Computational and Structural Biotechnology Journal 27 (2025) 78–88

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Review article

Evaluation of predictions of disordered binding regions in the CAID2 experiment

Fuhao Zhang [a,*], Lukasz Kurgan [b,*]

[a] College of Information Engineering, Northwest A & F University, China
[b] Department of Computer Science, Virginia Commonwealth University, Richmond, VA 23284, USA

ARTICLE INFO

Keywords:
Intrinsic disorder
Protein-protein binding
Protein-nucleic acids binding
Assessment
Prediction
Cross prediction

ABSTRACT

A large portion of the Intrinsically Disordered Regions (IDRs) in protein sequences interact with proteins, nucleic acids, and other types of ligands. Correspondingly, dozens of sequence-based predictors of binding IDRs were developed. A recently completed second community-based Critical Assessments of protein Intrinsic Disorder prediction (CAID2) evaluated 32 predictors of binding IDRs. However, CAID2 considered a rather narrow scenario by testing on 78 proteins with binding IDRs and not differentiating between different ligands, in spite that virtually all predictors target IDRs that interact with specific types of ligands. In that scenario, several intrinsic disorder predictors predict binding IDRs with accuracy equivalent to the best predictors of binding IDRs since large majority of IDRs in the 78 test proteins are binding. We substantially extended the CAID2's evaluation by using the entire CAID2 dataset of 348 proteins and considering several arguably more practical scenarios. We assessed whether predictors accurately differentiate binding IDRs from other types of IDRs and how they perform when predicting IDRs that interact with different ligand types. We found that intrinsic disorder predictors cannot accurately identify binding IDRs among other disordered regions, majority of the predictors of binding IDRs are ligand type agnostic (i.e., they cross predict binding in IDRs that interact with ligands that they do not cover), and only a handful of predictors of binding IDRs perform relatively well and generate reasonably low amounts of cross predictions. We also suggest a number of future research directions that would move this active field of research forward.

1. Introduction

Intrinsically disordered regions (IDRs) are segments in protein chains that lack a stable tertiary structure under physiological conditions and which typically form dynamic ensembles of conformations [1–3]. Proteins with IDRs facilitate a broad range of cellular functions [4–9], are particularly abundant in eukaryotic proteins [10–13], and localize across several subcellular compartments [14]. Many IDRs interact with partner molecules, which include proteins, nucleic acids, lipids, and a variety of small molecules [15–26]. The underlying conformational plasticity of binding IDRs differentiates them from the structured binding regions/sites, in particular allowing a single IDR to interact with multiple different partner molecules [27–30]. The binding IDRs are available in several databases that include DisProt [31], IDEAL [32], and MobiDB [33]. However, only a relatively small portion of binding IDRs has been characterized experimentally, prompting the development of computational predictors. Dozens of methods that predict binding IDRs in protein sequences were developed and published [34–42]. Majority of these tools focus on predicting molecular recognition features (MoRFs) [37,40], which are short IDRs (10–70 consecutive residues) that typically undergo disorder-to-structure transitions upon binding protein partners, i.e., they "morph" from disorder to order [43–45]. MoRF regions can remain partly or even fully disordered in the bound state [46,47], leading to the disorder-to-disorder transitions upon binding [48]. Moreover, the corresponding fuzzy complexes/assemblies extend to other types of binding IDRs, beyond MoRFs [49]. The first predictor of binding IDRs that was published in 2015, α-MoRFpred, targeted prediction of MoRFs that fold into structures that primarily consist of alpha-helices [50]. Moreover, over two dozen predictors that target other types of binding IDRs were published in recent years [40]. They include predictors of the protein-binding IDRs [51–53], lipid binding IDRs [54–56], and tools capable of predicting IDRs that interact with multiple types of ligands, such as proteins and nucleic acids [57–62].

Availability of the large number of predictors of binding IDRs

* Corresponding authors.
E-mail addresses: fhzhang@nwafu.edu.cn (F. Zhang), lkurgan@vcu.edu (L. Kurgan).

https://doi.org/10.1016/j.csbj.2024.12.009
Received 21 October 2024; Received in revised form 12 December 2024; Accepted 13 December 2024
Available online 17 December 2024

motivated inclusion of the corresponding evaluation in the large community-driven Critical Assessments of protein Intrinsic Disorder prediction (CAID) experiments: CAID1 in 2021 [63] and CAID2 in 2023 [64]. CAID experiments were conducted by independent evaluators (who exclude authors of the evaluated predictors) using large and blind test datasets (authors of predictors do not have access to the test data before the experiment), community accepted metrics, and predictors that were provided to the evaluators before the experiments started. Consequently, these evaluations are arguably more objective when compared to smaller-scale evaluations that are done when individual methods are published. The inaugural CAID1 evaluated 11 predictors of binding IDRs, while that number had grown to 32 in CAID2. The CAID2 experiment evaluated these methods on 78 disordered proteins that have binding IDRs, showing that several predictors, such as MoRFchibi [65], ENSHROUD, DeepDRPBind, and DeepDISObind [61], performed relatively well, achieving Area Under the ROC Curve (AUC) > 0.73 [64]. However, the evaluation of the binding IDR predictions in CAID2 suffers two key limitations. First, it focused on a relatively narrow scenario by evaluating predictors on the 78 proteins that are already known to have binding IDRs and where nearly 85 % of IDRs are binding. This potentially allows generic disorder predictors to correctly identify binding IDRs in these proteins and does not assess ability of predictors to discriminate between binding IDRs and other types of IDRs, e.g. entropic chains and post-translational modification display sites [31]. Second, CAID2 evaluated on binding IDRs without differentiating which ligand (s) they interact with, while predictors of binding IDRs typically target specific ligands types, such as proteins vs nucleic acids. This is problematic since different methods are penalized at different degrees, depending on the ligand type(s) they target, for not predicting interactions with the ligands they were not designed to predict, e.g., performance of predictors of nucleic acid binding IDRs was lowered when they (correctly) did not predict protein-binding IDRs that were included among the binding IDRs. Moreover, that analysis did not allow for quantifying and analyzing cross-predictions, which is when methods that target interactions with a specific ligand type also (cross-)predict interactions with other types of ligands. This is an important issue since literature shows that some predictors of protein-protein interactions cross predict protein-nucleic acids interactions and vice versa [66–68].

Motivated by the availability of the quality benchmark data and predictions from the CAID2 experiment, we evaluated predictions of binding IDRs by considering arguably more practical scenarios. First, we used the entire test dataset from CAID2, which covers the 78 proteins with binding IDRs and 270 other disordered proteins, to evaluate ability of 32 predictors of binding IDRs and a selection of 8 best disorder predictors to distinguish binding IDRs from other types of IDRs. Second, we assessed predictions separately for IDRs that interact with specific types of ligands and correspondingly divided the predictors by the ligands they target. This facilitated a more appropriate analysis where scope of the predictors was correctly matched with the type of the binding IDRs used for the evaluation and where we were able to quantify and compare the cross-predictions. Compared to the past studies of the cross predictions that focused on the predictions in structured (ordered) regions and for protein and nucleic acids interactions [66–68], we investigate predictions in IDRs and consider interactions with proteins, nucleic acids, and other (non-protein and non-nucleic acid) ligands. Our analysis offers a number of practical observations that summarize state-of-the-art in this active research area and which ultimately motivate the need for further efforts towards the development of more accurate predictors of binding IDRs.

2. Materials and methods

2.1. Selection of evaluated predictors

We evaluated all 32 binding IDR predictors and a selection of best disorder predictors that participated in CAID2 [64]. We divided the binding IDR predictors into four categories based on the ligands that they target:

1. 11 predictors of the protein binding IDRs: ENSHROUD-protein, MoRFchibi-web [65], DeepDRPBind-protein [69], MoRFchibi-light [65], DeepDISOBind-protein [61], OPAL [70,71], DRPBind-protein [69], DisoRDPbind-prot [59,60], ANCHOR2 [52], ProBiPred-protein, and DISOPRED3-bind [72].
2. 14 predictors of the nucleic acid binding IDRs: ENSHROUD-nucleic, DeepDISOBind-nucleic [61], DeepDRPBind-RNA [69], DeepDRPBind-DNA [69], DeepDRPBind_nuc [69], DRPBind_nuc [69], DRPBind-DNA [69], DRPBind-RNA [69], ProBiPred-nucleic, bindEmbed21IDR-idrNuc [73], bindEmbed21IDR-rawNuc [73], DisoRDPbind-rna [59,74], DisoRDPbind-dna [59,60], and DisoRDPbind-nuc [59,60].
3. 3 predictors of protein binding and nucleic acid binding IDRs: ENSHROUD-all, DeepDISOBind-all [61], and CLIP [57].
4. 4 predictors that cover a broad spectrum of ligands that include 2 that target prediction of ligand type agnostic (all) binding IDRs (DisoBindPred and AlphaFold-binding [42]) and 2 that cover predictions of IDRs that interact with nucleic acids, metal ions and small molecules (bindEmbed21IDR-idrGeneral [73] and bindEmbed21IDR-rawGeneral [73]).

We selected eight most accurate disorder predictors based on the CAID2 results: flDPnn2 [75], Dispredict3 [76], flDPlr2 [75], DisoPred, SPOT-Disorder2 [77], AlphaFold-rsa [78], SETH-0 [79], and PredIDR-long [80]. We identified them by combining the lists of the top three predictors selected based on two complementary metrics, AUC, and AUPRC (Area Under the Prediction-Recall Curve), that were produced using two test datasets, Disorder-NOX and Disorder-PDB (Figure 2 in Ref. [64]). Moreover, we also introduced a random baseline, which generates random number in the 0 and 1 range for each residue in the test set, which we used as a point of reference.

2.2. Test dataset

We used the full CAID2 test set that is available at https://caid.idpcentral.org/challenge, which we accessed on 27th of April 2023. This dataset covers 348 disordered proteins that include the 78 proteins with binding IDRs. It contains 287,020 residues, of which 37,101 are in IDRs and 8209 are in binding IDRs. We employed a two-step approach to annotate the ligand types for the binding IDRs, which we need to evaluate predictors on the ligand type specific binding IDRs. We considered three types of ligands: proteins, nucleic acids, and other ligands, motivated by the breakdown of the ligand types that are targeted by the 32 evaluated predictors of binding IDRs. First, we gathered Intrinsically Disordered Proteins Ontology (IDPO) annotations [81] from the DisProt database [31] for the binding IDRs in the CAID2 dataset; the corresponding proteins were matched to DisProt records based on identifiers available in the CAID2 dataset. We were able to identify protein binding IDRs based on annotations labeled as 'protein binding', 'SH3 domain binding', and 'calmodulin binding'. Similarly, we identified nucleic acids binding IDRs based on the 'mRNA binding', 'DNA binding' and 'RNA binding' terms. Finally, we categorized the binding IDRs that had 'ion binding,' 'phosphatidylinositol binding', 'metal ion binding', 'small molecule binding' and 'lipid binding' annotations as interacting with other ligands. In the second step, we used the InterProScan [82] to scan the protein sequences for functional annotations based on Gene Ontology (GO) terms [83]. We categorized these annotations into the protein binding, nucleic acid binding, and other binding classes. We combined them with the annotations that we collected in the first step. As a result, we identified 3841 protein binding residues, 455 nucleic acid binding residues, and 1580 residues interacting with the other partners. Altogether, we were able to annotate the ligand type for 72 % of residues in the binding IDRs. We also identified a subset of disordered residues

that are functionally annotated and where these annotations suggest that they are not associated with binding. We used these functionally verified non-binding residues when evaluating the ability of predictors to differentiate between binding IDRs and other types of IDRs (non-binding IDRs). Like for the annotations of ligand types, we combined IDPO annotations that we collected from DisProt (e.g., 'flexible linker') and the GO terms that we obtained using InterProScan. This approach allowed us to annotate 10,775 functionally verified non-binding residues. Table 1 summarizes the corresponding CAID2 test dataset that we enriched with the ligand types and the non-binding annotations.

2.3. Evaluation

The selected 40 methods predict numeric propensities for binding (or disorder propensities for the 8 disorder predictors) for each amino acid in a given protein sequence. We evaluated the quality of the predicted propensities using the two metrics that were utilized in recent comparative assessments [37,63,64,84–88]: the Area Under the ROC Curve (AUC) and the Area Under the Precision-Recall Curve (AUPRC). The propensities were used to produce binary state predictions (binding vs. non-binding) by using a threshold, i.e., residues with propensities $>$ threshold are assumed to bind while the remaining residues are assumed not to bind. We relied on the commonly used Matthews Correlation Coefficient (MCC) metric [63,84–88] to assess the accuracy of these binary state predictions. We calibrated the binary predictions between methods by using thresholds that generated the number of the predicted disordered binding residues (disordered residues for the 8 disorder predictors) that is equal to the number of the native disordered binding residues. This ensures that MCC scores can be directly compared across predictors. Moreover, since AUPRC and MCC scores are sensitive to the rate of disordered binding residues (positives), we performed sampling that equalizes this rate across different evaluations that use different collections of proteins and residues (i.e., which evaluate different scenarios). Specifically, we retained all residues in the binding IDRs and randomly selected an equal number of non-binding residues. This allowed us to directly compare results across the considered scenarios.

We also conducted tests of statistical significance of differences between results of predictors. These tests evaluate whether these differences are robust across different datasets. To do that, we randomly selected 50 % of the binding residues along with an equal number of the non-binding residues, repeated this sampling 10 times, and compared results across these 10 experiments. We used the Kolmogorov-Smirnov test to check whether the results followed a normal distribution (p-value < 0.05). For the normally distributed values, we used the paired t-test to evaluate significance; otherwise, we employed the Wilcoxon test. Moreover, we performed the Benjamini-Hochberg procedure that reduces probability of false positives by adjusting the significance level for multiple comparisons.

3. Results

3.1. Evaluation of predictions of binding IDRs on the entire CAID2 dataset

When compared to the evaluation scenario considered in CAID2, we used a larger dataset of 348 proteins (vs 78 proteins with binding IDRs in CAID2), provided results for all 32 predictors of binding IDRs (CAID2 article reports results for the 10 best predictors), 8 best disorder predictors (CAID2 does not test disorder predictors on binding IDRs), and a random baseline (CAID2 does not include a baseline). Table 2 shows results when performing sampling (balanced dataset with equal number of binding IDR residues and non-binding residues), which facilitates direct comparison with results for the other considered scenarios, while Supplementary Table S1 shows results without sampling, as it was done in CAID2. The coverage column shows the fraction of residues for which a given predictor produced results. Two methods, SPOT-Disorder2 and AlphaFold2-derived predictor, provided particularly low coverage at 64 % and 77 % of residues, respectively. The low coverage of SPOT-Disorder2 is because it can be only applied to sequences with less than 750 amino acids. The AlphaFold2 predictions were collected from AlphaFold Protein Structure Database [89] by searching for the UniProtKB accession number, and correspondingly proteins which were not found in that database were left unpredicted. The likely reason for the incomplete coverage for the other predictors is the fact that the CAID2 organizers limited the runtime to a maximum of 4 h per sequence with an upper limit of 24 CPU cores and 47 GB of RAM. They noted that "some methods crashed unexpectedly or did not provide an output in a reasonable time" [64].

Table 2 shows that about 1/3 of the binding IDR predictors achieved reasonably good levels of predictive performance, with AUC > 0.70, AUPRC > 0.70 and MCC > 0.35. Interestingly, 9 of these 11 well-performing methods target prediction of IDRs that interact with proteins. This can be explained by the fact that majority of the binding IDRs in CAID2 dataset interact with proteins, particularly when compared with the number of IDRs that interact with nucleic acids (Table 1). Correspondingly, the relatively poor performance of the virtually all binding IDRs predictors that target interactions with nucleic acids, some of which score worse than the random baseline, can be explained by the fact that in this scenario (as in CAID2) they were (incorrectly) expected to predict IDRs that bind proteins. Essentially, the low AUC and AUPRC values for these predictors mean that they successfully avoid predicting the majority of the binding IDRs that interact with proteins. We addressed this issue in Section 3.3 where we test binding IDR predictors using the matching types of binding IDRs. We found that the AlphaFold2-based predictor of binding IDRs fails to predict 33 % of residues in the test proteins and secures AUC and MCC scores that are significantly worse than the corresponding scores of the best binding IDR predictor, ENSHROUD-protein (p-value < 0.05). This is in line with recent works that similarly suggest that AlphaFold-based disorder predictors are outperformed by state-of-the-art predictors of disorder [85,90,91].

Furthermore, Table 2 shows that 6 of the 8 disorder predictors secure the same levels of performance as the best predictors of binding IDRs (AUC > 0.70, AUPRC > 0.70 and MCC > 0.35). This is consistent with observations based on the original CAID2's results for binding IDRs. It stems from the fact that the binding IDRs constitute majority of the disordered residues and thus the disorder predictors can accurately identify them among the large majority of the non-disordered residues. This might lead to a misleading conclusion that disorder predictors can be used to accurately identify binding IDRs, while in fact this stems from an arguably flawed testing scenario. We addressed this issue in the Section 3.2 where we evaluated the ability of the considered predictors to predict binding IDRs among disordered regions.

Table 1
Summary of the test dataset.

Number of proteins	Number of residues						
	Disordered protein-binding	Disordered nucleic acid-binding	Disordered other-binding	All disordered binding	Disordered non-binding	All disordered residues	All residues
348	3841	455	1580	8209 (2.8 %)	10,775 (3.8 %)	37,101 (12.9 %)	287,020

Table 2
Predictive performance on the CAID2 test dataset with 348 disordered proteins using sampling that considers all binding residues and an equal number of randomly selected non-binding residues. The predictors of the binding IDRs and the disorder predictors are sorted in the descending order of their AUC values. Predictors that secure the highest AUC, AUPRC, and MCC scores are shown in bold font. Statistical significance of differences is annotated using the *xy* symbols after the numerical value, where $x = \{+, -, =\}$ denotes comparison of a given method against the best predictor of binding IDRs, and $y = \{+, -, =\}$ denotes comparison of a given method against the best predictor of disorder, and +, =, and – indicate that the best predictor of binding IDRs (disorder predictor) is significantly better (p-value < 0.05), not significantly different (p-value ≥ 0.05), and significantly worse (p-value < 0.05) than the given method. Evaluation metrics and statistical tests of significance are explained in Section 2.3.

Predictor type	Type of ligand targeted	Predictor name	Coverage	AUC	AUPRC	MCC
Predictors of binding IDRs	Proteins	**ENSHROUD-protein**	1	**0.792** / +	**0.744** / +	0.461 = =
	Proteins, nucleic acids	**ENSHROUD-all**	1	0.791 = +	0.736 = +	**0.466** / =
	Proteins, nucleic acids	DeepDISOBind-all	1	0.782 + +	0.727 + +	0.443 + +
	Nucleic acids	ENSHROUD-nucleic	1	0.773 + +	0.718 + +	0.425 + +
	Proteins	DeepDISOBind-protein	1	0.771 + +	0.718 + +	0.446 + +
	Proteins	DRPBind-protein	1	0.769 + +	0.741 = +	0.404 + +
	Proteins	MoRFchibi-web	1	0.765 + +	0.737 = +	0.410 + +
	All ligand types (type agnostic)	AlphaFold-binding	0.77	0.765 + +	0.733 = +	0.380 + +
	Proteins	MoRFchibi-light	1	0.764 + +	0.733 = +	0.399 + +
	Nucleic acids	DeepDISOBind-nucleic	1	0.759 + +	0.689 + +	0.412 + +
	Proteins	DeepDRPBind-protein	0.99	0.757 + +	0.738 = +	0.408 + +
	Proteins	DisoRDPbind-prot	1	0.757 + +	0.725 + +	0.423 + +
	Proteins	ANCHOR2	1	0.734 + +	0.677 + +	0.392 + +
	Nucleic acids	DRPBind-DNA	1	0.730 + +	0.665 + +	0.374 + +
	Nucleic acids	DeepDRPBind-RNA	0.99	0.718 + +	0.659 + +	0.329 + +
	Nucleic acids	DeepDRPBind_nuc	0.99	0.705 + +	0.653 + +	0.312 + +
	Proteins	OPAL	1	0.691 + +	0.684 + +	0.291 + +
	Nucleic acids	DRPBind_nuc	1	0.678 + +	0.630 + +	0.314 + +
	Proteins, nucleic acids	CLIP	0.97	0.672 + +	0.682 + +	0.261 + +
	Nucleic acids	DeepDRPBind-DNA	0.99	0.642 + +	0.616 + +	0.207 + +
	All ligand types (type agnostic)	DisoBindPred	0.94	0.642 + +	0.623 + +	0.243 + +
	Nucleic acids	ProBiPred-nucleic	1	0.618 + +	0.642 + +	0.150 + +
	Nucleic acids	DRPBind-RNA	1	0.596 + +	0.584 + +	0.140 + +
	Proteins	DISOPRED3-bind	1	0.539 + +	0.531 + +	0.102 + +
	Nucleic acids	DisoRDPbind-dna	1	0.532 + +	0.541 + +	0.049 + +
	All ligand types (type agnostic)	random baseline	1	0.506 + +	0.508 + +	0.004 + +
	Nucleic acids	bindEmbed21IDR-idrNuc	1	0.496 + +	0.505 + +	−0.017 + +
	Nucleic acids	bindEmbed21IDR-rawNuc	1	0.495 + +	0.500 + +	−0.019 + +
	Proteins	ProBiPred-protein	1	0.481 + +	0.520 + +	−0.060 + +
	Nucleic acids, metal ions, small molecules	bindEmbed21IDR-idrGeneral	1	0.424 + +	0.471 + +	−0.143 + +
	Nucleic acids, metal ions, small molecules	bindEmbed21IDR-rawGeneral	1	0.416 + +	0.451 + +	−0.147 + +
	Nucleic acids	DisoRDPbind_nuc	1	0.410 + +	0.452 + +	−0.117 + +
	Nucleic acids	DisoRDPbind-rna	1	0.367 + +	0.429 + +	−0.208 + +
Predictors of IDRs		**DisoPred**	0.94	**0.830**- /	0.806- =	**0.494** = /
		Dispredict3	1	0.829- =	**0.816**- /	0.493 = =
		flDPlr2	0.99	0.809 = +	0.801- +	0.443 = +
		flDPnn2	1	0.797 = +	0.797-+	0.434 = +
		SPOT-Disorder2	0.64	0.773 + +	0.805- =	0.415 + +
		SETH−0	1	0.770 + +	0.714 + +	0.402 + +
		AlphaFold-rsa	0.77	0.724 + +	0.674 + +	0.337 + +
		PredIDR-long	1	0.709 + +	0.645 + +	0.321 + +

Lastly, results in Supplementary Table S1, where we did not perform sampling, are consistent with Table 2, showing that the same collections of about 1/3 predictors of binding IDRs and majority of the disorder predictors produce accurate results. The main difference is that AUPRC and MCC values reported in Supplementary Table S1 are smaller, although they sort predictors according to their predictive performance similarly to the sampling-based results in Table 2. This demonstrates that AUPRC and MCC scores are sensitive to the rate of binding residues (positives) and motivated the sampling that we utilized to facilitate side-by-side comparisons of the results that we report across Sections 3.1, 3.2, and 3.3.

Altogether, these results are consistent with the analysis that was completed in CAID2. The interesting result is that a few predictors of the protein binding IDRs performed well and were able to predict all test proteins. The two main issues are the potentially misleading observation that disorder predictors can accurately identify binding IDRs and the inconsistent evaluation of methods that target predictions of IDRs for specific ligand types. The predictors of the nucleic acid binding IDRs are particularly unfairly penalized given that only a small portion of the binding IDRs in this dataset interact with the nucleic acids.

3.2. Evaluation of predictions of binding IDRs in disordered regions

This analysis was motivated by the fact that the test scenario considered in Section 3.1 did not adequately assess the ability of methods to distinguish between binding IDRs and other types of disordered residues. Here, we evaluated predictors on the dataset that includes all binding IDRs and verified non-binding IDRs. The latter are functionally annotated IDRs that exclude ligand-binding regions. Table 3 summarizes results when performing sampling (balanced dataset with equal number of binding IDR residues and non-binding IDR residues), enabling side-by-side comparison with the other considered scenarios.

Table 3 reveals that predictive quality is modest, with the best predictors of binding IDRs securing AUC and AUPRC at about 0.65 and MCC at about 0.23. Based on the AUC and AUPRC metrics, the best-performing binding IDR predictor is DRPBind-protein. However, based on AUC its results are not significantly different from the results of three other methods, DeepDISOBind-all, ENSHROUD-all, and MoRFchibi-web (p-value ≥ 0.05). These four methods secured AUC > 0.65, AUPRC > 0.63 and MCC > 0.23. Compared to Table 2 that evaluates on the disordered and ordered residues, the top 10 binding IDR predictors in

Table 3
Predictive performance in the functionally annotated disordered regions in the CAID2 test dataset. This dataset includes binding IDRs and functionally verified non-binding IDRs, where the latter were randomly sampled to obtain the same number of residues as the number of residues in the binding IDRs. The predictors of the binding IDRs and the disorder predictors are sorted in the descending order of their AUC values. Predictors that secure the highest AUC, AUPRC, and MCC scores are shown in bold font. Statistical significance of differences is annotated using the *xy* symbols after the numerical value, where $x = \{+, -, =\}$ denotes comparison of a given method against the best predictor of binding IDRs, and $y = \{+, -, =\}$ denotes comparison of a given method against the best predictor of disorder, and +, =, and – indicate that the best predictor of binding IDRs (disorder predictor) is significantly better (p-value < 0.05), not significantly different (p-value ≥ 0.05), and significantly worse (p-value < 0.05) than the given method. Evaluation metrics and statistical tests of significance are explained in Section 2.3.

Predictor type	Type of ligand targeted	Predictor name	Coverage	AUC	AUPRC	MCC
Predictors of binding IDRs	Proteins	**DRPBind-protein**	1	**0.665**/ +	**0.672**/ =	0.233 = =
	Proteins, nucleic acids	**DeepDISOBind-all**	1	0.660 = +	0.640 + +	**0.243**/ =
	Proteins, nucleic acids	ENSHROUD-all	1	0.659 = +	0.652 + +	0.235 = =
	Proteins	MoRFchibi-web	1	0.653 = +	0.632 + +	0.232 = =
	Proteins	DeepDISOBind-protein	1	0.652 + +	0.627 + +	0.233 + =
	Proteins	ENSHROUD-protein	1	0.648 + +	0.643 + +	0.196 + +
	Nucleic acids	ENSHROUD-nucleic	1	0.645 + +	0.613 + +	0.224 = =
	Proteins	MoRFchibi-light	1	0.630 + +	0.602 + +	0.204 + +
	Nucleic acids	DeepDISOBind-nucleic	1	0.629 + +	0.592 + +	0.179 + =
	Proteins	DisoRDPbind-prot	1	0.615 + +	0.621 + +	0.135 + +
	Nucleic acids	DRPBind-DNA	1	0.612 + +	0.601 + +	0.148 + +
	Proteins	ANCHOR2	1	0.595 + +	0.557 + +	0.149 + +
	All ligand types (type agnostic)	AlphaFold-binding	0.81	0.591 + +	0.585 + +	0.124 + +
	Nucleic acids	ProBiPred-nucleic	1	0.589 + +	0.589 + +	0.125 + +
	Nucleic acids	DeepDRPBind-RNA	0.99	0.585 + +	0.534 + +	0.135 + +
	Proteins	DeepDRPBind-protein	0.99	0.583 + +	0.566 + +	0.126 + +
	Nucleic acids	DeepDRPBind_nuc	0.99	0.582 + +	0.555 + +	0.102 + +
	Nucleic acids	DRPBind_nuc	1	0.566 + +	0.542 + +	0.129 + +
	Nucleic acids	DRPBind-RNA	1	0.543 + +	0.553 + +	0.049 + +
	Nucleic acids	DeepDRPBind-DNA	0.99	0.541 + +	0.530 + +	0.052 + +
	Proteins	ProBiPred-protein	1	0.538 + +	0.531 + +	0.050 + +
	Nucleic acids	bindEmbed21IDR-idrNuc	1	0.537 + +	0.523 + +	0.066 + +
	Nucleic acids	bindEmbed21IDR-rawNuc	1	0.537 + +	0.522 + +	0.066 + +
	Nucleic acids	DisoRDPbind-dna	1	0.534 + +	0.529 + +	0.054 + +
	Proteins	OPAL	1	0.521 + +	0.545 + +	0.044 + +
	Nucleic acids, metal ions, small molecules	bindEmbed21IDR-idrGeneral	1	0.521 + +	0.509 + +	0.007 + +
	Nucleic acids, metal ions, small molecules	bindEmbed21IDR-rawGeneral	1	0.520 + +	0.504 + +	0.007 + +
	Proteins, nucleic acids	CLIP	1	0.510 + +	0.528 + +	−0.016 + +
	All ligand types (type agnostic)	DisoBindPred	0.99	0.504 + +	0.508 + +	0.008 + +
	All ligand types (type agnostic)	random baseline	1	0.497 + +	0.497 + +	−0.002 + +
	Nucleic acids	DisoRDPbind_nuc	1	0.477 + +	0.488 + +	−0.014 + +
	Proteins	DISOPRED3-bind	1	0.468 + +	0.490 + +	−0.078 + +
	Nucleic acids	DisoRDPbind-rna	1	0.454 + +	0.489 + +	−0.079 + +
Predictors of IDRs		**DisoPred**	0.95	**0.683**- /	**0.689**= /	**0.252**= /
		flDPlr2	0.99	0.590 + +	0.617 + +	0.051 + +
		Dispredict3	1	0.588 + +	0.626 + +	0.081 + +
		SETH−0	1	0.577 + +	0.567 + +	0.093 + +
		flDPnn2	1	0.564 + +	0.593 + +	0.026 + +
		SPOT-Disorder2	0.8	0.555 + +	0.539 + +	0.082 + +
		AlphaFold-rsa	0.81	0.552 + +	0.514 + +	0.087 + +
		PredIDR-long	1	0.525 + +	0.507 + +	0.088 + +

Table 3 showed consistent pattern of decline across the AUC, AUPRC, and MCC metrics. Specifically, their AUC decreased by between 0.13 and 0.18, while the largest declines in AUPRC and MCC reached 0.11 and 0.21, respectively. This can be explained by the fact that identifying binding IDRs among other types of IDRs is much harder (Table 3) when compared to differentiating them from ordered residues (Table 2). However, the former is the main objective for the predictors of binding IDRs since many predictors of disorder are already very capable of accurately discriminating disordered residues from ordered residues [63,64,92,93]. Our evaluation reveals that even if the location of IDRs in a given protein sequence is known (which can be done using accurate disorder predictors), accurately identifying which IDRs bind ligands remains rather challenging.

We also evaluated disorder predictors in the context of predicting binding IDRs in the CAID2 test dataset. In other words, we evaluated whether their disorder predictions can be used to differentiate the binding IDRs (positive labels) from the functionally verified non-binding IDRs (negative labels). As expected, Table 3 shows that the best disorder predictors cannot accurately find binding IDRs among disordered regions. Their AUCs are below 0.60 and MCCs are below 0.10, with the exception of DisoPred. To compare, 13 predictors of binding IDRs produce more accurate predictions, i.e., they secure AUC > 0.59 and MCC > 0.10. This is expected since these methods were not designed to distinguish between different types of disordered regions. The unusual results produced by DisoPred, which AUC is better than the AUCs of the binding IDR predictors, are difficult to explain. This predictor was not published and details of its design were not disclosed. One potential explanation why DisoPred is biased to predict binding IDRs, rather than predicting all IDRs equally, is that it uses predictions generated by DeepDISOBind as inputs. DeepDISOBind secures relatively good predictive quality for the prediction of binding IDRs (Table 3), and perhaps these inputs are partly transferred into the DisoPred's output.

Moreover, Tables 2 and 3 point to a consistent observation that nearly all of the top 10 binding IDR predictors target protein binding IDRs. This stems from the high proportion of the protein binding residues among the binding IDRs in the corresponding datasets (Table 1). On the other hand, predictors of nucleic acid binding IDRs populate the bottom of the list of the most accurate predictors in Tables 2 and 3. This is because they are evaluated on all binding IDRs while in fact they should not predict a large majority of them that do not interact with DNA or RNA. We addressed this limitation in the next section where we performed evaluations separately for IDRs that interact with different

types of ligands.

3.3. Evaluation of predictions of binding IDRs that interact with different types of ligands

We assessed the 40 predictors on three datasets that cover IDRs that interact with proteins, with nucleic acids, and with other types of ligands. Like in Section 3.2, we focus on identifying these three distinct types of binding IDRs from other functionally verified disordered non-binding residues. Table 4 reports the three AUC scores that were computed for the prediction of the protein binding IDRs, nucleic acid binding IDRs, and other ligand binding IDRs. Moreover, it divides the predictors by the type of ligand(s) they target to ease matching them with the corresponding AUC scores. Ideally, a given predictor would perform well for the binding IDRs that match its prediction target (e.g., predictors of protein binding IDRs should secure high AUCs for the evaluation on the protein binding IDRs) and secure low AUCs for the evaluations on the other types of binding IDRs (predictors of protein binding IDRs should not make accurate predictions for the IDRs that bind nucleic acids and other ligands). The high AUCs for the other types of the binding IDRs would mean that this predictor cross-predicts between ligands types. We used underline to identify the highest AUC that is statistically better (p-value < 0.05) than the other two AUCs for a given predictor, to facilitate the abovementioned analysis. Supplementary Table S2 provides the AUPRC, MCC, and coverage values.

We found that the following four predictors of protein binding IDRs

Table 4
Predictive performance measured with AUC for the prediction of protein, nucleic acid, and other ligand binding IDRs in the functionally annotated IDRs in the CAID2 test dataset. We tested on three datasets: 1) protein binding IDRs and functionally verified non-binding IDRs; 2) nucleic acid binding IDRs and functionally verified non-binding IDRs; and 3) other-binding IDRs and functionally verified non-binding IDRs; where the verified non-binding IDRs were randomly sampled to obtain the same number of residues as the number of residues in the corresponding binding IDRs. The predictors are sorted in the descending order of their AUC values for each group of methods and matching ligand type. Underline identifies the highest AUC that is statistically better (p-value < 0.05) than the other two AUCs for a given predictor. Statistical significance of differences is annotated using the {+ , -, = } symbols after the numerical AUC value, where we compare against the result/method shown in bold (best predictor for the matching ligand type); + , = , and – indicate that this best method is significantly better (p-value < 0.05), not significantly different (p-value ≥ 0.05), and significantly worse (p-value < 0.05) than the given method. Evaluation metrics and statistical tests of significance are explained in Section 2.3.

Type of predictor	Predictor name	AUC for predictions of different types of binding IDRs			Statistical significance of differences in AUCs obtained for different ligand types for a given predictor		
		protein binding	nucleic acid-binding	other ligands binding	p-value for protein vs nucleic acid	p-value for protein vs other ligands	p-value for nucleic acid vs other ligands
Protein binding IDRs	**DRPBind-protein**	**0.694**	0.542 +	0.658 -	0.00	0.00	0.00
	ENSHROUD-protein	0.671 +	0.543 +	0.639 -	0.00	0.00	0.00
	DisoRDPbind-prot	0.665 +	0.439 +	0.615 -	0.00	0.00	0.00
	MoRFchibi-web	0.662 +	0.594 +	0.584 =	0.00	0.00	0.30
	DeepDISOBind-protein	0.646 +	0.501 +	0.704 -	0.00	0.00	0.00
	MoRFchibi-light	0.637 +	0.596 +	0.551 =	0.01	0.00	0.01
	ANCHOR2	0.620 +	0.454 +	0.601 -	0.00	0.77	0.00
	DeepDRPBind-protein	0.539 +	0.462 +	0.593 -	0.00	0.00	0.00
	OPAL	0.526 +	0.477 +	0.427 +	0.01	0.00	0.01
	ProBiPred-protein	0.520 +	0.561 +	0.503 +	0.01	0.33	0.00
	DISOPRED3-bind	0.482 +	0.462 +	0.466 +	0.02	0.02	0.71
Nucleic acid binding IDRs	**DRPBind-RNA**	0.616 +	**0.791**	0.524 =	0.00	0.00	0.00
	DRPBind_nuc	0.609 +	0.705 +	0.519 +	0.00	0.00	0.00
	DeepDISOBind-nucleic	0.636 +	0.685 +	0.579 =	0.00	0.00	0.00
	ENSHROUD-nucleic	0.642 +	0.656 +	0.667 -	0.19	0.07	0.30
	DeepDRPBind-DNA	0.556 +	0.653 +	0.487 +	0.00	0.00	0.00
	DeepDRPBind_nuc	0.610 +	0.628 +	0.485 +	0.06	0.00	0.00
	bindEmbed21IDR-idrNuc	0.537 +	0.617 +	0.483 +	0.00	0.00	0.00
	bindEmbed21IDR-rawNuc	0.537 +	0.617 +	0.484 +	0.00	0.00	0.00
	DRPBind-DNA	0.631 +	0.615 +	0.523 +	0.28	0.00	0.00
	ProBiPred-nucleic	0.580 +	0.603 +	0.564 =	0.15	0.29	0.00
	DisoRDPbind-rna	0.540 +	0.572 +	0.294 +	0.00	0.00	0.00
	DeepDRPBind-RNA	0.625 +	0.560 +	0.458 +	0.00	0.00	0.00
	DisoRDPbind_nuc	0.537 +	0.557 +	0.384 +	0.26	0.00	0.00
	DisoRDPbind-dna	0.523 +	0.552 +	0.573 =	0.15	0.05	0.03
Protein and nucleic acid binding IDRs	ENSHROUD-all	0.688 =	0.572 +	0.641 -	0.00	0.00	0.00
	DeepDISOBind-all	0.652 +	0.556 +	0.717 -	0.00	0.00	0.00
	CLIP	0.552 +	0.589 +	0.369 +	0.00	0.00	0.00
Nucleic acid and other ligands (metal ions and small molecules) binding IDRs	bindEmbed21IDR-rawGeneral	0.507 +	0.476 +	0.520 +	0.03	0.24	0.01
	bindEmbed21IDR-idrGeneral	0.508 +	0.475 +	0.518 +	0.02	0.20	0.01
All ligand types (type agnostic) binding IDRs	**AlphaFold-binding**	0.583 +	0.470 +	**0.563**	0.00	0.15	0.00
	DisoBindPred	0.473 +	0.418 +	0.530 +	0.00	0.00	0.00
IDRs	DisoPred	0.695 =	0.526 +	0.725 -	0.00	0.01	0.00
	flDPlr2	0.605 +	0.519 +	0.597 -	0.00	0.47	0.00
	flDPnn2	0.593 +	0.511 +	0.565 =	0.00	0.05	0.00
	Dispredict3	0.581 +	0.507 +	0.614 -	0.00	0.00	0.00
	AlphaFold-rsa	0.581 +	0.394 +	0.538 +	0.04	0.04	0.00
	SETH−0	0.579 +	0.449 +	0.612 -	0.00	0.05	0.00
	SPOT-Disorder2	0.544 +	0.362 +	0.621 -	0.00	0.00	0.00
	PredIDR-long	0.514 +	0.538 +	0.558 =	0.01	0.00	0.02
Random baseline		0.492 +	0.500 +	0.517 +	0.62	0.62	0.77

provide reasonably accurate predictions of the protein binding IDRs: DRPBind-protein, ENSHROUD-protein, DisoRDPbind-prot, and MoRFchibi-web. These methods secured AUCs > 0.65 for the protein binding IDRs, which are statistically higher than their AUCs for the other two ligand types (p-value < 0.05; denoted with underline in Table 4). The most accurate DRPBind-protein method obtained AUC of 0.69 for the protein binding IDRs but also predicted many of the IDRs that interact with the other ligands (non-proteins and non-nucleic acids), given the corresponding AUC of 0.66. We also identified four relatively accurate predictors of the nucleic acid binding IDRs: DRPBind-RNA, DRPBind_nuc, DeepDISOBind-nucleic, and DeepDRPBind-DNA. These four tools similarly obtained AUCs > 0.65 for the nucleic acid binding IDRs with these AUCs being significantly better than their other two AUCs (p-value < 0.05; denoted with underline in Table 4). The ENSHROUD-all method, which predicts both protein and nucleic acids binding IDRs, performed well on the protein binding IDRs (AUC of 0.69) but struggled with the nucleic acid binding IDRs (AUC of 0.57). The results for the predictors of the other ligand binding IDRs and ligand type agnostic predictors are characterized by low levels of accuracy. These four methods secured AUCs < 0.6 across all test sets, including for their target ligands. Moreover, the only methods that secured AUCs > 0.65 for the other binding IDRs are DRPBind-protein, DeepDISOBind-protein, ENSHROUD-nucleic, and DeepDISOBind-all. The latter four methods target prediction of protein and/or nucleic acid binding IDRs, and so these results suggest that they cross predict between ligand types. Finally, we observed that the disorder predictors, with the exception of DisoPred, are not capable of accurately predicting binding IDRs across the three ligand types. This is expected since they were not designed for this purpose. Compared to Table 3, Table 4 provides further details for the unusual DisoPred's results, revealing that it is biased to predict protein and other ligand binding IDRs and it did not predict nucleic acid binding IDRs.

The results in Table 3 where we evaluated predictions in the ligand type agnostic manner (i.e., using all binding IDRs) suggests that the best-performing binding IDR predictors are DRPBind-protein, DeepDISOBind-all, ENSHROUD-all, and MoRFchibi-web. Using Table 4, we found that DRPBind-protein and MoRFchibi-web in fact performed well for their target ligand type (proteins). However, DeepDISOBind-all and ENSHROUD-all predicted nucleic acids binding IDRs poorly (AUCs $<$ 0.58) and incorrectly predicted other ligand binding IDRs (AUC of 0.72 and 0.64, respectively). Results in Table 4 also revealed several accurate predictors of the nucleic acid binding IDRs (DRPBind-RNA, DRPBind_nuc, DeepDISOBind-nucleic, and DeepDRPBind-DNA), which could not be identified using Table 3 due to the bias of the corresponding dataset towards the protein binding IDRs. This demonstrates the value of evaluation in Table 4 that properly matches the scope of the predictors with the underlying test data.

Altogether, our analysis suggests that protein and nucleic acid binding IDRs can be predicted modestly accurately by several current methods. However, none of the evaluated methods provided highly accurate predictions (AUCs are below 0.8). We also noted potential problems with cross predictions, where methods that target a particular ligand type predict IDRs that interact with other ligand types. We explore this further in Section 3.4.

3.4. Evaluation of cross predictions of ligand types for the prediction of binding IDRs

Since virtually all predictors of binding IDRs target interactions with specific ligand types (see Section 2.1), we evaluated whether their predictions are in fact ligand type-specific or whether they also predict interactions with the other ligand types as well. In other words, we investigated whether predictors are ligand type specific or agnostic to the ligand type. To this end, we calculated normalized cross prediction rate, which is defined as the cross-prediction rate (fraction of predicted binding residues for the "wrong" ligand types, i.e., ligand types that are not covered by a given predictor) that we normalized by dividing it by the sensitivity (fraction of predicted residues for the correct ligand type). For instance, to evaluate predictors of protein binding IDRs, we calculated fraction of binding residues that they predict among IDRs that interact with nucleic acids and other (non-protein and non-nucleic acid) ligands, and we divided it by the fraction of binding residues they predict among protein binding IDRs. Lower normalized cross prediction rate (i.e., ratio of the two fractions) implies lower levels of cross predictions, i.e., high predictive performance, with rate of 1 corresponding to a ligand-agnostic (random) predictor. We also calibrated the overall prediction rate across predictors by using the threshold that produces the number of predicted binding residues equal to the actual (native) number of binding residues. This way we were able to directly compare the normalized cross prediction rates between methods. We summarized these results in Fig. 1.

Table 4 suggests that four predictors of protein binding IDRs provided reasonably accurate predictions of the protein binding IDRs: DRPBind-protein, ENSHROUD-protein, DisoRDPbind-prot, and MoRFchibi-web. However, Fig. 1 reveals that the normalized cross prediction rates for two of them, DRPBind-protein and ENSHROUD-protein, are high and equal 0.85 and 0.88, respectively. This problem is also apparent in Table 4 where these two methods secured relatively high AUC ≥ 0.64 for the other ligand binding IDRs. To compare, the normalized cross prediction rates for the other two methods, DisoRDPbind-prot and MoRFchibi-web, are at 0.80 (Fig. 1). Correspondingly, their AUCs for the nucleic acid binding IDRs and the other ligand binding IDRs are below 0.62 (Table 4). In total, five predictors of protein binding IDRs secured the normalized cross prediction rates at about 0.80 (Fig. 1), which suggests that they provided relatively more ligand specific predictions than the other six predictors of protein binding IDRs. The four accurate predictors of the nucleic acid binding IDRs based on Table 4, DRPBind-RNA, DRPBind_nuc, DeepDRPBind-DNA, and DeepDISOBind-nucleic obtained the normalized cross prediction rates at 0.65, 0.78, 0.80, and 0.88, respectively (Fig. 1). The high cross prediction rate for DeepDISOBind-nucleic stem from the high AUC of 0.64 for the prediction of the protein binding IDRs (Table 4). In total, five predictors of nucleic acid binding IDRs produced normalized cross prediction rates < 0.80. By far the lowest normalized cross prediction rate of 0.44 was obtained by CLIP, which predicted 2.3 binding residues for the correct ligand types (proteins and nucleic acids) per 1 binding residue for the wrong ligand type (non-proteins and non-nucleic acids). However, this predictor secured low predictive performance for the protein and nucleic acid binding IDRs that it targets (AUCs of 0.55 and 0.59; Table 4). Lastly, Fig. 1 demonstrates that majority of the predictors obtained high normalized cross prediction rates at over 0.85 (18 methods out of 30) and some at 1 or higher (7 out of 30). These methods should be considered as ligand type agnostic since they predicted similar number of binding residues across IDRs that interact with different ligand types. This observation is in line with studies that investigated cross predictions in the structured (ordered regions) and which similarly showed that many methods are ligand type agnostic [66–68].

To sum up, we identified two predictors of the protein binding IDRs that generate relatively accurate predictions and cross prediction rates below 0.80, DisoRDPbind-prot and MoRFchibi-web. We also found three predictors of the nucleic acids binding IDRs, DRPBind-RNA, DRPBind_nuc, and DeepDRPBind-DNA, that share the same characteristic, with DRPBind-RNA producing relatively low cross prediction rate of 0.65. Moreover, our empirical analysis suggests that majority of the current predictors of binding IDRs generate ligand type agnostic results.

4. Summary and discussion

We assessed 32 predictors of binding IDRs and 8 leading disorder predictors on the CAID2 dataset by considering several practical scenarios. In particular, we evaluated these methods on the full CAID2 dataset, when differentiating binding IDRs from non-binding IDRs, and

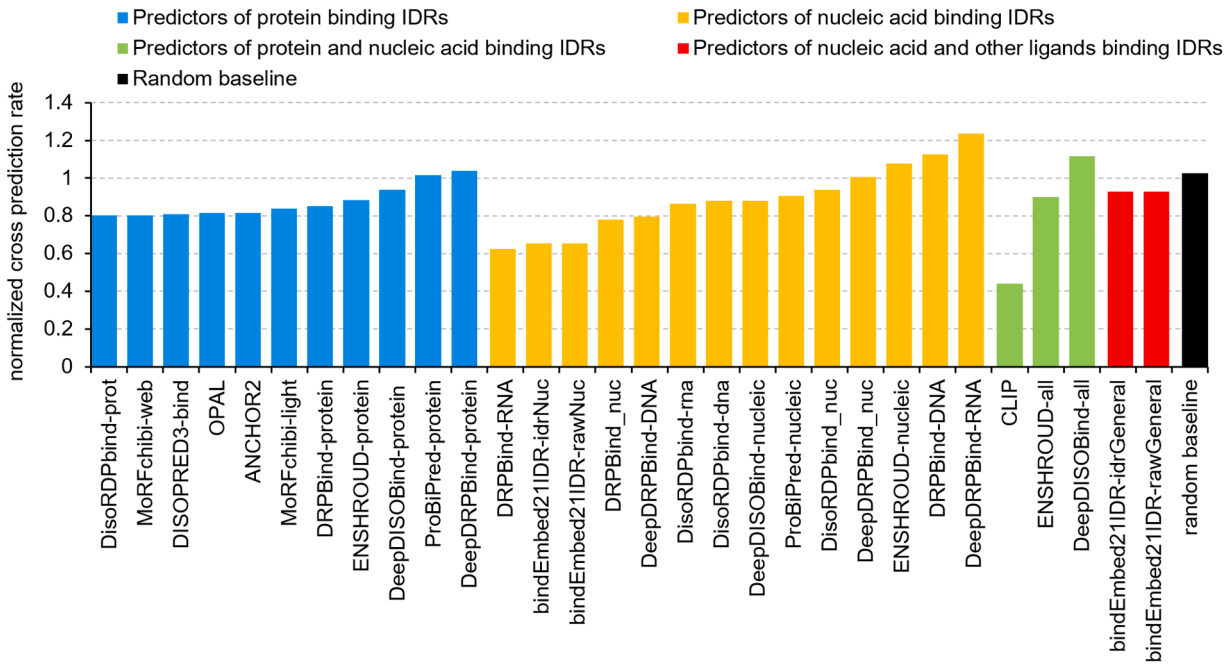

Fig. 1. The normalized cross prediction rates for the ligand specific predictors of binding IDRs using ligand type annotated IDRs in the CAID2 dataset. The predictors are sorted in the ascending order of their rates for each color-coded group of methods. The normalized cross prediction rate for predictors of protein binding IDRs (blue bars) was computed as the cross prediction rate among the nucleic acid and other ligand binding IDRs divided by the prediction rate among protein binding IDRs. The normalized cross prediction rate for predictors of nucleic acid binding IDRs (orange bars) was computed as the cross prediction rate among the protein and other ligand binding IDRs divided by the prediction rate among nucleic acid binding IDRs. The normalized cross prediction rate for predictors of nucleic acid and protein binding IDRs (green bars) was computed as the cross prediction rate among the other ligand binding IDRs divided by the prediction rate among nucleic acid and protein binding IDRs. The normalized cross prediction rate for predictors of nucleic acid and other binding IDRs (red bars) was computed as the cross prediction rate among the protein binding IDRs divided by the prediction rate among nucleic acid and other ligand binding IDRs. The overall prediction rates were calibrated across predictors by using the threshold that produces the number of predicted binding residues equal to the actual (native) number of binding residues. This facilitates directly comparison of the normalized cross prediction rates between predictors.

when making predictions for IDRs that interact with specific ligand type to properly match the scope of the considered predictors. We were also the first to quantify cross predictions between IDRs that bind different ligand types.

As expected, we found that disorder predictors cannot be used to accurately predict binding IDRs, with the exception of an unpublished DisoPred that seems to particularly focus on the prediction of the protein, metal ion, and other small ligand binding IDRs. This should be considered as an unintended behavior since this method participated in CAID2 as a predictor of generic disordered regions [64].

Our empirical analysis suggests that 13 predictors of binding IDRs perform rather poorly with AUCs < 0.60 for their matching type of binding IDRs (Table 4). Moreover, the evaluation of cross predictions reveals that majority of the current predictors of binding IDRs generate ligand type agnostic results (normalized cross prediction rates > 0.85; Fig. 1). On the other end of the spectrum we identified five relatively accurate predictors, including two predictors of the protein binding IDRs (DisoRDPbind-prot and MoRFchibi-web) and three predictors of the nucleic acids binding IDRs (DRPBind-RNA, DRPBind_nuc, and DeepDRPBind-DNA). These methods secured AUCs > 0.65 for the prediction of IDRs with the matching ligand type (Table 4) and the normalized cross prediction rates ≤ 0.80 (Fig. 1). We summarized practical aspects of these methods in Table 5 including their availability, runtime, coverage, and citations. They are available to the end users as web servers and/or standalone code and were able to predicts all proteins in CAID2. The two methods that were published several years ago (DisoRDPbind in 2015 and MoRFchibi in 2016), enjoy robust citations rate, with an annual average of about 18 citations. MoRFchibi was also recently highlighted in a tutorial article on the disorder prediction as one of the best predictors of binding IDRs [39]. Interestingly, these methods differ significantly in their runtime, from the very fast DisoRDPbind (1.2 s per protein), to modestly fast MoRFchibi (2 min and 41 s), to a rather slow DRPBind/DeepDRPBind (17 min and 26 s; 3 orders of magnitude slower than DisoRDPbind).

We also observed that 28 out of the 32 predictors of binding IDRs that participated in CAID2 focused on the protein and/or nucleic acids

Table 5
Key characteristics of the accurate predictors of binding IDRs including availability, runtime, coverage, and citations. Runtime was measured in CAID2 and was quantified in seconds for the prediction of 1000 amino acids long sequence, using the same hardware for all predictors [64]. The citations were collected from Google Scholar in October 2024; the annual citations are defined as the total number of citations divided by the number of years since publication.

Predictor name [reference]	URL (availability type)	Runtime	Coverage in CAID2	Citations	
				Total	Annual
DisoRDPbind [59]	http://biomine.cs.vcu.edu/servers/DisoRDPbind/(web server)	1.2 s	100 %	164	18.2
MoRFchibi-web [65]	https://gsponerlab.msl.ubc.ca/software/morf_chibi/(web server and standalone code)	141.2 s	100 %	146	18.2
DRPBind and DeepDRPBind [69]	https://github.com/roneshsharma/DNA-RNA-Protein-Binding/wiki (standalone code)	1046.8 s	100 %	1	1

binding regions. In contrast, only two methods addressed predictions for the other ligand types, bindEmbed21IDR-rawGeneral and bindEmbed21IDR-idrGeneral that target interactions with metal ions and small molecules. Similarly, there are only three published methods that predict IDRs that bind lipids [54–56]. One reason for this is that the corresponding amount of annotated binding IDRs (ground truth) was relatively low at the time when these methods were developed [94], making it difficult to properly design and test predictors. Since the amount of these ground truth annotations continues to grow [31], we anticipate that more tools that address prediction of interactions with lipids and small ligands will be developed and released. This is needed, particularly in the light of the low predictive quality produced by the bindEmbed21IDR predictors. Moreover, the methods that participated in CAID2 and the corresponding test datasets were geared towards interactions that involve induced folding. The prediction of the binding IDRs that are disordered in the bound state (disorder-to-disorder transition upon binding) may require more specialized test datasets and methods [48], such as FuzzPred [95].

Our results that suggest that many methods struggle to produce accurate results motivate further efforts towards the development of better predictors. We believe that these efforts would benefit from taking advantage of well-designed deep neural networks coupled with protein language models, which should be optimized to accurately identify binding IDRs among disordered regions and which should simultaneously minimize the cross predictions. This suggestion stems from recently published analyses that demonstrate that deep network-based predictors outperform other types of predictive models for the disorder prediction [96–98]. Among the accurate predictors of binding IDRs in Table 5, only DeepDRPBind utilizes a deep network, but this is a relatively dated convolutional network, while MoRFchibi and DisoRDPbind rely on simple Naïve Bayes and logistic regression models. The developers of new predictors should consider modern network topologies and training paradigms, such as transformers [99,100] and contrastive learning [101,102]. They can also utilize a broad selection of protein language models, which include ProtTrans [103], ESM2 [104], ESM-MSA [105], AminoBERT [106] and IDP-BERT [107], to produce high quality inputs for these neural networks. One of the particularly challenging aspects of predictive quality are the cross-predictions between DNA and RNA binding regions, given that DNA and RNA are relatively similar in their physicochemical nature. However, the future predictors could take advantage of the fact that these binding interfaces have a number of unique characteristics, such as their size and abundance of the π-interactions [108–111]. This information could be extracted from sequence-predicted 3D structures, providing a way to boost the ability to differentiate DNA vs RNA binding regions.

Lastly, prediction of the binding IDRs in protein chains should be followed by modeling structures of the corresponding complexes, given that these IDRs typically fold upon binding. The protein structures can be predicted when the native structures are missing [112,113] and numerous docking based that were designed for disordered proteins/regions are available, such as Flex-LZerD [114], IDP-LZerD [115,116], CABS-Dock [117] and AlphaFold-Multimer [118]. The recent release of AlphaFold3 that predicts structures of protein-ligand complexes is also notable [119]. However, authors of AlphaFold3 caution the users that their method generates "*spurious structural order (hallucinations) in disordered regions*" [119], diminishing its value in the context of modeling interactions for binding IDRs.

Funding

FZ was funded in part by the National Natural Science Foundation of China (62402392) and research funding from Northwest A&F University (Z1090124074). LK was funded in part by the National Science Foundation (DBI2146027 and IIS2125218) and the Robert J. Mattauch Endowment funds.

CRediT authorship contribution statement

Fuhao Zhang: Writing – original draft, Visualization, Validation, Investigation, Funding acquisition, Formal analysis, Data curation. **Lukasz Kurgan:** Writing – original draft, Visualization, Validation, Project administration, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of Competing Interest

Authors declare no conflict of interests.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2024.12.009.

References

[1] Oldfield C.J., , Introduction to intrinsically disordered proteins and regions, in N. Salvi, (Ed), Intrinsically Disordered Proteins, 2019, Academic Press. p. 1-34.
[2] Lieutaud P, et al. How disordered is my protein and what is its disorder for? A guide through the "dark side" of the protein universe. Intrinsically Disorded Proteins 2016;4(1):e1259708.
[3] Habchi J, et al. Introducing protein intrinsic disorder. Chem Rev 2014;114(13):6561–88.
[4] Dunker AK, et al. Function and structure of inherently disordered proteins. Curr Opin Struct Biol 2008;18(6):756–64.
[5] Xie HB, et al. Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. J Proteome Res 2007;6(5):1882–98.
[6] Berlow RB, Dyson HJ, Wright PE. Functional advantages of dynamic protein disorder. FEBS Lett 2015;589(19 Pt A):2433–40.
[7] Wright PE, Dyson HJ. Intrinsically disordered proteins in cellular signalling and regulation. Nat Rev Mol Cell Biol 2015;16(1):18–29.
[8] Zhou J, Zhao S, Dunker AK. Intrinsically disordered proteins link alternative splicing and post-translational modifications to complex cell signaling and regulation. J Mol Biol 2018;430(16):2342–59.
[9] Hahn S. Phase separation, protein disorder, and enhancer function. Cell 2018;175(7):1723–5.
[10] Xue B, Dunker AK, Uversky VN. Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. J Biomol Struct Dyn 2012;30(2):137–49.
[11] Peng Z, et al. Exceptionally abundant exceptions: comprehensive characterization of intrinsic disorder in all domains of life. Cell Mol Life Sci 2015;72(1):137–51.
[12] Trivedi R, Nagarajaram HA. Intrinsically disordered proteins: an overview. Int J Mol Sci 2022;23(22).
[13] Basile W, et al. Why do eukaryotic proteins contain more intrinsically disordered regions? PLoS Comput Biol 2019;15(7):e1007186.
[14] Zhao B, et al. IDPology of the living cell: intrinsic disorder in the subcellular compartments of the human cell. Cell Mol Life Sci 2020.
[15] Wang C, Uversky VN, Kurgan L. Disordered nucleiome: Abundance of intrinsic disorder in the DNA- and RNA-binding proteins in 1121 species from Eukaryota, Bacteria and Archaea. Proteomics 2016;16(10):1486–98.
[16] Zhao B, et al. Intrinsic disorder in human RNA-binding proteins. J Mol Biol 2021;433(21):167229.
[17] Peng Z, et al. More than just tails: intrinsic disorder in histone proteins. Mol Biosyst 2012;8(7):1886–901.
[18] Wu Z, et al. In various protein complexes, disordered protomers have large per-residue surface areas and area of protein-, DNA- and RNA-binding interfaces. FEBS Lett 2015;589(19 Pt A):2561–9.
[19] Dyson HJ. Roles of intrinsic disorder in protein-nucleic acid interactions. Mol Biosyst 2012;8(1):97–104.
[20] Meng F, et al. Compartmentalization and functionality of nuclear disorder: intrinsic disorder and protein-protein interactions in intra-nuclear compartments. Int J Mol Sci 2015;17(1).
[21] Varadi M, et al. Functional advantages of conserved intrinsic disorder in RNA-binding proteins. PLoS One 2015;10(10):e0139731.
[22] Peng Z, et al. A creature with a hundred waggly tails: intrinsically disordered proteins in the ribosome. Cell Mol Life Sci 2014;71(8):1477–504.
[23] Hu G, et al. Functional analysis of human hub proteins and their interactors involved in the intrinsic disorder-enriched interactions. Int J Mol Sci 2017;18(12).
[24] Jamecna D, Antonny B. Intrinsically disordered protein regions at membrane contact sites. Biochim Biophys Acta Mol Cell Biol Lipids 2021;1866(11):159020.
[25] Uversky VN. Intrinsically disordered proteins and novel strategies for drug discovery. Expert Opin Drug Discov 2012;7(6):475–88.
[26] Kjaergaard M, Kragelund BB. Functions of intrinsic disorder in transmembrane proteins. Cell Mol Life Sci 2017;74(17):3205–24.

[27] Patil A, Kinoshita K, Nakamura H. Domain distribution and intrinsic disorder in hubs in the human protein-protein interaction network. Protein Sci 2010;19(8):1461–8.
[28] Hu G, et al. Functional analysis of human hub proteins and their interactors involved in the intrinsic disorder-enriched interactions. Int J Mol Sci 2017;18(12).
[29] Oldfield CJ, et al. Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. BMC Genom 2008;9(1):S1.
[30] Hsu WL, et al. Exploring the binding diversity of intrinsically disordered proteins involved in one-to-many binding. Protein Sci 2013;22(3):258–73.
[31] Aspromonte MC, et al. DisProt in 2024: improving function annotation of intrinsically disordered proteins. Nucleic Acids Res 2024;52(D1):D434–41.
[32] Fukuchi S, et al. IDEAL in 2014 illustrates interaction networks composed of intrinsically disordered proteins and their binding partners. Nucleic Acids Res 2014;42(Database):D320–5.
[33] Piovesan D, et al. MobiDB: 10 years of intrinsically disordered proteins. Nucleic Acids Res 2023;51(D1):D438–44.
[34] Katuwawala A, Ghadermarzi S, Kurgan L. Computational prediction of functions of intrinsically disordered regions. Prog Mol Biol Transl Sci 2019;166:341–69.
[35] Meng F, Uversky VN, Kurgan L. Comprehensive review of methods for prediction of intrinsic disorder and its molecular functions. Cell Mol Life Sci 2017;74(17):3069–90.
[36] Varadi M, et al. Computational approaches for inferring the functions of intrinsically disordered proteins. Front Mol Biosci 2015;2:45.
[37] Katuwawala A, et al. Computational prediction of MoRFs, short disorder-to-order transitioning protein binding regions. Comput Struct Biotechnol J 2019;17:454–62.
[38] Barik A, Kurgan L. A comprehensive overview of sequence-based protein-binding residue predictions for structured and disordered regions. Protein Interact 2020:33–58.
[39] Kurgan L, et al. Tutorial: a guide for the selection of fast and accurate computational tools for the prediction of intrinsic disorder in proteins. Nat Protoc 2023;18(11):3157–72.
[40] Basu S, Kihara D, Kurgan L. Computational prediction of disordered binding regions. Comput Struct Biotechnol J 2023;21:1487–97.
[41] Tamburrini KC, et al. Predicting protein conformational disorder and disordered binding sites. Methods Mol Biol 2022;2449:95–147.
[42] Del Conte A, et al. CAID prediction portal: a comprehensive service for predicting intrinsic disorder and binding regions in proteins. Nucleic Acids Res 2023;51(W1):W62–9.
[43] Yan J, et al. Molecular recognition features (MoRFs) in three domains of life. Mol Biosyst 2016;12(3):697–710.
[44] Vacic V, et al. Characterization of molecular recognition features, MoRFs, and their binding partners. J Proteome Res 2007;6(6):2351–66.
[45] Mohan A, et al. Analysis of molecular recognition features (MoRFs). J Mol Biol 2006;362(5):1043–59.
[46] Wang D, et al. The importance of the compact disordered state in the fuzzy interactions between intrinsically disordered proteins. Chem Sci 2022;13(8):2363–77.
[47] Roterman I, Stapor K, Konieczny L. Engagement of intrinsic disordered proteins in protein-protein interaction. Front Mol Biosci 2023;10:1230922.
[48] Miskei M, et al. Sequence-based prediction of fuzzy protein interactions. J Mol Biol 2020;432(7):2289–303.
[49] Hatos A, et al. FuzDB: a new phase in understanding fuzzy interactions. Nucleic Acids Res 2022;50(D1):D509–17.
[50] Oldfield CJ, et al. Coupled folding and binding with α-Helix-forming molecular recognition elements. Biochemistry 2005;44(37):12454–70.
[51] Dosztanyi Z, Meszaros B, Simon I. ANCHOR: web server for predicting protein binding regions in disordered proteins. Bioinformatics 2009;25(20):2745–6.
[52] Meszaros B, Erdos G, Dosztanyi Z. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. Nucleic Acids Res 2018;46(W1):W329–37.
[53] Wong ETC, Gsponer J. Predicting protein-protein interfaces that bind intrinsically disordered protein regions. J Mol Biol 2019;431(17):3157–78.
[54] Katuwawala A, Zhao B, Kurgan L. DisoLipPred: accurate prediction of disordered lipid-binding residues in protein sequences with deep recurrent networks and transfer learning. Bioinformatics 2021;38(1):115–24.
[55] Basu S, Hegedus T, Kurgan L. CoMemMoRFPred: sequence-based prediction of MemMoRFs by combining predictors of intrinsic disorder, MoRFs and disordered lipid-binding regions. J Mol Biol 2023;435(21):168272.
[56] Dobson L, Tusnady GE. MemDis: predicting disordered regions in transmembrane proteins. Int J Mol Sci 2021;22(22).
[57] Peng Z, et al. CLIP: accurate prediction of disordered linear interacting peptides from protein sequences using co-evolutionary information. Brief Bioinf 2023;24(1).
[58] Monzon AM, et al. FLIPPER: predicting and characterizing linear interacting peptides in the protein data bank. J Mol Biol 2021;433(9):166900.
[59] Peng Z, Kurgan L. High-throughput prediction of RNA, DNA and protein binding regions mediated by intrinsic disorder. Nucleic Acids Res 2015;43(18):e121.
[60] Peng Z, et al. Prediction of disordered RNA, DNA, and protein binding regions using DisoRDPbind. Methods Mol Biol 2017;1484:187–203.
[61] Zhang F, et al. DeepDISOBind: accurate prediction of RNA-, DNA- and protein-binding intrinsically disordered residues with deep multi-task learning. Brief Bioinf 2022;23(1).
[62] Pang Y, Liu B. DisoFLAG: accurate prediction of protein intrinsic disorder and its functions using graph-based interaction protein language model. BMC Biol 2024;22(1):3.
[63] Necci M, et al. Critical assessment of protein intrinsic disorder prediction. Nat Methods 2021;18(5):472–81.
[64] Conte AD, et al. Critical assessment of protein intrinsic disorder prediction (CAID) - Results of round 2. Proteins 2023.
[65] Malhis N, Jacobson M, Gsponer J. MoRFchibi SYSTEM: software tools for the identification of MoRFs in protein sequences. Nucleic Acids Res 2016.
[66] Zhang J, Kurgan L. Review and comparative assessment of sequence-based predictors of protein-binding residues. Brief Bioinf 2018;19(5):821–37.
[67] Zhang J, et al. DNAgenie: accurate prediction of DNA-type-specific binding residues in protein sequences. Brief Bioinf 2021;22(6).
[68] Su H, et al. Improving the prediction of protein-nucleic acids binding residues via multiple sequence profiles and the consensus of complementary methods. Bioinformatics 2019;35(6):930–6.
[69] Sharma R, Tsunoda T, Sharma A. DRPBind: prediction of DNA, RNA and protein binding residues in intrinsically disordered protein sequences. bioRxiv 2023:2023. 03.20.533427.
[70] Sharma R, et al. OPAL: prediction of MoRF regions in intrinsically disordered protein sequences. Bioinformatics 2018;34(11):1850–8.
[71] Sharma R, et al. OPAL+ : length-specific MoRF prediction in intrinsically disordered protein sequences. Proteomics 2019;19(6):e1800058.
[72] Jones DT, Cozzetto D. DISOPRED3: precise disordered region predictions with annotated protein-binding activity. Bioinformatics 2015;31(6):857–63.
[73] Littmann M, et al. Protein embeddings and deep learning predict binding residues for various ligand classes. Sci Rep 2021;11(1).
[74] Oldfield CJ, Peng Z, Kurgan L. Disordered RNA-binding region prediction with DisoRDPbind. Methods Mol Biol 2020;2106:225–39.
[75] Wang K, et al. flDPnn2: accurate and fast predictor of intrinsic disorder in proteins. J Mol Biol 2024:168605.
[76] Ul Kabir MW, Hoque MT. DisPredict3.0: prediction of intrinsically disordered regions/ proteins using protein language model. Appl Math Comput 2024;472.
[77] Hanson J, et al. SPOT-disorder2: improved protein intrinsic disorder prediction by ensembled deep learning. Genom Proteom Bioinf 2019;17(6):645–56.
[78] Piovesan D, Monzon AM, Tosatto SCE. Intrinsic protein disorder and conditional folding in AlphaFoldDB. Protein Sci 2022;31(11):e4466.
[79] Ilzhofer D, Heinzinger M, Rost B. SETH predicts nuances of residue disorder from protein embeddings. Front Bioinf 2022;2:1019597.
[80] Han K-S, et al. PredIDR: accurate prediction of protein intrinsic disorder regions using deep convolutional neural network. bioRxiv 2024:2024. 07.24.604908.
[81] Quaglia F, et al. Exploring manually curated annotations of intrinsically disordered proteins with DisProt. Curr Protoc 2022;2(7):e484.
[82] Jones P, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics 2014;30(9):1236–40.
[83] The Gene Ontology C. The gene ontology resource: 20 years and still going strong. Nucleic Acids Res 2019;47(D1):D330–8.
[84] Zhang F, et al. DeepPRObind: modular deep learner that accurately predicts structure and disorder-annotated protein binding residues. J Mol Biol 2023:167945.
[85] Zhao B, Ghadermarzi S, Kurgan L. Comparative evaluation of AlphaFold2 and disorder predictors for prediction of intrinsic disorder, disorder content and fully disordered proteins. Comput Struct Biotechnol J 2023;21:3248–58.
[86] Wang K, et al. Assessment of disordered linker predictions in the CAID2 experiment. Biomolecules 2024;14(3).
[87] Necci M, et al. A comprehensive assessment of long intrinsic protein disorder from the DisProt database. Bioinformatics 2018;34(3):445–52.
[88] Monastyrskyy B, et al. Assessment of protein disorder region predictions in CASP10. Proteins 2014;82(2):127–37.
[89] Varadi M, et al. AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res 2022;50(D1):D439–44.
[90] Chakravarty D, Porter LL. AlphaFold2 fails to predict protein fold switching. Protein Sci 2022;31(6):e4353.
[91] Wilson CJ, Choy WY, Karttunen M. AlphaFold2: a role for disordered protein/region prediction? Int J Mol Sci 2022;23(9):4591.
[92] Basu S, Kurgan L. Taxonomy-specific assessment of intrinsic disorder predictions at residue and region levels in higher eukaryotes, protists, archaea, bacteria and viruses. Comput Struct Biotechnol J 2024;23:1968–77.
[93] Zhao B, Kurgan L. Deep learning in prediction of intrinsic disorder in proteins. Comput Struct Biotechnol J 2022;20:1286–94.
[94] Hatos A, et al. DisProt: intrinsic protein disorder annotation in 2020. Nucleic Acids Res 2020;48(D1):D269–76.
[95] Hatos A, et al. FuzPred: a web server for the sequence-based prediction of the context-dependent binding modes of proteins. Nucleic Acids Res 2023;51(W1):W198–206.
[96] Zhao B, Kurgan L. Deep learning in prediction of intrinsic disorder in proteins. Comput Struct Biotechnol J 2022;20:1286–94.
[97] Zhao B, Kurgan L. Machine learning for intrinsic disorder prediction. Mach Learn Bioinforma Protein Seq 2023:205–36.
[98] Zhao B, Kurgan L. Surveying over 100 predictors of intrinsic disorder in proteins. Expert Rev Proteom 2021;18(12):1019–29.
[99] Vaswani A, et al. Attention is all you need. 30 (Nips 2017) Adv Neural Inf Process Syst 2017;30.
[100] Tay Y., et al. Efficient Transformers: A Survey. arXiv:2009.06732. 2020.

[101] Gu Z, et al. Hierarchical graph transformer with contrastive learning for protein function prediction. Bioinformatics 2023;39(7).
[102] Lu AX, et al. Discovering molecular features of intrinsically disordered regions by using evolution for contrastive learning. PLoS Comput Biol 2022;18(6):e1010238.
[103] Elnaggar A, et al. ProtTrans: toward understanding the language of life through self-supervised learning. IEEE Trans Pattern Anal Mach Intell 2022;44(10):7112–27.
[104] Lin ZM, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. Science 2023;379(6637):1123–30.
[105] Rao R.M.Rao, R.M., et al., MSA transformer, in Proc. of the 38th Int Conf on Mach Learn, in (Eds) M. Marina and Z. Tong, PMLR: Proceedings of Machine Learning Research; 2021 pp. 8844–8856.
[106] Chowdhury R, et al. Single-sequence protein structure prediction using a language model and deep learning. Nat Biotechnol 2022;40(11):1617–23.
[107] Pang YH, Liu B. IDP-LM: Prediction of protein intrinsic disorder and disorder functions based on language models. Plos Comput Biol 2023;19(11).
[108] Kulandaisamy A, et al. Dissecting and analyzing key residues in protein-DNA complexes. J Mol Recognit 2018;31(4).
[109] Barik A, et al. Molecular architecture of protein-RNA recognition sites. J Biomol Struct Dyn 2015;33(12):2738–51.
[110] Wilson KA, Holland DJ, Wetmore SD. Topology of RNA-protein nucleobase-amino acid π-π interactions and comparison to analogous DNA-protein π-π contacts. RNA 2016;22(5):696–708.
[111] Wilson KA, et al. Anatomy of noncovalent interactions between the nucleobases or ribose and π-containing amino acids in RNA-protein complexes. Nucleic Acids Res 2021;49(4):2213–25.
[112] Jumper J, et al. Highly accurate protein structure prediction with AlphaFold. Nature 2021;596(7873):583–9.
[113] Tunyasuvunakool K, et al. Highly accurate protein structure prediction for the human proteome. Nature 2021;596(7873):590–6.
[114] Christoffer C, Kihara D. Modeling protein-nucleic acid complexes with extremely large conformational changes using Flex-LZerD. Proteomics 2023;23(17):e2200322.
[115] Peterson LX, et al. Modeling disordered protein interactions from biophysical principles. PLoS Comput Biol 2017;13(4):e1005485.
[116] Christoffer C, Kihara D. IDP-LZerD: software for modeling disordered protein interactions. Methods Mol Biol 2020;2165:231–44.
[117] Kurcinski M, et al. CABS-dock standalone: a toolbox for flexible protein-peptide docking. Bioinformatics 2019;35(20):4170–2.
[118] Bryant P, Noe F. Improved protein complex prediction with AlphaFold-multimer by denoising the MSA profile. PLoS Comput Biol 2024;20(7):e1012253.
[119] Abramson J, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature 2024;630(8016).