Comprehensive review of methods for prediction of intrinsic disorder and its molecular functions

Fanchi Meng[1], Vladimir Uversky[2,3,4], and Lukasz Kurgan[5*]

[1] Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Canada
[2] Department of Molecular Medicine and USF Health Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, Tampa, FL, USA
[3] Institute for Biological Instrumentation, Russian Academy of Sciences, Pushchino, Moscow Region, Russian Federation
[4] Department of Biology, Faculty of Science, King Abdulaziz University, Jeddah, Kingdom of Saudi Arabia
[5] Department of Computer Science, Virginia Commonwealth University, Richmond, USA.

* Corresponding author
Email: lkurgan@vcu.edu
Phone: 804-827-3986

Running title Review: prediction of intrinsic disorder and its functions

Abstract
Computational prediction of intrinsic disorder in protein sequences dates back to late 1970 and has flourished in the last two decades. We provide a brief historical overview and we review over 30 recent predictors of disorder. We are the first to also cover predictors of molecular functions of disorder, including 13 methods that focus on disordered linkers and disordered protein-protein, protein-RNA and protein-DNA binding regions. We overview their predictive models, usability and predictive performance. We highlight newest methods and predictors that offer strong predictive performance measured based on recent comparative assessments. We conclude that the modern predictors are relatively accurate, enjoy widespread use and many of them are fast. Their predictions are conveniently accessible to the end users, via webservers and databases that store pre-computed predictions for millions of proteins. However, research into methods that predict many not yet addressed functions of intrinsic disorder remains an outstanding challenge.

Key words intrinsic disorder; prediction; function of disordered proteins; protein-protein interactions; protein-RNA interactions; protein-DNA interactions; MoRF; SLiM.

Introduction

Intrinsic disorder in proteins manifests as a lack of stable tertiary structure and could be present along the entire protein chain or in specific regions. The corresponding intrinsically disorder proteins (IDPs) and intrinsically disordered regions (IDRs) form dynamic conformational ensembles. In other words, atomic coordinates of their residues and their dihedral angles vary largely over time, without a specific equilibrium [1,2]. IDPs and IDRs were shown to be abundant in nature [3]. According to estimates between 3 and 17% of eukaryotic proteins are fully disordered, depending on an organism [4], and about 30-50% of eukaryotic proteins have at least one long (≥ 30 consecutive residues) IDR [5,4,6,3]. These disordered proteins and regions are crucial for numerous cellular functions including regulation of transcription, translation [7-10] and cell signaling [11-15]. They were shown to be associated with various human diseases [16,17] and are being explored as potential targets for drug discovery [18,19].

Several databases were developed to store experimental annotations of disorder. The first and largest repository of the experimentally verified IDPs and IDRs is DisProt [20-22]. This resource was released over a decade ago, in 2005, by Prof. Dunker's group at the Indiana University. It contains manually curated IDRs together with the annotations of their functions, if available. The latest version 7.03 of DisProt contains 2167 IDRs from 803 protein chains, compared to 290 IDRs from 179 proteins from the earliest release of that database. Another source of experimentally verified IDPs is the IDEAL database [23]. This database was published in 2011, originally with 153 annotated proteins and has grown to 582 proteins in its latest version. While DisProt offers information on a larger set of disordered regions and a more complete set of functional annotations of disorder, IDEAL focuses on the annotation of interaction-driven functions. The latter database includes information on binding partners of IDPs and proteins with IDRs, illustrates them in a context of the protein-protein interaction networks, and includes annotations of domains. The Protein Data Bank (PDB) [24], which is the main source of the protein structures (ordered proteins), can be also used to extract experimental annotations of disorder. IDRs can be found in PDB as the regions that are missing in the X-ray crystal structures of proteins [25,20,26,27] or regions that are associated with high structural variability in the NMR models [28,29]. Although these repositories of the experimental annotations of disorder provide invaluable information to investigate disorder, they represent only a small fraction of sequences in nature.

Motivated by the high levels of abundance and functional importance of IDPs and IDRs, numerous computational methods were developed to predict disorder in protein sequences [30-33]. The predictive models that are used by these methods were computed and benchmarked using the experimental annotations of IDPs and IDRs from the abovementioned databases. These computational predictors are used to efficiently and accurately find disordered proteins and regions for the millions of proteins that lack experimental annotations. Given the large number and diversity of these methods, several relevant reviews and comparative studies were released in the last decade [30,34,32,33,35-40,27,31]. These articles covered most of the prediction methods and some related approaches (e.g. predictors of low complexity regions and flexible residues) dating back to 1994. We cover a similarly comprehensive set of methods including six newly released approaches that were not covered so far, and provide a more complete side-by-side comparison of their availability, usability, architecture and predictive performance. We highlight ten well-performing methods that were selected based on results from several large-scale comparative studies and six most recent methods and describe them in greater depth. We also discuss resources that provide access to predicted annotations of disorder and we are the first to comprehensively review a new group of methods that address prediction of various cellular functions of disordered regions and proteins.

Historical overview

Inspired by ref. [32], the development of predictors of IDPs and IDRs can be divided into three periods: the first generation (1979 to 2001), the second generation (2002 to 2006), and the third generation (2007 onwards).

The first generation predictors were released between 1979 and 2001, and during that time only a few methods were authored. The first method, which aims to predict lack of globular structure, was proposed in 1979 by Williams [41]. This approach was designed to identify proteins that form random coil conformations. However, this methods lacked a proper empirical validation when it was published and a recent evaluation showed that it provides relatively poor predictive performance [32]. The first well-tested IDP predictor was proposed in 1997 by Romero and colleagues [42]. It is based on a neural network model that uses a variety of physiochemical properties of the input protein chain including amino acid compositions, aromaticity, flexibility, hydropathy and hydrophobicity. Another early predictor was proposed by Uversky and co-workers in 2000 by using charge and hydropathy to find disordered proteins [43]. This idea was later implemented in the FoldIndex method [44].

This second generation methods were developed between 2002 and 2006. The defining features of this period are a rapid spike in the development efforts and use of relatively simple predictive models. The second generation methods include approaches that predict intrinsic disorder based solely on propensities/properties of amino acids of the input protein sequences, such as GlobPlot [45] and IUPred[46,47], and methods that utilize popular machine learning models, such as the PONDR family of predictors [48-53], DisEMBL [54] and DISOPRED [55]. One of new developments of this period was the introduction of the evolutionary profiles as the predictive inputs. These profiles are in the form of the position specific score matrix (PSSM) generated with PSI-BLAST. Several second generation methods including PONDR-VL3P [51], DISOPRED2 [6], PROFbval [56], DISpro [57] and NORSp [58], use this new type of the input. This is in contrast to the first generation methods that did not use this information.

The third generation methods were released after 2006. The main characteristics of these methods are the use of new or more sophisticated machine learning model and utilization of meta-predictors. Example methods that take advantage of more complex machine learning models include OnD-CRF [59] that applies a conditional random fields model, DNDisorder [60] that uses deep networks and boosting, and DISOPRED3 [61] that combines three machine learning models: support vector machine, neural network and nearest neighbor. The meta-predictors combine results generated by several individual prediction methods, either via a majority vote consensus or a separate predictive model. The main aim of the meta-predictors is to improve predictive performance when compared to their individual input predictors. Examples meta-methods include CSpritz [62], MetaDisorder [63], MFDp [64], DisMeta [65], and MFDp2 [66]. We also note that a few methods use structural modelling in the prediction, including PrDOS that utilizes structural templates [67] and DISOclust [68] that utilizes structural models.

Predictors of intrinsic disorder

We searched for the disorder predictors using a variety of sources including prior reviews [30,34,32,33,35-37,31], studies that assess and compare predictive performance of these methods [38-40,27], and manual search of PubMed with query “(((disorder[Title]) OR unstructured[Title]) AND prediction[Title]) AND protein”. Among over 70 resulting methods, we consider 32 predictors that are publically available as webservers or/and standalone software, that were published in reputable peer-reviewed scientific venues, and that were released as part of the second or third generation of predictors.

Table 1 summarizes availability and characteristics related to the convenience for the end users of the 32 methods, which are listed in a reverse chronological order. We show whether they are available as webservers, standalone packages or both and provide URLs of these resources. We also indicate whether their webservers accept batch submissions (multiple sequences) and whether their predictions could be considered high-throughput. The latter means that they finish a prediction in short amount of time, typically under 30 seconds per average length sequence. Consequently, these high-throughput methods can be used to perform predictions on a genomic scale. We found 12 such methods. They usually do not use computationally expensive evolutionary information as their input. Nearly half of the predictors (15 out of 32) are available as standalone software. This allows the end users to incorporate these methods into their own computational pipelines. All but one are implemented as webservers, which is convenient for a less computer savvy end users. To use a webserver, these users need just a modern web browser and Internet connection. Moreover, the webservers of five methods accept batch submission, which is useful when a user requires to run a large number of predictions, e.g., when predicting disorder for a particular family of proteins or in a particular proteome. The outputs generated by these methods could be binary (each residue in the input protein chain is classified as either disordered or structured) or numeric (propensity score that quantifies likelihood that a given residue is disordered). We note that all 32 methods output both binary values and propensity scores.

Apart from the availability and usability, we also summarize methodologies that are utilized by the selected 32 methods. Table 2 lists the various types of predictive models and inputs, and divides the predictors into four classes:

1) Scoring function-based methods. They compute propensity of disorder using a scoring function or formula based on selected physiochemical properties of the input amino acids, such as propensity to form structured and disordered regions, certain secondary structures and solvent accessibility. Examples include NORSp [58], GlobPlot [45] and IUPred [46,47].
2) Machine learning-based methods. The propensity for disorder is outputted from a classifier that is generated using a machine learning algorithm. This classifier utilizes the sequence and sequence-derived properties, such as evolutionary conservation, predicted secondary structure, predicted solvent accessibility, as its inputs. Example classifier types include neural network, support vector machine, regression, nearest neighbor, and conditional random field. Predictors in this class include DisEMBL [54], RONN [69], DeepCNF-D [70] and DISOPRED [55,6,71,61].
3) Meta-predictors. These methods use predictions of disorder, in some cases together with other sequence-derived properties, as the inputs to (re)predict disorder. This prediction is computed either via voting, which is typical for methods that use only the prediction of disorder as inputs (e.g, disCoP [72], MetaDisorder [63], metaPrDOS [73], DisMeta [65] and CSpritz [62]), or by using a classifier. The examples of the latter classifier-based consensuses are MD [74], MFDp [64] and MFDp2 [66] that use neural networks (MD) and support vector machines (both versions of MFDp).
4) Structure-based methods. Their predictive models use structural models, either predicted or in a form of structural templates. Examples are PrDOS [67] and Disoclust3 [75].

Majority of the more recent models are either meta-predictors or machine learning-based predictors. The most commonly used classifier in the latter class is the neural network. We also analyze various types of inputs that these methods use including type, physiochemical property or position of amino acids in the input protein sequence (AA), evolutionary conservation (EVO), predicted secondary structure (PSS), predicted solvent accessibility (PSA), and predicted disorder (PDIS). The most commonly used inputs is AA. The EVO input is often used by the machine learning methods. The use of PDIS has started only around 2008 because accurate predictions of disorder has become available at this time. Besides these inputs, some methods utilize other types of information including sequence alignment and predicted disorder content [66], predicted flexibility [64,76], predicted globular domains and torsional angles [64], and predicted residue-residue contacts [77]. The many available methods are diverse in terms of the predictive models and inputs that they use. This fact has motivated the development of the meta-predictors that exploit differences and complementarity between individual predictors to improve predictive performance [78,72].

Table 1. Availability and convenience of the selected 32 publically available disorder predictors. Batch submission refers to ability to submit multiple proteins using the webserver.

Name	Year last published	Ref.	Availability[1]	Batch submission (max # proteins)	High throughput	URL
Disoclust	2015	[79,75]	WS + SP	No		http://www.reading.ac.uk/bioinf/IntFOLD/IntFOLD2_form.html
DISOPRED	2015	[61,55,6,71]	WS + SP	No		http://bioinf.cs.ucl.ac.uk/psipred/?disopred=1
DeepCNF-D	2015	[70]	SP	No	Yes	http://ttic.uchicago.edu/~wangsheng/DeepCNF_D_package_v1.00.tar.gz
DisMeta	2014	[65]	WS	No		http://www-nmr.cabm.rutgers.edu/bioinformatics/disorder/
disCoP	2014	[72]	WS	Yes (up to 5)		http://biomine.cs.vcu.edu/servers/disCoP/
DNDisorder	2013	[60]	WS	No		http://iris.rnet.missouri.edu/dndisorder/
MFDp2	2013	[66]	WS	Yes (up to 100)		http://biomine.cs.vcu.edu/servers/MFDp2/
ESpritz	2012	[80]	WS + SP	Yes (no limit)	Yes	http://protein.bio.unipd.it/espritz/
MetaDisorder	2012	[63]	WS	No		http://iimcb.genesilico.pl/metadisorder/
SPINE-D	2012	[81]	WS + SP	No		http://sparks-lab.org/SPINE-D/
CSpritz	2011	[62]	WS	Yes (no limit)		http://protein.bio.unipd.it/cspritz/
IsUnstruct	2011	[82]	WS	No	Yes	http://bioinfo.protres.ru/IsUnstruct/
MFDp	2010	[64]	WS	Yes (up to 5)		http://biomine.cs.vcu.edu/servers/MFDp
PONDR-FIT	2010	[83]	WS	No	Yes	http://disorder.compbio.iupui.edu/metapredictor.php
MD	2009	[74]	WS + SP	No		https://ppopen.rostlab.org/
PreDisorder	2009	[84]	WS + SP	No		http://sysbio.rnet.missouri.edu/predisorder.html
metaPrDOS	2008	[73]	WS	No		http://prdos.hgc.jp/cgi-bin/meta/top.cgi
OnD-CRF	2008	[59]	WS	No		http://babel.ucmp.umu.se/ond-crf/
Norsnet	2007	[76]	WS + SP	No		https://ppopen.rostlab.org/
Ucon	2007	[77]	WS + SP	No		https://ppopen.rostlab.org/
PrDOS	2007	[67]	WS	No		http://prdos.hgc.jp/cgi-bin/top.cgi
PROFbval	2006	[56]	WS + SP	No		https://ppopen.rostlab.org/
PONDR-VSL2B	2006	[52,53]	WS + SP	No	Yes	http://www.dabi.temple.edu/disprot/predictor.php
FoldUnfold	2006	[85]	WS	No	Yes	http://bioinfo.protres.ru/ogu/
DISpro	2005	[57,86]	WS + SP	No		http://scratch.proteomics.ics.uci.edu/
FoldIndex	2005	[44]	WS	No	Yes	http://bioportal.weizmann.ac.il/fldbin/findex
IUPred	2005	[46,47]	WS + SP	No	Yes	http://iupred.enzim.hu/
RONN	2005	[69]	WS + SP	No	Yes	https://www.strubi.ox.ac.uk/RONN
PONDR-VL3	2005	[50,51]	WS	No	Yes	http://www.dabi.temple.edu/disprot/predictor.php
DisEMBL	2003	[54]	WS + SP	No	Yes	http://dis.embl.de/
GlobPlot	2003	[45]	WS	No	Yes	http://globplot.embl.de/
NORSp	2003	[58]	WS + SP	No		https://ppopen.rostlab.org/

[1] Availability: SP (standalone package); WS (webserver).

Table 2. Architectures of the selected 32 publically available disorder predictors.

Name	Class[1]	Predictive model[2]	Inputs[3]					
			AA	EVO	PSS	PSA	PDIS	Other inputs
Disoclust	SB	SF+consensus					X	Alignment of predicted folds.
DISOPRED	ML	SVM+NN+NNE	X	X				
DeepCNF-D	ML	Deep CNF	X	X	X	X		
DisMeta	Meta	Consensus					X	
disCoP	Meta	Regression					X	
DNDisorder	ML	DN+boosting	X	X	X	X		
MFDp2	Meta	SVM					X	Sequence alignment, predicted disorder content.
ESpritz	ML	NN	X	X				
MetaDisorder	Meta	Consensus			X		X	Predicted folds.
SPINE-D	ML	NN	X	X	X	X		
CSpritz	Meta	Consensus					X	
IsUnstruct	SF	SF						Energetic potential scores.
MFDp	Meta	SVM	X	X	X	X	X	Predicted flexibility, globular domains, and torsional angles.
PONDR-FIT	Meta	Consensus					X	
MD	Meta	NN	X	X	X	X	X	Local sequence profile, sequence complexity.
PreDisorder	ML	NN			X	X		Multiple sequence alignment profile.
metaPrDOS	Meta	Consensus					X	
OnD-CRF	ML	CRF	X		X			
Norsnet	ML	NN	X	X	X	X		Predicted flexibility.
Ucon	ML	NN						Predicted residue-residue contacts.
PrDOS	SB	SVM+templates		X				Structural templates.
PROFbval	ML	NN	X	X	X	X		Chain length.
PONDR-VSL2B	ML	SVM+LR	X					
FoldUnfold	SF	SF	X					
DISpro	ML	NN		X	X	X		
FoldIndex	SF	SF	X					
IUPred	SF	SF	X					Interaction energy.
RONN	ML	NN						Sequence alignment.
PONDR-VL3	ML	NN	X					Sequence complexity.
DisEMBL	ML	NN	X					
GlobPlot	SF	SF	X					
NORSp	SF	SF		X	X	X		Predicted membrane helices, coil-coil regions.

[1]Class: Meta (meta predictor); ML (machine learning-based method); SB (structure-based method); SF (scoring function-based method).

[2]Predictive model: CNF (convolutional neural fields); CRF (conditional random field); DN (deep neural network); LR (logistic regression); SF (scoring function); NN (neural network); NNE: (nearest neighbor); SVM (support vector machine).

[3]Inputs: AA (AA type, property, propensity and/or position); EVO (evolutionary information based on PSSM or HMM profile); PDIS (predicted disorder); PSA (predicted solvent accessibility); PSS (predicted secondary structure).

Predictive performance of predictors of intrinsic disorder

A key aspect of these predictors is their predictive performance, i.e., how well they predict the disordered and structured residues in the input protein sequence. The assessment of predictive performance is performed by comparing predicted disorder to native annotations of disorder for a set of proteins for which the native annotations are known; these proteins are typically dissimilar to the proteins that were used to derive predictors. Since predictions include the numeric propensities of disorder and binary values, they are accordingly accessed using different quality measures. The most widely used metric for the binary predictions is Matthews Correlation Coefficient (MCC), while the predicted propensities are usually evaluated with the Area Under receiver operating characteristic Curve (AUC). These two measures were used in the most recent Critical Assessment of protein Structure Prediction (CASP) experiments: CASP9 [40] and CASP10 [87], and in several recent empirical assessments of the disorder predictions [88,38,27]. The MCC is defined as

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}} \qquad (1)$$

where *TP* is the number of true positives (correctly predicted disordered residues), *FN* is the number of false negatives (native disorder residues predicted as structured residues), *FP* is the number of false positives (native structured residues predicted as disordered residues) and *TN* is the number of true negatives (correctly predicted structured residues). MCC values range between 0 that corresponds to random predictions and 1 for perfect predictions; negative values of MCC would indicate inverse predictions. The AUC is used to evaluate the propensities by considering a range of predictions with varying values of the true positive rates $TPR = TP/(TP + FN)$ and false positive rates $FPR = FP/(FP + TN)$. The propensity score is binarized using a set of thresholds that equal to a set of all unique values of the propensity. The residues associated with propensities above the threshold are assumed to be disordered and with propensities equals or lower than the threshold to be structured, and the corresponding TPR and FPR values that form the ROC curve are computed for each threshold. The area under the ROC curve typically ranges between 0.5 that corresponds to a random predictor and 1 for a perfect predictor.

Several large scale evaluations of the predictive performance of disorder predictors were published over the last quindecennial. Prediction of intrinsic disorder was included in the CASP starting with CASP5 in 2002 [89] and until CASP10 in 2012 [87]. CASP is a biannual event where predictions submitted by a large number of research groups across the world are evaluated on a blind dataset by an independent groups of assessors (the assessors do not participate in the event). The blind datasets typically include about 150 new proteins that could not be used to develop the participating predictors. The inclusion into CASP coincided with the start of the second generation period and was likely the driving factor of the rapid growth in the development of the disorder predictors. The latest CASP event that included prediction of the intrinsic disorder was CASP10 where 26 computer servers (methods that do not include any human intervention) and two human expert groups were evaluated, compared to a much smaller set of 6 groups that participated in CASP5. The two human expert groups in CASP 10 were outperformed by the computer models [87]. The highest MCC achieved in CASP10 was 0.53 and the highest AUC was 0.91, indicating that the modern predictors are characterized by strong predictive performance. One of interesting aspects that was assessed in CASP10 is a relation of the predictive quality and the length of the IDRs. Interestingly, predictions of long IDRs (over 30 consecutive residues in length) were found to be generally characterized by lower predictive performance when compared to the predictions of shorter regions [87]. Such differences in the predictive performance relative to the length of the IDRs motivate the development of methods, such as PONDR-VSL2 [53] and MFDp [64], that aim to improve predictive performance by specifically considering disordered regions that are either long or short. Finally, we note that the prediction of intrinsic disorder in CASP11 was cancelled due to a lack of a sufficient number of suitable protein targets.

Apart from CASP there were three major empirical assessments published in recent years [88,38,27]. The comparative review from 2012 by Peng and Kurgan includes 19 predictors that were tested on a dataset of nearly 500 proteins [38]. The second review by Cheng's group that was released in 2012 included 32 methods that were evaluated on 117 proteins [88]. The most recent study that was published in 2015 by Tosatto's group compared 14 predictors on a large set of 25 thousand proteins [27]. Table 3 summarizes the MCC and AUC values of the 23 out of the considered here 32 methods that were included in at least one of these four studies: CASP10 and the three comparative reviews. We report the best result across multiple versions of ESpritz, DisEMBL, and IUPred methods. The 10 methods that were ranked in the top three based on either the AUC or MCC score in at least one assessment are highlighted with bold font. According to these results, the most accomplished predictors include DISOPRED, MFDp, PONDR-VSL2B, and PrDOS that have secured top 3 finish in two assessments. Several other methods, such as ESpritz, PONDR-FIT, MD, PreDisorder, IUPred, and DisEMBL performed very well in one of the assessments. We observe that the predictive performance depends on the level of sophistication of the underlying predictive models. Typically, more complex models and meta-predictors offer stronger predictive performance but they also require longer runtime to generate the predictions. Examples are DISOPRED that used multiple machine learning models, MFDp and PONDR-VSL2B that are meta-predictors, and PrDOS that combines a modern machine learning model and structural templates. Overall, the AUC values (MCC values) range between 0.73 and 0.85 (MCC was not measured) in [88], 0.70 and 0.82 (0.18 and 0.45) in [38], 0.61 and 0.91 (0.24 and 0.53) in [87], and 0.61 and 0.81 (0.11 and 0.31) in [27]. The lower predictive performance in ref. [27] is attributed to the fact that this assessment included only high-throughput methods which typically trade predictive quality for the computational efficiency. The differences in the predictive quality in different studies stem from the use of different predictors and datasets but in general the range of values is comparable and the top performing methods secure consistently high scores. For example, DISOPRED has secured AUC (MCC) of at least 0.78 (0.41), MFDp at least 0.82 (0.45), and PrDOS at least 0.85 (0.53). We conclude that some of the current predictors of intrinsic disorder consistently provide high quality predictions with AUC > 0.8 and MCC > 0.4.

Table 3. Empirical evaluation of the selected disorder predictors based on results from comparative reviews [38,27,88] and CASP10 [87]. 9 of the 32 considered methods (DeepCNF-D, disCoP, DNDisorder, MFDp2, IsUnstruct, FoldUnfold, DISpro, PONDR-VL3, and NORSp) are not listed since they were not

included in these comparative studies. Methods ranked in the top three based on AUC or MCC and in at least one assessment are highlighted with bold font.

Name	AUC				MCC		
	[88]	[38]	[87]	[27]	[38]	[87]	[27]
Disoclust	0.79	0.78	0.82[1]		0.34	0.24[1]	
DISOPRED	**0.85[2]**	0.78[3]	**0.90[4]**		0.41[3]	**0.53[4]**	
DisMeta			0.69			0.46	
ESpritz			0.86[5]	0.78[7]		0.32[6]	**0.28[8]**
MetaDisorder	0.81[9]		0.84[9]			0.34[9]	
SPINE-D	0.83[10]						
CSpritz			0.83			0.32	
MFDp	0.82[11]	**0.82**	**0.89[12]**		**0.45**	**0.49[12]**	
PONDR-FIT		0.79			**0.42**		
MD		**0.82**			**0.44**		
PreDisorder	**0.85**		0.87[13]			0.40[13]	
metaPrDOS			0.88[14]			0.39[14]	
OnD-CRF	0.73		0.81[15]			0.31[15]	
Norsnet		0.74			0.34		
Ucon		0.74			0.31		
PrDOS	**0.85[16]**		**0.91[17]**			**0.53[17]**	
PROFbval		0.70			0.20		
PONDR-VSL2B		**0.79**		**0.81**	**0.40**		0.26
FoldIndex				0.61	0.28		0.11
IUPred		0.78[18]		**0.78[19]**	0.41[18]		**0.31[19]**
RONN		0.76		0.76	0.37		0.22
DisEMBL				**0.79[20]**	0.32[21]		**0.31[20]**
GlobPlot				0.63	0.18		0.12

[1] under group IntFOLD2; [2] under group DISOPRED3C; [3] result for DISOPRED2; [4] result for DISOPRED3; [5] under group ESpritz; [6] under group ESpritzv2; [7] result for ESpritz X-ray; [8] result for ESpritz NMR; [9] under group GSmetaDisorderMD; [10] under group ZHOU-SPINE-D; [11] under group biomine_DR_pdb; [12] under group biomine_dr_mixed; [13] under group MULTICOM-construct; [14] under group metaprdos2; [15] under group OnD-CRF2; [16] under group Prdos2; [17] under group Prdos-CNF; [18] IUPred for long IDRs; [19] IUPred for short IDRs; [20] DisEMBL-465; [21] DisEMBL-R.

Detailed summary of selected predictors of intrinsic disorder

We provide a detailed and structured summary of several selected methods. These methods include ten methods that secured the top three finish in at least one of the four assessments (methods shown in bold font in Table 3) and six most recent methods that were published after 2012: DNDisorder [60], MFDp2 [66], disCop [72], DisMeta [65], DeepCNF-D [70] and Disoclust3 [75]. We discuss these 16 methods in the chronological order. For each method, we introduce its authors, briefly overview its key architectural characteristics, and provide details about its inputs, outputs and availability.

DISEMBL (2003)

DISEMBL [54] was developed by Linding *et al.* at the European Molecular Biology Laboratory (EMBL). This method includes three predictive models, each implemented as a neural network, that focus on finding disordered residues and residues in disorder-like conformations: loops and coils defined by DSSP, hot loops (loops with high degree of mobility), and disordered residues defined as those that have missing coordinates (i.e., remark465) in the X-ray structures in PDB. The latter version has secured the second highest AUC (0.79) and MCC (0.31) in the recent assessment of disorder predictors by the Tosatto's group [27]. DISEMBL is a high throughput method and its predictions (based on the remark465 and hot loop versions) are included in the MobiDB database [90].
Input: SwissProt ID or a single raw (unformatted) amino acid sequence.
Output: Predicted propensities for disorder for each residue in the input sequence for each of the three models, formatted as plain text and in the CASP format (column-wise with the first column showing the amino acids, second showing the binary predictions and the third giving the propensities); a plot representing the propensity scores of being disorder for the three models. Binary prediction for each residue in the input sequence for each of the three models.
Availability: A webserver and a standalone package running on a Linux platform.
URL: http://dis.embl.de

IUPred (2005)

IUPred [47,46] was authored by Dosztányi *et al.* at the Hungarian Academy of Sciences. This predictor finds putative intrinsically disordered residues and regions using a scoring functions that estimates energy of inter-residue interactions and the fact that such energy differs between structured and unstructured regions. IUPred has two versions: short and long. The former was designed to predict missing residues in the X-ray structures while the long version was optimized to predict functionally relevant disordered segments. Although the underlying scoring function is relatively simple, this method offers good predictive performance and is very fast to compute. Based on the recent comparative review by the Tosatto's group [27], the version of IUPred that targets short regions secures third highest AUC (0.78) and the highest MCC (0.31) among the considered 14 high-throughput predictors. The predictions by both version of IUPred are included in the D^2P^2 database [91] and the MobiDB database [90].
Input: SwissProt ID, TrEMBL ID, or a single raw (unformatted) or FASTA formatted amino acid sequence.
Output: Predicted propensities for disorder for each residue in the input sequence for each of the two models; scores above 0.5 indicate that the corresponding residues is predicted as disordered. IUPred can also output plots that show structured regions and propensities of disorder for long and short regions.
Availability: A webserver and a standalone package running on a Linux platform.
URL: http://iupred.enzim.hu

PONDR-VSL2B (2006)

PONDR-VSL2B [53,52] was released by Obradovic *et al.* at the Temple University. This method is a part of a larger family of PONDR predictors of disorder. The VSL (various short long) suffix stands for fact that this method was built using disorder characterized by Various approaches (X-ray crystallography, NMR and circular dichroism) and to predict both Short and Long disordered regions. The 2B recognizes the fact that this is second of the two models, this one is based on SVM, while B indicates that this is a Baseline predictor that utilizes only the information derived from amino acid composition [53]. In contract, VSL2 (without "B") utilizes information derived from amino acid composition, PSSM generated with PSI-BLAST and predicted secondary structure. PONDR-VSL2B is much faster than PONDR-VSL2 (computations of PSSM and secondary structure are time consuming) and according to the test performed by the authors its accuracy is only 3% inferior to VSL2 [53]. In the evaluation by Kurgan's group [38], PONDR-VSL2B achieved the third highest AUC (0.79). Predictions by this method are included the D^2P^2 database [91]. Moreover, this is the first method that predicts IDRs of various length with similar predictive quality.
Input: A single raw (unformatted) amino acid sequence (up to 100 predictions per IP address per day; query sequence limited to up to 5000 residues).
Output: Predicted propensities and binary scores for each residue in the input sequence.
Availability: A webserver and standalone version running on a Linux platform.
URL: http://www.dabi.temple.edu/disprot/predictorVSL2.php

PrDOS (2007)

PrDOS [67] was developed by Ishida and Kinoshita at the University of Tokyo. This is a hybrid design that combines a machine learning model with a template-based approach. PrDOS uses an SVM model that takes the PSMM generated with PSI-BLAST run on the input protein chain as the input. The output by the SVM model is combined with results of a search for homologues in PDB. The final propensity for the intrinsic disorder is computed as a weighted average of the results from SVM and the homology search. This predictor offers one of the highest levels of predictive performance. PrDOS has secured the highest AUC (0.85) in the assessment by Cheng's group [88], and its new version based on conditional neural field has achieved the highest AUC (0.91) and the second highest MCC (0.53) in CASP10 [87]. However, the conditional neural field version is not available publically. Although PrDOS is not a high throughput method (it takes >1 minute to run it for a sequence), its predictions over multiple genomes are included in the D^2P^2 database [91].
Input: A single raw (unformatted) or FASTA-formatted amino acid sequence
Output: Predicted propensities and binary scores for each residue in the input sequence; a plot of the propensity scores.
Availability: A webserver.
URL: http://prdos.hgc.jp/cgi-bin/top.cgi

PreDisorder (2009)

PreDisorder [92-94] was created by Cheng *et al.* at the University of Missouri. This is a machine learning model based on a recursive neural network. The network utilizes a diverse set of inputs derived from the input sequence including multiple sequence alignment profiles, predicted secondary structure and predicted solvent accessibility. PreDisorder obtained the second highest AUC (0.82) in the comparative evaluation published in 2012 by the same group [88] and also performed well in the CASP8 experiment [95].

Input: A single raw (unformatted) amino acid sequence.
Output: Predicted propensities and binary scores for each residue in the input sequence.
Availability: A webserver.
URL: http://sysbio.rnet.missouri.edu/predisorder.html

MD (2009)

MD [74] is a meta predictor that was developed by Schlessinger *et al.* the Columbia University. This method relies on a neural network model that utilizes a large set of diverse inputs that are derived from the input protein chain. These inputs are divided into two types: 1) the outputs from four disorder predictors including NORSnet [76], IUPred [46], DISOPRED2 [71], and UCon [77]; and 2) other sequence-derived information including flexibility predicted with PROFbval [56], predicted secondary structure and solvent accessibility, amino acid composition, annotation of low complexity regions, sequence profiles, sequence length, estimated hydrophobicity and net-charge of the input protein, and estimated sequence energy. MD is tied with MFDp [64] for the highest AUC (0.82) in empirical evaluation by the Kurgan's group [38], and achieved the second highest MCC (0.44) in the same evaluation. This method is a part of a comprehensive PredictProtein Open platform [96] for the prediction of protein structure and function. PredictProtein Open offers predictions of intrinsic disorder and flexibility, disulphide bridges, effects of point mutations, gene ontology terms (functions), subcellular localization and binding sites.
Input: A single raw (unformatted) or FASTA-formatted amino acid sequence
Output: Predicted propensities and binary scores for each residue in the input sequence.
Availability: integrated into the PredictProtein Open webserver.
URL: https://ppopen.rostlab.org/

MFDp (2010)

MFDp [64] is a meta predictor designed by Mizianty *et al.* the University of Alberta. This method combines three SVMs that were trained to predict short IDRs (<30 consecutive residues), long IDRs (30 or more consecutive residues) and IDRs of all length. Each of the three SVMs uses a rich set of inputs that are categorized into two types: 1) the outputs from three predictors of intrinsic disorder, IUPred [46], DISOPRED2 [71] and DISOclust [68]; and 2) other sequence-derived information including the input sequence, PSSM profiles generated with PSI-BLAST, flexibility predicted with PROFbval [56], secondary structure predicted with PSIPRED [97], solvent accessibility and backbone dihedral torsion angles predicted with Real-SPINE3 [98], and globular domains predicted with IUPred. This method was shown to provide high levels of predictive performance. MFDp is tied with MD [74] for the highest AUC (0.82) in the empirical evaluation by the same group [38], and secured the highest MCC (0.45) in the same evaluation. Is also obtained third best AUC and MCC in CASP10 [87] and second best MCC in CASP9 [40].
Input: A single or multiple (≤ 5 sequences) FASTA-formatted amino acid sequence(s).
Output: Predicted propensities and binary scores for each residue in the input sequence.
Availability: A webserver.
URL: http://biomine.cs.vcu.edu/servers/MFDp

PONDR-FIT (2010)

PONDR-FIT [83] is a meta predictor that was authored by Xue *et al.* at the Indiana University. It combines outputs of six predictors of intrinsic disorder: PONDR-VLXT [48], PONDR-VSL2 [52,53], PONDR-VL3 [50], FoldIndex [44], IUPred [46] and TopIDP [99], using a neural network. The predictions from this

method are provided together with the results from the other PONDR methods: PONDR-VSL2B, PONDR-VL3 and PONDR-VLXT. PONDR-FIT achieved the third highest MCC (0.42) in the comparative evaluation from 2012 by the Kurgan's group [38].
Input: A single FASTA-formatted or EMBL-formatted amino acid sequence.
Output: Predicted propensities and binary scores for each residue in the input sequence; a plot of propensity scores where residues above the 0.5 cut-off are predicted as disordered.
Availability: A webserver.
URL: http://disorder.compbio.iupui.edu/metapredictor.php

ESpritz (2012)

ESpritz [80] was developed by Walsh *et al.* at the University of Padua. This machine learning predictor is based on a bidirectional recursive neural work. ESpritz has three versions that were trained using different sources of annotations of disorder based on X-ray crystals, NMR and the DisProt database. The NMR-based version of ESpritz secured the third highest MCC (0.28) in the evaluation from 2015 by the same group [27]. Each of the three versions has an option to be executed without the use of computationally expensive evolutionary profiles, which results in a very fast runtime (typically < 10s per sequence). Predictions generated by each of the three versions of ESpritz are included the D^2P^2 database [91] and the MobiDB database [90].
Input: A single or multiple FASTA-formatted amino acid sequence(s); the submission limit is less than 3000 proteins when pasted into online page; no limit if the proteins are uploaded in a file.
Output: Predicted propensities and binary scores for each residue in the input sequence; summary of disorder for input protein(s).
Availability: A webserver and a standalone package running on a Linux platform.
URL: http://protein.bio.unipd.it/espritz

DNDisorder (2013)

DNDisorder [60] was created by Eickholt and Cheng at the University of Missouri; the same research group also developed PreDisorder. The architecture of DNDisorder is an ensemble of deep neural networks and this is the first predictor that applied this type of a machine learning model. The inputs to these networks include information extracted from PSSM derived with PSI-BLAST, predicted solvent accessibility, predicted secondary structure, and Atchley factors [100]. The Atchley factors are five numeral values that quantify secondary structure, polarity, volume, codon diversity and electrostatic charge of amino acids. This predictor achieved relatively good AUC of 0.83 and 0.85 in CASP9 and CASP10, respectively [60].
Input: A single raw (unformatted) or FASTA-formatted amino acid sequence.
Output: Predicted propensities and binary scores for each residue in the input sequence.
Availability: A webserver.
URL: http://iris.rnet.missouri.edu/dndisorder

MFDp2 (2013)

MFDp2 [66] is a meta method that was designed by Mizianty *et al.* at the University of Alberta; the same research group that developed MFDp. It utilizes a novel architecture that includes three major components: disorder predictor MFDp [64], predictor of disordered content (i.e., overall amount of disorder in a whole protein) DisCon [101], and an alignment engine. DisCon was empirically shown to predict the disorder content more accurately than MFDp and several other disorder predictors [66,101]. The idea behind MFDp2 is to combine the predictions from MFDp with predictions using alignment against a database of

annotated disordered proteins and adjust these results so that they agree with the disorder content predicted with DisCon. MFDp2 was empirically shown in [66] to achieve relatively high AUC (0.86) and MCC (0.48) values on a benchmark dataset with 105 proteins.
Input: A single or multiple (up to 100) FASTA-formatted amino acid sequence(s).
Output: Predicted propensities and binary scores for each residue in the input sequence from MFDp2 and MFDp; disorder content predicted with DisCon; evolutionary conservation, secondary structure predicted with PSIPRED [97], and solvent accessibility predicted with Real-SPINE3[98] for each residue in the input sequence.
Availability: A webserver.
URL: http://biomine.cs.vcu.edu/servers/MFDp2

disCop (2014)

DisCop [72] is a meta predictor by Fan and Kurgan at the University of Alberta; this research group also developed MFDp and MFDp2. The defining feature of this meta method is that its input disorder predictors were selected empirically from a large set of 20 disorder predictors to maximize predictive performance. The selected seven methods include ESpritz (the DisProt and X-ray versions), CSpritz (the long disorder version), SPINE-D, DISOPRED2, MD and DISOclust. Their outputs are combined together using a regression model to produce a new disorder prediction that offers higher predictive performance compared to any of the 20 predictors [72]. DisCop was shown to achieve high values of AUC = 0.85 and MCC = 0.50 on a benchmark dataset with over 240 proteins [72].
Input: A single or multiple (up to 5) FASTA-formatted amino acid sequence(s).
Output: Predicted propensities and binary scores for each residue in the input sequence from disCop and MFDp; disorder content predicted with DisCon; evolutionary conservation, secondary structure predicted with PSIPRED [97], and solvent accessibility predicted with Real-SPINE3[98] for each residue in the input sequence
Availability: A webserver.
URL: http://biomine.cs.vcu.edu/servers/disCoP

DisMeta (2014)

DisMeta [65] is a meta predictor that was released by Huang *et al.* at the Rutgers University. This method implements consensus of eight disorder predictors: DISEMBL [54], DISOPRED2 [71], DISpro [57], FoldIndex [44], GlobPlot2 [45], IUPred [46], RONN [69] and PONDR-VSL2 [52,53]. A user can also select to generate the consensus prediction using a subset of these methods. To the best of our knowledge, DisMeta was not empirically evaluated neither by the authors in the corresponding publication or in other studies. This method has been used to select and prepare proteins for NMR and crystallization studies at the Northeastern Structural Genomic Consortium (NESG) [65].
Input: A single raw (unformatted) amino acid sequence or the NESG target ID.
Output: Predicted disordered residues from each selected input predictor and the consensus score for each residue in the input sequence; predicted secondary structure with PROFsec [102] and PSIPRED [97]; predicted secretion signal peptides with SignalP [103]; predicted transmembrane regions with TMHMM [104]; predicted low complexity regions with SEG [105]; predicted disordered protein-binding residues with ANCHOR [106].
Availability: A webserver.
URL: http://www-nmr.cabm.rutgers.edu/bioinformatics/disorder

DeepCNF-D (2015)

DeepCNF-D [70] was created by Wang *et al.* at the University of Chicago. This method utilizes weighted deep convolutional neural fields (CNF) as the machine learning model. This model uses physiochemical properties of the input amino acids and sequence-derived evolutionary information, predicted secondary structure and predicted solvent accessibility as its inputs. DeepCNF-D was evaluated by the authors on the CASP9 and CASP10 datasets and achieved relatively high AUC values (0.86 and 0.90) and MCC values (0.49 and 0.47) [70]; we note that this was done after the CASP experiments were concluded. This predictor has a high-throughput version that uses only the properties of amino acids as the input; the AUCs of that version are lower at 0.70 and 0.77 and MCCs at 0.40 and 0.43 when tested on the CASP9 and CASP10 datasets, respectively [70].
Input: A single FASTA-formatted amino acid sequence.
Output: Predicted propensities and binary scores for each residue in the input sequence.
Availability: A standalone software running on a Linux platform.
URL: http://ttic.uchicago.edu/~wangsheng/DeepCNF_D_package_v1.00.tar.gz

DISOPRED3 (2015)

DISOPRED3 [61] was authored by Jones and Cozzetto at the University College London. The first version of this method (DISOPRED) was published in 2003 [55], the second (DISOPRED2) around 2004 [6,71] and the latest third version in 2015 [61]. DISOPRED3 has a two layer design. The first layer uses three models to predict disorder which are next combined together in the second layer with a help of a neural network. The three models in the first layer include the SVM model from DISOPRED2, a new neural network model that aims to predict long disordered regions, and a nearest neighbor model that is used to predict disorder using a reference dataset of proteins annotated with disorder. Moreover, DISOPRED3 also predicts disordered protein-binding sites using an SVM-based model. This method provides very accurate predictions. It has secured the second highest AUC (0.90) and the highest MCC (0.53) in CASP10 [87]. DISOPRED3 and DISOPRED2 are now embedded into the PSIPRED platform [107] that also provides predictions of protein structure, membrane helices and topology of transmembrane helices, protein domains, and protein functions.
Input: A single raw (unformatted) or FASTA-formatted amino acid sequence, or the multiple sequence alignment of the input protein.
Output: Predicted propensities and binary scores for each residue in the input sequence; predicted binary scores and propensities for the disordered protein binding sites for each residue in the input sequence.
Availability: A webserver and a standalone software running on a Linux platform.
URL: http://bioinf.cs.ucl.ac.uk/psipred/?disopred=1

Disoclust3 (2015)

Disoclust3 [75] was released McGuffin *et al.* at the University of Reading. The first version of this method was published in 2008 [68]. This predictor is based on a premise that structured residues are conserved in three-dimensional space across multiple structural models, while the residues that vary in position or are missing across these models are likely to be disordered. Disoclust3 used ModFOLDclust2 [108] to identify residues with highly variable positions over multiple alternative structural models that are computed with the IntFOLD3-TS method. The results from the above approach are combined with the results generated with DISOPRED3 [61] to generate the final prediction. Disclust3 achieved AUC = 0.82 and MCC = 0.24 in CASP10 [87]. This predictor is embedded into the IntFOLD platform [75] that also provides prediction of tertiary structure, domain, binding sites and offers model quality assessment scores.

Input: A single raw (unformatted) amino acid sequence.
Output: Predicted propensities and binary scores for each residue in the input sequence.
Availability: A webserver and a standalone software package (Java environment required).
URL: http://www.reading.ac.uk/bioinf/IntFOLD/IntFOLD2_form.html

Databases of putative annotations of intrinsic disorder

Recent comparative reviews reveal that predictors of intrinsic disorder are relatively accurate. These predictions are used to guide experimental studies of disorder and to address practical problems in other areas, such as targets selection in structural genomics [109]. They were also used to analyze prevalence and functional characteristics of disorder on large scale across functionally related proteins [110,9,111] and in whole proteomes [3,19,112-114]. To this end, several databases of the putative annotations of IDPs and IDRs were developed to ease access to this information for the end users. Given that these resources provide access to putative disorder for large sets of proteins, they include results generated by high-throughput predictors of intrinsic disorder.

DICHOT [115] is the first such database. It provides predictions of intrinsic disorder for the human proteome. It includes 20,333 protein chains collected from the Swiss-Prot database [116], where the IDRs are predicted using DISOPRED2 [6] and CLADIST [117]. This resource is now superseded by the two more recent and much larger databases: MobiDB [90,118] and D^2P^2 [91]. MobiDB offers access the putative disorder generated by ten predictors: three versions of Espritz [80], two versions of IUPred [46], two versions of DisEMBL [54], GlobPlot [45], PONDR-VSL2b [52] and RONN [69]. The database also combines these 10 predictions into a consensus. Moreover, MobiDB includes experimental annotations of disorder collected from DisProt and PDB, the latter based on both X-ray and NMR structures. The current version 2.2 of MobiDB (version 2.0) covers over 80.37 million chains, which were obtained from the UniProtKB and Swiss-Prot resources [116]. Importantly, these putative annotations of disorder are also cross-referenced in UniProt [116]. D^2P^2 is the second large repository of predicted annotations of intrinsic disorder. It contains annotations generated with nine predictors: three versions of Espritz [80], two versions of IUPred [46], PV2 [119], PrDOS [67], PONDR-VSL2b [52] and PONDR-VLXT [48]. It also links to the experimental annotations of disorder from DisProt and IDEAL and includes putative annotations of disordered protein binding regions computed with ANCHOR [120,106]. The current version of D^2P^2 contains annotations for 10.43 million proteins from 1,765 proteomes across all kingdoms of life. The main difference between MobiDB and D^2P^2 is that the former provides annotations for arguably largest possible set of currently known proteins, while the latter provides the annotations for all complete proteomes. Both MobiDB and D^2P^2 include a number of secondary annotations to put the putative disorder in the structural and functional context. For example, MobiDB includes information about organism a given protein comes from, subcellular location, annotations of functions, post-translational modifications, domains, secondary structure, and protein interactions. D^2P^2 includes the source organism and a comprehensive annotation of domains and post-translational modifications.

Predictors of functions of intrinsic disorder

IDPs and IDRs are involved in a wide repertoire of cellular functions. In recent years progress has been made to develop methods that predict these functions from the protein sequences. In contrast to the predictors of intrinsic disorder, these methods find a subset of IDRs that carry out a specific function. The current predictors of functions of disorder address primarily binding-related functions that include interactions of IDRs with proteins, DNAs and RNAs. This is motivated by an observation that these binding-related functions are the most prevalent functions carried out by IDRs. Based on the experimental data from DisProt, 74% of the over 1000 functionally annotated IDRs in DisProt interact with proteins, DNAs, RNAs, metals and lipids. The protein-protein binding is the most populated function, with over 450 annotated IDRs in DisProt.

The predictors of the most populated disordered protein binding regions are categorized into three classes. The first class are the methods that predict generic disordered protein binding regions which include ANCHOR [120,106] and disoRDPbind [121]. The second class focuses on a specific type of protein binding regions called molecular recognition features (MoRFs). MoRFs are protein binding regions located within IDRs that include at least five consecutive residues and which undergo disorder-to-order transitions upon binding to their protein partners [122,123]. There are several predictors of MoRFs including alpha-MoRFpred [124,125], MoRFpred [126], MFSPSSMpred [127], MoRFChiBi [128], MoRFChiBiWeb [129], fMoRFpred [130], retro-MoRF [131] and DISOPRED3 [61]. The third category of methods aims to predict short linear sequence motifs (SLiMs). SLiMs are conserved in the sequence and their length typically ranges between 3 and 10 consecutive amino acids [132]. They mediate protein-protein interactions and although they are primarily disordered, about 20% of them are located in globular protein domains [129]. The currently experimentally annotated SLiMs can be obtained from the Eukaryotic Linear Motif (ELM) resource [133] and they can be predicted with the help of the SLiMpred [134] and PepBindPred [135] methods.

So far only one predictor, disoRDPbind [121], which considers IDRs that bind to other types of ligands was developed. This method combines three predictive models that provide putative annotations of the disordered protein-, DNA- and RNA-binding residues. Just recently, the first method that addresses prediction of a function of intrinsic disorder that is not related to binding was released. The DFLpred method [136] predicts disordered flexible linker regions, elements that serve as linkers/spacers in multi-domain proteins or between structured constituents within protein domains. The disordered flexible linkers differ from linkers in three aspects. They are characterized by lack of defined structure, are longer (avg length of 25 residues) and could be localized both within and between domains, for instance to link structured elements within a domain. These linkers constitute the most populated in DisProt type of the non-binding function of IDRs that accounts for about 9% of all functionally annotated disordered regions.

Table 4 lists summarizes availability and features related to the use user convenience of the abovementioned 13 predictors of cellular functions of disorder, which are listed in the reverse chronological order. Most of these methods, except for alpha-MoRFpred and retro-MoRF, are provided to the end users as convenient to use webservers. Moreover, five methods: ANCHOR, MFSPSSMpred, MoRFCHiBi, MoRFCHiBiWeb and DISOPRED3 are available as standalone packages. This option is useful for users who would want to include them in other predictive pipelines. The table also indicates whether the webservers accept batch submissions (i.e., multiple sequences in a single request) and whether their predictions are high-throughput (they are computed quickly, typically in under 30 seconds, for an average length sequence). Several predictors, such as ANCHOR, disoRDPbind, MoRFChiBi, fMoRFpred, and DFLpred, are very fast and can be used to perform predictions on the whole proteome scale. Four methods are that available online, including MoRFpred, DisoRDPbind, fMoRFpred, and DFLpred, offer an option to perform batch predictions to facilitate large-scale applications over protein families or whole proteomes. Moreover, predictions from ANCHOR for over 10 million proteins are already included in the D^2P^2 database.

Table 5 discusses architectures of the 13 predictors. Similar to the predictors of intrinsic disorder, it divides these models into four classes based on the predictive models and inputs that they use:

1) Scoring function-based methods. These approaches input properties computed directly from the protein sequence, such as sequence alignment and propensity for intra-chain interactions and binding, as well as the propensity for intrinsic disorder into a scoring function to predict disordered protein binding regions. The two methods in this category are retro-MoRF [131] and ANCHOR [120,106]
2) Machine learning-based methods. This the largest by far category includes nine methods: alpha-MoRFpred [124,125], SLiMpred method [134], MoRFpred [126], MFSPSSMpred [127], disoRDPbind [121], DISOPRED3 [61], fMoRFpred [130], MoRFChiBi [128] and DFLpred [136]. They compute propensity for a specific function utilizing machine learning classifiers. Inputs for these classifiers are generated directly from the sequence and from the sequence-derived properties, such as evolutionary conservation, putative secondary structure and putative solvent accessibility. While these architectural details are similar to the predictors of intrinsic disorder, these methods also frequently use multiple sequence alignment and putative annotations of disordered residues. The machine learning-based methods predict protein, RNA and DNA binding regions as well as the disordered flexible linkers.
3) Meta-predictors which include the MoRFChiBiWeb method [129]. This predictor uses sequence alignment and a Bayesian approach to combine MoRFChiBi, Espritz, and sequence conservation profiles to (re)predict MoRF regions. Benchmarks performed by the authors of MoRFChiBiWeb reveal that it is more accurate than MoRFChiBi but it also requires longer runtime.
4) Structure-based methods that include PepBindPred [135]. This method relies on the structure of the protein that binds to a disordered region to generate molecular docking scores that are processed with a machine learning model to predict SLiMs.

Majority of the predictors of the cellular functions of disorder are based on machine learning models. In contrast to the predictors of intrinsic disorder that primarily use neural networks, these methods most often adopt machine learning models in the form of support vector machines. Moreover, nine out of the 13 methods use putative annotations of intrinsic disorder as one of their inputs, which is motivated by the fact that these functional regions are located in IDRs.

Table 4. Availability, convenience, and architecture of the 12 predictors of functions of disorder. Batch submission refers to ability to submit multiple proteins using a webserver.

Method	Year last published	Ref.	Predictive target	Availability[1]	Batch submission	High throughput	URL
DFLpred	2016	[136]	linkers	WS	Yes	Yes	http://biomine.cs.vcu.edu/servers/DFLpred/
MoRFCHiBiWeb	2016	[129]	protein binding	WS + SP	No		http://morf.chibi.ubc.ca:8080/mcw/index.xhtml
fMoRFpred	2015	[130]	protein binding	WS	Yes	Yes	http://biomine.cs.vcu.edu/servers/fMoRFpred/
DISOPRED3	2015	[61]	protein binding	WS + SP	No		http://bioinf.cs.ucl.ac.uk/disopred
MoRFCHiBi	2015	[128]	protein binding	WS + SP	No	Yes	http://morf.chibi.ubc.ca:8080/mcw/index.xhtml
disoRDPbind	2015	[121]	protein, RNA, DNA binding	WS	Yes	Yes	http://biomine.cs.vcu.edu/servers/DisoRDPbind/
PepBindPred	2013	[135]	protein binding	WS	No		http://bioware.ucd.ie/~compass/biowareweb/Server_pages/pepbindpred.php
MFSPSSMpred	2013	[127]	protein binding	WS + SP	No		http://webapp.yama.info.waseda.ac.jp/fang/MoRFs.php
MoRFpred	2012	[126]	protein binding	WS	Yes		http://biomine.cs.vcu.edu/servers/MoRFpred/
SLiMPred	2012	[134]	protein-binding	WS	No		http://bioware.ucd.ie/~compass/biowareweb//Server_pages/slimpred.php
retro-MoRFs	2010	[131]	protein-binding	N/A	N/A	N/A	N/A
ANCHOR	2009	[120,106]	protein binding	WS + SP	No	Yes	http://anchor.enzim.hu
alpha-MoRFpred	2007	[124,125]	protein binding	N/A	N/A	N/A	N/A

[1] Availability: WS (webserver); SP (standalone package).

Table 5. Architectures of the 12 predictors of functions of disorder.

Name	Class[1]	Predictive model[2]	Inputs[3]					
			AA	EVO	PSS	PSA	PDIS	Other inputs
DFLpred	ML	LR	X				X	Propensity for secondary structure estimated from sequence.
MoRFCHiBiWeb	Meta	Bayes	X	X			X	
fMoRFpred	ML	SVM	X		X		X	
DISOPRED3	ML	SVM	X	X				
MoRFCHiBi	ML	SVM	X					
disoRDPbind	ML	LR	X		X		X	Multiple sequence alignment, sequence complexity.
PepBindPred	SB	NN			X		X	Docking scores.
MFSPSSMpred	ML	SVM		X				
MoRFpred	ML	SVM	X	X		X	X	Predicted B-factors, multiple sequence alignment.
SLiMPred	ML	NN	X		X	X	X	Predicted structural motifs and domains.
retro-MoRFs	SF	SF	X				X	Multiple sequence alignment.
ANCHOR	SF	SF	X					Propensity for disorder, intra-chain interactions, and binding.
alpha-MoRFpred	ML	NN	X		X		X	Three disorder predictors are used.

[1]Class: Meta (meta predictor); ML (machine learning-based method); SB (structure-based method); SF (scoring function-based method).
[2]Predictive model: Bayes (Bayesian rule); LR (logistic regression); NN (neural network); SF (scoring function); SVM (support vector machine).
[3]Inputs: AA (AA type, property, propensity and/or position); EVO (evolutionary information based on PSSM or HMM profile); PDIS (predicted disorder); PSA (predicted solvent accessibility); PSS (predicted secondary structure).

Figure 1 visualizes predictions generated by several methods that find disordered protein binding regions (Figure 1A) and disordered DNA binding regions (Figure 1B) and compares them to the location of the corresponding native annotations of disordered protein and DNA binding residues, respectively. The predictors produce putative numeric propensities for a given function and the corresponding binary values that denote which disordered residues are predicted to bind proteins and DNA. The propensity scores are included at the top of each panel while the corresponding binary predictions are shown as horizontal lines at the bottom. The binary predictions are generated from the propensities by applying a threshold suggested by the authors, i.e., residues with propensities higher than the threshold are predicted as binding. We caution the reader that these results should not be assumed as typical and representative of the predictive performance of the corresponding methods, but rather they are used to illustrate how to use and interpret these predictions.

Figure 1A shows predictions of disordered protein binding regions from DisoRDPbind and putative MoRF regions generated by MoRFpred, ANCHOR, MFSPSSMpred, and MoRFCHiBiWeb for anophelin protein (DisProt ID: DP00824). This protein has a MoRF region between positions 54 and 83 [137]. We also include native annotations of IDR in this protein (positions 23 to 83) and the corresponding predictions of intrinsic disorder generated with IUPred. The predicted disordered residues are in good agreement with the native disordered residues (gray and black/light red horizontal lines at the bottom of Figure 1A that denote putative and native annotations, respectively) and they can be used to filter out results from the methods that predict the functional disordered regions. This allows us to eliminate the false predictions from ANCHOR, MoRFpred and MoRFCHiBiWeb near the N-terminus. Interestingly, both MoRFpred and ANCHOR accurately find the protein binding region at the C terminus (green and pink horizontal lines at the bottom of Figure 1A). The other three methods, MFSPSSMpred (violet horizontal line), MoRFCHiBiWeb (dark red horizontal line) and DisoRDPbind (orange horizontal line) identify the entire IDRs as protein binding. The latter method aims to find generic disordered protein binding regions, rather than the MoRF region that is present in this protein, and this is likely why its results are less accurate. However, all five methods correctly suggest presence of the disordered protein binding region in this protein demonstrating that their outputs can be used for a practical purpose.

Figure 1B illustrates predictions of disordered DNA binding regions from DisoRDPbind for Thymine-DNA glycosylase protein (DisProt ID: DP00719). This protein includes disordered DNA binding region between positions 51 and 111 [138] denoted by the light red horizontal line at the bottom of Figure 1B and two disordered regions (positions 1 to 111, and positions 340 to 410). Like in the above example, the putative disordered regions produced with IUPred are in relatively good agreement with the native disordered regions (gray and black/light red horizontal lines at the bottom of Figure 1B that denote putative and native annotations, respectively). Using these putative annotations of IDRs as a filter, the predictions from DisoRDPbind (dark red horizontal lines) point to the correct location of the native disordered DNA binding region. Once again, our example reveals that use of the putative annotations of disorder in tandem with the putative annotations of disordered binding regions leads to an accurate hypothesis that suggest location of the native DNA binding region.

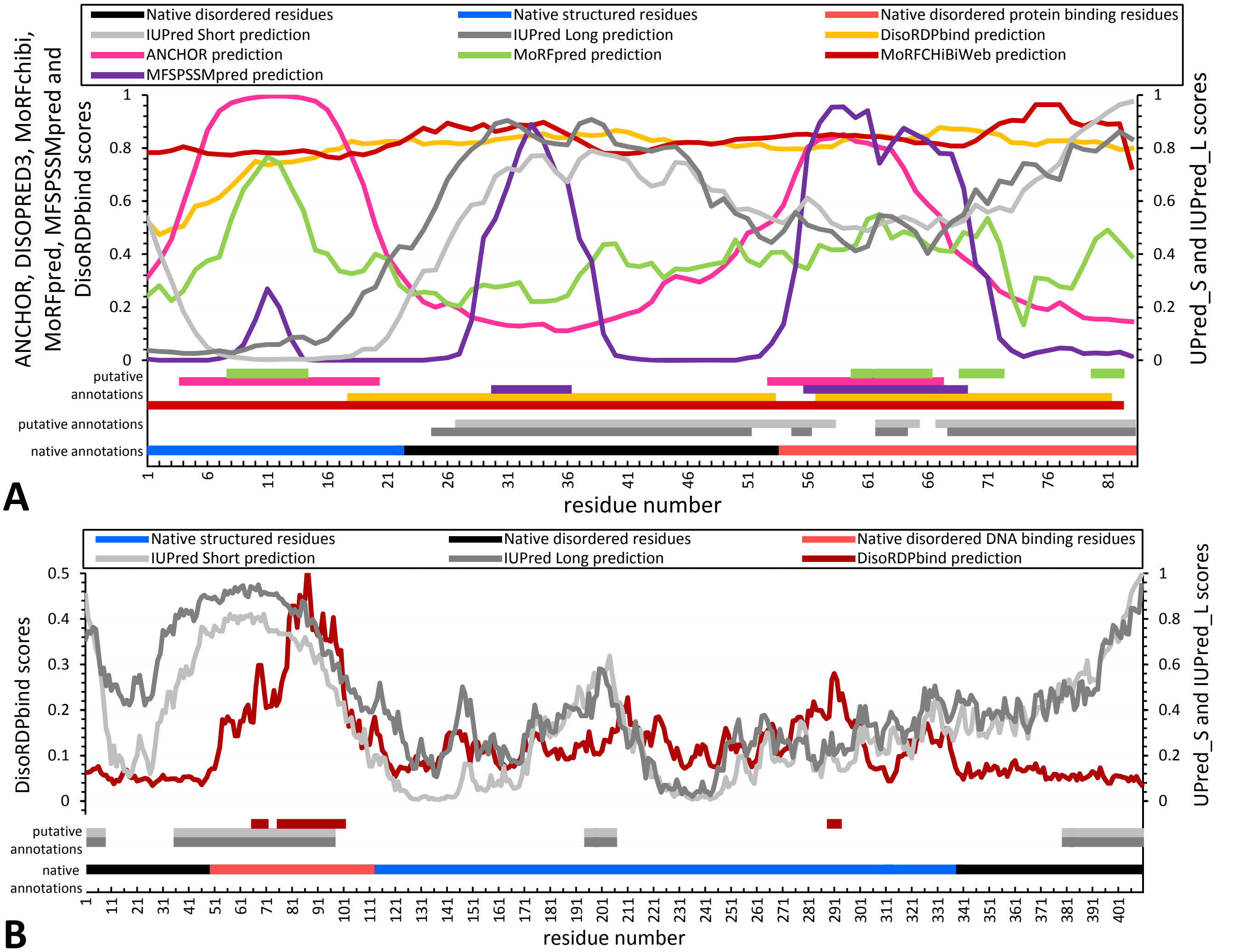

Figure 1. Putative and native annotations of disordered protein and DNA binding regions. Panel A shows the anophelin protein (DisProt ID: DP00824) that includes disordered protein binding region between positions 54 and 83 (denoted by the light red horizontal line at the bottom). This panel also gives putative disordered regions generated by IUPred (dark and light gray lines that correspond to the predictions by the IUPred that predicts long and short disordered regions), putative disordered protein binding regions output by DisoRDPbind (orange lines) and putative MoRF regions produced by MoRFpred (green), ANCHOR (pink), MFSPSSMpred (violet), and MoRFCHiBiWeb (dark red). Panel B shows the Thymine-DNA glycosylase protein (DisProt ID: DP00719) that includes disordered DNA binding region between positions 51 and 111 (denoted by the light red horizontal line at the bottom). It also visualizes putative disordered regions provided by IUpred (dark and light gray lines that correspond to the predictions by the IUPred that predicts long and short disordered regions) and putative disordered DNA binding regions predicted by DisoRDPbind (dark red lines).

The putative annotations of disorder function that were generated with the considered in this article computational tools have already found numerous practical applications. For example, both ANCHOR and fMoRFpred were applied on a large scale to identify and characterize putative disordered protein-protein binding regions in 736 [120] and 868 [130] complete proteomes, respectively. Similarly, SLiMPred has been applied to identify putative peptide binding motifs in the human proteome [134] and disoRDPbind has been applied to predict RNA/DNA binding residues in the proteomes of human, mouse, worm and fruit fly [121]. Moreover, ANCHOR was used to characterize disordered protein-binding regions in the nuclear proteins in mouse [139], in the human spliceosome [140] and the human autophagy proteins [141]. MoRFpred was utilized to analyze such regions in the proteomes of the dengue [142] and hepatitis C [112] viruses, in the ribosomal proteins [9], and in plants [143]. Putative disordered protein-binding regions that were identified with ANCHOR and MoRFpred were also used to functionally characterize histones [10].

Detailed summary of selected predictors of functions of intrinsic disorder

We provide a detailed and structured summary of a few selected methods. They include the first and only method that predicts disordered flexible linkers, DFLpred, the first-of-its-kind predictor of disordered DNA and RNA binding regions, DisoRDPbind, two most cited predictors of disordered protein binding regions that are available online: ANCHOR and MoRFpred. Their citations counts in Google Scholar as of December 2016 are 267 and 124, respectively. We also overview the most recent predictor of protein binding regions, MoRFChiBiWeb. We review these five methods in the chronological order.

ANCHOR (2009)

ANCHOR [120,106] was developed by Dosztányi *et al.* at the Hungarian Academy of Sciences (currently at the Eötvös Loránd University). The design of this method was inspired by the approach used in the IUPred method. ANCHOR finds disordered protein binding regions by using a scoring function that combines three hallmarks of these regions. It identifies regions that are likely to be disordered, which do not form intrachain interactions to fold on their own, and which are likely to gain stabilizing energy through an interaction with a structured protein. The calculation of the score using this function is quick and this predictor is available as an easy to use webserver and a standalone software. The speed, availability and the fact that this is the first method that predicts disordered protein binding regions contribute to the popularity of this methodology. We note that the earlier alpha-MoRFpred finds only the disordered protein binding regions that fold into helical confirmation upon binding.
Input: SwissProt/TrEMBL ID or a single raw (unformatted) amino acid sequence.
Output: Predicted propensities for disordered protein binding for each residue in the input sequence. A plot representing the propensity scores from ANCHOR and IUPred; the latter should be used to filter predictions from ANCHOR. Binary prediction for each residue in the input sequence.
Availability: A webserver and a standalone package running on a Linux platform.
URL: http://anchor.enzim.hu/

MoRFpred (2012)

MoRFpred [126] was released by Disfani *et al.* at the University of Alberta. MoRFpred predicts MoRF regions. It uses a support vector machine model that takes a set of 24 custom-designed numerical features generated from evolutionary profiles, selected physiochemical properties of amino acids and predicted disorder, relative solvent accessibility and B-factors as its inputs. The features aggregate this information for a given predicted residue and its neighbors in the input protein chain. The predictions from this machine learning model are combined with predictions based on sequence alignment against a large database of proteins annotated with native MoRF regions. This fairly sophisticated design results in an accurate prediction of MoRFs [126] but it also requires long runtime. A typical prediction with this method takes anywhere from about one to several minutes for a single protein chain.
Input: A single or multiple ($\leq$ 5 sequences) FASTA-formatted amino acid sequence(s).

Output: Predicted propensities and binary scores for each residue in the input sequence.
Availability: A webserver.
URL: http://biomine.cs.vcu.edu/servers/MoRFpred/

DisoRDPbind (2015)

DisoRDPbind [121] was authored by Peng and Kurgan at the University of Alberta. This is the first methodology that addressed simultaneous prediction of multiple functions of disordered regions. It predicts disordered protein, RNA and DNA binding regions. This predictor uses 7, 11 and 7 custom-designed features computed from selected physiochemical properties of input amino acids, sequence complexity estimated at the residue level, and putative intrinsic disorder and secondary structure as inputs. These three sets of inputs are processed by three logistic regression models to generate predictions of disordered protein, RNA and DNA binding regions. The results generated by each of these three predictors are merged with the corresponding annotations transferred via sequence alignment from a dataset of proteins annotated with native disordered protein, RNA and DNA binding regions. Empirical assessment shows that DisoRDPbind's outputs predictions that complement structured DNA and RNA binding residues generated by several representative methods [121]. DisoRDPbind generates predictions for a single protein in about 2 seconds compared to ANCHOR that is approximately two times faster [121]. However, these predictions include all three functions in contrast to ANCHOR that generates solely the protein binding predictions.
Input: A single or multiple (≤ 5000 sequences) FASTA-formatted amino acid sequence(s).
Output: Predicted propensities and binary scores for each of the three functions for each residue in the input sequence.
Availability: A webserver.
URL: http://biomine.cs.vcu.edu/servers/DisoRDPbind/

MoRFChiBiWeb (2016)

MoRFChiBiWeb [129] is a meta method that was designed by Malhis *et al.* at the University of British Columbia. It predicts MoRF regions. The meta design means that it re-predicts putative annotations of MoRF regions computed by the MoRFChiBi method to improve predictive performance. The architecture of this method is based on combining three inputs: the results generated by MoRFChiBi, putative intrinsic disorder produced by Espritz and sequence alignment profiles generated with PSI-BLAST. MoRFChiBiWeb is using Bayes rule to combine the three inputs. The improved predictive performance of MoRFChiBiWeb comes at an expense of a longer runtime when compared to MoRFChiBi, primarily due to the high computational cost associated with the calculation of the alignment profiles. MoRFChiBiWeb takes over 30 seconds to process a single, average length protein sequence compared to MoRFChiBi that need about one second. The webserver also offers results generated by MoRFChiBi_light, a fast MoRF predictor (one second per protein) that combines MoRFChiBi and the Espritz method.
Input: A single FASTA-formatted amino acid sequence.
Output: Predicted propensities for each residue in the input sequence. A plot representing the propensity scores from MoRFChiBiWeb, MoRFChiBi, and MoRFChiBiWeb_light.
Availability: An HTML webserver, a RESTful webserver and a standalone package.
URL: http://morf.chibi.ubc.ca:8080/mcw/index.xhtml

DFLpred (2016)

DFLpred [136] was created by Meng and Kurgan at the Virginia Commonwealth University. This method predicts disordered flexible linker regions. The predictive model is based on the machine learning architecture. Propensity for disordered flexible linkers for each residue in the input protein sequence is computed using a logistic regression model that takes four numerical inputs. These inputs quantify tendency of the predicted residue and its neighbors in the sequence to form structured domains, disordered regions, helical conformations and turns. This predictor has low runtime. A single average length protein chain can be predicted in 0.1 second while prediction of an entire human proteome takes less than one hour on a modern desktop computer.
Input: A single or multiple (≤ 5000 sequences) FASTA-formatted amino acid sequence(s).
Output: Predicted propensities and binary scores for each residue in the input sequence.
Availability: A webserver.
URL: http://biomine.cs.vcu.edu/servers/DFLpred/

Conclusions and future directions

It has been nearly four decades since the first predictor of intrinsic disorder was introduced. The efforts to develop methods that predict IDRs and IDPs span three periods, each advancing this area to new heights. Modern predictors are characterized by use of meta-approaches and sophisticated predictive models that are typically derived using machine learning algorithms. Predictive performance of these methods was evaluated in several large comparative studies which revealed that some of them provide accurate predictions, with AUC > 0.7 and MCC > 0.4. They are also conveniently available to the end users as webservers and/or standalone software. The webservers are particularly attractive for less technically savvy users. They can be accessed via all major web browsers, perform computations on the server side, automate the prediction process, and provide the results in an easy to understand format. Moreover, the end users nowadays can also access and search pre-computed putative annotations of disorder via several large databases, including DICHOT, MobiDB and D^2P^2. These databases store predictions of multiple methods for virtually all currently known proteins.

Compared with the prediction of IDRs and IDPs, computational prediction of functions of disorder is in early stages. These predictors primarily focus on the binding-related functions, such as disordered protein-protein, protein-RNA and protein-DNA binding. Most of these methods rely on the machine learning models and are provided as fast and user-friendly webservers. Although 13 of these methods were already released, the development of models that address prediction of other functions of disorder remains an outstanding and pressing challenge. The version 6.0.2 of the Disprot database, the main source of the functionally annotated IDPs, lists close to 40 cellular functions that have been assigned to about 1200 IDRs. To date, only 11 of them can be predicted with the currently available methods. Another interesting and unexplored subject is related to an observation that IDRs are implicated in moonlighting [144]. In contrast to the structured moonlighting proteins that carry out multiple cellular function through inclusion of multiple domains [145,146], a single IDR can carry out multiple functions by itself. Our analysis reveals that 37% of the functionally annotated IDRs in the Disprot database carry out more than one distinct function. Given such high prevalence, computational characterization and prediction of these disordered moonlighting regions should be pursued.

Acknowledgements

This work was supported in part by the National Science Foundation (NSF) grant 1617369 and the Qimonda Research Chair to L.K., and a scholarship from the China Scholarship Council to F.M.

References

1. Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradović Z (2002) Intrinsic Disorder and Protein Function†. Biochemistry 41 (21):6573-6582.
2. Habchi J, Tompa P, Longhi S, Uversky VN (2014) Introducing Protein Intrinsic Disorder. Chemical Reviews 114 (13):6561-6588.
3. Peng Z, Yan J, Fan X, Mizianty MJ, Xue B, Wang K, Hu G, Uversky VN, Kurgan L (2015) Exceptionally abundant exceptions: comprehensive characterization of intrinsic disorder in all domains of life. Cell Mol Life Sci 72 (1):137-151.
4. Dunker AK, Obradovic Z, Romero P, Garner EC, Brown CJ (2000) Intrinsic protein disorder in complete genomes. Genome Inform Ser Workshop Genome Inform 11:161-171.
5. Xue B, Dunker AK, Uversky VN (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 (2):137-149.
6. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J Mol Biol 337 (3):635-645.
7. Liu J, Perumal NB, Oldfield CJ, Su EW, Uversky VN, Dunker AK (2006) Intrinsic disorder in transcription factors. Biochemistry 45 (22):6873-6888.
8. Fuxreiter M, Tompa P, Simon I, Uversky VN, Hansen JC, Asturias FJ (2008) Malleable machines take shape in eukaryotic transcriptional regulation. Nat Chem Biol 4 (12):728-737.
9. Peng Z, Oldfield CJ, Xue B, Mizianty MJ, Dunker AK, Kurgan L, Uversky VN (2014) A creature with a hundred waggly tails: intrinsically disordered proteins in the ribosome. Cell Mol Life Sci 71 (8):1477-1504.
10. Peng Z, Mizianty MJ, Xue B, Kurgan L, Uversky VN (2012) More than just tails: intrinsic disorder in histone proteins. Mol Biosyst 8 (7):1886-1901.
11. Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol 6 (3):197-208.
12. Xie H, Vucetic S, Iakoucheva LM, Oldfield CJ, Dunker AK, Uversky VN, Obradovic Z (2007) Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. J Proteome Res 6 (5):1882-1898.
13. Galea CA, Wang Y, Sivakolundu SG, Kriwacki RW (2008) Regulation of cell division by intrinsically unstructured proteins: intrinsic flexibility, modularity, and signaling conduits. Biochemistry 47 (29):7598-7609.
14. Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN (2005) Flexible nets. The roles of intrinsic disorder in protein interaction networks. FEBS J 272 (20):5129-5148.
15. Uversky VN, Oldfield CJ, Dunker AK (2005) Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. J Mol Recognit 18 (5):343-384.
16. Uversky VN, Oldfield CJ, Dunker AK (2008) Intrinsically disordered proteins in human diseases: introducing the D2 concept. Annu Rev Biophys 37:215-246.
17. Midic U, Oldfield CJ, Dunker AK, Obradovic Z, Uversky VN (2009) Unfoldomics of human genetic diseases: illustrative examples of ordered and intrinsically disordered members of the human diseasome. Protein and peptide letters 16 (12):1533-1547.
18. Cheng Y, LeGall T, Oldfield CJ, Mueller JP, Van YY, Romero P, Cortese MS, Uversky VN, Dunker AK (2006) Rational drug design via intrinsically disordered protein. Trends Biotechnol 24 (10):435-442.
19. Hu G, Wu Z, Wang K, Uversky VN, Kurgan L (2015) Untapped potential of disordered proteins in current druggable human proteome. Current drug targets.

20. Sickmeier M, Hamilton JA, LeGall T, Vacic V, Cortese MS, Tantos A, Szabo B, Tompa P, Chen J, Uversky VN, Obradovic Z, Dunker AK (2007) DisProt: the Database of Disordered Proteins. Nucleic Acids Research 35 (suppl 1):D786-D793.
21. Vucetic S, Obradovic Z, Vacic V, Radivojac P, Peng K, Iakoucheva LM, Cortese MS, Lawson JD, Brown CJ, Sikes JG, Newton CD, Dunker AK (2005) DisProt: a database of protein disorder. Bioinformatics 21 (1):137-140.
22. Piovesan D, Tabaro F, Micetic I, Necci M, Quaglia F, Oldfield CJ, Aspromonte MC, Davey NE, Davidovic R, Dosztanyi Z, Elofsson A, Gasparini A, Hatos A, Kajava AV, Kalmar L, Leonardi E, Lazar T, Macedo-Ribeiro S, Macossay-Castillo M, Meszaros A, Minervini G, Murvai N, Pujols J, Roche DB, Salladini E, Schad E, Schramm A, Szabo B, Tantos A, Tonello F, Tsirigos KD, Veljkovic N, Ventura S, Vranken W, Warholm P, Uversky VN, Dunker AK, Longhi S, Tompa P, Tosatto SC (2016) DisProt 7.0: a major update of the database of disordered proteins. Nucleic Acids Res.
23. Fukuchi S, Sakamoto S, Nobe Y, Murakami SD, Amemiya T, Hosoda K, Koike R, Hiroaki H, Ota M (2012) IDEAL: Intrinsically Disordered proteins with Extensive Annotations and Literature. Nucleic Acids Research 40 (D1):D507-D511.
24. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. Nucleic Acids Research 28 (1):235-242.
25. Tompa P (2002) Intrinsically unstructured proteins. Trends in Biochemical Sciences 27 (10):527-533.
26. Tompa P (2012) Intrinsically disordered proteins: a 10-year recap. Trends in Biochemical Sciences 37 (12):509-516.
27. Walsh I, Giollo M, Di Domenico T, Ferrari C, Zimmermann O, Tosatto SCE (2015) Comprehensive large-scale assessment of intrinsic protein disorder. Bioinformatics 31 (2):201-208.
28. Martin AJM, Walsh I, Tosatto SCE (2010) MOBI: a web server to define and visualize structural mobility in NMR protein ensembles. Bioinformatics 26 (22):2916-2917.
29. Ota M, Koike R, Amemiya T, Tenno T, Romero PR, Hiroaki H, Dunker AK, Fukuchi S (2013) An assignment of intrinsically disordered regions of proteins based on NMR structures. J Struct Biol 181 (1):29-36.
30. Ferron F, Longhi S, Canard B, Karlin D (2006) A practical overview of protein disorder prediction methods. Proteins: Structure, Function, and Bioinformatics 65 (1):1-14.
31. Deng X, Eickholt J, Cheng J (2012) A comprehensive overview of computational protein disorder prediction methods. Mol Biosyst 8 (1):114-121.
32. He B, Wang K, Liu Y, Xue B, Uversky VN, Dunker AK (2009) Predicting intrinsic disorder in proteins: an overview. Cell Res 19 (8):929-949.
33. Dosztányi Z, Mészáros B, Simon I (2010) Bioinformatical approaches to characterize intrinsically disordered/unstructured proteins. Briefings in bioinformatics 11 (2):225-243.
34. Dosztányi Z, Tompa P (2008) Prediction of Protein Disorder. In: Kobe B, Guss M, Huber T (eds) Structural Proteomics, vol 426. Methods in Molecular Biology™. Humana Press, pp 103-115.
35. Pentony M, Ward J, Jones D (2010) Computational Resources for the Prediction and Analysis of Native Disorder in Proteins. In: Hubbard SJ, Jones AR (eds) Proteome Bioinformatics, vol 604. Methods in Molecular Biology™. Humana Press, pp 369-393.
36. Atkins J, Boateng S, Sorensen T, McGuffin L (2015) Disorder Prediction Methods, Their Applicability to Different Protein Targets and Their Usefulness for Guiding Experimental Studies. International Journal of Molecular Sciences 16 (8):19040.
37. Li J, Feng Y, Wang X, Li J, Liu W, Rong L, Bao J (2015) An Overview of Predictors for Intrinsically Disordered Proteins over 2010–2014. International Journal of Molecular Sciences 16 (10):23446.
38. Peng ZL, Kurgan L (2012) Comprehensive comparative assessment of in-silico predictors of disordered regions. Curr Protein Pept Sci 13 (1):6-18.
39. Monastyrskyy B, Kryshtafovych A, Moult J, Tramontano A, Fidelis K (2014) Assessment of protein disorder region predictions in CASP10. Proteins 82 Suppl 2:127-137.
40. Monastyrskyy B, Fidelis K, Moult J, Tramontano A, Kryshtafovych A (2011) Evaluation of disorder predictions in CASP9. Proteins 79 Suppl 10:107-118.
41. Williams RJP (1979) THE CONFORMATION PROPERTIES OF PROTEINS IN SOLUTION. Biological Reviews 54 (4):389-437.

42. Romero P, Obradovic Z, Kissinger C, Villafranca JE, Dunker AK Identifying disordered regions in proteins from amino acid sequence. In: Neural Networks,1997., International Conference on, 9-12 Jun 1997 1997. pp 90-95 vol.91.
43. Uversky VN, Gillespie JR, Fink AL (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins: Structure, Function, and Bioinformatics 41 (3):415-427.
44. Prilusky J, Felder CE, Zeev-Ben-Mordehai T, Rydberg EH, Man O, Beckmann JS, Silman I, Sussman JL (2005) FoldIndex©: a simple tool to predict whether a given protein sequence is intrinsically unfolded. Bioinformatics 21 (16):3435-3438.
45. Linding R, Russell RB, Neduva V, Gibson TJ (2003) GlobPlot: exploring protein sequences for globularity and disorder. Nucleic Acids Research 31 (13):3701-3708.
46. Dosztányi 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 (16):3433-3434.
47. Dosztányi Z, Csizmók V, Tompa P, Simon I (2005) The Pairwise Energy Content Estimated from Amino Acid Composition Discriminates between Folded and Intrinsically Unstructured Proteins. Journal of Molecular Biology 347 (4):827-839.
48. Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK (2001) Sequence complexity of disordered protein. Proteins: Structure, Function, and Bioinformatics 42 (1):38-48.
49. Vucetic S, Brown CJ, Dunker AK, Obradovic Z (2003) Flavors of protein disorder. Proteins: Structure, Function, and Bioinformatics 52 (4):573-584.
50. Obradovic Z, Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK (2003) Predicting intrinsic disorder from amino acid sequence. Proteins: Structure, Function, and Bioinformatics 53 (S6):566-572.
51. Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK, Obradovic Z (2005) Optimizing long intrinsic disorder predictors with protein evolutionary information. Journal of bioinformatics and computational biology 3 (1):35-60.
52. Obradovic Z, Peng K, Vucetic S, Radivojac P, Dunker AK (2005) Exploiting heterogeneous sequence properties improves prediction of protein disorder. Proteins: Structure, Function, and Bioinformatics 61 (S7):176-182.
53. Peng K, Radivojac P, Vucetic S, Dunker AK, Obradovic Z (2006) Length-dependent prediction of protein intrinsic disorder. BMC Bioinformatics 7 (1):208.
54. Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ, Russell RB (2003) Protein Disorder Prediction: Implications for Structural Proteomics. Structure 11 (11):1453-1459.
55. Jones DT, Ward JJ (2003) Prediction of disordered regions in proteins from position specific score matrices. Proteins: Structure, Function, and Bioinformatics 53 (S6):573-578.
56. Schlessinger A, Yachdav G, Rost B (2006) PROFbval: predict flexible and rigid residues in proteins. Bioinformatics 22 (7):891-893.
57. Cheng J, Sweredoski M, Baldi P (2005) Accurate Prediction of Protein Disordered Regions by Mining Protein Structure Data. Data Min Knowl Disc 11 (3):213-222.
58. Liu J, Rost B (2003) NORSp: predictions of long regions without regular secondary structure. Nucleic Acids Research 31 (13):3833-3835.
59. Wang L, Sauer UH (2008) OnD-CRF: predicting order and disorder in proteins using [corrected] conditional random fields. Bioinformatics 24 (11):1401-1402.
60. Eickholt J, Cheng J (2013) DNdisorder: predicting protein disorder using boosting and deep networks. BMC Bioinformatics 14 (1):1-10.
61. Jones DT, Cozzetto D (2015) DISOPRED3: precise disordered region predictions with annotated protein-binding activity. Bioinformatics 31 (6):857-863.
62. Walsh I, Martin AJM, Di Domenico T, Vullo A, Pollastri G, Tosatto SCE (2011) CSpritz: accurate prediction of protein disorder segments with annotation for homology, secondary structure and linear motifs. Nucleic Acids Research 39 (suppl 2):W190-W196.
63. Kozlowski LP, Bujnicki JM (2012) MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins. BMC Bioinformatics 13 (1):1-11.
64. Mizianty MJ, Stach W, Chen K, Kedarisetti KD, Disfani FM, Kurgan L (2010) Improved sequence-based prediction of disordered regions with multilayer fusion of multiple information sources. Bioinformatics 26 (18):i489-i496.

65. Huang YJ, Acton TB, Montelione GT (2014) DisMeta: a meta server for construct design and optimization. Methods Mol Biol 1091:3-16.
66. Mizianty MJ, Peng Z, Kurgan L (2013) MFDp2-Accurate predictor of disorder in proteins by fusion of disorder probabilities, content and profiles. Intrinsically Disordered Proteins 1 (1):e24428.
67. Ishida T, Kinoshita K (2007) PrDOS: prediction of disordered protein regions from amino acid sequence. Nucleic Acids Research 35 (suppl 2):W460-W464.
68. McGuffin LJ (2008) Intrinsic disorder prediction from the analysis of multiple protein fold recognition models. Bioinformatics 24 (16):1798-1804.
69. Yang ZR, Thomson R, McNeil P, Esnouf RM (2005) RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. Bioinformatics 21 (16):3369-3376.
70. Wang S, Weng S, Ma J, Tang Q (2015) DeepCNF-D: Predicting Protein Order/Disorder Regions by Weighted Deep Convolutional Neural Fields. International Journal of Molecular Sciences 16 (8):17315.
71. Ward JJ, McGuffin LJ, Bryson K, Buxton BF, Jones DT (2004) The DISOPRED server for the prediction of protein disorder. Bioinformatics 20 (13):2138-2139.
72. Fan X, Kurgan L (2014) Accurate prediction of disorder in protein chains with a comprehensive and empirically designed consensus. J Biomol Struct Dyn 32 (3):448-464.
73. Ishida T, Kinoshita K (2008) Prediction of disordered regions in proteins based on the meta approach. Bioinformatics 24 (11):1344-1348.
74. Schlessinger A, Punta M, Yachdav G, Kajan L, Rost B (2009) Improved Disorder Prediction by Combination of Orthogonal Approaches. PLoS ONE 4 (2):e4433.
75. McGuffin LJ, Atkins JD, Salehe BR, Shuid AN, Roche DB (2015) IntFOLD: an integrated server for modelling protein structures and functions from amino acid sequences. Nucleic Acids Research 43 (W1):W169-W173.
76. Schlessinger A, Liu J, Rost B (2007) Natively Unstructured Loops Differ from Other Loops. PLoS Comput Biol 3 (7):e140.
77. Schlessinger A, Punta M, Rost B (2007) Natively unstructured regions in proteins identified from contact predictions. Bioinformatics 23 (18):2376-2384.
78. Peng Z, Kurgan L (2012) On the complementarity of the consensus-based disorder prediction. Pac Symp Biocomput:176-187.
79. McGuffin LJ (2008) Intrinsic disorder prediction from the analysis of multiple protein fold recognition models. Bioinformatics 24 (16):1798-1804.
80. Walsh I, Martin AJM, Di Domenico T, Tosatto SCE (2012) ESpritz: accurate and fast prediction of protein disorder. Bioinformatics 28 (4):503-509.
81. Zhang T, Faraggi E, Xue B, Dunker AK, Uversky VN, Zhou Y (2012) SPINE-D: Accurate Prediction of Short and Long Disordered Regions by a Single Neural-Network Based Method. Journal of biomolecular structure & dynamics 29 (4):799-813.
82. Michail Yu L, Oxana VG (2011) The Ising model for prediction of disordered residues from protein sequence alone. Physical Biology 8 (3):035004.
83. Xue B, Dunbrack RL, Williams RW, Dunker AK, Uversky VN (2010) PONDR-FIT: A meta-predictor of intrinsically disordered amino acids. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics 1804 (4):996-1010.
84. Deng X, Eickholt J, Cheng J (2009) PreDisorder: ab initio sequence-based prediction of protein disordered regions. BMC Bioinformatics 10 (1):436.
85. Galzitskaya OV, Garbuzynskiy SO, Lobanov MY (2006) Prediction of natively unfolded regions in protein chains. Mol Biol 40 (2):298-304.
86. Hecker J, Yang JY, Cheng J (2008) Protein disorder prediction at multiple levels of sensitivity and specificity. BMC Genomics 9 (1):1-7.
87. Monastyrskyy B, Kryshtafovych A, Moult J, Tramontano A, Fidelis K (2014) Assessment of protein disorder region predictions in CASP10. Proteins 82 (0 2):127-137.
88. Deng X, Eickholt J, Cheng J (2012) A comprehensive overview of computational protein disorder prediction methods. Molecular BioSystems 8 (1):114-121.
89. Melamud E, Moult J (2003) Evaluation of disorder predictions in CASP5. Proteins 53 Suppl 6:561-565.

90. Potenza E, Domenico TD, Walsh I, Tosatto SCE (2015) MobiDB 2.0: an improved database of intrinsically disordered and mobile proteins. Nucleic Acids Research 43 (D1):D315-D320.
91. Oates ME, Romero P, Ishida T, Ghalwash M, Mizianty MJ, Xue B, Dosztányi Z, Uversky VN, Obradovic Z, Kurgan L, Dunker AK, Gough J (2013) D2P2: database of disordered protein predictions. Nucleic Acids Research 41 (D1):D508-D516.
92. Deng X, Eickholt J, Cheng J (2009) PreDisorder: ab initio sequence-based prediction of protein disordered regions. BMC Bioinformatics 10 (1):1-6.
93. Hecker J, Yang JY, Cheng JL (2008) Protein disorder prediction at multiple levels of sensitivity and specificity. Bmc Genomics 9.
94. Cheng JL, Sweredoski MJ, Baldi P (2005) Accurate prediction of protein disordered regions by mining protein structure data. Data Min Knowl Disc 11 (3):213-222.
95. Noivirt-Brik O, Prilusky J, Sussman JL (2009) Assessment of disorder predictions in CASP8. Proteins 77 Suppl 9:210-216.
96. Yachdav G, Kloppmann E, Kajan L, Hecht M, Goldberg T, Hamp T, Hönigschmid P, Schafferhans A, Roos M, Bernhofer M, Richter L, Ashkenazy H, Punta M, Schlessinger A, Bromberg Y, Schneider R, Vriend G, Sander C, Ben-Tal N, Rost B (2014) PredictProtein—an open resource for online prediction of protein structural and functional features. Nucleic Acids Research 42 (W1):W337-W343.
97. Buchan DWA, Minneci F, Nugent TCO, Bryson K, Jones DT (2013) Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic Acids Research 41 (W1):W349-W357.
98. Faraggi E, Xue B, Zhou Y (2009) Improving the prediction accuracy of residue solvent accessibility and real-value backbone torsion angles of proteins by guided-learning through a two-layer neural network. Proteins: Structure, Function, and Bioinformatics 74 (4):847-856.
99. Campen A, Williams RM, Brown CJ, Meng J, Uversky VN, Dunker AK (2008) TOP-IDP-scale: a new amino acid scale measuring propensity for intrinsic disorder. Protein and peptide letters 15 (9):956-963.
100. Atchley WR, Zhao J, Fernandes AD, Drüke T (2005) Solving the protein sequence metric problem. Proceedings of the National Academy of Sciences of the United States of America 102 (18):6395-6400.
101. Mizianty MJ, Zhang T, Xue B, Zhou Y, Dunker AK, Uversky VN, Kurgan L (2011) In-silico prediction of disorder content using hybrid sequence representation. BMC Bioinformatics 12 (1):1-16.
102. Rost B, Sander C (1994) Combining evolutionary information and neural networks to predict protein secondary structure. Proteins 19 (1):55-72.
103. Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007) Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protocols 2 (4):953-971.
104. Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305 (3):567-580.
105. Wootton JC (1994) Non-globular domains in protein sequences: Automated segmentation using complexity measures. Computers & Chemistry 18 (3):269-285.
106. Dosztányi Z, Mészáros B, Simon I (2009) ANCHOR: web server for predicting protein binding regions in disordered proteins. Bioinformatics 25 (20):2745-2746.
107. Buchan DW, Minneci F, Nugent TC, Bryson K, Jones DT (2013) Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic Acids Res 41 (Web Server issue):W349-357.
108. McGuffin LJ, Roche DB (2010) Rapid model quality assessment for protein structure predictions using the comparison of multiple models without structural alignments. Bioinformatics 26 (2):182-188.
109. Oldfield CJ, Xue B, Van YY, Ulrich EL, Markley JL, Dunker AK, Uversky VN (2013) Utilization of protein intrinsic disorder knowledge in structural proteomics. Biochim Biophys Acta 1834 (2):487-498.
110. Varadi M, Zsolyomi F, Guharoy M, Tompa P (2015) Functional Advantages of Conserved Intrinsic Disorder in RNA-Binding Proteins. PLoS One 10 (10):e0139731.
111. Peng Z, Xue B, Kurgan L, Uversky VN (2013) Resilience of death: intrinsic disorder in proteins involved in the programmed cell death. Cell Death Differ 20 (9):1257-1267.
112. Fan X, Xue B, Dolan PT, LaCount DJ, Kurgan L, Uversky VN (2014) The intrinsic disorder status of the human hepatitis C virus proteome. Mol Biosyst 10 (6):1345-1363.

113. Xue B, Mizianty MJ, Kurgan L, Uversky VN (2012) Protein intrinsic disorder as a flexible armor and a weapon of HIV-1. Cell Mol Life Sci 69 (8):1211-1259.
114. Pentony MM, Jones DT (2010) Modularity of intrinsic disorder in the human proteome. Proteins 78 (1):212-221.
115. Fukuchi S, Hosoda K, Homma K, Gojobori T, Nishikawa K (2011) Binary classification of protein molecules into intrinsically disordered and ordered segments. BMC Structural Biology 11 (1):1-10.
116. Consortium TU (2010) The Universal Protein Resource (UniProt) in 2010. Nucleic Acids Research 38 (suppl 1):D142-D148.
117. Fukuchi S, Homma K, Minezaki Y, Gojobori T, Nishikawa K (2009) Development of an accurate classification system of proteins into structured and unstructured regions that uncovers novel structural domains: its application to human transcription factors. BMC Structural Biology 9 (1):1-13.
118. Di Domenico T, Walsh I, Martin AJM, Tosatto SCE (2012) MobiDB: a comprehensive database of intrinsic protein disorder annotations. Bioinformatics 28 (15):2080-2081.
119. Ghalwash MF, Dunker AK, Obradovic Z (2012) Uncertainty analysis in protein disorder prediction. Molecular BioSystems 8 (1):381-391.
120. Mészáros B, Simon I, Dosztányi Z (2009) Prediction of Protein Binding Regions in Disordered Proteins. PLoS Comput Biol 5 (5):e1000376.
121. Peng Z, Kurgan L (2015) High-throughput prediction of RNA, DNA and protein binding regions mediated by intrinsic disorder. Nucleic Acids Research.
122. Yan J, Dunker AK, Uversky VN, Kurgan L (2016) Molecular recognition features (MoRFs) in three domains of life. Mol Biosyst 12 (3):697-710.
123. Mohan A, Oldfield CJ, Radivojac P, Vacic V, Cortese MS, Dunker AK, Uversky VN (2006) Analysis of molecular recognition features (MoRFs). J Mol Biol 362 (5):1043-1059.
124. Oldfield CJ, Cheng Y, Cortese MS, Brown CJ, Uversky VN, Dunker AK (2005) Comparing and Combining Predictors of Mostly Disordered Proteins†. Biochemistry 44 (6):1989-2000.
125. Cheng Y, Oldfield CJ, Meng J, Romero P, Uversky VN, Dunker AK (2007) Mining α-Helix-Forming Molecular Recognition Features with Cross Species Sequence Alignments†. Biochemistry 46 (47):13468-13477.
126. Disfani FM, Hsu W-L, Mizianty MJ, Oldfield CJ, Xue B, Dunker AK, Uversky VN, Kurgan L (2012) MoRFpred, a computational tool for sequence-based prediction and characterization of short disorder-to-order transitioning binding regions in proteins. Bioinformatics 28 (12):i75-i83.
127. Fang C, Noguchi T, Tominaga D, Yamana H (2013) MFSPSSMpred: identifying short disorder-to-order binding regions in disordered proteins based on contextual local evolutionary conservation. BMC Bioinformatics 14 (1):1-14.
128. Malhis N, Gsponer J (2015) Computational identification of MoRFs in protein sequences. Bioinformatics 31 (11):1738-1744.
129. Malhis N, Jacobson M, Gsponer J (2016) MoRFchibi SYSTEM: software tools for the identification of MoRFs in protein sequences. Nucleic Acids Res.
130. Yan J, Dunker AK, Uversky VN, Kurgan L (2015) Molecular recognition features (MoRFs) in three domains of life. Molecular BioSystems.
131. Xue B, Dunker AK, Uversky VN (2010) Retro-MoRFs: Identifying Protein Binding Sites by Normal and Reverse Alignment and Intrinsic Disorder Prediction. International Journal of Molecular Sciences 11 (10):3725-3747.
132. Van Roey K, Uyar B, Weatheritt RJ, Dinkel H, Seiler M, Budd A, Gibson TJ, Davey NE (2014) Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation. Chem Rev 114 (13):6733-6778.
133. Dinkel H, Van Roey K, Michael S, Kumar M, Uyar B, Altenberg B, Milchevskaya V, Schneider M, Kuhn H, Behrendt A, Dahl SL, Damerell V, Diebel S, Kalman S, Klein S, Knudsen AC, Mader C, Merrill S, Staudt A, Thiel V, Welti L, Davey NE, Diella F, Gibson TJ (2016) ELM 2016--data update and new functionality of the eukaryotic linear motif resource. Nucleic Acids Res 44 (D1):D294-300.
134. Mooney C, Pollastri G, Shields DC, Haslam NJ (2012) Prediction of Short Linear Protein Binding Regions. Journal of Molecular Biology 415 (1):193-204.

135. Khan W, Duffy F, Pollastri G, Shields DC, Mooney C (2013) Predicting Binding within Disordered Protein Regions to Structurally Characterised Peptide-Binding Domains. PLoS ONE 8 (9):e72838.
136. Meng F, Kurgan L (2016) DFLpred: High-throughput prediction of disordered flexible linker regions in protein sequences. Bioinformatics 32 (12):i341-i350.
137. Figueiredo AC, de Sanctis D, Gutierrez-Gallego R, Cereija TB, Macedo-Ribeiro S, Fuentes-Prior P, Pereira PJ (2012) Unique thrombin inhibition mechanism by anophelin, an anticoagulant from the malaria vector. Proc Natl Acad Sci U S A 109 (52):E3649-3658.
138. Smet-Nocca C, Wieruszeski JM, Chaar V, Leroy A, Benecke A (2008) The thymine-DNA glycosylase regulatory domain: residual structure and DNA binding. Biochemistry 47 (25):6519-6530.
139. Meng F, Na I, Kurgan L, Uversky VN (2016) Compartmentalization and Functionality of Nuclear Disorder: Intrinsic Disorder and Protein-Protein Interactions in Intra-Nuclear Compartments. Int J Mol Sci 17 (1).
140. Korneta I, Bujnicki JM (2012) Intrinsic disorder in the human spliceosomal proteome. PLoS Comput Biol 8 (8):e1002641.
141. Mei Y, Su M, Soni G, Salem S, Colbert CL, Sinha SC (2014) Intrinsically disordered regions in autophagy proteins. Proteins 82 (4):565-578.
142. Meng F, Badierah RA, Almehdar HA, Redwan EM, Kurgan L, Uversky VN (2015) Unstructural biology of the Dengue virus proteins. FEBS J 282 (17):3368-3394.
143. Marin M, Ott T (2014) Intrinsic disorder in plant proteins and phytopathogenic bacterial effectors. Chem Rev 114 (13):6912-6932.
144. Tompa P, Szász C, Buday L (2005) Structural disorder throws new light on moonlighting. Trends in Biochemical Sciences 30 (9):484-489.
145. Jeffery CJ (1999) Moonlighting proteins. Trends in Biochemical Sciences 24 (1):8-11.
146. Khan Ishita K, Kihara D (2014) Computational characterization of moonlighting proteins. Biochemical Society Transactions 42 (6):1780-1785.