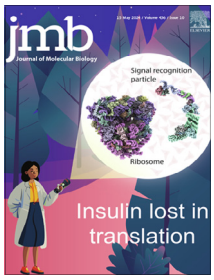

fIDPnn2: Accurate and Fast Predictor of Intrinsic Disorder in Proteins

Kui Wang [1], Gang Hu [1], Sushmita Basu [2] and Lukasz Kurgan [2,*]

*1 - **NITFID,** School of Statistics and Data Science, LPMC and KLMDASR, Nankai University, Tianjin, China*
*2 - **Department of Computer Science,** Virginia Commonwealth University, Richmond, VA, USA*

***Correspondence to Lukasz Kurgan:** lkurgan@vcu.edu (L. Kurgan)*
https://doi.org/10.1016/j.jmb.2024.168605
Edited by Michael Sternberg

Abstract

Prediction of the intrinsic disorder in protein sequences is an active research area, with well over 100 predictors that were released to date. These efforts are motivated by the functional importance and high levels of abundance of intrinsic disorder, combined with relatively low amounts of experimental annotations. The disorder predictors are periodically evaluated by independent assessors in the Critical Assessment of protein Intrinsic Disorder prediction (CAID) experiments. The recently completed CAID2 experiment assessed close to 40 state-of-the-art methods demonstrating that some of them produce accurate results. In particular, fIDPnn2 method, which is the successor of fIDPnn that performed well in the CAID1 experiment, secured the overall most accurate results on the Disorder-NOX dataset in CAID2. fIDPnn2 implements a number of improvements when compared to its predecessor including changes to the inputs, increased size of the deep network model that we retrained on a larger training set, and addition of an alignment module. Using results from CAID2, we show that fIDPnn2 produces accurate predictions very quickly, modestly improving over the accuracy of fIDPnn and reducing the runtime by half, to about 27 s per protein. fIDPnn2 is freely available as a convenient web server at http://biomine.cs.vcu.edu/servers/fIDPnn2/.

Introduction

Bioinformatics studies suggest that proteins with intrinsically disorder regions (IDRs)[1–3] are common in nature, especially in eukaryotes where they account for over 30% of proteins[4–6]. These proteins are implicated in pathogenesis of several human diseases,[7–10] which explains growing interest in utilizing them as drug targets[11–15]. To date, only a few thousand IDRs were characterized experimentally[16], prompting the need to develop computational methods that accurately predict intrinsic disorder in the broadly available protein sequences[17–23]. Predictive quality of the intrinsic disorder predictors was evaluated in several comparative surveys[24–28] and community-driven assessments that include CASP (Critical Assessment of protein Structure Prediction), between CASP5[29] and CASP10[30], and more recently CAID (Critical Assessment of protein Intrinsic Disorder).[31–32] The CASP and CAID assessments were done by independent assessors (who exclude authors of the assessed predictors) on relatively large and blind test datasets (the authors had no access to these datasets before the assessment), applying community-accepted metrics, and in the case of CAID using predictors that were deposited to the assessors. This arguably makes these assessments more reliable and objective when compared to the other evaluations that are typically done by authors of tools on already available test datasets.

Please cite tts article as: K. Wang, G. Hu, S. Basu, et al., fIDPnn2: Accurate and Fast Predictor of Intrinsic Disorder in ProteinsJournal of Molecular Biology, https://doi.org/10.1016/j.jmb.2024.168605

The first CAID has shown that several methods, such as flDPnn[33], rawMSA[34], ESpritz[35], DisoMine[36], and SPOT-Disorder2,[37] generate accurate disorder predictions with AUC (area under the ROC curve) $\geq$ 0.76 when evaluated on the DisProt dataset.[16,38] Our flDPnn predictor secured the highest AUC of 0.81 and a commentary on the CAID assessment noted that "*flDPnn is at least an order of magnitude faster than its competitors, and it succeeded on all sequences*".[39] Given this success, we developed a new version, flDPnn2, which implements multiple improvements to the original method that make it modestly more accurate and substantially faster. These improvements include changes to the input profile and features, slightly increased size of the deep network model, retraining on larger training and validation datasets, and addition of an alignment module. We submitted flDPnn2 along with flDPnn to the second CAID experiment, where flDPnn2 secured favourable results on the DisProt-NOX dataset.[31] The results in CAID2[31] suggest that flDPnn2 is a state-of-the-art predictor that supersedes flDPnn and other available methods, offering very accurate and fast predictions of the intrinsic disorder.

Materials and Methods

Datasets

The dataset collection for flDPnn2 follows the process that we used for flDPnn (see Supplement for details) but using a much larger and publicly available collection of the disorder-annotated proteins from the version 9.1 of DisProt that we downloaded in March 2022[16]. We randomly divided these 2,227 sequences into the training dataset (1,527 sequences), validation dataset (200 sequences), and test dataset (500 sequences). We designed the new predictive model using the training and validation datasets and we compared alternative designs of flDPnn2 on the test dataset. The ultimate comparative assessment against the state-of-the-art relies on the independent evaluation done by the assessors in the CAID2 experiment, which was done after we completed the design and validation of our predictor. We rely on the results on the DisProt-NOX dataset from CAID2,[31] which is composed of 210 protein sequences that have 19.5% of intrinsically disordered residues. This dataset, together with the corresponding predictions from the tools that participated in the CAID2 experiment, are available at https://caid.idpcentral.org/challenge#Data.

Evaluation

Disorder predictors produce two outputs for each amino acid in a given protein sequence: a real-valued propensity for the intrinsic disorder and a binary state (disordered vs. structured). The states are typically generated from the propensities where amino acids with propensities greater than a given threshold predicted as disordered, and otherwise they are assumed to be structured. For each considered predictor, we set the threshold to the value that produces predictions which match the native disorder content in the DisProt-NOX dataset. This calibrates the binary predictions across methods allowing us to directly compare their results. Inspired by the CAID2 assessment,[31] we evaluate the binary predictions with F1 and MCC (Matthews correlation coefficient) and we use AUC (area under the receiver-operating characteristic (ROC) curve) and AUPRC (area under the precision-recall curve) to assess the propensities. Since the disordered residues are in minority (19.5% of amino acids), we follow related studies[40–45] and compute lowAUCratio and lowPRCratio. These two ratios are defined as the AUC and AUPRC values for the parts of the curves where the amount of the predicted disorder is equal or lower than the actual amount of disorder (i.e., disorder is not over-predicted) divided by the corresponding AUC and AUPRC of a random predictor, respectively. This way, these ratios quantify the rate of improvement over the random result. We provide a more detailed discussion of these metrics in the Supplement.

Moreover, we assess statistical significance of differences between predictors, focusing on comparing other methods against flDPnn2. We evaluate whether differences are consistent over a range of different test sets by performing 20 random selections of 50% of proteins from the DisProt-NOX dataset. We evaluate significance of differences over these 20 paired results using the student's *t*-test if the measurements are normal, as evaluated with the Anderson-Darling test at 0.05 significance[46]; otherwise we use the Wilcoxon test. This procedure is consistent with recent related works.[40–42,44,47–48] We assume that the difference is statistically significant if the resulting *p*-value < 0.01.

The flDPnn2 model

Similar to flDPnn, the flDPnn2 model produces disorder predictions in three steps (Figure 1): 1) compute a sequence profile from the input sequence; 2) encode features from the profile; and 3) process the features via a deep neural network and an alignment module to generate putative propensities of disorder.

The sequence profile extends the input sequence with several sequence-derived/predicted structural and functional characteristics that are relevant to the disorder prediction. We extract this information from the input chain utilizing popular and fast bioinformatics tools that include (Figure 1): positions specific scoring matrix generated with PSI-BLAST[49] from the Swiss-Prot database (i.e., we use a small database to ensure that this calculation is fast), which we use to generate entropy-

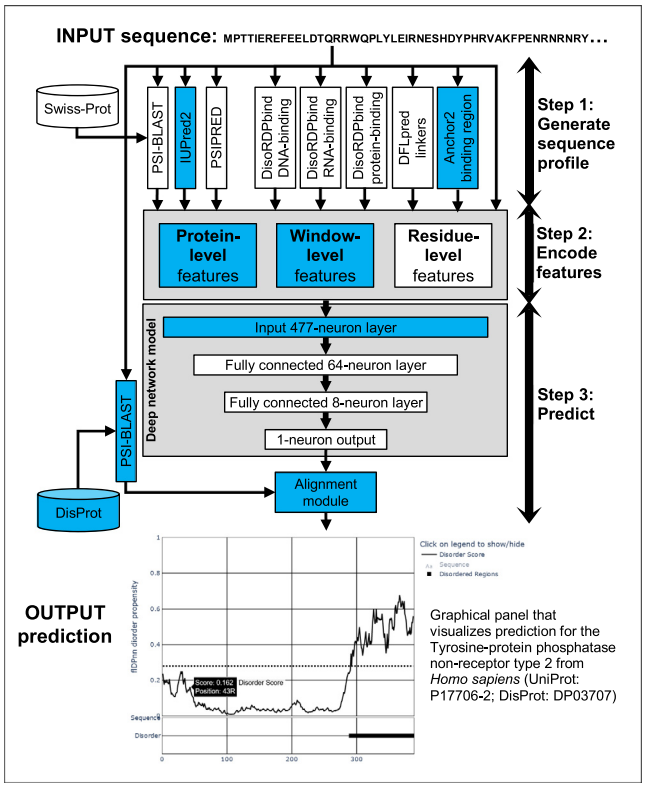

Figure 1. Architecture of the flDPnn2 model. The blue elements denote major changes when compared to the flDPnn model.

based conservation scores,[50] initial disorder prediction generated by IUPred2,[51] which we refine and improve using our model; secondary structure predicted with the single-sequence version of PSIPRED[52]; disordered DNA binding, RNA binding and protein binding propensities predicted with DisoRDPbind[53–54] and ANCHOR2,[51] and disordered linkers predicted by DFLpred.[55] The major improvements when compared to flDPnn are the removal of now outdated fMoRFpred,[56] replacement of IUPred[57] with newer and more accurate IUPred2,[51] and addition of a fast and accurate ANCHOR2.[21,51]

In the second step, we use the profile to compute three feature sets: residue-level, window-level and protein-level. We re-use the residue-level features from flDPnn and introduce additional window-level and protein-level features. We calculate the residue-level and window-level features with the popular sliding window approach, where the amino acid in the middle of a small sequence segment (window) is predicted based on the information from all amino acids in that segment. The residue-level features correspond to the profile values for individual residues in a small window of size 5. The window-level features correspond to the average over two window sizes, 5 and 21 residues. The length of the longer window doubles the size of the shortest disordered regions that are available in the DisProt database (length of 10)[16,38] plus one that is for the middle residue. Similarly, window of size 5 is motivated by the minimal disordered region size that was used in CASP assessments (length of 4)[30] plus one that is for the middle residue. Moreover, we compute features that contrast the average-based values from the sub-segments of 11 residues in the middle of the 21-residues long window against the two flanking segments of 5 residues on both sides of the middle sub-segment. This approach aims to detect changes among nearby residues, such as transition from a putative helix to a putative coil segment or from a putative non-binding to a binding region, and was used in the past to improve predictions of related characteristics of intrinsic disorder.[58–60] The use of the longer, 21-residues long window and the flanking regions are new additions to the flDPnn2 model. We utilize the protein-level features to quantify the overall bias of a given protein, such as propensity to be fully disordered or having a large number of binding regions. Like in flDPnn, the protein-level features include the sequence-average of the profile values, sequence length, and the distance of the residue being predicted to each sequence terminus. We also add the minimum of the two distances to the terminus (to detect proximity to either terminus) and average of the 30% of the highest and the 30% of the lowest profile scores. The latter two

averages attempt to detect large structural and functional regions, such as long helices or long binding segments.

We input the three feature sets into a deep feed-forward neural network. We used the same network type in flDPnn, however, in flDPnn2 we increase the size of the first layer by 50%, from 318 nodes to 477, driven by the inclusion of the additional window- and protein-level features. Moreover, proper training of the larger network is possible due to the substantially larger size of the training dataset. The input layer is followed by a dropout layer with 0.2 dropout rate, hidden layer with 64 nodes, dropout layer with 0.2 dropout rate, second hidden layer with 8 nodes, and the output layer with one node that generates the propensities for disorder. The dropout layers are meant to reduce overfitting into the training dataset.[61] Similar to flDPnn, we apply the ReLU activation functions for all neurons except for the output neuron that relies on the sigmoid function to properly scale the output propensities.

Finally, we add an alignment module that combines the outputs produced by the deep network and an alignment-based prediction using PSI-BLAST, with an objective to improve performance when predicting sequences that are similar to the proteins with experimentally annotated disorder in an alignment dataset. We investigate the most similar protein from the alignment dataset and if it is sufficiently similar (i.e., sequence identity >60%) then we use the resulting alignment to augment the predictions from the network; otherwise we use the network prediction. We consider four ways to combine predictions generated by the network with the alignment. We explain and compare these four options with the results obtained without the alignment module in the Supplement. This experiment performs predictions on the test dataset when using the training proteins as the alignment dataset. We find that the use of the alignment provides modest improvements in predictive performance for each of the four variants of the alignment module, with the best variant producing AUC = 0.811 when compared to AUC = 0.805 when the alignment module is excluded (Suppl. Figure S1). The modest magnitude can be explained by the fact that the DisProt proteins that we use to derive training and test datasets share low levels of sequence similarity. The corresponding best variant averages the network prediction with the score of 1 for residues that are aligned and annotated as disordered, and otherwise it uses the propensity generated by the network. The flDPnn2 implements the alignment module using this design and the alignment dataset that consists of the DisProt database from March 2022. This alignment dataset is small and thus the computational cost of the inclusion of the alignment module is minimal.

We also use the test dataset to quantify impact of the main improvements implemented in flDPnn2 on the predictive performance (Suppl. Figure S1). We find that replacement of IUPred with IUPred2 and the inclusion of ANCHOR2 lead to an increase in AUC from 0.801 to 0.811, while the removal of fMoRFpred does not reduce predictive performance while saving a substantial amount of runtime. A more detailed discussion of these results is included in the Supplement. Altogether, flDPnn2 extends the original flDPnn model by improving the sequence profile and features, correspondingly increasing the size of the deep neural network, adding the alignment module, retraining the network on a much larger training dataset with 1,527 proteins, and streamlining implementation of this design. These improvements reduce the runtime and modestly increase the predictive performance.

Results

Comparative assessment of predictive performance

We compare flDPnn2 against the state-of-the-art disorder predictors using results from the CAID2 experiment on the DisProt-NOX dataset.[31] This assessment was performed independently by the CAID2 assessors using predictors that were deposited by the corresponding authors. The results in Table 1 show that flDPnn2 secures the highest values of AUPRC = 0.596, lowPRCratio = 4.1, lowAUCratio = 6.7, F1 = 0.526 and MCC = 0.414, and the second highest AUC = 0.838. These values suggest that the flDPnn2′s predictions are accurate, with lowPRCratio and lowAUCratio revealing that they are 4.1 and 6.7 times better than a random prediction. The improvements offered by flDPnn2 are statistically significant for AUPRC and lowPRCratio when compared to all 36 other predictors (p-value < 0.01). For the lowAUCratio, flDPnn2 is not statistically different than flDPnn but statistically better than the other methods (p-value < 0.01). For AUC, flDPnn2, flDPnn and Dispredict3 are not statistically different and statistically more accurate than the other methods (p-value < 0.01). We note that flDPnn2 is able to make predictions for all test proteins (coverage of 100%), compared to some other tools that cannot be used to predict certain sequences (e.g. short, very long or having non-standard amino acids). When compared to its predecessor, flDPnn2 generates modestly more accurate predictions with AUPRC of 0.596 (flDPnn2) vs. 0.581 (flDPnn), AUC of 0.838 vs. 0.835, F1 of 0.526 vs. 0.521, and MCC of 0.414

Table 1 Comparative assessment of predictive performance on the DisProt-NOX dataset from the CAID2 experiment. We report averages computed over the 20 repetitions using 50% of the DisProt-NOX dataset. We sort methods by their AUPRC values. Bold font identifies the best results for a given performance metric. Statistical significance of differences between flDPnn2 and each of the other 36 predictors is denoted by "+" if flDPnn2 is statistically better (p-value < 0.01) and by "=" if the difference is not statistically significant (p-value $\geq$ 0.01). We provide details of the assessment protocol in the "Evaluation" section in the Supplement. The coverage quantifies the fraction of proteins for which a given tool was able to produce predictions. Predictors that lack reference were not published.

Predictor [reference]	Coverage [%]	AUPRC	lowPRCratio	AUC	lowAUCratio	F1	MCC
flDPnn2	100.0	**0.596**	**4.07**	0.838	**6.73**	**0.526**	**0.414**
flDPnn[33]	100.0	0.581^{+}	3.95^{+}	0.835$^{=}$	6.53$^{=}$	0.521$^{=}$	0.409$^{=}$
Dispredict3	100.0	0.579^{+}	3.91^{+}	**0.842**$^{=}$	6.31^{+}	0.517$^{=}$	0.404$^{=}$
SPOT-Disorder2[37]	82.9	0.544^{+}	2.09^{+}	0.777^{+}	2.89^{+}	0.506$^{=}$	0.303^{+}
DisoPred	92.4	0.503^{+}	3.21^{+}	0.824^{+}	4.96^{+}	0.484^{+}	0.364^{+}
rawMSA[34]	100.0	0.476^{+}	3.25^{+}	0.790^{+}	4.75^{+}	0.464^{+}	0.338^{+}
ESpritz-D[35]	100.0	0.475^{+}	3.13^{+}	0.806^{+}	4.10^{+}	0.442^{+}	0.311^{+}
IDP-Fusion[62]	96.7	0.468^{+}	2.97^{+}	0.818^{+}	4.49^{+}	0.463^{+}	0.339^{+}
DisoMine[36]	100.0	0.466^{+}	3.07^{+}	0.799^{+}	3.92^{+}	0.434^{+}	0.301^{+}
DeepIDP-2L[63]	100.0	0.458^{+}	2.94^{+}	0.801^{+}	4.24^{+}	0.461^{+}	0.335^{+}
SETH-1[64]	100.0	0.403^{+}	2.39^{+}	0.781^{+}	3.45^{+}	0.434^{+}	0.301^{+}
pyHCA[65]	100.0	0.396^{+}	2.36^{+}	0.779^{+}	3.29^{+}	0.419^{+}	0.283^{+}
AlphaFold-RSA[66–68]	82.4	0.396^{+}	1.93^{+}	0.748^{+}	2.56^{+}	0.414^{+}	0.245^{+}
SETH-0[64]	100.0	0.378^{+}	1.79^{+}	0.735^{+}	2.32^{+}	0.389^{+}	0.212^{+}
PreDisorder[69]	95.7	0.376^{+}	1.79^{+}	0.729^{+}	2.22^{+}	0.393^{+}	0.218^{+}
SPOT-Disorder[70]	100.0	0.370^{+}	1.98^{+}	0.776^{+}	2.54^{+}	0.373^{+}	0.227^{+}
SPOT-Disorder-Single[71]	100.0	0.368^{+}	2.00^{+}	0.775^{+}	2.66^{+}	0.388^{+}	0.244^{+}
IUPred3[72]	100.0	0.367^{+}	2.09^{+}	0.758^{+}	2.77^{+}	0.403^{+}	0.264^{+}
Dispredict2[73]	96.2	0.364^{+}	2.29^{+}	0.645^{+}	2.33^{+}	0.348^{+}	0.163^{+}
AIUPred	100.0	0.363^{+}	1.92^{+}	0.770^{+}	2.64^{+}	0.402^{+}	0.263^{+}
AUCpred-profile[74]	99.5	0.363^{+}	3.58^{+}	0.767^{+}	2.56^{+}	0.000^{+}	0.000^{+}
MobiDB-lite[75]	100.0	0.361^{+}	3.24^{+}	0.748^{+}	3.19^{+}	0.364^{+}	0.236^{+}
IsUnstruct[76]	100.0	0.355^{+}	2.15^{+}	0.750^{+}	2.63^{+}	0.375^{+}	0.228^{+}
VSL2[77]	100.0	0.346^{+}	2.02^{+}	0.746^{+}	2.48^{+}	0.365^{+}	0.216^{+}
ESpritz-X[35]	100.0	0.341^{+}	1.93^{+}	0.743^{+}	2.45^{+}	0.363^{+}	0.214^{+}
AUCpred-no-profile[74]	100.0	0.337^{+}	3.43^{+}	0.743^{+}	2.35^{+}	0.307^{+}	0.178^{+}
PredIDR-long	100.0	0.335^{+}	1.90^{+}	0.743^{+}	2.39^{+}	0.339^{+}	0.184^{+}
RONN[78]	100.0	0.335^{+}	2.01^{+}	0.724^{+}	2.47^{+}	0.356^{+}	0.208^{+}
Metapredict[79]	100.0	0.334^{+}	1.83^{+}	0.759^{+}	1.86^{+}	0.328^{+}	0.172^{+}
PredIDR-short	100.0	0.333^{+}	1.91^{+}	0.738^{+}	2.44^{+}	0.347^{+}	0.195^{+}
AlphaFold-pLDDT[80]	82.4	0.329^{+}	1.36^{+}	0.698^{+}	1.54^{+}	0.322^{+}	0.128^{+}
ESpritz-N[35]	100.0	0.328^{+}	1.85^{+}	0.726^{+}	2.36^{+}	0.358^{+}	0.208^{+}
DisEMBL-dis465[81]	100.0	0.310^{+}	1.96^{+}	0.680^{+}	2.35^{+}	0.341^{+}	0.187^{+}
s2D	100.0	0.291^{+}	1.72^{+}	0.677^{+}	2.06^{+}	0.316^{+}	0.156^{+}
DISOPRED3[82]	100.0	0.291^{+}	2.26^{+}	0.696^{+}	1.61^{+}	0.275^{+}	0.116^{+}
FoldUnfold[83]	100.0	0.281^{+}	3.47^{+}	0.680^{+}	1.91^{+}	0.000^{+}	0.000^{+}
DisEMBL-disHL[81]	100.0	0.274^{+}	1.75^{+}	0.635^{+}	2.01^{+}	0.299^{+}	0.135^{+}

vs. 0.409. In the case of the AUPRC and lowPRCratio the improvements offered by flDPnn2 are statistically significant (p-value < 0.01). While being modest in magnitude, these improvements are consistent across all performance metrics.

Suppl. Figure S2 shows the ROC and precision-recall curves for the top five tools. We find that while flDPnn2 has modestly lower AUC when compared to Dispredict3, this is because Dispredict3 performs well for the part of the curve with FPR > 0.15 where the disorder is overpredicted. The higher value of lowAUCratio for flDPnn2 when compared to Dispredict3 (Table 1) means that the former tool performs better for the part of the curve where intrinsic disorder is not overpredicted (Suppl. Figure S2B).

Interestingly, the AlphaFold-RSA's predictions are substantially more accurate than the AlphaFold-pLDDT variant (AUPRC of 0.396 vs. 0.329, AUC of 0.748 vs. 0.698), which agrees with the recent analyses.[27,67–68] Moreover, the better AlphaFold-RSA is significantly less accurate than several state-of-the art disorder predictors, which is also in line with a recently published study[27]. Moreover, we find that the top six methods (i.e., flDPnn2, flDPnn, Dispredict3, SPOT-Disorder2, DisoPred, and rawMSA) rely on deep learning models, suggesting that these models are more accurate than other types of predictors. A similar conclusion was reached in a recent analysis that compared deep learning-based vs. other types of the intrinsic disorder predictors.[28]

We perform an additional analysis to investigate whether flDPnn2 performs similarly well on the DisProt-NOX proteins that share low, below 25% sequence similarity with the proteins in its alignment dataset (i.e., DisProt from March 2022) and its training dataset, which is a subset of the alignment dataset. More specifically, we use BLASTClust[49] to cluster the combined set of sequences from the alignment and the DisProt-NOX datasets at 25% similarity and select clusters that do not include the alignment proteins. The resulting sequence similarity reduced subset of the DisProt-NOX dataset includes 193 proteins and is available on the flDPnn2's page at https://biomine.cs.vcu.edu/servers/flDPnn2/. We summarize results on this dataset in Suppl. Table S1, with the corresponding ROC and precision-recall curves for the top five tools in Suppl. Figure S3. Overall, these results produce consistent observations when compared to the results on the complete DisProt-NOX dataset. We find that flDPnn2 secures the highest values of AUPRC of 0.614, lowPRCratio of 3.48, lowAUCratio of 5.81, F1 of 0.540 and MCC of 0.407 and the second highest AUC of 0.820, but the difference to the highest AUC of 0.823 for Dispredict3 is not statistically significant. It is also consistently and modestly better than flDPnn, and its AUPRC, lowPRCratio, lowAUCratio, and MCC are statistically better than the results of all other predictors (p-value < 0.01). The AUPRC values of flDPnn2, flDPnn and Dispredict3 (i.e., predictors that secure the highest AUCs and AUPRCs on the DisProt-NOX dataset) for the complete DisProt-NOX dataset are 0.596, 0.581 and 0.579 vs. 0.614, 0.592, and 0.594 on the sequence similarity reduced subset of DisProt-NOX, respectively. The same comparison for AUC produces 0.838, 0.835, and 0.842 (complete dataset) vs. 0.820, 0.817, and 0.823 (similarity reduced dataset), and for MCC is produces 0.414, 0.409, and 0.404 (complete dataset) vs. 0.407, 0.395, and 0.389 (similarity reduced dataset), respectively. These results suggest that flDPnn2 produces accurate results for the proteins that share low levels of similarity with proteins in its training and alignment datasets. They quantify performance when the alignment module is not used, given the insufficient levels of similarity to the alignment proteins, and suggest that this module did not provide measurable benefits for flDPnn2 in the CAID2 experiment. This means that the modest improvements over the original flDPnn are due to the changes in the deep network model. Interestingly, in spite of the fact that the sequence similarity was reduced only for flDPnn2, this predictor still maintains similar levels of differences in predictive performances when compared with the other methods. We note that we could not analyze results on the sequence similarity reduced subset of the DisProt-NOX dataset for the other disorder predictors since some of them, including Dispredict3, are unpublished and their training datasets are not available.

Runtime

The CAID2 assessors measured the runtimes of the predictors using the same hardware environment (i.e., the methods were deposited with and run by the assessors). The runtime values, which are measured in seconds per 1,000 residues long sequence, are available at https://caid.idpcentral.org/challenge#Runtimes. Figure 2 summarizes the runtime against the AUPRC values to visualize trade-offs between predictive performance and execution time. We find that the fastest methods, which predict in under 1 s per protein, provide limited levels of predictive performance, with the most accurate Espritz-D securing AUPRC of 0.475. On the other end of the runtime spectrum, by far the slowest option is AlphaFold that takes 71,690 s (i.e., close to 20 h). To compare, the top three most accurate tools are between 260 times faster (Dispredict3) and 2,600 times faster (flDPnn2) than AlphaFold. This suggests that some modern disorder predictors are both faster and more accurate in the context of the disorder prediction when compared to AlphaFold. The fastest among the most accurate methods is flDPnn2, which takes 26.6 s, and is 2.4 times faster than its predecessor, flDPnn, which takes 63.7 s. Dispredict3, which is another tool that is comparably accurate to flDPnn2, secures AUPRC = 0.579 vs. 0.596 for flDPnn2, while needing 275.4 s vs. 26.6 s (i.e., over 10 times slower). Altogether, these results suggest that flDPnn2 is relatively fast and substantially faster than the other very accurate disorder predictors, including flDPnn.

Prediction of fully disordered proteins

Similar to the CAID experiment[32], we apply the 37 methods to predict fully disordered proteins, i.e., proteins with a significant amount of disorder. These proteins have unique cellular functions[84–85] and are particularly difficult to predict accurately.[86] Like in CAID, we consider several classifications of the fully disordered proteins based on varying amounts of disorder content (DC), defined as a fraction of the intrinsically disordered residues, including 1, 0.9 and 0.8, i.e., proteins with 100%, over 90% and over 80% of disordered residues are assumed to be fully disordered. This is meant to investigate whether the results are robust to these different definitions. We use the average value of the predicted amino acid-level propensities to make prediction of the fully disordered proteins and summarize these results on the DisProt-NOX dataset in Suppl. Table S2. The most accurate predictions of fully disordered proteins are generated by flDPnn2,

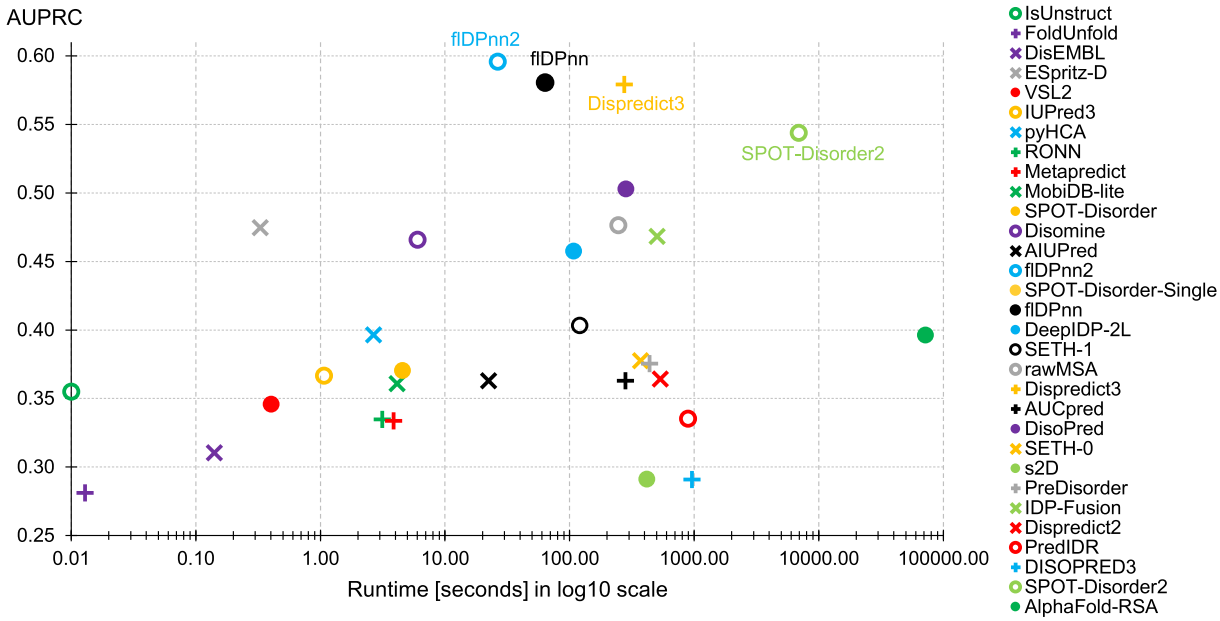

Figure 2. Analysis of runtime, quantified in seconds per a sequence of 1000 residues long (shown in log10 scale), based on the results from the CAID2 experiment. We report the highest AUPRC value for the tools that have multiple variants that share the same runtime, such as AlphaFold and DisEMBL.

followed by flDPnn and Dispredict3, which is consistent with the assessment of the intrinsic disorder predictions (Table 1). We note that the margin of the improvement of flDPnn2 vs. flDPnn is rather substantial, with the average AUPRC of 0.761 vs. 0.713 and the average AUC of 0.921 vs. 0.910, and these improvements are consistent across the three definitions of the fully disordered proteins. This suggests that flDPnn2 provides more accurate predictions of the fully disordered proteins when contrasted with its predecessor. Moreover, the Pearson correlation coefficients between the AUPRC values of the 37 predictors for each pair of the definitions (i.e., column-wise correlations in Suppl. Table S2 between DC = 1 and DC > 0.9, between DC = 1 and DC > 0.8, and between DC > 0.9 and DC > 0.8) range between 0.96 and 0.99, and for AUC between 0.98 and 0.99. These high correlations imply that the results are highly consistent across the different definitions of the fully disordered proteins, which agrees with the observations in CAID.[32]

Web server

We release flDPnn2 as a convenient web server at https://biomine.cs.vcu.edu/servers/flDPnn2/ to maximize its impact.[87] Given the relatively low runtime of flDPnn2, the server supports batch predictions of up to 50 FASTA-formatted protein sequences per job. The prediction process is fully automated and done on the server side. The server provides results in multiple formats that include raw predictions and an interactive graphical interface, which we detail in the Supplement.

CRediT authorship contribution statement

Kui Wang: Writing – original draft, Validation, Software, Resources, Methodology, Funding acquisition, Formal analysis, Data curation. **Gang Hu:** Writing – original draft, Validation, Resources, Methodology, Funding acquisition, Formal analysis, Data curation. **Sushmita Basu:** Writing – original draft, Software, Resources. **Lukasz Kurgan:** Writing – original draft, Visualization, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

DATA AVAILABILITY

The data is available publically and details are described in the article

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported in part by the National Science Foundation (grants DBI 2146027 and IIS 2125218) and the Robert J. Mattauch Endowment funds to L.K., and the National Natural Science Foundation of China (grant 92370128) to G.H. and K.W.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmb.2024.168605.

Received 3 January 2024;
Accepted 4 May 2024;
Available online xxxx

References

1. Habchi, J., Tompa, P., Longhi, S., Uversky, V.N., (2014). Introducing protein intrinsic disorder. *Chem Rev.* **114**, 6561–6588.
2. Lieutaud, P., Ferron, F., Uversky, A.V., Kurgan, L., Uversky, V.N., Longhi, S., (2016). How disordered is my protein and what is its disorder for? A guide through the "dark side" of the protein universe. *Intrinsically Disord Proteins.* **4**, e1259708.
3. Oldfield, C.J., Uversky, V.N., Dunker, A.K., Kurgan, L., (2019). Introduction to intrinsically disordered proteins and regions. In: Salvi, N. (Ed.), *Intrinsically Disordered Proteins*. Academic Press, pp. 1–34.
4. Peng, Z., Yan, J., Fan, X., Mizianty, M.J., Xue, B., Wang, K., et al., (2015). Exceptionally abundant exceptions: comprehensive characterization of intrinsic disorder in all domains of life. *Cell. Mol. Life Sci.* **72**, 137–151.
5. Xue, B., Dunker, A.K., Uversky, V.N., (2012). Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. *J. Biomol. Struct. Dyn.* **30**, 137–149.
6. Oates, M.E., Romero, P., Ishida, T., Ghalwash, M., Mizianty, M.J., Xue, B., et al., (2013). D(2)P(2): database of disordered protein predictions. *Nucleic Acids Res.* **41**, D508–D516.
7. Uversky, V.N., Dave, V., Iakoucheva, L.M., Malaney, P., Metallo, S.J., Pathak, R.R., et al., (2014). Pathological unfoldomics of uncontrolled chaos: intrinsically disordered proteins and human diseases. *Chem. Rev.* **114**, 6844–6879.
8. Rajagopalan, K., Mooney, S.M., Parekh, N., Getzenberg, R.H., Kulkarni, P., (2011). A majority of the cancer/testis antigens are intrinsically disordered proteins. *J. Cell. Biochem.* **112**, 3256–3267.
9. Uversky, V.N., (2015). Intrinsically disordered proteins and their (disordered) proteomes in neurodegenerative disorders. *Front. Aging Neurosci.* **7**, 18.
10. Gadhave, K., Gehi, B.R., Kumar, P., Xue, B., Uversky, V. N., Giri, R., (2020). The dark side of Alzheimer's disease: unstructured biology of proteins from the amyloid cascade signaling pathway. *Cell. Mol. Life Sci.* **77**, 4163–4208.
11. Santofimia-Castano, P., Rizzuti, B., Xia, Y., Abian, O., Peng, L., Velazquez-Campoy, A., et al., (2020). Targeting intrinsically disordered proteins involved in cancer. *Cell. Mol. Life Sci.* **77**, 1695–1707.
12. Ambadipudi, S., Zweckstetter, M., (2015). Targeting intrinsically disordered proteins in rational drug discovery. *Expert Opin. Drug Discov.*, 1–13.
13. Hu, G., Wu, Z., Wang, K., Uversky, V.N., Kurgan, L., (2016). Untapped potential of disordered proteins in current druggable human proteome. *Curr. Drug Targets* **17**, 1198–1205.
14. Tenchov, R., Zhou, Q.A., (2022). Intrinsically disordered proteins: perspective on COVID-19 infection and drug discovery. *ACS Infect. Dis.* **8**, 422–432.
15. Su, B.G., Henley, M.J., (2021). Drugging fuzzy complexes in transcription. *Front. Mol. Biosci.* **8**, 795743.
16. Aspromonte, M.C., Nugnes, M.V., Quaglia, F., Bouharoua, A., DisProt, C., Tosatto, S.C.E., et al., (2023). DisProt in 2024: improving function annotation of intrinsically disordered proteins. *Nucleic Acids Res.*.
17. Liu, Y., Wang, X., Liu, B., (2019). A comprehensive review and comparison of existing computational methods for intrinsically disordered protein and region prediction. *Brief. Bioinform.* **20**, 330–346.
18. Meng, F., Uversky, V.N., Kurgan, L., (2017). Comprehensive review of methods for prediction of intrinsic disorder and its molecular functions. *Cell. Mol. Life Sci.* **74**, 3069–3090.
19. Zhao, B., Kurgan, L., (2021). Surveying over 100 predictors of intrinsic disorder in proteins. *Expert Rev. Proteomics* **18**, 1019–1029.
20. Zhao, B., & Kurgan, L., (2023). Machine Learning for Intrinsic Disorder Prediction. Machine Learning in Bioinformatics of Protein Sequences. 205–236.
21. Kurgan, L., Hu, G., Wang, K., Ghadermarzi, S., Zhao, B., Malhis, N., et al., (2023). Tutorial: a guide for the selection of fast and accurate computational tools for the prediction of intrinsic disorder in proteins. *Nature Protoc.* **18**, 3157–3172.
22. Punta, M., Simon, I., Dosztanyi, Z., (2015). Prediction and analysis of intrinsically disordered proteins. *Methods Mol. Biol.* **1261**, 35–59.
23. He, B., Wang, K., Liu, Y., Xue, B., Uversky, V.N., Dunker, A.K., (2009). Predicting intrinsic disorder in proteins: an overview. *Cell Res.* **19**, 929–949.
24. Katuwawala, A., Kurgan, L., (2020). Comparative assessment of intrinsic disorder predictions with a focus on protein and nucleic acid-binding proteins. *Biomolecules*, 10.
25. Necci, M., Piovesan, D., Dosztanyi, Z., Tompa, P., Tosatto, S.C.E., (2018). A comprehensive assessment of long intrinsic protein disorder from the DisProt database. *Bioinformatics* **34**, 445–452.
26. Walsh, I., Giollo, M., Di Domenico, T., Ferrari, C., Zimmermann, O., Tosatto, S.C., (2015). Comprehensive large-scale assessment of intrinsic protein disorder. *Bioinformatics* **31**, 201–208.
27. Zhao, B., Ghadermarzi, S., Kurgan, L., (2023). Comparative evaluation of AlphaFold2 and disorder predictors for prediction of intrinsic disorder, disorder content and fully disordered proteins. *Comput. Struct. Biotechnol. J.* **21**, 3248–3258.
28. Zhao, B., Kurgan, L., (2022). Deep learning in prediction of intrinsic disorder in proteins. *Comput. Struct. Biotechnol. J.* **20**, 1286–1294.
29. Melamud, E., Moult, J., (2003). Evaluation of disorder predictions in CASP5. *Proteins* **53**, 561–565.
30. Monastyrskyy, B., Kryshtafovych, A., Moult, J., Tramontano, A., Fidelis, K., (2014). Assessment of protein disorder region predictions in CASP10. *Proteins* **82**, 127–137.
31. Conte, A.D., Mehdiabadi, M., Bouhraoua, A., Miguel Monzon, A., Tosatto, S.C.E., Piovesan, D., (2023). Critical assessment of protein intrinsic disorder prediction (CAID) – Results of round 2. *Proteins*.

32. Necci, M., Piovesan, D., Predictors, C., DisProt, C., Tosatto, S.C.E., (2021). Critical assessment of protein intrinsic disorder prediction. *Nature Methods* **18**, 472–481.
33. Hu, G., Katuwawala, A., Wang, K., Wu, Z., Ghadermarzi, S., Gao, J., et al., (2021). flDPnn: accurate intrinsic disorder prediction with putative propensities of disorder functions. *Nature Commun.* **12**, 4438.
34. Mirabello, C., Wallner, B., (2019). rawMSA: end-to-end deep learning using raw multiple sequence alignments. *PLoS One* **14**
35. Walsh, I., Martin, A.J.M., Di Domenico, T., Tosatto, S.C.E., (2012). ESpritz: accurate and fast prediction of protein disorder. *Bioinformatics* **28**, 503–509.
36. Orlando, G., Raimondi, D., Codice, F., Tabaro, F., Vranken, W., (2022). Prediction of disordered regions in proteins with recurrent neural networks and protein dynamics. *J. Mol. Biol.* **434**, 167579.
37. Hanson, J., Paliwal, K.K., Litfin, T., Zhou, Y., (2019). SPOT-Disorder 2: improved protein intrinsic disorder prediction by ensembled deep learning. *Genomics Proteomics Bioinformatics* **17**, 645–656.
38. Hatos, A., Hajdu-Soltesz, B., Monzon, A.M., Palopoli, N., Alvarez, L., Aykac-Fas, B., et al., (2020). DisProt: intrinsic protein disorder annotation in 2020. *Nucleic Acids Res.* **48**, D269–D276.
39. Lang, B., Babu, M.M., (2021). A community effort to bring structure to disorder. *Nature Methods* **18**, 454–455.
40. Zhang, J., Basu, S., Kurgan, L., (2023). HybridDBRpred: improved sequence-based prediction of DNA-binding amino acids using annotations from structured complexes and disordered proteins. *Nucleic Acids Res.*.
41. Yan, J., Kurgan, L., (2017). DRNApred, fast sequence-based method that accurately predicts and discriminates DNA- and RNA-binding residues. *Nucleic Acids Res.* **45**, e84.
42. Zhang, J., Ghadermarzi, S., Kurgan, L., (2020). Prediction of protein-binding residues: dichotomy of sequence-based methods developed using structured complexes versus disordered proteins. *Bioinformatics* **36**, 4729–4738.
43. Zhang, J., Kurgan, L., (2018). Review and comparative assessment of sequence-based predictors of protein-binding residues. *Brief. Bioinform.* **19**, 821–837.
44. Zhang, J., Ghadermarzi, S., Katuwawala, A., Kurgan, L., (2021). DNAgenie: accurate prediction of DNA-type-specific binding residues in protein sequences. *Brief. Bioinform.* **22**, bbab336.
45. Basu, S., Hegedus, T., Kurgan, L., (2023). CoMemMoRFPred: sequence-based prediction of MemMoRFs by combining predictors of intrinsic disorder, MoRFs and disordered lipid-binding regions. *J. Mol. Biol.* **435**, 168272.
46. Nelson, L.S., (1998). The anderson-darling test for normality. *J. Qual. Technol.* **30**, 298.
47. Yan, J., Friedrich, S., Kurgan, L., (2016). A comprehensive comparative review of sequence-based predictors of DNA- and RNA-binding residues. *Brief. Bioinform.* **17**, 88–105.
48. 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. *Brief. Bioinform.* **23**, bbab521.
49. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., et al., (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402.
50. Wang, K., Samudrala, R., (2006). Incorporating background frequency improves entropy-based residue conservation measures. *BMC Bioinf.* **7**, 385.
51. Meszaros, B., Erdos, G., Dosztanyi, Z., (2018). IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* **46**, W329–W337.
52. Buchan, D.W.A., Jones, D.T., (2019). The PSIPRED protein analysis workbench: 20 years on. *Nucleic Acids Res.* **47**, W402–W407.
53. Peng, Z., Kurgan, L., (2015). High-throughput prediction of RNA, DNA and protein binding regions mediated by intrinsic disorder. *Nucleic Acids Res.* **43**, e121.
54. Peng, Z., Wang, C., Uversky, V.N., Kurgan, L., (2017). Prediction of disordered RNA, DNA, and protein binding regions using DisoRDPbind. *Methods Mol. Biol.* **1484**, 187–203.
55. Meng, F., Kurgan, L., (2016). DFLpred: High-throughput prediction of disordered flexible linker regions in protein sequences. *Bioinformatics* **32**, i341–i350.
56. Yan, J., Dunker, A.K., Uversky, V.N., Kurgan, L., (2016). Molecular recognition features (MoRFs) in three domains of life. *Mol. Biosyst.* **12**, 697–710.
57. Dosztanyi, Z., Csizmok, V., Tompa, P., Simon, I., (2005). IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **21**, 3433–3434.
58. Disfani, F.M., Hsu, W.L., Mizianty, M.J., Oldfield, C.J., Xue, B., Dunker, A.K., et al., (2012). MoRFpred, a computational tool for sequence-based prediction and characterization of short disorder-to-order transitioning binding regions in proteins. *Bioinformatics* **28**, i75–i83.
59. Peng, Z., Li, Z., Meng, Q., Zhao, B., Kurgan, L., (2023). CLIP: accurate prediction of disordered linear interacting peptides from protein sequences using co-evolutionary information. *Brief. Bioinform.* **24**.
60. Peng, Z., Xing, Q., Kurgan, L., (2020). APOD: accurate sequence-based predictor of disordered flexible linkers. *Bioinformatics* **36**, i754–i761.
61. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., (2014). Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**, 1929–1958.
62. Tang, Y.J., Yan, K., Zhang, X., Tian, Y., Liu, B., (2023). Protein intrinsically disordered region prediction by combining neural architecture search and multi-objective genetic algorithm. *BMC Biol.* **21**, 188.
63. Tang, Y.J., Pang, Y.H., Liu, B., (2022). DeepIDP-2L: protein intrinsically disordered region prediction by combining convolutional attention network and hierarchical attention network. *Bioinformatics* **38**, 1252–1260.
64. Ilzhofer, D., Heinzinger, M., Rost, B., (2022). SETH predicts nuances of residue disorder from protein embeddings. *Front Bioinform.* **2**, 1019597.
65. Bitard-Feildel, T., Callebaut, I., (2018). HCAtk and pyHCA: a toolkit and python API for the hydrophobic cluster analysis of protein sequences. *bioRxiv*.
66. Akdel, M., Pires, D.E.V., Pardo, E.P., Janes, J., Zalevsky, A.O., Meszaros, B., et al., (2022). A structural biology community assessment of AlphaFold2 applications. *Nature Struct. Mol. Biol.* **29**, 1056–1067.
67. Wilson, C.J., Choy, W.Y., Karttunen, M., (2022). AlphaFold2: a role for disordered protein/region prediction? *Int. J. Mol. Sci.* **23**, 4591.

68. Piovesan, D., Monzon, A.M., Tosatto, S.C.E., (2022). Intrinsic protein disorder and conditional folding in AlphaFoldDB. *Protein Sci.* **31**, e4466.
69. Deng, X., Eickholt, J., Cheng, J., (2009). PreDisorder: ab initio sequence-based prediction of protein disordered regions. *BMC Bioinf.* **10**, 436.
70. Hanson, J., Yang, Y.D., Paliwal, K., Zhou, Y.Q., (2017). Improving protein disorder prediction by deep bidirectional long short-term memory recurrent neural networks. *Bioinformatics* **33**, 685–692.
71. Hanson, J., Paliwal, K., Zhou, Y., (2018). Accurate single-sequence prediction of protein intrinsic disorder by an ensemble of deep recurrent and convolutional architectures. *J. Chem. Inf. Model.* **58**, 2369–2376.
72. Erdos, G., Pajkos, M., Dosztanyi, Z., (2021). IUPred3: prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation. *Nucleic Acids Res.* **49**, W297–W303.
73. Iqbal, S., Hoque, M.T., (2016). Estimation of position specific energy as a feature of protein residues from sequence alone for structural classification. *PLoS One* **11**
74. Wang, S., Ma, J.Z., Xu, J.B., (2016). AUCpreD: proteome-level protein disorder prediction by AUC-maximized deep convolutional neural fields. *Bioinformatics* **32**, 672–679.
75. Necci, M., Piovesan, D., Dosztanyi, Z., Tosatto, S.C.E., (2017). MobiDB-lite: fast and highly specific consensus prediction of intrinsic disorder in proteins. *Bioinformatics* **33**, 1402–1404.
76. Lobanov, M.Y., Galzitskaya, O.V., (2011). The Ising model for prediction of disordered residues from protein sequence alone. *Phys. Biol.* **8**, 035004.
77. Peng, K., Radivojac, P., Vucetic, S., Dunker, A.K., Obradovic, Z., (2006). Length-dependent prediction of protein intrinsic disorder. *BMC Bioinf.* **7**, 208.
78. Yang, Z.R., Thomson, R., McNeil, P., Esnouf, R.M., (2005). RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics* **21**, 3369–3376.
79. Emenecker, R.J., Griffith, D., Holehouse, A.S., (2021). Metapredict: a fast, accurate, and easy-to-use predictor of consensus disorder and structure. *Biophys. J.* **120**, 4312–4319.
80. Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Zidek, A., et al., (2021). Highly accurate protein structure prediction for the human proteome. *Nature* **596**, 590–596.
81. Linding, R., Jensen, L.J., Diella, F., Bork, P., Gibson, T.J., Russell, R.B., (2003). Protein disorder prediction: implications for structural proteomics. *Structure* **11**, 1453–1459.
82. Jones, D.T., Cozzetto, D., (2015). DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics* **31**, 857–863.
83. Galzitskaya, O.V., Garbuzynskiy, S.O., Lobanov, M.Y., (2006). FoldUnfold: web server for the prediction of disordered regions in protein chain. *Bioinformatics* **22**, 2948–2949.
84. Deiana, A., Forcelloni, S., Porrello, A., Giansanti, A., (2019). Intrinsically disordered proteins and structured proteins with intrinsically disordered regions have different functional roles in the cell. *PLoS One* **14**, e0217889.
85. Necci, M., Piovesan, D., Tosatto, S.C., (2016). Large-scale analysis of intrinsic disorder flavors and associated functions in the protein sequence universe. *Protein Sci.* **25**, 2164–2174.
86. Zhao, B., Kurgan, L., (2022). Compositional bias of intrinsically disordered proteins and regions and their predictions. *Biomolecules* **12**.
87. Song, J.N., Kurgan, L., (2023). Availability of web servers significantly boosts citations rates of bioinformatics methods for protein function and disorder prediction. *Bioinformat. Adv.* **3.**