlphaFold predictions with an ipTM lower than the cutoff. The binary contact map was zero-padded to the maximum length (100 residues). The interface residues and coarse-grained predictions were derived from the CG 1 contact map as mentioned previously (See Inputs and Outputs for training).

Predictions on disordered residues alone

Disordered residues were identified based on DisProt, IDEAL, and MobiDB annotations (See Creating dataset for Disobind). To evaluate the predictions on these disordered residues alone, we obtain predictions from Disobind and AF2 as before and apply a binary mask to ignore interface residues (contacts) not involving disordered residues (disordered residue pairs). We only consider confident predictions from AF2 having pLDDT scores higher than 70 and PAE values lower than 5. No ipTM cutoff is used.

Disobind+AF2 predictions

The Disobind+AF2 predictions were obtained by combining the corresponding predictions from Disobind and AF2 using a logical OR operation. We ignore low-confidence AF2 predictions having a pLDDT less than 70 and a PAE greater than 5. No ipTM cutoff is used.

Code availability

The Disobind model and scripts to create datasets, train the model, use Disobind/Disobind+AF2, and perform analysis are available at GitHub https://github.com/isblab/disobind.

Data availability

Data is deposited to Zenodo at 10.5281/zenodo.14504763. The deposition contains all datasets (train, dev , IDT, OOD) along with the input sequences and target contact maps, AF2 and AF3 predictions for the OOD test set, and the PDB structures, SIFTS mappings, PDB API files, and Uniprot Sequences used to create the datasets.

Acknowledgement

We thank ISB Lab members Shreyas Arvindekar, Muskaan Jindal, Omkar Golatkar, and Mubashira KP for their useful comments on the manuscript. Molecular graphics images were produced using the UCSF Chimera and UCSF ChimeraX packages from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR001081, NIH R01-GM129325, and National Institute of Allergy and Infectious Diseases).

Author contribution

Conceptualization: K.M., S.V.
Reading and synthesis: K.M., S.V.
Data curation: K.M.
Methodology: K.M., V.U., S.V.
Writing – original draft: K.M.
Writing – review & editing: K.M., V.U., S.V.
Resources, Supervision, Funding acquisition: S.V.

Funding

This work has been supported by the following grants: Department of Atomic Energy (DAE) TIFR grant RTI 4006, Department of Science and Technology (DST) SERB grant SPG/2020/000475, and Department of Biotechnology (DBT) grant BT/PR40323/BTIS/137/78/2023 from the Government of India to S.V.

Conflict of Interest declaration

None declared.

References

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., Bodenstein, S. W., Evans, D. A., Hung, C.-C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., … Jumper, J. M. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, *630*(8016), 493–500. https://doi.org/10.1038/s41586-024-07487-w

Akey, C. W., Singh, D., Ouch, C., Echeverria, I., Nudelman, I., Varberg, J. M., Yu, Z., Fang, F., Shi, Y., Wang, J., Salzberg, D., Song, K., Xu, C., Gumbart, J. C., Suslov, S., Unruh, J., Jaspersen, S. L., Chait, B. T., Sali, A., … Rout, M. P. (2022). Comprehensive structure and functional adaptations of the yeast nuclear pore complex. *Cell*, *185*(2), 361-378.e25. https://doi.org/10.1016/j.cell.2021.12.015

Alderson, T. R., Pritišanac, I., Kolarić, Đ., Moses, A. M., & Forman-Kay, J. D. (2023). Systematic identification of conditionally folded intrinsically disordered regions

by AlphaFold2. *Proceedings of the National Academy of Sciences*, *120*(44), e2304302120. https://doi.org/10.1073/pnas.2304302120

Arvindekar, S., Jackman, M. J., Low, J. K. K., Landsberg, M. J., Mackay, J. P., & Viswanath, S. (2022). Molecular architecture of nucleosome remodeling and deacetylase sub-complexes by integrative structure determination. *Protein Science*, *31*(9), e4387. https://doi.org/10.1002/pro.4387

Arvindekar, S., Pathak, A. S., Majila, K., & Viswanath, S. (2024). Optimizing representations for integrative structural modeling using Bayesian model selection. *Bioinformatics*, *40*(3), btae106. https://doi.org/10.1093/bioinformatics/btae106

Aspromonte, M. C., Nugnes, M. V., Quaglia, F., Bouharoua, A., DisProt Consortium, Tosatto, S. C. E., & Piovesan, D. (2024). DisProt in 2024: Improving function annotation of intrinsically disordered proteins. *Nucleic Acids Research*, *52*(D1), D434–D441. https://doi.org/10.1093/nar/gkad928

Bepler, T., & Berger, B. (2019). *Learning protein sequence embeddings using information from structure* (arXiv:1902.08661). arXiv. https://doi.org/10.48550/arXiv.1902.08661

Bepler, T., & Berger, B. (2021). Learning the protein language: Evolution, structure, and function. *Cell Systems*, *12*(6), 654-669.e3. https://doi.org/10.1016/j.cels.2021.05.017

Bret, H., Gao, J., Zea, D. J., Andreani, J., & Guerois, R. (2024). From interaction networks to interfaces, scanning intrinsically disordered regions using AlphaFold2. *Nature Communications*, *15*(1), 597. https://doi.org/10.1038/s41467-023-44288-7

Christoffer, C., & Kihara, D. (2020). IDP-LZerD: Software for Modeling Disordered

Protein Interactions. In D. Kihara (Ed.), *Protein Structure Prediction* (pp. 231–244). Springer US. https://doi.org/10.1007/978-1-0716-0708-4_13

Dai, B., & Bailey-Kellogg, C. (2021). Protein interaction interface region prediction by geometric deep learning. *Bioinformatics*, *37*(17), 2580–2588. https://doi.org/10.1093/bioinformatics/btab154

del Alamo, D., Sala, D., Mchaourab, H. S., & Meiler, J. (2022). Sampling alternative conformational states of transporters and receptors with AlphaFold2. *eLife*, *11*, e75751. https://doi.org/10.7554/eLife.75751

Dominguez, C., Boelens, R., & Bonvin, A. M. J. J. (2003). HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *Journal of the American Chemical Society*, *125*(7), 1731–1737. https://doi.org/10.1021/ja026939x

Edmunds, N. S., Genc, A. G., & McGuffin, L. J. (2024). Benchmarking of AlphaFold2 accuracy self-estimates as indicators of empirical model quality and ranking: A comparison with independent model quality assessment programmes. *Bioinformatics*, *40*(8), btae491. https://doi.org/10.1093/bioinformatics/btae491

Elnaggar, A., Heinzinger, M., Dallago, C., Rehawi, G., Wang, Y., Jones, L., Gibbs, T., Feher, T., Angerer, C., Steinegger, M., Bhowmik, D., & Rost, B. (2022). ProtTrans: Toward Understanding the Language of Life Through Self-Supervised Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(10), 7112–7127. IEEE Transactions on Pattern Analysis and Machine Intelligence. https://doi.org/10.1109/TPAMI.2021.3095381

Escobedo, N., Saldaño, T., Mac Donagh, J., Sawicki, L. R., Palopoli, N., Alberti, S. F., Fornasari, M. S., & Parisi, G. (2024). Revealing Missing Protein–Ligand

Interactions Using AlphaFold Predictions. *Journal of Molecular Biology*, *436*(23), 168852. https://doi.org/10.1016/j.jmb.2024.168852

Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J., Ronneberger, O., Bodenstein, S., Zielinski, M., Bridgland, A., Potapenko, A., Cowie, A., Tunyasuvunakool, K., Jain, R., Clancy, E., … Hassabis, D. (2022). *Protein complex prediction with AlphaFold-Multimer* (p. 2021.10.04.463034). bioRxiv. https://doi.org/10.1101/2021.10.04.463034

Fichó, E., Reményi, I., Simon, I., & Mészáros, B. (2017). MFIB: A repository of protein complexes with mutual folding induced by binding. *Bioinformatics*, *33*(22), 3682–3684. https://doi.org/10.1093/bioinformatics/btx486

Fonin, A. V., Darling, A. L., Kuznetsova, I. M., Turoverov, K. K., & Uversky, V. N. (2018). Intrinsically disordered proteins in crowded milieu: When chaos prevails within the cellular gumbo. *Cellular and Molecular Life Sciences: CMLS*, *75*(21), 3907–3929. https://doi.org/10.1007/s00018-018-2894-9

Fukuchi, S., Amemiya, T., Sakamoto, S., Nobe, Y., Hosoda, K., Kado, Y., Murakami, S. D., Koike, R., Hiroaki, H., & Ota, M. (2014). IDEAL in 2014 illustrates interaction networks composed of intrinsically disordered proteins and their binding partners. *Nucleic Acids Research*, *42*(D1), D320–D325. https://doi.org/10.1093/nar/gkt1010

Gainza, P., Sverrisson, F., Monti, F., Rodolà, E., Boscaini, D., Bronstein, M. M., & Correia, B. E. (2020). Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nature Methods*, *17*(2), Article 2. https://doi.org/10.1038/s41592-019-0666-6

Ghafouri, H., Lazar, T., Del Conte, A., Tenorio Ku, L. G., PED Consortium, Tompa, P.,

Tosatto, S. C. E., & Monzon, A. M. (2024). PED in 2024: Improving the community deposition of structural ensembles for intrinsically disordered proteins. *Nucleic Acids Research*, *52*(D1), D536–D544. https://doi.org/10.1093/nar/gkad947

Guo, Z., Liu, J., Skolnick, J., & Cheng, J. (2022). Prediction of inter-chain distance maps of protein complexes with 2D attention-based deep neural networks. *Nature Communications*, *13*(1), 6963. https://doi.org/10.1038/s41467-022-34600-2

Heinzinger, M., Weissenow, K., Sanchez, J. G., Henkel, A., Mirdita, M., Steinegger, M., & Rost, B. (2024). Bilingual language model for protein sequence and structure. *NAR Genomics and Bioinformatics*, *6*(4), lqae150. https://doi.org/10.1093/nargab/lqae150

Jahn, L. R., Marquet, C., Heinzinger, M., & Rost, B. (2024). Protein embeddings predict binding residues in disordered regions. *Scientific Reports*, *14*(1), 13566. https://doi.org/10.1038/s41598-024-64211-4

Janson, G., & Feig, M. (2024). *Transferable deep generative modeling of intrinsically disordered protein conformations*. https://doi.org/10.1101/2024.02.08.579522

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

Kraus, A., Hoyt, F., Schwartz, C. L., Hansen, B., Artikis, E., Hughson, A. G., Raymond, G. J., Race, B., Baron, G. S., & Caughey, B. (2021).

High-resolution structure and strain comparison of infectious mammalian prions. *Molecular Cell*, *81*(21), 4540-4551.e6. https://doi.org/10.1016/j.molcel.2021.08.011

Lee, C. Y., Hubrich, D., Varga, J. K., Schäfer, C., Welzel, M., Schumbera, E., Djokic, M., Strom, J. M., Schönfeld, J., Geist, J. L., Polat, F., Gibson, T. J., Keller Valsecchi, C. I., Kumar, M., Schueler-Furman, O., & Luck, K. (2024). Systematic discovery of protein interaction interfaces using AlphaFold and experimental validation. *Molecular Systems Biology*, *20*(2), 75–97. https://doi.org/10.1038/s44320-023-00005-6

Lin, P., Tao, H., Li, H., & Huang, S.-Y. (2023). Protein–protein contact prediction by geometric triangle-aware protein language models. *Nature Machine Intelligence*, *5*(11), 1275–1284. https://doi.org/10.1038/s42256-023-00741-2

Lindorff-Larsen, K., & Kragelund, B. B. (2021). On the Potential of Machine Learning to Examine the Relationship Between Sequence, Structure, Dynamics and Function of Intrinsically Disordered Proteins. *Journal of Molecular Biology*, *433*(20), 167196. https://doi.org/10.1016/j.jmb.2021.167196

Magana, P. G., & Kovalevskiy, G. O. (2024). *AlphaFold—A practical guide*. https://www.ebi.ac.uk/training/online/courses/alphafold

Majila, K., Arvindekar, S., Jindal, M., & Viswanath, S. (2024). *Frontiers in integrative structural biology: Modeling disordered proteins and utilizing in situ data* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2407.00566

Mansoor, S., Baek, M., Park, H., Lee, G. R., & Baker, D. (2024). Protein Ensemble Generation Through Variational Autoencoder Latent Space Sampling. *Journal of Chemical Theory and Computation*, *20*(7), 2689–2695. https://doi.org/10.1021/acs.jctc.3c01057

Mészáros, B., Simon, I., & Dosztányi, Z. (2009). Prediction of Protein Binding Regions in Disordered Proteins. *PLOS Computational Biology*, *5*(5), e1000376. https://doi.org/10.1371/journal.pcbi.1000376

Miskei, M., Antal, C., & Fuxreiter, M. (2017). FuzDB: Database of fuzzy complexes, a tool to develop stochastic structure-function relationships for protein complexes and higher-order assemblies. *Nucleic Acids Research*, *45*(D1), D228–D235. https://doi.org/10.1093/nar/gkw1019

Miskei, M., Horvath, A., Vendruscolo, M., & Fuxreiter, M. (2020). Sequence-Based Prediction of Fuzzy Protein Interactions. *Journal of Molecular Biology*, *432*(7), 2289–2303. https://doi.org/10.1016/j.jmb.2020.02.017

Mosalaganti, S., Obarska-Kosinska, A., Siggel, M., Taniguchi, R., Turoňová, B., Zimmerli, C. E., Buczak, K., Schmidt, F. H., Margiotta, E., Mackmull, M.-T., Hagen, W. J. H., Hummer, G., Kosinski, J., & Beck, M. (2022). AI-based structure prediction empowers integrative structural analysis of human nuclear pores. *Science*, *376*(6598), eabm9506. https://doi.org/10.1126/science.abm9506

Noid, W. G. (2013). Perspective: Coarse-grained models for biomolecular systems. *The Journal of Chemical Physics*, *139*(9), 090901. https://doi.org/10.1063/1.4818908

Oldfield, C. J., & Dunker, A. K. (2014). Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions. *Annual Review of Biochemistry*, *83*(Volume 83, 2014), 553–584. https://doi.org/10.1146/annurev-biochem-072711-164947

Olsen, J. G., Teilum, K., & Kragelund, B. B. (2017). Behaviour of intrinsically disordered proteins in protein–protein complexes with an emphasis on

fuzziness. *Cellular and Molecular Life Sciences*, *74*(17), 3175–3183. https://doi.org/10.1007/s00018-017-2560-7

Omidi, A., Møller, M. H., Malhis, N., Bui, J. M., & Gsponer, J. (2024). AlphaFold-Multimer accurately captures interactions and dynamics of intrinsically disordered protein regions. *Proceedings of the National Academy of Sciences*, *121*(44), e2406407121. https://doi.org/10.1073/pnas.2406407121

O'Reilly, F. J., Graziadei, A., Forbrig, C., Bremenkamp, R., Charles, K., Lenz, S., Elfmann, C., Fischer, L., Stülke, J., & Rappsilber, J. (2023). Protein complexes in cells by AI-assisted structural proteomics. *Molecular Systems Biology*, *19*(4), e11544. https://doi.org/10.15252/msb.202311544

Pasani, S., Menon, K. S., & Viswanath, S. (2024). The molecular architecture of the desmosomal outer dense plaque by integrative structural modeling. *Protein Science*, *33*(12), e5217. https://doi.org/10.1002/pro.5217

Piovesan, D., Del Conte, A., Mehdiabadi, M., Aspromonte, M. C., Blum, M., Tesei, G., von Bülow, S., Lindorff-Larsen, K., & Tosatto, S. C. E. (2024). MOBIDB in 2025: Integrating ensemble properties and function annotations for intrinsically disordered proteins. *Nucleic Acids Research*, gkae969. https://doi.org/10.1093/nar/gkae969

Pittala, S., & Bailey-Kellogg, C. (2020). Learning context-aware structural representations to predict antigen and antibody binding interfaces. *Bioinformatics (Oxford, England)*, *36*(13), 3996–4003. https://doi.org/10.1093/bioinformatics/btaa263

Rantos, V., Karius, K., & Kosinski, J. (2022). Integrative structural modeling of macromolecular complexes using Assembline. *Nature Protocols*, *17*(1), Article

1. https://doi.org/10.1038/s41596-021-00640-z

Rives, A., Meier, J., Sercu, T., Goyal, S., Lin, Z., Liu, J., Guo, D., Ott, M., Zitnick, C. L., Ma, J., & Fergus, R. (2021). Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proceedings of the National Academy of Sciences*, *118*(15), e2016239118. https://doi.org/10.1073/pnas.2016239118

Rogers, J. R., Nikolényi, G., & AlQuraishi, M. (2023). Growing ecosystem of deep learning methods for modeling protein–protein interactions. *Protein Engineering, Design and Selection*, *36*, gzad023. https://doi.org/10.1093/protein/gzad023

Rose, Y., Duarte, J. M., Lowe, R., Segura, J., Bi, C., Bhikadiya, C., Chen, L., Rose, A. S., Bittrich, S., Burley, S. K., & Westbrook, J. D. (2021). RCSB Protein Data Bank: Architectural Advances Towards Integrated Searching and Efficient Access to Macromolecular Structure Data from the PDB Archive. *Journal of Molecular Biology*, *433*(11), 166704. https://doi.org/10.1016/j.jmb.2020.11.003

Rossi, D., & Zlotnik, A. (2000). The Biology of Chemokines and their Receptors. *Annual Review of Immunology*, *18*(Volume 18, 2000), 217–242. https://doi.org/10.1146/annurev.immunol.18.1.217

Ruff, K. M., & Pappu, R. V. (2021). AlphaFold and Implications for Intrinsically Disordered Proteins. *Journal of Molecular Biology*, *433*(20), 167208. https://doi.org/10.1016/j.jmb.2021.167208

Russel, D., Lasker, K., Webb, B., Velázquez-Muriel, J., Tjioe, E., Schneidman-Duhovny, D., Peterson, B., & Sali, A. (2012). Putting the Pieces Together: Integrative Modeling Platform Software for Structure Determination of Macromolecular Assemblies. *PLoS Biology*, *10*(1), e1001244.

https://doi.org/10.1371/journal.pbio.1001244

Schad, E., Fichó, E., Pancsa, R., Simon, I., Dosztányi, Z., & Mészáros, B. (2018). DIBS: A repository of disordered binding sites mediating interactions with ordered proteins. *Bioinformatics (Oxford, England)*, *34*(3), 535–537. https://doi.org/10.1093/bioinformatics/btx640

Sepuru, K. M., Nair, V., Prakash, P., Gorfe, A. A., & Rajarathnam, K. (2020). Long-Range Coupled Motions Underlie Ligand Recognition by a Chemokine Receptor. *iScience*, *23*(12), 101858. https://doi.org/10.1016/j.isci.2020.101858

Si, Y., & Yan, C. (2021). Improved protein contact prediction using dimensional hybrid residual networks and singularity enhanced loss function. *Briefings in Bioinformatics*, *22*(6), bbab341. https://doi.org/10.1093/bib/bbab341

Sigurdson, C. J., Bartz, J. C., & Glatzel, M. (2019). Cellular and Molecular Mechanisms of Prion Disease. *Annual Review of Pathology: Mechanisms of Disease*, *14*(Volume 14, 2019), 497–516. https://doi.org/10.1146/annurev-pathmechdis-012418-013109

Singh, J., Litfin, T., Singh, J., Paliwal, K., & Zhou, Y. (2022). SPOT-Contact-LM: Improving single-sequence-based prediction of protein contact map using a transformer language model. *Bioinformatics*, *38*(7), 1888–1894. https://doi.org/10.1093/bioinformatics/btac053

Steinegger, M., & Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology*, *35*(11), 1026–1028. https://doi.org/10.1038/nbt.3988

Tompa, P., Davey, N. E., Gibson, T. J., & Babu, M. M. (2014). A million peptide motifs for the molecular biologist. *Molecular Cell*, *55*(2), 161–169. https://doi.org/10.1016/j.molcel.2014.05.032

Uversky, V. N. (2013). The most important thing is the tail: Multitudinous functionalities of intrinsically disordered protein termini. *FEBS Letters*, *587*(13), 1891–1901. https://doi.org/10.1016/j.febslet.2013.04.042

Velankar, S., Dana, J. M., Jacobsen, J., van Ginkel, G., Gane, P. J., Luo, J., Oldfield, T. J., O'Donovan, C., Martin, M.-J., & Kleywegt, G. J. (2013). SIFTS: Structure Integration with Function, Taxonomy and Sequences resource. *Nucleic Acids Research*, *41*(D1), D483–D489. https://doi.org/10.1093/nar/gks1258

Verburgt, J., Zhang, Z., & Kihara, D. (2022). Multi-level analysis of intrinsically disordered protein docking methods. *Methods*, *204*, 55–63. https://doi.org/10.1016/j.ymeth.2022.05.006

Wang, D., Wang, Y., Evans, L., & Tiwary, P. (2024). From Latent Dynamics to Meaningful Representations. *Journal of Chemical Theory and Computation*, *20*(9), 3503–3513. https://doi.org/10.1021/acs.jctc.4c00249

Wayment-Steele, H. K., Ojoawo, A., Otten, R., Apitz, J. M., Pitsawong, W., Hömberger, M., Ovchinnikov, S., Colwell, L., & Kern, D. (2024). Predicting multiple conformations via sequence clustering and AlphaFold2. *Nature*, *625*(7996), 832–839. https://doi.org/10.1038/s41586-023-06832-9

Wright, P. E., & Dyson, H. J. (1999). Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. *Journal of Molecular Biology*, *293*(2), 321–331. https://doi.org/10.1006/jmbi.1999.3110

Yin, R., Feng, B. Y., Varshney, A., & Pierce, B. G. (2022). Benchmarking AlphaFold for protein complex modeling reveals accuracy determinants. *Protein Science: A Publication of the Protein Society*, *31*(8), e4379. https://doi.org/10.1002/pro.4379

Zhang, F., Zhao, B., Shi, W., Li, M., & Kurgan, L. (2022). DeepDISOBind: Accurate

prediction of RNA-, DNA- and protein-binding intrinsically disordered residues with deep multi-task learning. *Briefings in Bioinformatics*, *23*(1), bbab521. https://doi.org/10.1093/bib/bbab521

Zhang, J., Durham, J., & Qian Cong. (2024). Revolutionizing protein–protein interaction prediction with deep learning. *Current Opinion in Structural Biology*, *85*, 102775. https://doi.org/10.1016/j.sbi.2024.102775