The methods and tools for intrinsic disorder prediction and their application to systems medicine

Lukasz Kurgan
Department of Computer Science, Virginia Commonwealth University
401 West Main Street, Room E4225, Richmond, Virginia 23284, USA
Email: lkurgan@vcu.edu
Phone: (804) 827-3986

Min Li
School of Computer Science and Engineering, Central South University, Changsha, 410083, P.R. China

Yaohang Li
Department of Computer Science, Old Dominion University
4700 Elkhorn Ave, Suite 3300, Norfolk, VA 23529, USA

Keywords (10 to 15)
DISOPRED; Drug targets; Human diseases; Intrinsic disorder; Intrinsically disordered proteins; Intrinsically disordered regions; MFDp; Prediction; PrDOS, Proteins; Protein function; Viral proteomes

Abstract

While as many as 50% of eukaryotic proteins are estimated to have intrinsically disordered regions (IDRs), only about 30 thousand out of 130 million of currently sequenced proteins are experimentally annotated with disorder. This large and growing annotation gap can be reduced with computational sequence-based predictors of IDRs. This chapter motivates and illustrates computational prediction of IDRs and surveys popular methods. We discuss several practical facets concerning their availability, impact, outputs and predictive performance. Results published in the most recent CASP experiment that included disorder prediction assessment, CASP10, shows that IDRs can be predicted with high accuracy. The top three performers in CASP10 (PrDOS-CNF, DISOPRED3 and MFDp) secure AUC ≥ 0.89. Moreover, disorder predictors enjoy strong citation profiles, with three tools that are cited at least 50 times annually. We also delineate impact of IDR predictions on the systems medicine field, focusing on new frontiers to treat and understand human diseases. We show that annotation and functional understanding of IDRs assist with deciphering mechanisms of viral infections and the resultant immune responses, facilitate improved molecular-level understanding of a wide range of human diseases, and aid the development of novel drug targets and therapeutics.

1 Introduction

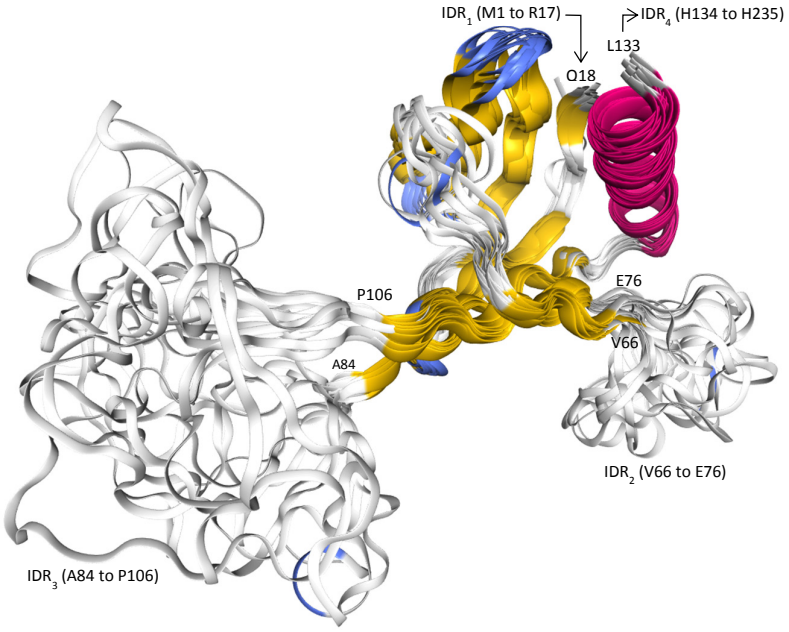
Intrinsically disordered regions (IDRs) lack a stable tertiary structure in isolation and they exist as heterogeneous ensemble of conformations [1,2]. A given protein may have one or many IDRs, and in some cases entire protein chains are intrinsically disordered. The term intrinsically disordered refers to an intrinsic property of the protein sequence and was originally coined by X-ray crystallographers to describe unresolved portions of a protein structure, which are due to highly fluctuating positions with the crystal lattice. These regions were assigned with many other names in the literature including floppy, pliable, rheomorphic, mobile, partially folded, natively denatured, natively unfolded, natively disordered, intrinsically unstructured, intrinsically denatured, and intrinsically unfolded [2].

IDRs form dynamic conformational ensembles, which means that atomic coordinates of their residues and their dihedral angles vary substantially over time and populations, often without a specific equilibrium [3]. Figure 1 shows two example IDRs in the structure of the chloride conductance regulatory (ICln) protein (UniProt ID: P35521): V66 to E76 region and A84 to P106 region. The top of the figure is a cartoon view of 15 superimposed NMR structures of this protein. The structured/ordered regions are where the NMR structures converge to the same conformation, compared to the two IDRs that are characterized by highly structurally variable conformations that are composed of coils. This proteins also includes two other IDRs, one at each terminus (M1 to R17 region and H134 to H235 region), where the structure could not be resolved.

Proteins with IDRs are highly abundant in nature. Recent computational studies estimate that between 3% and 17% of eukaryotic proteins are fully disordered, depending on an organism, and that about 30-50% of eukaryotic proteins have at least one long IDR with ≥ 30 consecutive amino acids [4,5]. Majority of eukaryotic proteins have both structured regions and IDRs, and both types of these regions implement functions that these protein have in a variety of cellular contexts. Interestingly, while some IDRs perform their function when remaining disordered other IDRs undergo a disorder-to-order transition upon binding their physiological partner(s) [6]. The structural plasticity of IDRs allows them to interact with numerous and structurally distinct targets. This is also why inclusion of IDRs is one of the features of hub proteins that interact with large number of proteins in protein-protein interaction networks [7,8]. As such, proteins with IDRs are crucial for cellular functions that involve multiple partner interactions, such as molecular assembly, molecular recognition, signal transduction, cell cycle and cell death regulation as well as transcription and translation [6,9-19]. Moreover, on an applied side, proteins with IDRs are associated with various human diseases [20] and constitute attractive drug targets [21,22].

Several databases provide access to experimentally annotated IDRs. They include DisProt [23], the database of curated and functionally annotated IDRs, IDEAL [24], which features information about binding partners of IDRs, and Protein Data Bank (PDB) [25], where IDRs correspond to the regions with missing coordinates in crystal structures and the highly structurally diverse regions in the NMR structures, see Figure 1 for the example of the latter conformational diversity. However, these databases cover only a very small portion of the proteins sequences in nature. Current version 7.05 of DisProt covers 803 proteins [23], most recent version 20 of the IDEAL resource includes 913 proteins [24], and a recent article

estimated the number of proteins with IDRs that can be collected from PDB to be 25,833 [26]. These are rather small protein sets compared to the number of currently known proteins sequences that are included in the UniProt resource [27], which as of January 21, 2019 already includes 139.7 million proteins. The large and continually growing annotation gap can be alleviated with computational methods that provide accurate prediction of IDRs in the input protein sequences.




```
Residue number 1        10        20        30        40        50        60        70
Sequence       MSFLKSFPPPGSAEGLRQQQPETEAVLNGKGLGTGTLYIAESRLSWLDGSGLGFSLEYPTISLHAVSRDLNAYPREHLY
DSSP           XXXXXXXXXXXXXXXX-B-SSEEEEBTTB---EEEEEEETTEEEEEETTTEEEEESS---SEEE--SS-SS--S--EE
Native IDRs    DDDDDDDDDDDDDDDD                                                  DDDDDDDDDD
Putative IDRs  DDDDDDDDDDDDDDDDDDDDDDDDDDDDDD
ID propensity  77789888777766665555444444443333322222221111111111111121111111111112222222222222

Residue number 80        90        100       110       120       130       140       150
Sequence       VMVNAKFGEESKESVAEEEDSDDDVEPIAEFRFVPSDKSALEAMFTAMCECQALHPDPEDEDSDDYDGEEYDVEAHEQGQ
DSSP           EEEE----SS-SSS---S--SS--S-SEEEEEEEES-HHHHHHHHHHHHHHHH-XXXXXXXXXXXXXXXXXXXXXXXX
Native IDRs        DDDDDDDDDDDDDDDDDDDDDDD                          DDDDDDDDDDDDDDDDDDDDDDDDD
Putative IDRs    DDDDDDDDDDDDDDDDDDDDDDDDDDD             DDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDD
ID propensity  233445578999999999999998775544333322222333337554556677888999999999999999999998

Residue number 160       170       180       190       200       210       220       230
Sequence       GDIPTFYTYEEGLSHLTAEGQATLERLEGMLSQSVSSQYNMAGVRTEDSTRDYEDGMEVDTTPTVAGQFEDADVDH
DSSP           XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
Native IDRs    DDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDD
Putative IDRs  DDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDD
ID propensity  898888878876776777777667777778888899999999999999999999999999999999999999999999
```

Figure 1. Native intrinsic disorder and putative disorder for the chloride conductance regulatory (ICln) protein (UniProt ID: P35521).

The top of the figure is a cartoon view of 15 superimposed NMR structures obtained from PDB that cover sequence positions Q18 to L133 (PDB ID: 1ZYI). The structures are colored by the secondary structures assigned with the DSSP method where β structures are in golden, helices in red, and coils in white and blue. Bottom of the figure shows the protein sequence together with the residue-level annotations of secondary structure and native and predicted disorder. Detailed assignment of the secondary structure is provided in the "DSSP" line where B is an isolated β-strand residue, E is an extended strand that forms β sheets, G is 3_{10} helix, H is α-helix, I is π helix, S is a bend, T is a hydrogen bonded turn, – denotes residue without specific secondary and tertiary structure, and X means that this residues is missing in the structure. Ordered regions are where all NMR structures converge to the same conformation. There are four IDRs that are annotated with D in the "Native IDRs" line: IDR_1 (residues M1 to R17), IDR_2 (V66 to E76), IDR_3 (A84 to P106), IDR_4 (H134 to H235). IDR_1 and IDR_4 are located at the sequence termini and were annotated as disordered based on REMARK 465 in the PDB structure. IDR_2 and IDR_3 were annotated using the DisProt database (DisProt ID: DP000717), and they correspond to the two regions of conformational ensembles in the above structure, i.e., regions of highly variable structures composed of

coils. The putative disorder was generated with the MFDp method. The "Putative IDRs" line shows binary predictions where D denotes residues predicted as disordered. The "ID propensity" line gives the putative propensities for disorder that range between 1 and 9, with higher value denoting higher likelihood for disorder.

This chapter summarizes popular predictors of IDRs and outlines their applications in the area of systems medicine. Section 2 explains key concepts related to the prediction of IDRs, overviews and categorizes current predictors, and provides practical details for ten popular methods. It also explains outputs that disorder predictors produce and comments on their predictive performance. Section 3 discusses relevance of the intrinsic disorder in the context of systems medicine. In particular, it points to the relevance of IDRs to the understanding and treatments of viral infections, conformational diseases, and cancers. It also remarks on the suitability of disordered proteins as drug targets. Section 4 summarizes the chapter, discusses future research directions and suggests further readings.

2 Methods and tools for prediction of intrinsic disorder

The functional and biophysical properties of IDRs are intrinsic to their sequences. The amino acid composition and conservation of IDRs are distinct from those of the structured regions, and these differences underlie the development of computational predictors of IDRs in protein sequences. These methods were developed using the currently limited number of proteins with experimentally annotated IDRs and can be applied to provide accurate prediction of IDRs in protein sequences that currently lack these annotations. They rely on predictive models that are typically generated using algorithms that optimize parameters and topology of these models to minimize error between their outputs and the native annotation of IDRs for a set of training proteins. After the optimization is finished, the models are empirically tested on set aside (during the model training process) proteins to evaluate their predictive performance. Several community-driven assessments of these models, including the CASP (Critical Assessment of protein Structure Prediction) experiments, show that these models offer accurate predictions [28,29]; we discuss these results later in this chapter.

2.1 Overview of the disorder prediction

The field of computational disorder prediction goes back four decades [30,31]. The first predictor was published in 1979 and aimed to predict random coil conformations [32]. Two notable early methods were developed by Romero, Obradovic and Dunker in 1997 [33] and by Uversky and Fink in 2000 [34]. Inclusion of the disorder prediction in the biannual CASP experiments from the CASP5 in 2002 [35] until CASP10 in 2012 [29] has spurred the development efforts. In total over 60 disorder predictors were developed so far. A complete list of these methods can be compiled from several relevant surveys [36,37,31,38,1].

The disorder predictors are typically categorized into two broad types: 1) methods for which predictive models were produced using machine learning algorithms; and 2) *ab-initio* models. The former approaches utilize predictive models that are parametrized to maximize predictive on disorder-annotated training datasets. This training can be performed using a variety of machine learning algorithms, such as neural networks, support vector machines, and regression, to name a few. The latter models are derived from fundamental biophysical principles that are known to differentiate between the disordered and ordered

regions. They typically take form of scoring functions. A couple of representative examples from this group are GlobPlot [39] and IUPred [40,41]. The machine learning-based group of method is substantially larger with several illustrative examples that include PONDR predictors [42-47], RONN [48], DisEMBL [49], DISpro [50,51], Disoclust3 [52], DISOPRED [53,54], SPINE-D [55], DeepCNF-D [56], SPOT-Disorder [57], PrDOS [58], and SPOT-Disorder-Single [59]. Recent designs of disorder predictors rely on meta-architectures which combine outputs produced by several predictors, either via a majority vote consensus or a separate predictive model. They are motivated by empirical studies showing that such meta-predictors improve predictive performance when compared to the results produced by their input single predictors [60,61]. Example meta-predictors include MD [62], MetaDisorder [63], disCoP [61], DisMeta [64], CSpritz [65], MFDp [66-68], MFDp2 [69], DISOPRED3 [70], and ESpritz [71].

2.2 Popular disorder predictors

We provide details for ten popular disorder predictors that can be accessed and used online. The popularity of these predictors is quantified based on their citation data shown in Table 1. The ten predictors secure relatively high annual citations counts that range between 15 and 110. Two predictors, IUPred and DisEMBL, have received over 1100 citations each. The median total number of citations across the ten methods is 197, suggesting that they are heavily utilized by the community. The ten highly cited predictors include several methods that were originally developed in early 2000s including DISOPRED, DisEMBL, PONDR and IUPred, and a few methods that were published after 2010, such as Espritz, SPINE-D, disCoP, and SPOT-Disorder. They can be accessed via the websites listed in Table 1.

Table 2 summarizes information concerning predictive architecture and performance for the ten disorder predictors. These computational tools rely a wide range of different predictive inputs. All 10 methods use the inputs generated directly from the protein sequence, which typically include composition, physiochemical properties, propensity for disorder and position of amino acids. Six of the 10 methods use sequence-derived evolutionary profiles, which are usually generated with the PSI-BLAST algorithm [72]. A few tools also utilize structural properties such as solvent accessibility, secondary structure, backbone angles and B-factors that are predicted from the input sequence. The three meta-predictors (MFDp, ESpritz and disCoP) by definition use the predicted disorder as one of their inputs. PrDOS also performs alignment to a database of template proteins. Table 2 also reveals that all but one of the 10 methods utilize machine learning models. The only one popular *ab-initio* predictor is IUPred. The high popularity of this tool (Table 1) is driven in part by its short runtime resulting from the use of a simple predictive model and sequence only-based input. Table 2 shows that the most frequently used machine learning models are neural networks and support vector machines. Overall, this analysis reveals that the described here methods make use of a wide range of predictive architectures.

Table 1. Availability and citation counts for the ten popular disorder predictors that are available online.

Name	References	Year most cited article published	Number of citations	Annual number of citations	URL
IUPred	[40,41,73]	2005	1533	109.5	http://dis.embl.de/
DisEMBL	[49]	2003	1121	70.0	http://bioinf.cs.ucl.ac.uk/psipred/?disopred=1
DISOPRED	[70,53,54]	2015	201	50.2	http://www.pondr.com/
PONDR	[43-47]	2006	583	44.8	https://iupred2a.elte.hu/
PrDOS	[58]	2007	392	32.7	http://prdos.hgc.jp/
ESpritz	[71]	2012	193	27.6	http://biomine.cs.vcu.edu/servers/MFDp
SPOT-Disorder	[57]	2017	39	19.3	http://protein.bio.unipd.it/espritz/
disCoP	[61]	2014	91	18.2	http://sparks-lab.org/SPINE-D/
SPINE-D	[55]	2012	117	16.7	http://biomine.cs.vcu.edu/servers/disCoP/
MFDp	[66,67]	2010	131	14.6	http://sparks-lab.org/server/SPOT-disorder/

Citations were obtained from Google Scholar on January 21, 2019. To avoid duplicate citations, the most cited publication for tools that were published multiple times was used. Methods are sorted by their annual number of citations, which is calculated by dividing the number of citations by the number of years since the corresponding article was published, which is given in the "Year most cited article published" column.

Table 2. Predictive architecture and relative predictive performance for the ten popular disorder predictors that are available online.

Name	References	Inputs	Model type	ML model used	Meta-predictor	Rank in [29]	Rank in [36]	Rank in [74]
DisEMBL	[49]	SEQ	ML	NN	No	Unavailable	14	18
DISOPRED	[70,53,54]	SEQ ECP	ML	NN	No	2	5	15
PONDR	[43-47]	SEQ PDI	ML	SVM+LR	No	Unavailable	3	13
IUPred	[40,41,73]	SEQ	ab-initio	Not applicable	No	Unavailable	6	5
PrDOS	[58]	SEQ ECP ALT	ML	SVM	No	1	Unavailable	Unavailable
MFDp	[66,67]	SEQ ECP PSS PSA PBF PDI	ML	SVM	Yes	3	1	19
ESpritz	[71]	SEQ ECP PDI	ML	NN	Yes	12	Unavailable	1
SPINE-D	[55]	SEQ ECP PSS PSA	ML	NN	No	9	Unavailable	3
disCoP	[61]	SEQ PDI	ML	LR	Yes	Unavailable	Unavailable	Unavailable
SPOT-Disorder	[57]	SEQ ECP PSS PSA	ML	NN	No	Unavailable	Unavailable	Unavailable

Methods are sorted by the date of their first publication. The "Inputs" column gives a list of predictive inputs that include SEQ (composition, physiochemical properties, propensity for disorder and position of amino acids); ECP (evolutionary conservation and PSSM/HMM profile); ALT (alignment to template proteins); PSA (putative solvent accessibility); PSS (putative secondary structure and/or backbone angles); PBF (putative B-factors); and PDI (putative disorder);. The "Model type" column categorizes predictors into machine learning (ML) vs. ab-initio models. The "ML algorithm" column specifies type of the ML algorithm used: LR (logistic regression), NN (neural network) and SVM (support vector machine). The last three columns provide ranking in three recent assessments of predictive performance [29,36,74].

The last three columns in Table 2 overview relative predictive performance of the ten popular methods. This analysis is based on three recent comparative assessments that include CASP10 [29], one of the largest recent assessments that covered 19 methods [36], and a recent assessment that focuses on the prediction of long IDRs [74]. We have to combine results from multiple assessments since each of them relies on a different benchmark dataset and covers a different subset of methods. The ten popular methods include several tools that secure top three rank in at least one of the comparative studies. They are DISOPRED (second in [29]), PONDR (third in [36]), PrDOS (first in [29]), MFDp (first in [36] and third in [29]), ESpritz (first in [74]), and SPINE-D (third in [74]). The predictive performance is quantified and further discussed in Section 2.4. Based on the above analysis, we recommend several methods for the end users. They include the top three predictors from the CASP10 assessment: PrDOS, DISOPRED, and MFDp, the best method to predict long IDRs, ESpritz, and the most cited and fast IUPred.

We also note that pre-computed disorder predictions can be conveniently accessed and downloaded from databases, such as MobiDB [75] and D^2P^2 [76]. These databases provide access to the results generated by several methods without the need to find and run the algorithms. D^2P^2 includes predictions from nine methods for over 10 million proteins from 1765 complete genomes of 1256 distinct organisms. MobiDB comprehensively covers the complete set of about 130 million proteins from UniProt and includes predictions for 10 methods.

2.3 Outputs generated by the disorder predictors

The disorder predictors generate two types of outputs for each amino acid in the input protein sequence: a real-value propensity for disordered conformation and/or a binary prediction (disordered vs. ordered). As example prediction generated with the webserver of the MFDp method [66-68], which is available at http://biomine.cs.vcu.edu/servers/MFDp/, is shown at the bottom of Figure 1 in the "Putative IDR" (binary prediction) and "ID propensity" lines (real-value putative propensity). The binary prediction line shows MPDp predicts three IDRs: M1 to K30 region at the N-terminus, N83 to E109 region, and M123 to H235 region at the C-terminus. They nicely coincide with three of the native IDRs (M1 to R17, A84 to P106, and H134 to H235 regions), while the other native IDR (V66 to E76) was missed by this predictor. The propensity scores provide context for the binary predictions. The predictions of IDRs that have higher scores (5 or above) are assumed to be more reliable. Similarly, residues that are predicted with low scores (1 or below) are assumed to be likely structured. In contrast, amino acids with the putative propensities between 2 and 4 are assumed to be less accurately predicted. The "missed" regions (V66 to E76) is actually associated with the putative propensities = 2 that suggest that the prediction may not be reliable there.

We also note recent efforts to provide quality assessment score that accompany the disorder predictions [77,78]. These scores are used to more accurately annotate residues that are correctly predicted, when compared to using the putative propensities. The QUARTER tool that offers this functionality for ten disorder predictors is available at http://biomine.cs.vcu.edu/servers/QUARTER/ [78].

2.4 Predictive quality of the disorder predictors

Various predictors of intrinsic disorder utilize different training datasets, different information extracted from the input sequence (Table 2), and a variety of different types of predictive models (Table 2) [31,38]. This leads to different predictions for the same input sequence, with some being more accurate than others.

The most popular measure of the predictive quality for the putative propensities is the area under the ROC curve (AUC). AUC ranges between 0.5 (equivalent to random predictions) and 1 (always correct predictions). The binary predictions of disorder are typically assessed using the Matthews correlation coefficient (MCC) that ranges between -1 and 1. Predictions with MCC = 0 are equivalent to a random result, while larger and positive values of MCC correspond to better (more correlated with native annotations of disorder) predictions.

Figure 2 compares predictive performance for the top 10 predictors that have the highest AUC values in the CASP6 and CASP10 experiments. The figure shows the values of the MCC and AUC measures. We do not use the CASP5 results, the first time the disorder predictors was assessed in CASP, since this assessment included only six method and did not include an adequate set of measures of predictive performance (e.g., neither AUC nor MCC were quantified). CASP10 is the last time disorder was evaluated as part of the CASP experiment. The figure shows that majority of the best predictors at the time of CASP6, which is in 2004, offered modest levels of predictive quality, with AUCs ranging between 0.65 and 0.88 and with MCCs between 0.14 and 0.41. The results at CASP10 in 2013 demonstrate that much progress has been made. The AUCs of the top 10 predictors in CASP10 range from 0.86 to 0.91 and MCCs from 0.34 to 0.53. In essence, the best from CASP6 offer predictive performance that is equivalent to the methods at the bottom of the top 10 list in CASP10. This analysis also suggests that the current methods, in particular the top 3 performers at CASP10 (PrDOS-CNF, DISOPRED3 and MFDp) offer very accurate predictions.

The top three disorder predictors in CASP10 rely on the machine learning-derived predictive models. PrDOS-CNF is a new version of the popular PrDOS predictor [58] that uses first order conditional neural field model geared for analysis of linear chains. The inputs to this model include a 27-residues long sliding window that covers the amino acid chain and the corresponding position specific scoring matrix (PSSM) generated with PSI-BLAST [72]. DISOPRED3 [79] is a meta-predictor that combines predictions generated by three older and complementary versions of the DISOPRED methods [53,54] that utilize different machine learning models: neural network, support vector machine and nearest neighbour. The results produced by these methods are processed with a small neural network that has one hidden layer. MFDp [66-68] is also a meta-predictor but it pools results generated by nine disorder predictors: DISOPRED2 [80], IUPred [81], MD [62], Norsnet [82], Ucon [83], SPINE-D [84], GlobPlot [85], DisEMBL [86], and PreDisorder [87]. Besides these inputs, it also uses information extracted from the HMM-based substitution matrix generated with HHpred [88] and putative secondary structure, solvent accessibility and B-factors. These inputs are processed with a simple logistic regression model that outputs propensities for intrinsic disorder. The high predictive quality of these methods can be attributed to either

the use of a sophisticated predictive model, such as the conditional neural field, or the use of a well-designed or comprehensive consensus.

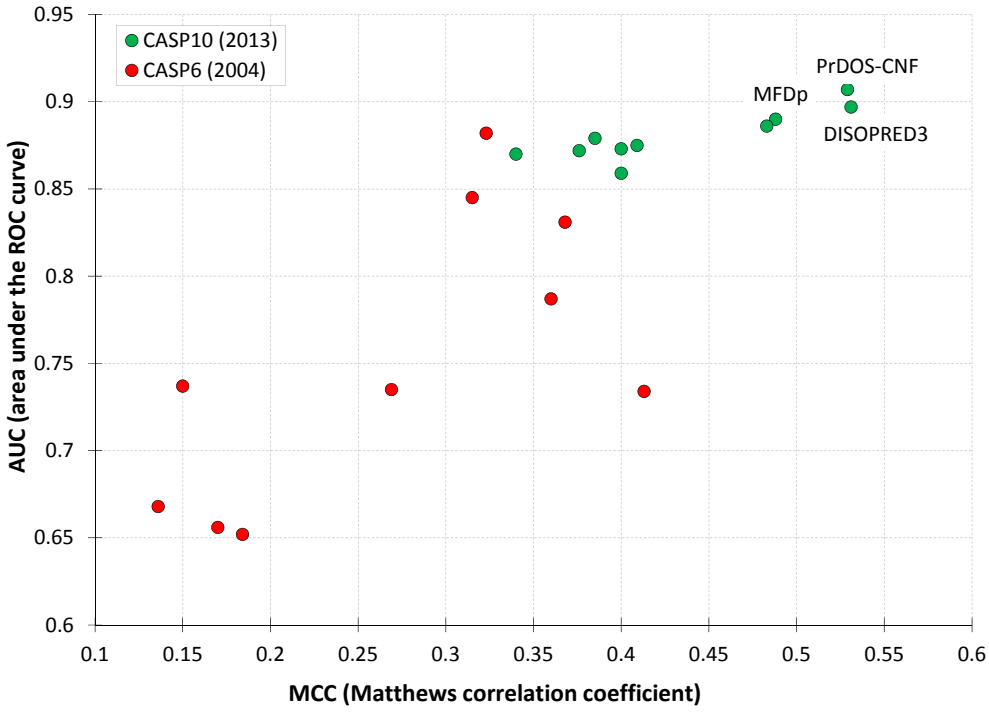


Figure 2. Comparison of the predictive performance for the top 10 predictors from CASP6 and CASP10 experiments.

The predictive performance is quantified with AUC (*y*-axis) for the putative propensities and with MCC (*x*-axis) for the putative binary predictions. The top three methods in the CASP10 experiment are named in the top right corner.

3 Importance of intrinsic disorder predictions for systems medicine

Intrinsic disorder is important for several areas of applied research that are intimately associated with the systems medicine efforts. As one example, recent computational and experimental studies demonstrate that viruses rely heavily on protein with IDRs [89]. Viruses typically have small genomes that code for as few as a dozen proteins that, at the same time, have to interact with many different elements of the host organism, such as nucleic acids, proteins and membranes. Recent computational studies that are supported by accurate predictions of IDRs suggest that many viral proteins contains IDRs or are fully disordered and that a lot of these proteins are involved in the protein-protein, protein-RNA and protein-DNA binding events [90-92]. Consequently, viral proteins are capable of performing multiple interactions through their IDRs to exert the multiple concomitant biological effects. Thus, targeting IDRs within viral proteins to impair critical protein-protein protein-nucleic acids interactions could constitute an appealing antiviral strategy [89]. Interestingly, a recent bioinformatics analysis suggests that antiviral innate immune response of viral hosts also utilizes IDRs [93]. IDRs were found to be common among human antiviral proteins, including major players involved in controlling and regulating the innate antiviral immunity. Proteins with IDRs are engaged in protein-protein

and protein-nucleic acids interactions and are enriched in post-translational modification sites [93], helping us to successfully overcome viral invasions. Correspondingly, we note the availability of several computational tools that accurately predict IDRs that are specifically involved in the protein-protein interactions [73,94-96] and protein-nucleic acids binding [97,94].

Dysfunction of proteins with IDRs is associated with a wide range of human diseases [20]. The main consequences of protein misfolding that leads to the development of various conformational diseases are aggregation and pathological fibrillation (amyloidogenesis), which are fundamentally associated with proteins that are enriched in IDRs. Examples of these conformational maladies are the Alzheimer's disease, Down's syndrome, polyQ diseases, prion diseases, Parkinson's disease, and dementia [98]. Moreover, since proteins with IDRs carry regulatory and signalling functions that often rely on molecular interactions, their misregulation, misinteractions, and missignaling are also linked to several types of cancers, diabetes, and cardiovascular diseases [20]. The support for these observations comes from numerous experimentally characterized disease-related proteins, such as AFP, p53 and BRCA-1 that are involved in cancers, and Aβ and tau proteins that are associated with the Alzheimer's disease, as well as and from comprehensive bioinformatics studies that rely on high-quality predictions of IDRs [20,99]. A more comprehensive understanding of the functional significance of IDRs, which can be supported with the currently available predictive tools, would allow us to gain further insight into molecular-level underpinnings of these diseases.

The disorder-rich proteins such as α-synuclein, tau, p53, and BRCA-1, are attractive targets for drugs modulating protein-protein interactions [20]. This type of drug targeting is a relatively new paradigm that aims to expand the current druggable protein targets by designing new classes of therapeutic agents [100,101]. A recent comprehensive computational analysis of the druggable human proteins that covers a dozen drug classes and close to 20 major classes of drug targets reveals a strong bias towards structured protein targets [102]. This finding is related to the use of rational drug design techniques that rely on protein structures to model protein-drug interactions. Given the substantial enrichment of the intrinsic disorder in the human proteome [4,5], importance of disorder for protein-protein interactions [20], and current bias towards structured drug targets, it is inevitable that proteins with IDRs will raise up on the list of prospective drug targets. Thus, novel strategies for drug discovery efforts that target these proteins are being developed [21,22] and they will undoubtedly benefit from the availability of accurate predictors of IDRs.

4 Summary and further readings

Intrinsic disorder is abundant in proteins and crucial for numerous cellular functions associated with molecular assembly and recognition, signalling, regulation, transcription and translation, to name just a few examples. However, majority of the disordered protein regions remain to be discovered and functionally deciphered. These annotation efforts can be effectively supported with the current predictors of disordered regions and disorder functions [31]. We survey several well-cited methods for the prediction of IDRs in protein sequences, focusing on practical aspects related to their availability, impact, outputs and

predictive performance. We also demonstrate that the leading methods provide accurate predictions. Moreover, we discuss impact of these predictions on the systems medicine field. We postulate that a more comprehensive knowledge of disorder will open new frontiers for systems medicine. It will allow us to decipher mechanisms underlying viral infections and the corresponding immune responses, attain a more complete understanding of several human diseases, and will contribute to elucidation of novel drug targets and therapeutics.

While there are many disorder predictors and modern methods offer accurate results, there is still room for further improvements and development. One potential avenue is to apply deep neural network models that were recently shown to provide promising results in several related bioinformatics area [103,104]. Three recent examples of such disorder predictors are DeepCNF-D that utilizes weighted deep convolutional neural fields [56], SPOT-Disorder that uses deep recurrent neural network [57] and SPOT-Disorder-Single that applies an ensemble of deep recurrent and convolutional neural networks [59]. Another, arguably more impactful research direction is the development of methods that target prediction of specific types of functional IDRs. Several of these tools were released in recent years including methods that predict protein-binding IDRs [73,94-96,79,105-109], nucleic acids-binding IDRs [97,94] and disordered linker regions [110]. Lastly, IDRs are known to be multifunctional [111]. While as many as 37% of the functionally annotated IDRs in the Disprot database have multiple functions [31], so far only one predictor of these regions, DMRpred [112], was developed. Numerous other functions of IDRs have no associated predictive tools and some of the currently covered function would benefit from availability of more accurate predictors. Thus, we anticipate further growth in the area.

Readers interested in additional information would benefit from a recently published in-depth survey of predictors of disorder and disorder functions [31]. We also recommend the special issue on “Intrinsically Disordered Proteins” that was published in the *Chemical Reviews* journal in 2014, which includes comprehensive reviews on topics related to importance of IDRs in human diseases [113], in viral proteomes [89] and in protein-protein interactions [16].

Acknowledgment

This research was supported in part by the National Science Foundation (grant 1617369), the National Natural Science Foundation of China (grants 61728211 and 61832019), the 111 Project (grant B18059), and the Robert J. Mattauch Endowment funds.

References

1. Lieutaud P, Ferron F, Uversky AV, Kurgan L, Uversky VN, Longhi S (2016) How disordered is my protein and what is its disorder for? A guide through the "dark side" of the protein universe. Intrinsically Disord Proteins 4 (1):e1259708.
2. Dunker AK, Babu MM, Barbar E, Blackledge M, Bondos SE, Dosztányi Z, Dyson HJ, Forman-Kay J, Fuxreiter M, Gsponer J, Han K-H, Jones DT, Longhi S, Metallo SJ, Nishikawa K, Nussinov R, Obradovic Z, Pappu RV, Rost B, Selenko P, Subramaniam V, Sussman JL, Tompa P, Uversky VN

(2013) What's in a name? Why these proteins are intrinsically disordered. Intrinsically Disordered Proteins 1 (1):e24157.
3. Habchi J, Tompa P, Longhi S, Uversky VN (2014) Introducing Protein Intrinsic Disorder. Chemical Reviews 114 (13):6561-6588.
4. 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.
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. Yan J, Dunker AK, Uversky VN, Kurgan L (2016) Molecular recognition features (MoRFs) in three domains of life. Mol Biosyst 12 (3):697-710.
7. Hu G, Wu Z, Uversky VN, Kurgan L (2017) Functional Analysis of Human Hub Proteins and Their Interactors Involved in the Intrinsic Disorder-Enriched Interactions. Int J Mol Sci 18 (12).
8. Haynes C, Oldfield CJ, Ji F, Klitgord N, Cusick ME, Radivojac P, Uversky VN, Vidal M, Iakoucheva LM (2006) Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. Plos Computational Biology 2 (8):890-901.
9. Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol 6 (3):197-208.
10. 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.
11. Liu J, Perumal NB, Oldfield CJ, Su EW, Uversky VN, Dunker AK (2006) Intrinsic disorder in transcription factors. Biochemistry 45 (22):6873-6888.
12. 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.
13. 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.
14. Wang C, Uversky VN, Kurgan L (2016) Disordered nucleiome: Abundance of intrinsic disorder in the DNA- and RNA-binding proteins in 1121 species from Eukaryota, Bacteria and Archaea. Proteomics 16 (10):1486-1498.
15. Meng F, Na I, Kurgan L, Uversky VN (2015) Compartmentalization and Functionality of Nuclear Disorder: Intrinsic Disorder and Protein-Protein Interactions in Intra-Nuclear Compartments. Int J Mol Sci 17 (1).
16. Fuxreiter M, Toth-Petroczy A, Kraut DA, Matouschek A, Lim RY, Xue B, Kurgan L, Uversky VN (2014) Disordered proteinaceous machines. Chem Rev 114 (13):6806-6843.
17. Na I, Meng F, Kurgan L, Uversky VN (2016) Autophagy-related intrinsically disordered proteins in intra-nuclear compartments. Mol Biosyst 12 (9):2798-2817.
18. Uversky AV, Xue B, Peng Z, Kurgan L, Uversky VN (2013) On the intrinsic disorder status of the major players in programmed cell death pathways. F1000Res 2:190.
19. 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.
20. Uversky VN, Oldfield CJ, Dunker AK (2008) Intrinsically disordered proteins in human diseases: introducing the D2 concept. Annu Rev Biophys 37:215-246.
21. 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.
22. Uversky VN (2012) Intrinsically disordered proteins and novel strategies for drug discovery. Expert Opin Drug Discov 7 (6):475-488.
23. 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 D1:D219-D227.

24. Fukuchi S, Amemiya T, Sakamoto S, Nobe Y, Hosoda K, Kado Y, Murakami SD, Koike R, Hiroaki H, Ota M (2014) IDEAL in 2014 illustrates interaction networks composed of intrinsically disordered proteins and their binding partners. Nucleic Acids Res 42 (Database issue):D320-325.
25. 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.
26. Walsh I, Giollo M, Di Domenico T, Ferrari C, Zimmermann O, Tosatto SC (2015) Comprehensive large-scale assessment of intrinsic protein disorder. Bioinformatics 31 (2):201-208.
27. The UniProt C (2017) UniProt: the universal protein knowledgebase. Nucleic Acids Res 45 (D1):D158-D169.
28. Jin Y, Dunbrack RL, Jr. (2005) Assessment of disorder predictions in CASP6. Proteins 61 Suppl 7:167-175.
29. 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.
30. 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.
31. Meng F, Uversky VN, Kurgan L (2017) Comprehensive review of methods for prediction of intrinsic disorder and its molecular functions. Cell Mol Life Sci 74 (17):3069-3090.
32. Williams RJP (1979) THE CONFORMATION PROPERTIES OF PROTEINS IN SOLUTION. Biological Reviews 54 (4):389-437.
33. 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.
34. 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.
35. Melamud E, Moult J (2003) Evaluation of disorder predictions in CASP5. Proteins 53 Suppl 6:561-565.
36. Peng ZL, Kurgan L (2012) Comprehensive comparative assessment of in-silico predictors of disordered regions. Curr Protein Pept Sci 13 (1):6-18.
37. 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.
38. Meng F, Uversky V, Kurgan L (2017) Computational Prediction of Intrinsic Disorder in Proteins. Curr Protoc Protein Sci 88:2 16 11-12 16 14.
39. Linding R, Russell RB, Neduva V, Gibson TJ (2003) GlobPlot: exploring protein sequences for globularity and disorder. Nucleic acids research 31 (13):3701-3708.
40. 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.
41. 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.
42. 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.
43. Vucetic S, Brown CJ, Dunker AK, Obradovic Z (2003) Flavors of protein disorder. Proteins: Structure, Function, and Bioinformatics 52 (4):573-584.
44. 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.

45. Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK, Obradovic Z (2005) Optimizing long intrinsic disorder predictors with protein evolutionary information. J Bioinform Comput Biol 3 (1):35-60.
46. 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.
47. Peng K, Radivojac P, Vucetic S, Dunker AK, Obradovic Z (2006) Length-dependent prediction of protein intrinsic disorder. BMC bioinformatics 7 (1):208.
48. 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.
49. Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ, Russell RB (2003) Protein Disorder Prediction: Implications for Structural Proteomics. Structure (London, England : 1993) 11 (11):1453-1459.
50. 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.
51. Hecker J, Yang JY, Cheng J (2008) Protein disorder prediction at multiple levels of sensitivity and specificity. BMC Genomics 9 (1):1-7.
52. 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.
53. 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.
54. 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.
55. 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.
56. 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.
57. Hanson J, Yang Y, Paliwal K, Zhou Y (2017) Improving protein disorder prediction by deep bidirectional long short-term memory recurrent neural networks. Bioinformatics 33 (5):685-692.
58. Ishida T, Kinoshita K (2007) PrDOS: prediction of disordered protein regions from amino acid sequence. Nucleic acids research 35 (suppl 2):W460-W464.
59. Hanson J, Paliwal KK, Zhou Y (2018) Accurate Single-Sequence Prediction of Protein Intrinsic Disorder by an Ensemble of Deep Recurrent and Convolutional Architectures. J Chem Inf Model.
60. Peng Z, Kurgan L (2012) On the complementarity of the consensus-based disorder prediction. Pac Symp Biocomput:176-187.
61. 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.
62. Schlessinger A, Punta M, Yachdav G, Kajan L, Rost B (2009) Improved disorder prediction by combination of orthogonal approaches. PLoS One 4 (2):e4433.
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. Huang YJ, Acton TB, Montelione GT (2014) DisMeta: a meta server for construct design and optimization. Methods Mol Biol 1091:3-16.
65. 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.
66. 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.

67. Mizianty MJ, Uversky V, Kurgan L (2014) Prediction of intrinsic disorder in proteins using MFDp2. Methods Mol Biol 1137:147-162.
68. Mizianty MJ, Peng ZL, Kurgan L (2013) MFDp2: Accurate predictor of disorder in proteins by fusion of disorder probabilities, content and profiles. Intrinsically Disordered Proteins 1 (1):e24428.
69. 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.
70. Jones DT, Cozzetto D (2015) DISOPRED3: precise disordered region predictions with annotated protein-binding activity. Bioinformatics 31 (6):857-863.
71. Walsh I, Martin AJM, Di Domenico T, Tosatto SCE (2012) ESpritz: accurate and fast prediction of protein disorder. Bioinformatics 28 (4):503-509.
72. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25 (17):3389-3402.
73. Meszaros B, Erdos G, Dosztanyi Z (2018) IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. Nucleic Acids Res 46 (W1):W329-W337.
74. Necci M, Piovesan D, Dosztanyi Z, Tompa P, Tosatto SCE (2017) A comprehensive assessment of long intrinsic protein disorder from the DisProt database. Bioinformatics.
75. Piovesan D, Tabaro F, Paladin L, Necci M, Micetic I, Camilloni C, Davey N, Dosztanyi Z, Meszaros B, Monzon AM, Parisi G, Schad E, Sormanni P, Tompa P, Vendruscolo M, Vranken WF, Tosatto SCE (2018) MobiDB 3.0: more annotations for intrinsic disorder, conformational diversity and interactions in proteins. Nucleic Acids Res 46 (D1):D471-D476.
76. 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.
77. Wu Z, Hu G, Wang K, Kurgan L (2017) Exploratory Analysis of Quality Assessment of Putative Intrinsic Disorder in Proteins. 6th International Conference on Artificial Intelligence and Soft Computing, vol LNAI 10245. Zakopane, Poland.
78. Hu G, Wu Z, Oldfield C, Wang C, Kurgan L (2018) Quality Assessment for the Putative Intrinsic Disorder in Proteins. Bioinformatics.
79. Jones DT, Cozzetto D (2015) DISOPRED3: precise disordered region predictions with annotated protein-binding activity. Bioinformatics 31 (6):857-863.
80. 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.
81. Dosztanyi Z, Csizmok V, Tompa P, Simon I (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics 21 (16):3433-3434.
82. Schlessinger A, Liu J, Rost B (2007) Natively unstructured loops differ from other loops. PLoS Comput Biol 3 (7):e140.
83. Schlessinger A, Punta M, Rost B (2007) Natively unstructured regions in proteins identified from contact predictions. Bioinformatics 23 (18):2376-2384.
84. 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. J Biomol Struct Dyn 29 (4):799-813.
85. Linding R, Russell RB, Neduva V, Gibson TJ (2003) GlobPlot: Exploring protein sequences for globularity and disorder. Nucleic Acids Res 31 (13):3701-3708.
86. 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.
87. Deng X, Eickholt J, Cheng J (2009) PreDisorder: ab initio sequence-based prediction of protein disordered regions. BMC Bioinformatics 10:436.
88. Soding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res 33 (Web Server issue):W244-248.

89. Xue B, Blocquel D, Habchi J, Uversky AV, Kurgan L, Uversky VN, Longhi S (2014) Structural disorder in viral proteins. Chem Rev 114 (13):6880-6911.
90. 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.
91. 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.
92. 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.
93. Xue B, Uversky VN (2014) Intrinsic disorder in proteins involved in the innate antiviral immunity: another flexible side of a molecular arms race. J Mol Biol 426 (6):1322-1350.
94. Peng Z, Kurgan L (2015) High-throughput prediction of RNA, DNA and protein binding regions mediated by intrinsic disorder. Nucleic Acids Res 43 (18):e121.
95. Disfani FM, Hsu WL, 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):i75-83.
96. Malhis N, Jacobson M, Gsponer J (2016) MoRFchibi SYSTEM: software tools for the identification of MoRFs in protein sequences. Nucleic Acids Res.
97. Peng Z, Wang C, Uversky VN, Kurgan L (2017) Prediction of Disordered RNA, DNA, and Protein Binding Regions Using DisoRDPbind. Methods Mol Biol 1484:187-203.
98. Uversky VN (2014) The triple power of D(3): protein intrinsic disorder in degenerative diseases. Front Biosci (Landmark Ed) 19:181-258.
99. Uversky VN, Oldfield CJ, Midic U, Xie H, Xue B, Vucetic S, Iakoucheva LM, Obradovic Z, Dunker AK (2009) Unfoldomics of human diseases: linking protein intrinsic disorder with diseases. BMC Genomics 10 Suppl 1:S7.
100. Makley LN, Gestwicki JE (2013) Expanding the Number of 'Druggable' Targets: Non-Enzymes and Protein-Protein Interactions. Chem Biol Drug Des 81 (1):22-32.
101. Modell AE, Blosser SL, Arora PS (2016) Systematic Targeting of Protein-Protein Interactions. Trends Pharmacol Sci 37 (8):702-713.
102. Hu G, Wu Z, Wang K, Uversky VN, Kurgan L (2016) Untapped Potential of Disordered Proteins in Current Druggable Human Proteome. Current drug targets 17 (10):1198-1205.
103. Cao C, Liu F, Tan H, Song D, Shu W, Li W, Zhou Y, Bo X, Xie Z (2018) Deep Learning and Its Applications in Biomedicine. Genomics Proteomics Bioinformatics 16 (1):17-32.
104. Min S, Lee B, Yoon S (2017) Deep learning in bioinformatics. Brief Bioinform 18 (5):851-869.
105. Malhis N, Gsponer J (2015) Computational identification of MoRFs in protein sequences. Bioinformatics 31 (11):1738-1744.
106. Yan J, Dunker AK, Uversky VN, Kurgan L (2015) Molecular recognition features (MoRFs) in three domains of life. Molecular BioSystems.
107. Dosztányi Z, Mészáros B, Simon I (2009) ANCHOR: web server for predicting protein binding regions in disordered proteins. Bioinformatics 25 (20):2745-2746.
108. Mooney C, Pollastri G, Shields DC, Haslam NJ (2012) Prediction of Short Linear Protein Binding Regions. Journal of Molecular Biology 415 (1):193-204.
109. 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.
110. Meng F, Kurgan L (2016) DFLpred: High-throughput prediction of disordered flexible linker regions in protein sequences. Bioinformatics 32 (12):i341-i350.
111. Tompa P, Szász C, Buday L (2005) Structural disorder throws new light on moonlighting. Trends in Biochemical Sciences 30 (9):484-489.
112. Meng F, Kurgan L (2018) High-throughput prediction of disordered moonlighting regions in protein sequences. Proteins.

113. Uversky VN, Dave V, Iakoucheva LM, Malaney P, Metallo SJ, Pathak RR, Joerger AC (2014) Pathological unfoldomics of uncontrolled chaos: intrinsically disordered proteins and human diseases. Chem Rev 114 (13):6844-6879.

Glossary

AUC: Area under the receiver operating curve which is used to assess predictive performance for numerical (real-valued) predictions.

CASP: Critical Assessment of protein Structure Prediction experiment, a structural-bioinformatics community initiative to assess predictive performance of tool for the prediction of protein structure and disorder

Intrinsic disorder: Lack a stable tertiary structure in protein which typically occurs in specific regions in protein sequence and which is manifested as a highly diverse conformation ensemble.

MCC: Matthews correlation coefficient which is used to assess predictive performance for binary predictions.

Meta-predictor: Predictive architecture that combines outputs produced by several predictors, typically using a majority vote consensus or a separate predictive model.

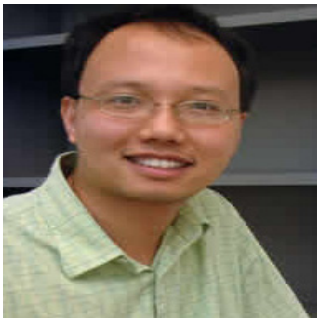
Relevant Websites

IUPred: https://iupred2a.elte.hu/
DisEMBL: http://dis.embl.de/
DISOPRED3: http://bioinf.cs.ucl.ac.uk/psipred/?disopred=1
PONDR: http://www.pondr.com/
PrDOS: http://prdos.hgc.jp/
Espritz: http://protein.bio.unipd.it/espritz/
SPOT-Disorder: http://sparks-lab.org/server/SPOT-disorder/
disCoP: http://biomine.cs.vcu.edu/servers/disCoP/
SPINE-D: http://sparks-lab.org/SPINE-D/
MFDp: http://biomine.cs.vcu.edu/servers/MFDp

Biography and Photo

Lukasz Kurgan received the Ph.D. degree in computer science from the University of Colorado at Boulder in 2003. He is the Robert J. Mattauch Endowed Professor and vice-Chair in the Computer Science Department at the Virginia Commonwealth University. He is Associate Editor-in-Chief of *Biomolecules*, structural bioinformatics area Editor for *BMC Bioinformatics* and serves on the editorial boards of several other journals including *International Journal of Molecular Sciences* and *Current Protein and Peptide Science*. He has published over 140 journal articles in venues that include *Chemical Reviews*, *Nucleic Acids Research*, *Bioinformatics*, *Science Signaling*, *Cellular and Molecular Life Sciences*, and *Briefings in Bioinformatics*. His research interests are in structural bioinformatics of proteins and RNAs, intrinsic disorder, structural genomics, and protein-ligand/drug interactions. His research was funded by NSF, CIHR, NSERC, and Alberta Cancer Foundation. He is a fellow of AIMBE, fellow of the Kosciuszko Foundation Collegium of Eminent Scientists, and a senior member of ACM.

Min Li received the Ph.D. degree in Computer Science from the Central South University, China, in 2008. She is currently the vice dean and a Professor at the School of Computer Science and Engineering, Central South University, Changsha, Hunan, P.R. China. Her research interests include computational biology, systems biology and bioinformatics. She has published over 80 technical papers in refereed journals including Bioinformatics, IEEE/ACM Transactions on Computational Biology and Bioinformatics, Proteomics, and in peer-reviewed conference proceedings such as BIBM, GIW and ISBRA.



Yaohang Li received the B.S. degree in Computer Science and Engineering from the South China University of Technology in 1997, and the M.S. and Ph.D. degrees in computer science from Florida State University, Tallahassee, FL, USA, in 2000 and 2003, respectively. He is an Associate Professor in Computer Science at the Old Dominion University, Norfolk, VA, USA. His research interests are in protein structure modeling, computational biology, bioinformatics, Monte Carlo methods, big data algorithms, and parallel and distributive computing.