Surveying over 100 Predictors of Intrinsic Disorder in Proteins

Bi Zhao[1*], Lukasz Kurgan[1*]

[1]Department of Computer Science, Virginia Commonwealth University, Richmond, Virginia, United States

*Corresponding authors: Lukasz Kurgan (lkurgan@vcu.edu) and Bi Zhao (zhaob4@vcu.edu)

Abstract

Introduction: Intrinsic disorder prediction field develops, assesses and deploys computational predictors of disorder in protein sequences and constructs and disseminates databases of these predictions. Over 40 years of research resulted in the release of numerous resources.

Areas covered: We identify and briefly summarize the most comprehensive to date collection of over 100 disorder predictors. We focus on their predictive models, availability and predictive performance. We categorize and study them from a historical point of view to highlight informative trends.

Expert opinion: We find a consistent trend of improvements in predictive quality as newer and more advanced predictors are developed. The original focus on machine learning methods has shifted to meta-predictors in early 2010s, followed by a recent transition to deep learning. The use of deep learners will continue in foreseeable future given recent and convincing success of these methods. Moreover, a broad range of resources that facilitate convenient collection of accurate disorder predictions is available to users. They include web servers and standalone programs for disorder prediction, servers that combine prediction of disorder and disorder functions, and large databases of pre-computed predictions. We also point to the need to address the shortage of accurate methods that predict disordered binding regions.

Keywords: Database; deep learning; intrinsic disorder; intrinsically disordered regions; machine learning; prediction; predictive performance; protein function.

1 Introduction

Intrinsically disordered proteins (IDPs) contain both structured domains and intrinsically disordered regions (IDRs) or could be fully disordered, i.e., composed of one sequence-wide IDR [1]. IDRs are defined by the lack of stable secondary and tertiary structures under physiological conditions [2-4]. They are instrumental for a wide range of cellular functions including molecular assembly, molecular recognition, cell cycle regulation, transcription, translation, and signal transduction [5-11]. Bioinformatics studies suggest that intrinsic disorder is abundant across all kingdoms of life and virus, with substantially higher levels in eukaryotes [12-14]. Studies approximate that around one-third of eukaryotic proteins have long IDRs, i.e., regions with 30 or more consecutive disordered residues [12-15].

Several databases, such as DisProt [16], PDB [17], IDEAL [18], DIBS [19], and MFIB [20], store the experimentally established annotations of IDRs and/or their associated functions. However, the amount of these annotations is relatively low. For instances, the current version 8.3 of DisProt covers about 2 thousand IDPs [16] and a recent study extracted about 25 thousand IDPs from PDB [21]. This is just a miniscule fraction of the 220 million protein sequences that are available in the 2021_03 release the UniProt database [22]. The large gap between the huge number of protein sequences and the small number of experimentally annotated IDPs motivates the development of computational predictors of IDRs. These methods identify putative disordered residues and regions in an input protein sequence.

Many disorder predictors have been developed to date [23,24]. Consequently, numerous surveys that summarize the field of the computational disorder prediction were published [2,23-32]. These works provide historical overview, categorize and describe disorder predictors and, in some cases, compare their predictive performance. The most comprehensive of these surveys cover as many as 70 [23], 55 [25] and 45 predictors [24]. Moreover, about a dozen large-scale comparative studies were carried out to assess predictive performance of the intrinsic

disorder predictors [21,33-42]. These studies include several community assessments where predictors are evaluated on blind test datasets (i.e., datasets that were not available to the authors of the predictors) by independent assessors who do not take part in the competitions. The community assessments include Critical Assessment of Structure Prediction (CASP) between CASP5 to CASP10 [37-42] and Critical Assessment of Intrinsic Protein Disorder (CAID) [36]. The largest of these assessments reported results for 28 disorder predictors in CASP10 [40] and 32 in CAID [36]. Our objective is to compile and summarize the most complete list of disorder predictors to date. We identify 103 methods and provide an updated historical perspective on the progression of the efforts to develop these methods that covers a recent infusion of the deep learning models. We summarize the predictive models that these methods utilize and comment on their availability to end users. Moreover, we analyze results of the two largest and most recent community assessments, CASP 10 and CAID, to study whether and how the predictive performance of disorder predictors changed over the years as new methods have become available.

Table 1. Summary of 67 predictors of intrinsic disorder that are available to the users. The methods are sorted in the chronological order of their publication.

Predictor name	Year published	Reference[1]	Algorithms used[2]	Meta prediction[3]	Availability[4]
SEG	1994	[43]	SSM	No	SP+WS
PONDR CaN-XT	1997	[44]	SNN	No	WS
PONDR XL1	1997	[44]	SNN	No	WS
PONDR VL-XT	2001	[45]	SNN	No	WS
DisEMBL-REM465	2003	[46]	SNN	No	SP+WS
DisEMBL-HL	2003	[46]	SNN	No	SP+WS
DisEMBL-COIL	2003	[46]	SNN	No	SP+WS
GlobPlot	2003	[47]	SSM	No	SP+WS
NORSp	2003	[48]	SSM	No	WS
DISOPRED	2003	[49]	SNN	No	SP
DISOPRED2	2004	[12]	SVM+SNN	No	SP
DISpro	2005	[50,51]	SNN	No	SP+WS
FoldIndex	2005	[52]	SSM	No	WS
IUPred-long	2005	[53,54]	SSM	No	SP+WS
IUPred-short	2005	[53,54]	SSM	No	SP+WS
PONDR VL3	2005	[55]	SNN	No	WS
PONDR VL3H	2005	[55]	SNN	No	WS
PONDR VL3E	2005	[55]	SNN	No	WS
RONN (JRONN)	2005	[56]	SNN	No	WS
PONDR VSL1	2005	[57]	RM	No	WS
PROFbval	2006	[58]	SNN	No	SP+WS
PONDR VSL2B	2006	[57,59]	SVM	No	WS
PONDR VSL2P	2006	[57,59]	SVM	No	WS
NORSnet	2007	[60]	SNN	No	SP+WS
PrDOS (PrDOS2)	2007	[61]	SVM	No	WS
Pdisorder	2007	No	SNN	No	SP
UCON	2007	[60]	SNN	No	SP+WS
DISOclust	2008	[62]	SSM	No	SP+WS
DRIPPRED	2008	No	SSM	No	SP
metaPrDOS (metaPrDOS2)	2008	[63]	SVM	Yes	WS
MD	2009	[64]	SNN	Yes	SP+WS
MFDp	2010	[65]	SVM	Yes	WS
PONDR FIT	2010	[66]	SNN	Yes	WS
Cspritz	2011	[67]	SNN	Yes	WS
DISOclust3 (IntFOLD)	2011	[68]	CS	No	SP+WS
IsUnstruct	2011	[69]	SSM	No	SP+WS
Espritz-D	2012	[70]	SNN	No	SP+WS
Espritz-N	2012	[70]	SNN	No	SP+WS
Espritz-X	2012	[70]	SNN	No	SP+WS
GSmetaDisorderMD	2012	[71,72]	CS	Yes	WS
GSmetadisorder	2012	[71,72]	CS	Yes	WS
GSmetaserver	2012	[71,72]	CS	Yes	WS
Gsmetadisorder3D	2012	[71,72]	CS	No	WS
SPINE-D	2012	[73]	SNN	No	SP
MFDp2	2013	[74,75]	SVM	Yes	WS
DisMeta	2014	[76]	SSM	Yes	WS
disCoP	2014	[77,78]	RM	Yes	WS
DynaMine	2014	[79,80]	RM	Yes	SP+WS
PON-Diso	2014	[81]	RF	No	WS
DISOPRED3	2015	[82]	SVM+SNN+NN	No	SP+WS
DisPredict (DisPredict2)	2016	[83]	SVM	No	SP

Predictor name	Year published	Reference[1]	Algorithms used[2]	Meta prediction[3]	Availability[4]
MobiDB-lite	2017	[84]	CS	Yes	WS
SPOT-Disorder1	2017	[85]	DNN(R)	No	SP+WS
IUpred2A-long	2018	[86]	SSM	No	SP+WS
IUpred2A-short	2018	[86]	SSM	No	SP+WS
pyHCA	2018	No	SSM	No	SP
SPOT-Disorder-Single	2018	[87]	DNN(C+R)	No	SP+WS
rawMSA	2019	[88]	DNN(C+R)	No	SP
SPOT-Disorder2	2019	[89]	DNN(C+R)	No	SP+WS
DisoMine	2020	No	DNN(R)	No	WS
ODiNPred	2020	[90]	SNN	No	WS
IDP-Seq2Seq	2020	[91]	DNN(R)	No	WS
flDPnn	2021	[92]	DNN(F)	No	SP+WS
flDPlr	2021	[92]	RM+RF	No	SP+WS
IUPred3	2021	[93]	SSM	No	SP+WS
RFPR-IDP	2021	[94]	DNN(C+R)	No	WS
Metapredict	2021	[95]	DNN(R)	Yes	SP

[1]"No" means that a given predictor was not published in a peer-reviewed journal.
[2]Algorithms used: "SSM" (Sequence scoring function-based methods); Machine learning methods including "SVM" (support vector machine), "DT" (decision tree), "RF" (random forest), "CS" (consensus score), "RM" (regression model), "CRF" (conditional random field), "NN" (nearest neighbor), "BC" (Bayesian classifier), "RBFN" (radial basis function networks), "SNN" (shallow neural network); and "DNN" (deep neural network) where "C" stands for convolutional network, "R" for recurrent network, and "F" for feed-forward network.
[3]Meta prediction refers to the predictors that use outputs of multiple disorder predictors as input.
[4]Availability: released as "SP" (standalone program), "WS" (web server).

2 Intrinsic Disorder Prediction

We search for the disorder predictors using a variety of sources including databases that provide access to disorder predictions, MobiDB [96] and D^2P^2 [97], community assessments of the disorder predictors, such as CASP 5, CASP 6, CASP 7, CASP 8, CASP 9, CASP 10 and CAID [36-42], 12 previously published surveys of disorder predictors [21,23-28,31,33-35], and a manual search of relevant articles collected from PubMed using the "(disorder[Title]) AND (prediction[Title]) AND protein" query. This extensive search produced a list of 103 disorder predictors [12,43-47,49-74,76,77,79,80,82-95,98-127]. This list includes four unpublished methods that participated in CASP and/or CAID assessments.

We summarize these predictors in Table 1, which focuses on the 67 methods that are currently available to users, and Suppl. Table S1, which covers 36 methods that were published but are presently unavailable. These tables identify when and where these tools were published and summarize key characteristics of the underlying predictive models, such as the type of an algorithm used and whether meta-prediction is utilized.

2.1 Predictive models

The predictive models cover a broad spectrum of alternatives. We classify predictors into four categories based on the type of predictive models that they utilize: (1) sequence scoring function-based methods; (2) machine learning approaches; (3) deep learning methods; and (4) meta-predictors. This classification is inspired by past surveys [23,24,26,28,31], but with the addition of the deep learning category that is fueled by the most recent developments. The sequence scoring function-based methods aggregate information extracted from the input protein sequence and sequence-derived structural and evolutionary information using additive and/or weighted functions. Some of these functions are grounded in physical principles governing protein folding processes. Representative examples include FoldIndex [52], IUPred [53,54], and IUPred2A [86], and IUPred3 [94]. The machine learning predictors rely on more sophisticated predictive models that are trained from data using a variety of machine learning algorithms. These algorithms include support vector machine [61,65,74,83,105], regression [77,79,128], conditional random field [113,121,124,125], random forest [81,129], nearest neighbor [106,109], Bayesian classifier [130], radial basis function networks [104,131], and shallow neural networks [46,55,66,114]. Two or more models generated by different machine learning algorithms are sometimes used in tandem with the aim to improve predictive performance when compared to using one model [82,123,126]. Examples of popular machine learning methods include DisEMBL [46], DISOPRED [49], PONDR [55], PrDOS [61] and DISOPRED3 [82]. The deep learning methods rely on deep neural networks, which is a specific type of machine learning algorithms. The deep neural networks differ from shallow networks by inclusion of multiple

hidden layers. They also typically use more advanced types of neurons and connections. We separate deep networks from the other machine learning methods since many recent methods use these types of predictive models. These disorder predictors rely on several different types of deep networks including feed-forward networks, recurrent networks, convolutional networks, and their hybrids that combine convolutional and recurrent topologies (Table 1 and Suppl. Table S1). The focus on the deep networks stems from the fact that they provide favorable levels of predictive performance when compared with the other categories of disorder predictors. In particular, the best performing methods that participated in the most recent CAID competition [36,132], which include flDPnn [92], SPOT-Disorder2 [89], RawMSA [88] and AUCpred [123], rely on the deep learning models. We highlight the diversity of the deep network types that they utilize including feed-forward by flDPnn, convolutional by AUCpred, and a hybrid by SPOT-Disorder2 and RawMSA. The meta-predictors apply multiple disorder predictions as inputs to (re)predict disorder with the underlying aim to improve predictive performance when compared to the input predictions. The development of these methods was motivated by the availability of many diverse machine learning and sequence scoring function-based methods that could be used as the inputs. Moreover, several empirical studies show that well-designed meta predictors indeed produce predictions that outperform their inputs [77,84,133]. A few illustrative examples in this category include metaPrDOS [63], MFDp [65], Cspritz [67], GSmetadisorder [71,72], MFDp2 [74,75], disCoP [77,78], and MobiDB-lite [84]. We note that some meta-predictors use machine learning models to process their inputs (e.g., metaPrDOS [63] and MFDp [65]), which means that they belong in both categories of disorder predictors. Moreover, a few methods, such as PrDOS[61], DISOclust[62] and Gsmetadisorder3D[72], utilize homology modelling to predict disorder.

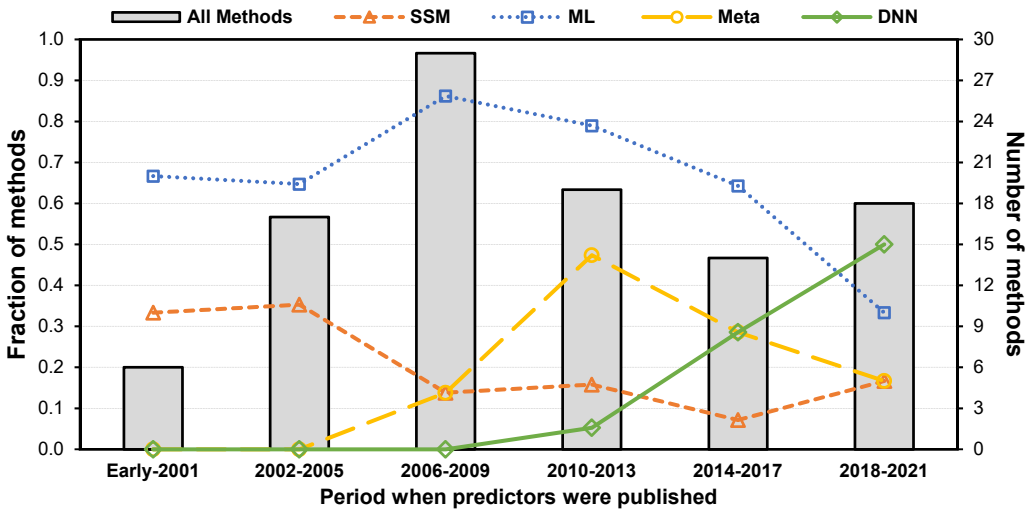

Figure 1. Distribution of disorder predictors categorized by their predictive models over time. The *x*-axis represents the periods when the predictors were published. The left/right *y*-axis gives the fraction/number of methods in the given time period. The light grey bars represent the number of all disorder predictors. Color-coded lines show fractions of different categories of predictors: orange with triangle markers for SSM (sequence scoring function-based methods); blue with square markers for ML (machine learning methods); yellow with circle markers for meta (meta-predictors); and green with diamond maskers for DNN (deep neural networks).

Among the 103 disorder predictors 19 use the sequence scoring function-based models, 70 are machine learning methods, and 14 predictors rely on the deep learning. Moreover, there are 20 meta-predictors. **Figure 1** summarizes major trends in popularity of these four categories of methods over time. Only six predictors were released before 2002, with four that rely on the machine learning models, specifically shallow neural networks, and two that apply sequence scoring function-based models. On average five new methods were developed annually after 2002, with a sharp peak between 2006 and 2009. The least popular type of methods are the scoring function-based models, although new methods of that type are developed consistently over the years. We note a relatively steady stream of the most numerous class of the machine learning methods, with a modest decline in the recent years. The first meta-predictors were published in 2008, and their development was fueled by the availability of over 40 disorder predictors at that time. However, after a spike in popularity of meta-predictors in early 2010s, with nearly half of all predictors developed in 2010-2013 period in this category, the

number of these methods is in a steady decline. The first deep learning method was released in 2013 and gradually more of these methods are being developed in the recent years. The deep learners became the most popular category of disorder predictors in the last three years. This is likely fueled by the overall popularity of deep learners in the area of structural bioinformatics of protein [134,135] and the success of deep learners in the recent assessments of the disorder predictions [31,33,36]. To summarize, recent years have produced a substantial shift in the development efforts where the decreasing numbers of meta-predictors are coupled with the rising popularity of the deep neural networks.

2.2 Availability

Table 1 summarizes 67 disorder predictors that are currently available to the users and explains how these methods are shared, i.e., as web servers and/or standalone programs. The web servers are an attractive option for less computer savvy users and those who perform predictions in an ad hoc manner. They are accessible via a web site where users input their sequences and collect the corresponding disorder predictions. The entire prediction process is done on the server side, which means that users do not need to invest in any hardware or software to collect the results. On the other hand, the standalone programs are more suited for bioinformaticians who perform predictions regularly and/or who aim to embed these predictions into larger bioinformatics platforms. There programs must be downloaded and have to be installed and run on the user's hardware. Our analysis reveals that nearly two-thirds of predictors are available as web servers and/or standalone software. The other one-third of predictors, which we summarize in **Supplementary Table S1**, were not released at the time of publication or were released but were unavailable as of August 2021 when we tested the access. We also provide URLs of the available predictors in **Supplementary Table S2.**

2.3 Historical perspective

Previous surveys define three distinct periods in the development of the disorder predictors [23,25]. We expand this timeline to four periods to recognize recent advances that resulted in the development of the deep learners. The ***first-generation*** methods were released between 1979 and 2001. Only a few methods were developed during that time. The first method, which was published in 1979 targets prediction of random coil conformations in protein sequences [98]. However, due to limited availability of data on IDRs at that time, this method could not be sufficiently tested at the time of publication. Later evaluations show that its predictive performance is relatively poor [25]. The first machine learning-based method was proposed by Romero, Obradovic, and Dunker in 1997 [99]. It uses a shallow neural network model that relies on physical and chemical properties of protein sequences. Interestingly, some of these early methods are still available online as webservers and/or standalone code (**Table 1** and **Supplementary Table S2**).

The ***second-generation*** methods date between 2002 and 2006. The key characteristic of this period is the rapid intensification of the efforts to develop disorder predictors. Moreover, most of these methods are based on relatively simple predictive models, often relying on the sequence scoring functions and shallow neural networks. The key innovation was the use of evolutionary profiles that are generated from the position specific score matrix (PSSM) produced by PSI-BLAST from the input protein sequences [128,136]. A significant event during the period of time was the inclusion of the disorder prediction assessment into CASP5 in 2003 [41]. This resulted in popularization of this predictive area [25]. Some of the prominent second-generation methods include sequence scoring-function-based models, such as GlobPlot [47], FoldUnfold [102,103], and IUPred [53,54], and machine learning methods including PONDR family of disorder predictors [45,55,57,59], DISOPRED [49], DisEMBL [46], and RONN [56].

The ***third-generation*** predictors were released between 2007 and 2015. The defining characteristics of this period is the introduction of the meta-predictors. Nearly all meta-predictors, 16 out of the total of 20, were published during that timeframe. Popular third-generation meta-predictors include MFDp [65], PONDR-FIT [66], and CSpritz [67], GSmetadisorder [71,72] and DisCoP [77,78]. The disorder predictions continued to be biannually evaluated in the CASP experiments [38-40,42]. This resulted in a steady influx of new methods, totaling 48 for this time period. However, the inclusion of disorder predictions in CASP ended with CASP10 that was held in 2012 [40].

The ***fourth-generation*** period started in 2016. Its defining feature is the rapid acceleration in the development of the deep learning models. While the first deep learning method was developed in 2013 [118], these efforts intensified in the last five years. About half of the fourth-generation disorder predictors, 11 out of 23, rely on the deep neural networks. This is the largest of the four categories of the predictive models during this time period (**Figure 1**). A few representative deep learners include flDPnn that utilizes deep feed-forward topology [92]; AUCpred which integrates conditional random field with a convolutional network [123]; SPOT-Disorder predictors that rely on the recurrent networks [85,87,89]; and RawMSA which combines convolutional and bidirectional recurrent networks [88]. As we mention above, the strong focus on designing deep networks was motivated by their popularity and ability to produce accurate results, culminated in a convincing success by this class of methods in the most recent CAID community assessment [36,132].

2.4 Related resources

Table 1 shows that users can secure disorder predictions by using webservers and standalone applications. Another arguably more convenient alternative is to collect pre-computed disorder predictions from one of the three currently available databases: Database of Disorder Protein Predictions (D^2P^2) [97], MobiDB [137], and DescribePROT [138]. Each database provides access to results produced by several disorder predictors for large collections of proteins ranging from 1.35 million proteins in DescribePROT, 10.43 million proteins in D^2P^2, to 189.52 million proteins in MobiDB. The main advantage of these resources is the instantaneous access to the pre-computed predictions generated by several different methods (i.e., users do not have to wait for the predictions to complete) and the corresponding consensus prediction. However, webservers or standalone applications must be used when users want to predict sequences that are not included in a given database.

Another useful resource is the recently developed DisorderEd PredictIon CenTER (DEPICTER) [139], a large-scale webserver that produces predictions of disorder and disorder functions. The latter predictions identify IDRs that share the same molecular function, such as binding to proteins or nucleic acids. DEPICTER produces disorder predictions from UPred-short [54], IUPred-long [54] and SPOT-Disorder-Single [87] that are accompanied by the predictions of disordered linkers by DFLpred [140], disordered moonlighting regions by DMRpred [141], and disordered regions that interact with proteins and nucleic acids by fMoRFpred [142], DisoRDPbind [143] and ANCHOR2 [86]. While disorder function predictions are outside of the scope of this survey, several recent reviews offer in-depth treatment of this topic [23,144,145]. We highlight the fact that assessment of the methods that predict binding IDRs (i.e, disordered regions that interact with proteins, DNA, RNA and small ligands) was introduced in the CAID experiment [36]. This experiment identified ANCHOR2 [86], DisoRDPbind [143,146,147] and MoRFCHiBi [148] as the top-three most accurate predictors of binding IDRs.

3 Predictive Performance of Intrinsic Disorder Predictors

One of the key aspects of disorder predictors is their predictive performance. These methods produce two types of outputs: numeric propensities that quantifies likelihood that a given residue in the input protein is disordered, and binary value that categorizes this residue as either disordered or structured. The predictive quality of the propensities is typically evaluated with the Area Under receiver operating characteristic Curve (AUC), while Matthews Correlation Coefficient (MCC) is used to assess the binary values [21,23,24,27,33,36,40,42]. We note that these metrics were used in the most recent community assessments [36,40]. The MCC is defined as:

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP) * (TP + FN) * (TN + FP) * (TN + FN)}}$$

where TP (true positive) and TN (true negative) represent the number of correctly predicted disordered and structured residues, respectively, while FP (false positive) and FN (false negative) quantify the number of misclassified structured and disordered residues, respectively.

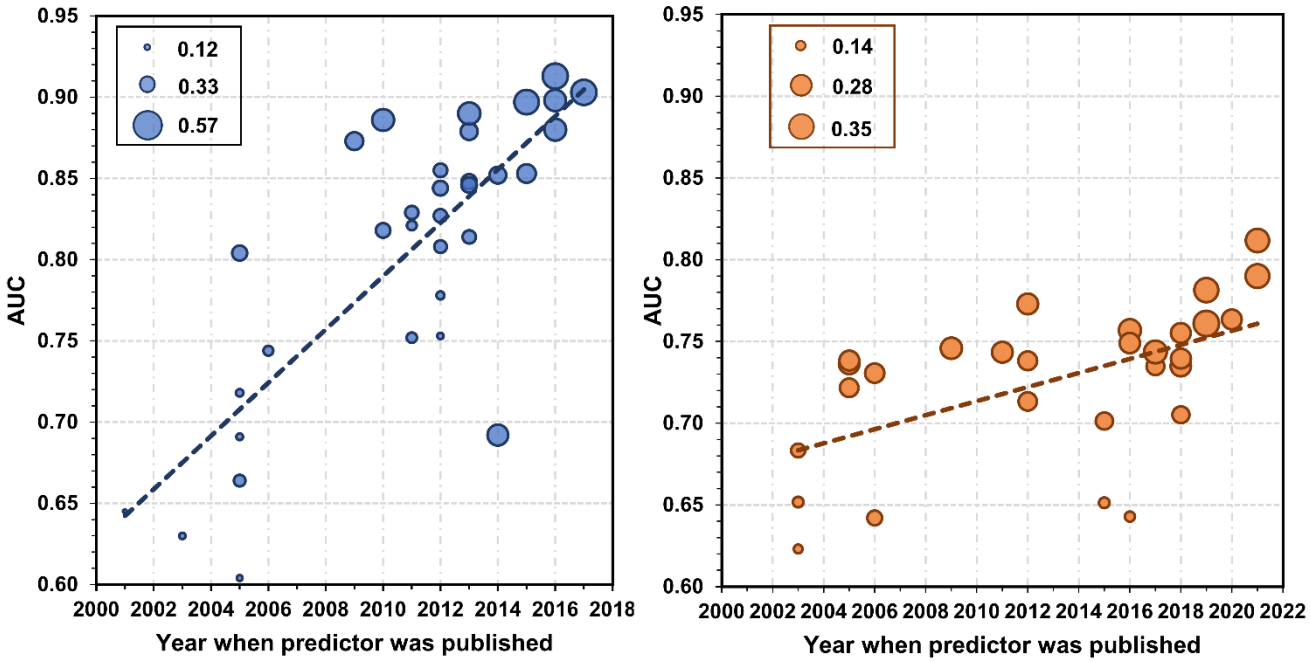

Figure 2. Relation between year of publication and predictive performance measured in CASP10 (panel A in blue) and in CAID (panel B in orange) for disorder predictors. The predictive performance is quantified with AUC (y-axis) and MCC (size of the markers). Figure legends show the size of markers for the minimal, median, and maximal MCC values. The dashed lines correspond to a linear fit into the year vs. AUC data.

We use results produced in the two most recent community assessments, CASP10 [40] and CAID [36], to study whether and how the predictive performance has changed over the years, as newer and more sophisticated predictors were developed. We consider the two assessments since they rely on complementary sources of the ground truth data: CASP10 from PDB [17] while CAID from DisProt [16]. We collect the AUC and MCC values directly from the corresponding CASP10 and CAID articles [36,40] and from other published works that reported results on these benchmark datasets. The numerical values of AUC and MCC are listed in **Supplementary Table S3** for CASP10 and **Supplementary Table S4** for CAID. **Figure 2** visualizes the relation between the predictive performance and the year of publication of the corresponding disorder predictors. Panel 2A focuses on the results of 33 predictors that were assessed on the CASP10 data while panel 2B on the CAID results of 29 disorder predictors. The dashed lines show linear fit that approximates the relation between the AUC values and the year of publication. This analysis suggests that more recently published predictors offer on average higher values of AUC. Furthermore, we observe that larger AUC values are usually accompanied by the larger sizes of markers, which correspond to higher MCC values. We quantify the relations between AUC or MCC values and the year of publication using Pearson correlation coefficients (PCCs). PCCs for the CASP10 dataset results, which we compute using data from **Supplementary Table S1**, are 0.78 for AUC and 0.75 for MCC. The corresponding PCCs for the CAID data (**Supplementary Table S2**) are 0.55 and 0.50. The linear trends combined with the correlations suggest that the predictive performance trends upwards as newer methods are released. These improvements are consistent across both assessments and both measures of predictive quality. The more recent CAID assessment identifies several accurate methods, all of which rely on deep learning [36,132]. They include flDPnn [92], SPOT-Disorder2 [89], RawMSA [88] and AUCpred [123]. We note that flDPnn additionally offers prediction of molecular functions for the putative IDRs that it generates, covering predictions of linkers and IDRs that interact with proteins, DNA and RNA [92]. Moreover, flDPnn produces these predictions very quickly, in a few seconds per proteins, which is at least an order of magnitude faster than the other three methods [36].

4 Expert opinion

Disorder prediction is an active field of research characterized by steady progress fueled by use of recent innovations, such as meta and deep learning, vibrant research community and deep historical roots. We identify 103 disorder predictors were developed over the last four decades. This provides us with ample amount of historical data to identify interesting trends and patterns.

The field of disorder prediction has a strong tradition of community assessments where large collections of methods are evaluated by independent assessors on blind benchmark datasets. Our empirical analysis of results from two most recent community assessments, CASP10 [40] and CAID [36], shows a clear trend of improvements in predictive quality as newer and more advanced predictors are being developed over time. We find that the original focus on machine learning methods has shifted towards the meta-predictors in early 2010s, which was followed by a transition to the deep learning methods in the last 5 years. We anticipate that the current strong focus on the development of deep learning models will continue in a foreseeable future. This is motivated by the recent and rather convincing success of deep learners in the CAID experiment [36]. The entire collection of the top performing methods in that experiment relies on the deep learners [132], including flDPnn that applies deep feed-forward network [92], SPOT-Disorder2 and RawMSA that combine convolutional and recurrent networks [88,89], and AUCpred that utilizes convolutional network [123]. A few other well-performing predictors in this assessment, such as SPOT-Disorder1, SPOT-Disorder-Single and DisoMine, also apply a variety of different deep networks (**Table 1**) [36]. The diversity of the underlying deep network types suggests that there is no one best type of neural network for this predictive problem. Results published in the recent flDPnn article suggest that predictive performance can be strengthened by innovating the predictive inputs of the deep networks [92]. Possible options include development of extended sequences profiles that cover relevant sequence-derived protein characteristics and formulation of protein-level features that express an overall bias of a given input sequence, for instance to be disordered or structured. Innovating both of these aspects, deep network topologies and predictive inputs, is likely to bring further progress in predictive performance.

An often-overlooked factor in this research area is the availability of a broad spectrum of resources that facilitate convenient collection of accurate disorder predictions. We find that about two-thirds of disorder predictors are available as web servers and/or standalone programs. This is a much higher rate of availability then in other related areas, such as prediction of protein-binding and RNA-binding residues where the availability was estimated to be at around 40% [149,150]. This perhaps stems from maturity of the intrinsic disorder prediction field that has over 40 years-long history compared to a much shorter timeframe of the efforts towards the prediction of protein-binding and RNA-binding residues [149,150]. Moreover, users can conveniently obtain disorder and disorder function predictions with the DEPICTER server [139] and collect pre-computed disorder predictions from several large databases, such as MobiDB [137], D^2P^2 [97], and DescribePROT [138]. Availability of this large collection of facilities should ensure that the disorder predictions will continue to make significant impact in related areas of research, such as systems medicine [32], drug design [151-155], and target selection for structural genomics [46,156,157].

While we show that modern disorder predictors offer high-quality results, prediction of disordered binding regions remains a challenge. There are well over a dozen predictors of disordered protein-binding regions [145], including recent methods, such as FLIPPER [158] and SPOT-MoRF [159]. However, recent CAID assessment concludes that "*disordered binding regions remain hard to predict*" and that the corresponding predictors should be substantially improved in the future [36]. This is particularly acute for the prediction of the disordered protein-binding regions [33]. We also highlight a shortage of methods that predict disordered regions which interact with nucleic acids, with only two such methods that were released to date, DisoRDPbind [143,147] and DeepDISOBind [160]. Similarly, there is only one method capable of predicting IDRs that bind lipids, DisoLipPred [161]. This points to the need to develop novel tools that address prediction of different functional flavors of IDRs.

Article highlights

- Over 100 disorder predictors were developed to date
- The original focus on machine learning methods has shifted to meta-predictors in early 2010s, followed by a transition to deep learning predictors since 2016
- Empirical analysis shows that newer and more advanced predictors provide more accurate results when compared to older methods
- The focus on the development of the deep learning predictors will continue in a foreseeable future motivated by the recent and convincing success of these methods in the recent CAID experiment
- Users have access to a broad range of resources that facilitate convenient collection of accurate disorder predictions

Declaration of interest

The authors declare no conflicts of interest.

Funding

This research was funded in part by the National Science Foundation (grant 2125218) and the Robert J. Mattauch Endowment funds to L.K.

References

1. Dunker AK, Babu MM, Barbar E, et al. What's in a name? Why these proteins are intrinsically disordered. Intrinsically Disordered Proteins. 2013 2013/01/01;1(1):e24157.
2. Lieutaud P, Ferron F, Uversky AV, et al. How disordered is my protein and what is its disorder for? A guide through the "dark side" of the protein universe. Intrinsically Disord Proteins. 2016;4(1):e1259708.
3. Oldfield CJ, Uversky VN, Dunker AK, et al. Introduction to intrinsically disordered proteins and regions. Intrinsically Disordered Proteins: Dynamics, Binding, and Function. 2019:1-34.
4. Habchi J, Tompa P, Longhi S, et al. Introducing protein intrinsic disorder. Chem Rev. 2014 Jul 9;114(13):6561-88.
5. Uversky VN, Oldfield CJ, Dunker AK. Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. J Mol Recognit. 2005 Sep-Oct;18(5):343-384.
6. Liu J, Perumal NB, Oldfield CJ, et al. Intrinsic disorder in transcription factors. Biochemistry. 2006 Jun 6;45(22):6873-88.
7. Peng Z, Oldfield CJ, Xue B, et al. A creature with a hundred waggly tails: intrinsically disordered proteins in the ribosome. Cell Mol Life Sci. 2014 Apr;71(8):1477-504.
8. Meng F, Na I, Kurgan L, et al. Compartmentalization and Functionality of Nuclear Disorder: Intrinsic Disorder and Protein-Protein Interactions in Intra-Nuclear Compartments. Int J Mol Sci. 2015 Dec 25;17(1).
9. Zhao B, Katuwawala A, Uversky VN, et al. IDPology of the living cell: intrinsic disorder in the subcellular compartments of the human cell. Cellular and molecular life sciences : CMLS. 2021 Mar;78(5):2371-2385.
10. Babu MM. The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease. Biochem Soc Trans. 2016 Oct 15;44(5):1185-1200.
11. Peng ZL, Mizianty MJ, Xue B, et al. More than just tails: intrinsic disorder in histone proteins. Molecular bioSystems. 2012;8(7):1886-1901.
12. Ward JJ, Sodhi JS, McGuffin LJ, et al. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J Mol Biol. 2004 Mar 26;337(3):635-45.

13. 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.
14. Peng Z, Yan J, Fan X, et al. Exceptionally abundant exceptions: comprehensive characterization of intrinsic disorder in all domains of life. Cell Mol Life Sci. 2015 Jan;72(1):137-51.
15. Peng Z, Mizianty MJ, Kurgan L. Genome-scale prediction of proteins with long intrinsically disordered regions. Proteins. 2014 Jan;82(1):145-58.
16. Hatos A, Hajdu-Soltesz B, Monzon AM, et al. DisProt: intrinsic protein disorder annotation in 2020. Nucleic Acids Res. 2020 Jan 8;48(D1):D269-D276.
17. Le Gall T, Romero PR, Cortese MS, et al. Intrinsic disorder in the Protein Data Bank. J Biomol Struct Dyn. 2007 Feb;24(4):325-42.
18. Fukuchi S, Amemiya T, Sakamoto S, et al. IDEAL in 2014 illustrates interaction networks composed of intrinsically disordered proteins and their binding partners. Nucleic Acids Research. 2014 Jan;42(D1):D320-D325.
19. Schad E, Ficho E, Pancsa R, et al. DIBS: a repository of disordered binding sites mediating interactions with ordered proteins. Bioinformatics. 2018 Feb 1;34(3):535-537.
20. Ficho E, Remenyi I, Simon I, et al. MFIB: a repository of protein complexes with mutual folding induced by binding. Bioinformatics. 2017 Nov 15;33(22):3682-3684.
21. Walsh I, Giollo M, Di Domenico T, et al. Comprehensive large-scale assessment of intrinsic protein disorder. Bioinformatics. 2015 Jan 15;31(2):201-8.
22. UniProt C. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res. 2021 Jan 8;49(D1):D480-D489.
23. Meng F, Uversky VN, Kurgan L. Comprehensive review of methods for prediction of intrinsic disorder and its molecular functions. Cell Mol Life Sci. 2017 Sep;74(17):3069-3090.
24. Liu Y, Wang X, Liu B. A comprehensive review and comparison of existing computational methods for intrinsically disordered protein and region prediction. Brief Bioinform. 2019 Jan 18;20(1):330-346.
25. He B, Wang K, Liu Y, et al. Predicting intrinsic disorder in proteins: an overview. Cell Res. 2009 Aug;19(8):929-49.
26. Meng F, Uversky V, Kurgan L. Computational Prediction of Intrinsic Disorder in Proteins. Curr Protoc Protein Sci. 2017 Apr 3;88:2 16 1-2 16 14.
27. Deng X, Eickholt J, Cheng J. A comprehensive overview of computational protein disorder prediction methods. Mol Biosyst. 2012 Jan;8(1):114-21.
28. Li J, Feng Y, Wang X, et al. An Overview of Predictors for Intrinsically Disordered Proteins over 2010-2014. Int J Mol Sci. 2015 Sep 29;16(10):23446-62.
29. Pryor EE, Jr., Wiener MC. A critical evaluation of in silico methods for detection of membrane protein intrinsic disorder. Biophys J. 2014 Apr 15;106(8):1638-49.
30. Dosztanyi Z, Meszaros B, Simon I. Bioinformatical approaches to characterize intrinsically disordered/unstructured proteins. Brief Bioinform. 2010 Mar;11(2):225-43.
31. Katuwawala A, Oldfield CJ, Kurgan L. Accuracy of protein-level disorder predictions. Brief Bioinform. 2020 Sep 25;21(5):1509-1522.
32. Kurgan L, Li M, Li Y. The Methods and Tools for Intrinsic Disorder Prediction and their Application to Systems Medicine. In: Wolkenhauer O, editor. Systems Medicine. Oxford: Academic Press; 2021. p. 159-169.
33. Katuwawala A, Kurgan L. Comparative Assessment of Intrinsic Disorder Predictions with a Focus on Protein and Nucleic Acid-Binding Proteins. Biomolecules. 2020 Dec 4;10(12).
34. Necci M, Piovesan D, Dosztanyi Z, et al. A comprehensive assessment of long intrinsic protein disorder from the DisProt database. Bioinformatics. 2018 Feb 1;34(3):445-452.
35. Peng ZL, Kurgan L. Comprehensive comparative assessment of in-silico predictors of disordered regions. Curr Protein Pept Sci. 2012 Feb;13(1):6-18.
36. Necci M, Piovesan D, Predictors C, et al. Critical assessment of protein intrinsic disorder prediction. Nat Methods. 2021 May;18(5):472-481.

37. Jin Y, Dunbrack RL, Jr. Assessment of disorder predictions in CASP6. Proteins. 2005;61 Suppl 7:167-75.
38. Bordoli L, Kiefer F, Schwede T. Assessment of disorder predictions in CASP7. Proteins. 2007;69 Suppl 8:129-36.
39. Noivirt-Brik O, Prilusky J, Sussman JL. Assessment of disorder predictions in CASP8. Proteins. 2009;77 Suppl 9:210-6.
40. Monastyrskyy B, Kryshtafovych A, Moult J, et al. Assessment of protein disorder region predictions in CASP10. Proteins. 2014 Feb;82 Suppl 2:127-37.
41. Melamud E, Moult J. Evaluation of disorder predictions in CASP5. Proteins. 2003;53 Suppl 6:561-5.
42. Monastyrskyy B, Fidelis K, Moult J, et al. Evaluation of disorder predictions in CASP9. Proteins. 2011;79 Suppl 10:107-18.
43. Wootton JC. Non-globular domains in protein sequences: automated segmentation using complexity measures. Comput Chem. 1994 Sep;18(3):269-85.
44. Romero, Obradovic, Dunker K. Sequence Data Analysis for Long Disordered Regions Prediction in the Calcineurin Family. Genome Inform Ser Workshop Genome Inform. 1997;8:110-124.
45. Romero P, Obradovic Z, Li X, et al. Sequence complexity of disordered protein. Proteins. 2001 Jan 1;42(1):38-48.
46. Linding R, Jensen LJ, Diella F, et al. Protein disorder prediction: implications for structural proteomics. Structure. 2003 Nov;11(11):1453-9.
47. Linding R, Russell RB, Neduva V, et al. GlobPlot: Exploring protein sequences for globularity and disorder. Nucleic Acids Res. 2003 Jul 1;31(13):3701-8.
48. Liu J, Rost B. NORSp: Predictions of long regions without regular secondary structure. Nucleic Acids Res. 2003 Jul 1;31(13):3833-5.
49. Jones DT, Ward JJ. Prediction of disordered regions in proteins from position specific score matrices. Proteins. 2003;53 Suppl 6:573-8.
50. Hecker J, Yang JY, Cheng JL. Protein disorder prediction at multiple levels of sensitivity and specificity. Bmc Genomics. 2008;9.
51. Cheng JL, Sweredoski MJ, Baldi P. Accurate prediction of protein disordered regions by mining protein structure data. Data Min Knowl Disc. 2005 Nov;11(3):213-222.
52. Prilusky J, Felder CE, Zeev-Ben-Mordehai T, et al. FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. Bioinformatics. 2005 Aug 15;21(16):3435-8.
53. Dosztanyi Z, Csizmok V, Tompa P, et al. The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. J Mol Biol. 2005 Apr 8;347(4):827-39.
54. Dosztanyi Z, Csizmok V, Tompa P, et al. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics. 2005 Aug 15;21(16):3433-4.
55. Peng K, Vucetic S, Radivojac P, et al. Optimizing long intrinsic disorder predictors with protein evolutionary information. J Bioinform Comput Biol. 2005 Feb;3(1):35-60.
56. Yang ZR, Thomson R, McNeil P, et al. RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. Bioinformatics. 2005 Aug 15;21(16):3369-76.
57. Obradovic Z, Peng K, Vucetic S, et al. Exploiting heterogeneous sequence properties improves prediction of protein disorder. Proteins. 2005;61 Suppl 7:176-82.
58. Schlessinger A, Yachdav G, Rost B. PROFbval: predict flexible and rigid residues in proteins. Bioinformatics. 2006 Apr 01;22(7):891-3.
59. Peng K, Radivojac P, Vucetic S, et al. Length-dependent prediction of protein intrinsic disorder. BMC Bioinformatics. 2006;7:208.
60. Schlessinger A, Punta M, Rost B. Natively unstructured regions in proteins identified from contact predictions. Bioinformatics. 2007 Sep 15;23(18):2376-84.

61. Ishida T, Kinoshita K. PrDOS: prediction of disordered protein regions from amino acid sequence. Nucleic Acids Res. 2007 Jul;35(Web Server issue):W460-4.
62. McGuffin LJ. Intrinsic disorder prediction from the analysis of multiple protein fold recognition models. Bioinformatics. 2008 Aug 15;24(16):1798-804.
63. Ishida T, Kinoshita K. Prediction of disordered regions in proteins based on the meta approach. Bioinformatics. 2008 Jun 1;24(11):1344-8.
64. Schlessinger A, Punta M, Yachdav G, et al. Improved disorder prediction by combination of orthogonal approaches. PLoS One. 2009;4(2):e4433.
65. Mizianty MJ, Stach W, Chen K, et al. Improved sequence-based prediction of disordered regions with multilayer fusion of multiple information sources. Bioinformatics. 2010 Sep 15;26(18):i489-96.
66. Xue B, Dunbrack RL, Williams RW, et al. PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. Biochim Biophys Acta. 2010 Apr;1804(4):996-1010.
67. Walsh I, Martin AJ, Di Domenico T, et al. CSpritz: accurate prediction of protein disorder segments with annotation for homology, secondary structure and linear motifs. Nucleic Acids Res. 2011 Jul;39(Web Server issue):W190-6.
68. Roche DB, Buenavista MT, Tetchner SJ, et al. The IntFOLD server: an integrated web resource for protein fold recognition, 3D model quality assessment, intrinsic disorder prediction, domain prediction and ligand binding site prediction. Nucleic Acids Res. 2011 Jul;39(Web Server issue):W171-6.
69. Lobanov MY, Galzitskaya OV. The Ising model for prediction of disordered residues from protein sequence alone. Phys Biol. 2011 Jun;8(3):035004.
70. Walsh I, Martin AJ, Di Domenico T, et al. ESpritz: accurate and fast prediction of protein disorder. Bioinformatics. 2012 Feb 15;28(4):503-9.
71. Kozlowski LP, Bujnicki JM. MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins. BMC Bioinformatics. 2012 May 24;13:111.
72. Kurowski MA, Bujnicki JM. GeneSilico protein structure prediction meta-server. Nucleic Acids Res. 2003 Jul 1;31(13):3305-7.
73. Zhang T, Faraggi E, Xue B, et al. SPINE-D: accurate prediction of short and long disordered regions by a single neural-network based method. J Biomol Struct Dyn. 2012;29(4):799-813.
74. Mizianty MJ, Peng Z, Kurgan L. MFDp2: Accurate predictor of disorder in proteins by fusion of disorder probabilities, content and profiles. Intrinsically Disord Proteins. 2013 Jan-Dec;1(1):e24428.
75. Mizianty MJ, Uversky V, Kurgan L. Prediction of intrinsic disorder in proteins using MFDp2. Methods Mol Biol. 2014;1137:147-62.
76. Huang YJ, Acton TB, Montelione GT. DisMeta: a meta server for construct design and optimization. Methods in molecular biology (Clifton, NJ). 2014;1091:3-16.
77. Fan X, Kurgan L. Accurate prediction of disorder in protein chains with a comprehensive and empirically designed consensus. J Biomol Struct Dyn. 2014;32(3):448-64.
78. Oldfield CJ, Fan X, Wang C, et al. Computational Prediction of Intrinsic Disorder in Protein Sequences with the disCoP Meta-predictor. Methods Mol Biol. 2020;2141:21-35.
79. Cilia E, Pancsa R, Tompa P, et al. From protein sequence to dynamics and disorder with DynaMine. Nat Commun. 2013;4:2741.
80. Cilia E, Pancsa R, Tompa P, et al. The DynaMine webserver: predicting protein dynamics from sequence. Nucleic Acids Res. 2014 Jul;42(Web Server issue):W264-70.
81. Ali H, Urolagin S, Gurarslan O, et al. Performance of protein disorder prediction programs on amino acid substitutions. Hum Mutat. 2014 Jul;35(7):794-804.
82. Jones DT, Cozzetto D. DISOPRED3: precise disordered region predictions with annotated protein-binding activity. Bioinformatics. 2015 Mar 15;31(6):857-63.
83. Iqbal S, Hoque MT. DisPredict: A Predictor of Disordered Protein Using Optimized RBF Kernel. PloS one. 2015;10(10):e0141551.
84. Necci M, Piovesan D, Dosztanyi Z, et al. MobiDB-lite: fast and highly specific consensus prediction of intrinsic disorder in proteins. Bioinformatics. 2017 May 01;33(9):1402-1404.

85. Hanson J, Yang Y, Paliwal K, et al. Improving protein disorder prediction by deep bidirectional long short-term memory recurrent neural networks. Bioinformatics. 2017 Mar 1;33(5):685-692.
86. 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 Jul 2;46(W1):W329-W337.
87. Hanson J, Paliwal K, Zhou Y. Accurate Single-Sequence Prediction of Protein Intrinsic Disorder by an Ensemble of Deep Recurrent and Convolutional Architectures. J Chem Inf Model. 2018 Nov 26;58(11):2369-2376.
88. Mirabello C, Wallner B. rawMSA: End-to-end Deep Learning using raw Multiple Sequence Alignments. PloS one. 2019;14(8):e0220182.
89. Hanson J, Paliwal KK, Litfin T, et al. SPOT-Disorder2: Improved Protein Intrinsic Disorder Prediction by Ensembled Deep Learning. Genomics Proteomics Bioinformatics. 2019 Dec;17(6):645-656.
90. Dass R, Mulder FAA, Nielsen JT. ODiNPred: comprehensive prediction of protein order and disorder. Sci Rep. 2020 Sep 8;10(1):14780.
91. Tang YJ, Pang YH, Liu B. IDP-Seq2Seq: identification of intrinsically disordered regions based on sequence to sequence learning. Bioinformatics. 2021 Jan 29;36(21):5177-5186.
92. Hu G, Katuwawala A, Wang K, et al. flDPnn: Accurate intrinsic disorder prediction with putative propensities of disorder functions. Nature communications. 2021 Jul 21;12(1):4438.
93. Erdos G, Pajkos M, Dosztanyi Z. IUPred3: prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation. Nucleic Acids Res. 2021 Jul 2;49(W1):W297-W303.
94. Liu Y, Wang X, Liu B. RFPR-IDP: reduce the false positive rates for intrinsically disordered protein and region prediction by incorporating both fully ordered proteins and disordered proteins. Briefings in bioinformatics. 2021 Mar 22;22(2):2000-2011.
95. Emenecker RJ, Griffith D, Holehouse AS. Metapredict: a fast, accurate, and easy-to-use predictor of consensus disorder and structure. Biophys J. 2021 Oct 19;120(20):4312-4319.
96. Piovesan D, Necci M, Escobedo N, et al. MobiDB: intrinsically disordered proteins in 2021. Nucleic Acids Res. 2021 Jan 8;49(D1):D361-D367.
97. Oates ME, Romero P, Ishida T, et al. D(2)P(2): database of disordered protein predictions. Nucleic Acids Res. 2013 Jan;41(Database issue):D508-16.
98. Williams RJ. The conformation properties of proteins in solution. Biol Rev Camb Philos Soc. 1979 Nov;54(4):389-437.
99. Romero P, Obradovic Z, Kissinger C, et al. Identifying disordered regions in proteins from amino acid sequence. 1997 Ieee International Conference on Neural Networks, Vols 1-4. 1997:90-95.
100. Liu JF, Rost B. NORSp: predictions of long regions without regular secondary structure. Nucleic Acids Research. 2003 Jul 1;31(13):3833-3835.
101. Bau D, Martin AJ, Mooney C, et al. Distill: a suite of web servers for the prediction of one-, two- and three-dimensional structural features of proteins. BMC Bioinformatics. 2006 Sep 5;7:402.
102. Galzitskaya OV, Garbuzynskiy SA, Lobanov MY. Prediction of natively unfolded regions in protein chain. Mol Biol+. 2006 Mar-Apr;40(2):341-348.
103. Galzitskaya OV, Garbuzynskiy SO, Lobanov MY. FoldUnfold: web server for the prediction of disordered regions in protein chain. Bioinformatics. 2006 Dec 1;22(23):2948-9.
104. Su CT, Chen CY, Ou YY. Protein disorder prediction by condensed PSSM considering propensity for order or disorder. BMC Bioinformatics. 2006 Jun 23;7:319.
105. Vullo A, Bortolami O, Pollastri G, et al. Spritz: a server for the prediction of intrinsically disordered regions in protein sequences using kernel machines. Nucleic Acids Res. 2006 Jul 1;34(Web Server issue):W164-8.
106. Yang MQ, Yang JY. IUP: Intrinsically unstructured protein predictor - A software tool for analyzing polypeptide sequences. Bibe 2006: Sixth Ieee Symposium on Bioinformatics and Bioengineering, Proceedings. 2006:3-+.

107. Hirose S, Shimizu K, Kanai S, et al. POODLE-L: a two-level SVM prediction system for reliably predicting long disordered regions. Bioinformatics. 2007 Aug 15;23(16):2046-53.
108. Shimizu K, Hirose S, Noguchi T. POODLE-S: web application for predicting protein disorder by using physicochemical features and reduced amino acid set of a position-specific scoring matrix. Bioinformatics. 2007 Sep 1;23(17):2337-2338.
109. Shimizu K, Muraoka Y, Hirose S, et al. Predicting mostly disordered proteins by using structure-unknown protein data. BMC bioinformatics. 2007 Mar 6;8.
110. Su CT, Chen CY, Hsu CM. iPDA: integrated protein disorder analyzer. Nucleic Acids Res. 2007 Jul;35(Web Server issue):W465-72.
111. Campen A, Williams RM, Brown CJ, et al. TOP-IDP-scale: a new amino acid scale measuring propensity for intrinsic disorder. Protein Pept Lett. 2008;15(9):956-63.
112. Lieutaud P, Canard B, Longhi S. MeDor: a metaserver for predicting protein disorder. BMC Genomics. 2008 Sep 16;9 Suppl 2:S25.
113. Wang L, Sauer UH. OnD-CRF: predicting order and disorder in proteins using [corrected] conditional random fields. Bioinformatics. 2008 Jun 1;24(11):1401-2.
114. Deng X, Eickholt J, Cheng J. PreDisorder: ab initio sequence-based prediction of protein disordered regions. BMC Bioinformatics. 2009 Dec 21;10:436.
115. Hirose S, Shimizu K, Noguchi T. POODLE-I: disordered region prediction by integrating POODLE series and structural information predictors based on a workflow approach. In Silico Biol. 2010;10(3):185-91.
116. Terashi G, Oosawa M, Nakamura Y, et al. United3D: a protein model quality assessment program that uses two consensus based methods. Chem Pharm Bull (Tokyo). 2012;60(11):1359-65.
117. Becker J, Maes F, Wehenkel L. On the encoding of proteins for disordered regions prediction. PloS one. 2013;8(12):e82252.
118. Eickholt J, Cheng J. DNdisorder: predicting protein disorder using boosting and deep networks. BMC bioinformatics. 2013 Mar 6;14:88.
119. Sormanni P, Camilloni C, Fariselli P, et al. The s2D method: simultaneous sequence-based prediction of the statistical populations of ordered and disordered regions in proteins. Journal of molecular biology. 2015 Feb 27;427(4):982-996.
120. Wang S, Weng S, Ma J, et al. DeepCNF-D: Predicting Protein Order/Disorder Regions by Weighted Deep Convolutional Neural Fields. Int J Mol Sci. 2015 Jul 29;16(8):17315-30.
121. Wang Z, Yang Q, Li T, et al. DisoMCS: Accurately Predicting Protein Intrinsically Disordered Regions Using a Multi-Class Conservative Score Approach. PloS one. 2015;10(6):e0128334.
122. Iqbal S, Hoque MT. Estimation of Position Specific Energy as a Feature of Protein Residues from Sequence Alone for Structural Classification. PloS one. 2016 Sep 2;11(9).
123. Wang S, Ma J, Xu J. AUCpreD: proteome-level protein disorder prediction by AUC-maximized deep convolutional neural fields. Bioinformatics. 2016 Sep 1;32(17):i672-i679.
124. Liu YM, Wang XL, Liu B. IDP-CRF: Intrinsically Disordered Protein/Region Identification Based on Conditional Random Fields. International Journal of Molecular Sciences. 2018 Sep;19(9).
125. Liu Y, Chen S, Wang X, et al. Identification of Intrinsically Disordered Proteins and Regions by Length-Dependent Predictors Based on Conditional Random Fields. Mol Ther Nucleic Acids. 2019 Sep 6;17:396-404.
126. Malysiak-Mrozek B, Baron T, Mrozek D. Spark-IDPP: high-throughput and scalable prediction of intrinsically disordered protein regions with Spark clusters on the Cloud. Cluster Comput. 2019 Jun;22(2):487-508.
127. Cameron M, Williams HE, Cannane A. Improved gapped alignment in BLAST. IEEE/ACM Trans Comput Biol Bioinform. 2004 Jul-Sep;1(3):116-29.
128. Altschul SF, Madden TL, Schaffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997 Sep 1;25(17):3389-402.
129. Han PF, Zhang XZ, Feng ZP. Predicting disordered regions in proteins using the profiles of amino acid indices. Bmc Bioinformatics. 2009 Jan 30;10.

130. Bulashevska A, Eils R. Using Bayesian multinomial classifier to predict whether a given protein sequence is intrinsically disordered. Journal of Theoretical Biology. 2008 Oct 21;254(4):799-803.
131. Su CT, Chen CY, Hsu CM. iPDA: integrated protein disorder analyzer. Nucleic Acids Research. 2007 Jul;35:W465-W472.
132. Lang B, Babu MM. A community effort to bring structure to disorder. Nat Methods. 2021 May;18(5):454-455.
133. Peng Z, Kurgan L. On the complementarity of the consensus-based disorder prediction. Pac Symp Biocomput. 2012:176-87.
134. Torrisi M, Pollastri G, Le Q. Deep learning methods in protein structure prediction. Comput Struct Biotechnol J. 2020;18:1301-1310.
135. Pakhrin SC, Shrestha B, Adhikari B, et al. Deep Learning-Based Advances in Protein Structure Prediction. Int J Mol Sci. 2021 May 24;22(11).
136. Hu G, Kurgan L. Sequence Similarity Searching. Curr Protoc Protein Sci. 2019 Feb;95(1):e71.
137. Piovesan D, Tabaro F, Paladin L, et al. MobiDB 3.0: more annotations for intrinsic disorder, conformational diversity and interactions in proteins. Nucleic Acids Res. 2018 Jan 4;46(D1):D471-D476.
138. Zhao B, Katuwawala A, Oldfield CJ, et al. DescribePROT: database of amino acid-level protein structure and function predictions. Nucleic Acids Res. 2021 Jan 8;49(D1):D298-D308.
139. Barik A, Katuwawala A, Hanson J, et al. DEPICTER: Intrinsic Disorder and Disorder Function Prediction Server. J Mol Biol. 2020 May 15;432(11):3379-3387.
140. Meng F, Kurgan L. DFLpred: High-throughput prediction of disordered flexible linker regions in protein sequences. Bioinformatics. 2016 Jun 15;32(12):i341-i350.
141. Meng F, Kurgan L. High-throughput prediction of disordered moonlighting regions in protein sequences. Proteins. 2018 Oct;86(10):1097-1110.
142. Yan J, Dunker AK, Uversky VN, et al. Molecular recognition features (MoRFs) in three domains of life. Mol Biosyst. 2016 Mar;12(3):697-710.
143. Peng Z, Kurgan L. High-throughput prediction of RNA, DNA and protein binding regions mediated by intrinsic disorder. Nucleic Acids Res. 2015 Oct 15;43(18):e121.
144. Katuwawala A, Ghadermarzi S, Kurgan L. Computational prediction of functions of intrinsically disordered regions. Prog Mol Biol Transl Sci. 2019;166:341-369.
145. Katuwawala A, Peng Z, Yang J, et al. Computational Prediction of MoRFs, Short Disorder-to-order Transitioning Protein Binding Regions. Comput Struct Biotechnol J. 2019;17:454-462.
146. Oldfield CJ, Peng Z, Kurgan L. Disordered RNA-Binding Region Prediction with DisoRDPbind. Methods Mol Biol. 2020;2106:225-239.
147. Peng Z, Wang C, Uversky VN, et al. Prediction of Disordered RNA, DNA, and Protein Binding Regions Using DisoRDPbind. Methods Mol Biol. 2017;1484:187-203.
148. Malhis N, Jacobson M, Gsponer J. MoRFchibi SYSTEM: software tools for the identification of MoRFs in protein sequences. Nucleic Acids Res. 2016 May 12.
149. Zhang J, Kurgan L. Review and comparative assessment of sequence-based predictors of protein-binding residues. Brief Bioinform. 2018 Sep 28;19(5):821-837.
150. Wang K, Hu G, Wu Z, et al. Comprehensive Survey and Comparative Assessment of RNA-Binding Residue Predictions with Analysis by RNA Type. International Journal of Molecular Sciences. 2020;21(18):6879.
151. Hu G, Wu Z, Wang K, et al. Untapped Potential of Disordered Proteins in Current Druggable Human Proteome. Current drug targets. 2016 Jul 22;17(10):1198-205.
152. Hosoya Y, Ohkanda J. Intrinsically Disordered Proteins as Regulators of Transient Biological Processes and as Untapped Drug Targets. Molecules. 2021 Apr 7;26(8).
153. Biesaga M, Frigole-Vivas M, Salvatella X. Intrinsically disordered proteins and biomolecular condensates as drug targets. Curr Opin Chem Biol. 2021 Jun;62:90-100.
154. Ambadipudi S, Zweckstetter M. Targeting intrinsically disordered proteins in rational drug discovery. Expert Opin Drug Discov. 2016;11(1):65-77.

155. Uversky VN. Intrinsically disordered proteins and novel strategies for drug discovery. Expert Opin Drug Discov. 2012 Jun;7(6):475-88.
156. Hu G, Wang K, Song J, et al. Taxonomic Landscape of the Dark Proteomes: Whole-Proteome Scale Interplay Between Structural Darkness, Intrinsic Disorder, and Crystallization Propensity. Proteomics. 2018 Sep 10:e1800243.
157. Oldfield CJ, Xue B, Van YY, et al. Utilization of protein intrinsic disorder knowledge in structural proteomics. Biochim Biophys Acta. 2013 Feb;1834(2):487-98.
158. Monzon AM, Bonato P, Necci M, et al. FLIPPER: Predicting and Characterizing Linear Interacting Peptides in the Protein Data Bank. J Mol Biol. 2021 Feb 27;433(9):166900.
159. Hanson J, Litfin T, Paliwal K, et al. Identifying molecular recognition features in intrinsically disordered regions of proteins by transfer learning. Bioinformatics. 2020 Feb 15;36(4):1107-1113.
160. Zhang F, Zhao B, Shi W, et al. DeepDISOBind: accurate prediction of RNA-, DNA- and protein-binding intrinsically disordered residues with deep multi-task learning. Brief Bioinform. 2021.
161. 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 Sep 6.