Overview Update: Computational Prediction of Intrinsic Disorder in Proteins

Vladimir N. Uversky,[1] and Lukasz Kurgan[2]

[1]Department of Molecular Medicine and USF Health Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, Tampa, FL, USA

[2]Department of Computer Science, Virginia Commonwealth University, Richmond, USA

There are over 100 computational predictors of intrinsic disorder. These methods predict amino acid-level propensities for disorder directly from protein sequences. The propensities can be used to annotate putative disordered residues and regions. This unit provides a practical and holistic introduction to the sequence-based intrinsic disorder prediction. We define intrinsic disorder, explain format of computational prediction of disorder, and identify and describe several accurate predictors. We also introduce recently released databases of intrinsic disorder predictions and use an illustrative example to provide insights into how predictions should be interpreted and combined. Lastly, we summarize key experimental methods that can be used to validate computational predictions.

Keywords: intrinsic disorder ● intrinsically disordered protein ● prediction ● deep learning ● experimental validation

INTRODUCTION

Though many proteins have a well-defined tertiary structure, a large portion of any studied proteome includes intrinsically disordered proteins (IDPs), which range from proteins that are entirely disordered to proteins containing ordered domains and one or more intrinsically disordered regions (IDRs). IDRs lack structure under physiological conditions and take the form of dynamic conformational ensembles (A. Keith Dunker et al., 2013; Habchi, Tompa, Longhi, & Uversky, 2014; Christopher J. Oldfield, Uversky, Dunker, & Kurgan, 2019; van der Lee et al., 2014). Several bioinformatics studies demonstrated that IDPs are abundant in nature, with disordered amino acids comprising 19% of amino acids in eukaryotic proteins, 6% in bacterial proteins, and 4% in archaeal proteins (Z. Peng et al., 2015). Moreover, 30% to 50% of eukaryotic proteins have at least one long intrinsically disordered region (IDR; ≥30 consecutive amino acids (Ward, Sodhi, McGuffin, Buxton, & Jones, 2004; Xue, Dunker, & Uversky, 2012; Yan, Mizianty, Filipow, Uversky, & Kurgan, 2013). IDPs are crucial for many diverse cellular functions (Xie et al., 2007), such as transcription and translation (J. Liu et al., 2006; Z. Peng et al., 2014; Z. L. Peng, Mizianty, Xue, Kurgan, & Uversky, 2012; Staby et al., 2017; Toth-Petroczy et al., 2008), protein-protein interactions (Fuxreiter et al., 2014; Hu, Wu, Uversky, & Kurgan, 2017; V. N. Uversky, 2015c; Vacic et al., 2007; Yan, Dunker, Uversky, & Kurgan, 2016), protein-nucleic acids interactions (Varadi, Zsolyomi, Guharoy, & Tompa, 2015; C. Wang, Uversky, & Kurgan, 2016; Zhao, Katuwawala, Oldfield, Hu, et al., 2021), cell signaling (Bondos, Dunker, & Uversky, 2022; Mitrea & Kriwacki, 2013; V. N. Uversky, Oldfield, & Dunker, 2005), and phase separation (Ibrahim et al., 2023; V. N. Uversky, 2017). Disordered proteins also underly dark proteomes,

which are collections of proteins that are not amenable to experimental structure determination (Hu, Wang, Song, Uversky, & Kurgan, 2018; Kulkarni & Uversky, 2018; V. N. Uversky, 2018). IDPs are linked to various human diseases (Anbo, Sato, Okoshi, & Fukuchi, 2019; Babu, 2016; Kulkarni & Uversky, 2019; V. N. Uversky et al., 2014; V. N. Uversky, Oldfield, & Dunker, 2008) and they were suggested to be attractive targets for drug discovery efforts (Biesaga, Frigole-Vivas, & Salvatella, 2021; A. K. Dunker & Uversky, 2010; Hu, Wu, Wang, Uversky, & Kurgan, 2016; Metallo, 2010; Tsafou, Tiwari, Forman-Kay, Metallo, & Toretsky, 2018; V. N. Uversky, 2012).

There are several databases of IDPs. The largest and oldest database of manually curated and functionally annotated IDRs is DisProt (Piovesan et al., 2017; Quaglia et al., 2022; Sickmeier et al., 2007; Vucetic et al., 2005). It includes about 2500 IDPs and provides functional information for several hundred IDRs. Information on experimentally annotated IDRs can be also collected from the IDEAL database (Fukuchi et al., 2014), which covers information about binding partners of IDPs, and Protein Data Bank (PDB) (Burley et al., 2021), where they correspond to amino acids with missing coordinates in crystal structures or highly flexible residues in nuclear magnetic resonance (NMR) structures (DeForte & Uversky, 2016; Monzon et al., 2020). However, these repositories cover only a small fraction of sequences in nature, given that recent version of the UniProt resource provides access to over 230 million protein sequences (UniProt, 2023). Interestingly, sequences of IDRs are very different when compared to the sequences of structured regions and proteins. Their compositional bias includes enrichment in charged and polar amino acids and depletion in hydrophobic residues (A. K. Dunker et al., 2001; Theillet et al., 2013; V. N. Uversky, 2013, 2015b; V. N. Uversky & Dunker, 2010; V. N. Uversky, Gillespie, & Fink, 2000; R. M. Williams et al., 2001; Yan, Cheng, Kurgan, & Uversky, 2020; B. Zhao & L. Kurgan, 2022a). This bias inspired the development of the TOP-IDP scale that quantifies propensities of amino acids for the disordered state (Campen et al., 2008). The huge number of proteins that lack disorder annotations and the intrinsic compositional bias of IDRs motivate development of computational predictors of disorder. These methods are designed, trained and validated using the experimentally annotated IDRs and IDPs and can be used to predict intrinsic disorder directly from protein sequences. Over 100 disorder predictors have been already developed (Zhao & Kurgan, 2021). Several recent studies review and comparatively assess disorder predictors (Katuwawala & Kurgan, 2020; Katuwawala, Oldfield, & Kurgan, 2020; Kurgan, 2022; Y. Liu, Wang, & Liu, 2019; F. Meng, V. Uversky, & L. Kurgan, 2017a; F. Meng, V. N. Uversky, & L. Kurgan, 2017b; Necci, Piovesan, Dosztanyi, Tompa, & Tosatto, 2018; Necci, Piovesan, Predictors, DisProt, & Tosatto, 2021; Z. L. Peng & L. Kurgan, 2012; Walsh et al., 2015; B. Zhao & L. Kurgan, 2022b; Zhao & Kurgan, 2023b). These studies survey and categorize large collections of methods, analyze and compare their predictive quality, and identify potential future research directions. Disorder predictors are often categorized into four broad classes based on predictive models that they use (Kurgan, 2022; Meng et al., 2017a):

- Scoring function-based methods. These methods compute propensity for intrinsic disorder utilizing a function that takes physiochemical properties of individual amino acid in protein sequences as its inputs. Examples methods in this group include NORSp (Jinfeng Liu & Rost, 2003), GlobPlot (Linding, Russell, Neduva, & Gibson, 2003), IUPred (Dosztányi, Csizmok, Tompa, & Simon, 2005; Dosztányi, Csizmók, Tompa, & Simon, 2005), IUPred2A (Meszaros, Erdos, & Dosztanyi, 2018) and IUPred3 (Erdos, Pajkos, & Dosztanyi, 2021).
- Machine learning-based methods. They produce the disorder propensity using predictive models that are generated by machine learning algorithms, such as neural networks, support vector machines, and random forests. Inputs to these models typically include physiochemical properties of amino acid, evolutionary conservation, and sequence-derived characteristics of the input sequences, such as putative secondary structure and solvent accessibility. Examples

are DisEMBL (Linding, Jensen, et al., 2003), DISOPRED (Jones & Cozzetto, 2015; Jones & Ward, 2003), ESpritz (Walsh, Martin, Di Domenico, & Tosatto, 2012), SPINE-D (T. Zhang et al., 2012), AUCpred (S. Wang, Ma, & Xu, 2016), SPOT-Disorder (Hanson, Paliwal, Litfin, & Zhou, 2019; J. Hanson, Y. Yang, K. Paliwal, & Y. Zhou, 2017), rawMSA (Mirabello & Wallner, 2019), and flDPnn(Hu et al., 2021). Recently published machine learning methods nearly exclusively rely on deep neural networks, primarily motivated by the fact that deep networks produce the most accurate disorder predictions (Bi Zhao & Lukasz Kurgan, 2022)

- Meta/consensus methods. These approaches combine predictions of multiple predictors of intrinsic disorder to provide improved accuracy when compared to using predictors individually. They include MFDp (M. J. Mizianty, Peng, & Kurgan, 2013; Marcin J. Mizianty et al., 2010; M. J. Mizianty, Uversky, & Kurgan, 2014), MetaDisorder (Kozlowski & Bujnicki, 2012), PONDR-FIT (Xue, Dunbrack, Williams, Dunker, & Uversky, 2010). CSpritz (Walsh et al., 2011), DisCoP (Fan & Kurgan, 2014; C. J. Oldfield, Fan, Wang, Dunker, & Kurgan, 2020), and MobiDB-lite (Necci, Piovesan, Dosztanyi, & Tosatto, 2017).
- Hybrid methods. These predictors combine the machine learning approaches with structural modelling, typically using template-based structure predictions. Examples of representative methods in this category are PrDOS (Ishida & Kinoshita, 2007) and Disoclust3 (McGuffin, Atkins, Salehe, Shuid, & Roche, 2015).

This overview complements the current surveys by providing a practical guide to the prediction of intrinsic disorder from protein sequences. It explains the format of the computational prediction of disorder, uses an example prediction to illustrate how to use and understand results produced by disorder predictors, identifies and summarizes a few arguably most accurate and useful predictors, introduces databases of disorder predictions, and discusses experimental methods that can be used to validate the disorder predictions.

PREDICTION OF INTRINSIC DISORDER FROM SEQUENCE

Computational predictors of intrinsic disorder use protein sequence as their sole input. They automate the entire prediction process and generate putative propensity for intrinsic disorder for every amino acid in the input sequence. Typically, this propensity is a positive numeric score where a low value denotes high propensity for a structured conformation and a high value denotes high propensity for the intrinsic disorder. The numeric propensity is usually accompanied by a binary prediction, where an amino acid is categorized as either structured or disordered. The binary prediction is typically derived from the propensity such that the disorder is predicted when the propensity is higher than a pre-defined threshold. While generally propensity values generated by different methods are bound to the unit interval, the thresholds differ across predictors.

We illustrate predictions of intrinsic disorder using the nucleoprotein (also known as nucleocapsid protein, NC, or protein N) from the SARS-CoV-2 virus (UniProt entry: P0DTC9). N protein is one of the major viral proteins playing several significant roles in transcription, and virion assembly of coronaviruses (McBride, van Zyl, & Fielding, 2014). This structurally heterogeneous multidomain RNA-binding protein is found inside the viral envelope, where it binds to and stabilizes the viral genomic RNA forming a ribonucleoprotein (RNP) core required for the RNA encapsidation during viral particle assembly (Chang, Hou, Chang, Hsiao, & Huang, 2014; Chang et al., 2009; Saikatendu et al., 2007). The self-association of the N protein is also responsible for the formation of a shell, the capsid, which protects the genetic material from external agents. The 419 amino acid-long N protein of SARS-CoV-2 shows a high sequence identity of 88.76% and 89.74% with N proteins of Bat CoV and Human SARS N proteins,

respectively. Similar to its SARS-CoV homologue (Chang et al., 2006), this protein includes two functional domains known as N- and C-terminal domains, or NTD and CTD respectively, that are responsible for RNA binding (NTD) and homo-dimerization (CTD). Earlier bioinformatics analysis revealed that the N proteins of coronaviruses contain the highest levels of intrinsic disorder, with N proteins from SARS-COV-2 Human SARS CoV, and Bat CoV being characterized by the mean predicted intrinsic disorder content of 64.91%, 71.09%, and 65.80%, respectively (Giri et al., 2021). Bioinformatics analysis also revealed the presence of three long intrinsically disordered regions in the polypeptide chain, which are believed to be responsible for an intricate mechanism that leads to the regulation of the formation of the RNP complex. They are also engaged in many interactions with other viral proteins or host proteins, as has been demonstrated for the homologous nucleocapsid protein of the CoV that causes SARS (Chang et al., 2014; Giri et al., 2021).

```
>NCAP_SARS2

M,S,D,N,G,P,Q,N,Q,R,N,A,P,R,I,T,F,G,G,P,S,D,S,T,G,S,N,Q,N,G,E,R,S,G,A,R,S,K,Q,R,R,P,Q,G,L,P,N,N,T,A,S,W,F,T,A,L,T,Q,H,G,K,E,D,L,K,F,P,R,
G,Q,G,V,P,I,N,T,N,S,S,P,D,D,Q,I,G,Y,Y,R,R,A,T,R,R,I,R,G,G,D,G,K,M,K,D,L,S,P,R,W,Y,F,Y,Y,L,G,T,G,P,E,A,G,L,P,Y,G,A,N,K,D,G,I,I,W,V,A,T,E,
G,A,L,N,T,P,K,D,H,I,G,T,R,N,P,A,N,N,A,A,I,V,L,Q,L,P,Q,G,T,T,L,P,K,G,F,Y,A,E,G,S,R,G,G,S,Q,A,S,S,R,S,S,S,R,S,R,N,S,S,R,N,S,T,P,G,S,S,R,G,T,S,
P,A,R,M,A,G,N,G,G,D,A,A,L,A,L,L,L,L,D,R,L,N,Q,L,E,S,K,M,S,G,K,G,Q,Q,Q,Q,G,Q,T,V,T,K,K,S,A,A,E,A,S,K,K,P,R,Q,K,R,T,A,T,K,A,Y,N,V,T,Q,
A,F,G,R,R,G,P,E,Q,T,Q,G,N,F,G,D,Q,E,L,I,R,Q,G,T,D,Y,K,H,W,P,Q,I,A,Q,F,A,P,S,A,S,A,F,F,G,M,S,R,I,G,M,E,V,T,P,S,G,T,W,L,T,Y,T,G,A,I,K,L,D,
D,K,D,P,N,F,K,D,Q,V,I,L,L,N,K,H,I,D,A,Y,K,T,F,P,P,T,E,P,K,K,D,K,K,K,K,A,D,E,T,Q,A,L,P,Q,R,Q,K,K,Q,Q,T,V,T,L,L,P,A,A,D,L,D,D,F,S,K,Q,L,Q,
Q,S,M,S,S,A,D,S,T,Q,A

1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,
1,1,1,0,0,0,0,0,1,0,1,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,1,1,1,1,0,0,1,1,1,1,1,0,1,0,0,0,0,0,0,1,1,1,1,
1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,
1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,0,1,0,1,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,1,1,1,1,1,1,1,1,1,1,1,1

0.502,0.482,0.47,0.413,0.392,0.426,0.546,0.439,0.413,0.482,0.528,0.528,0.49,0.574,0.496,0.535,0.606,0.69,0.555,0.477,0.502,0.506,0.497,0.523,0.504,0.496,0.4
56,0.5,0.48,0.426,0.475,0.531,0.475,0.474,0.563,0.547,0.511,0.52,0.542,0.552,0.577,0.588,0.59,0.585,0.561,0.576,0.54,0.633,0.625,0.69,0.648,0.66,0.713,0.683,
0.665,0.628,0.555,0.478,0.54,0.552,0.534,0.524,0.531,0.572,0.636,0.593,0.717,0.679,0.765,0.708,0.675,0.619,0.686,0.588,0.609,0.449,0.573,0.545,0.566,0.516,
0.542,0.545,0.656,0.532,0.678,0.634,0.72,0.58,0.491,0.277,0.261,0.232,0.224,0.216,0.351,0.287,0.415,0.355,0.412,0.369,0.219,0.139,0.131,0.12,0.112,0.107,0.0
86,0.09,0.077,0.07,0.08,0.087,0.106,0.118,0.146,0.179,0.214,0.265,0.234,0.202,0.178,0.27,0.168,0.177,0.138,0.12,0.126,0.101,0.085,0.073,0.082,0.089,0.1,0.11
9,0.129,0.153,0.158,0.187,0.181,0.216,0.228,0.419,0.519,0.513,0.567,0.537,0.583,0.487,0.583,0.447,0.492,0.403,0.401,0.258,0.291,0.444,0.35,0.397,0.311,0.39
7,0.25,0.354,0.256,0.214,0.17,0.176,0.185,0.25,0.397,0.516,0.453,0.517,0.535,0.581,0.575,0.631,0.579,0.55,0.502,0.65,0.642,0.638,0.604,0.589,0.552,0.572,0.5
79,0.605,0.599,0.596,0.598,0.586,0.588,0.56,0.575,0.562,0.596,0.576,0.576,0.558,0.554,0.56,0.572,0.559,0.582,0.584,0.619,0.624,0.632,0.574,0.587,0.484,0.51
1,0.476,0.325,0.256,0.187,0.197,0.186,0.168,0.175,0.155,0.168,0.18,0.186,0.206,0.232,0.213,0.207,0.238,0.326,0.411,0.494,0.496,0.502,0.46,0.708,0.743,0.648,
0.623,0.626,0.67,0.595,0.557,0.537,0.525,0.588,0.602,0.531,0.524,0.526,0.539,0.536,0.59,0.524,0.574,0.531,0.537,0.52,0.554,0.657,0.685,0.64,0.575,0.522,0.52
5,0.578,0.486,0.563,0.491,0.434,0.397,0.445,0.468,0.617,0.701,0.598,0.493,0.522,0.535,0.524,0.54,0.531,0.564,0.541,0.578,0.671,0.642,0.565,0.34,0.281,0.212,
0.332,0.194,0.375,0.231,0.284,0.173,0.238,0.357,0.111,0.094,0.12,0.118,0.108,0.101,0.072,0.078,0.091,0.08,0.111,0.083,0.05,0.047,0.056,0.083,0.088,0.084,0.0
81,0.078,0.062,0.059,0.049,0.053,0.066,0.052,0.059,0.043,0.042,0.035,0.045,0.047,0.055,0.062,0.057,0.043,0.047,0.06,0.09,0.107,0.134,0.132,0.144,0.129,0.12
5,0.136,0.154,0.146,0.135,0.106,0.12,0.116,0.103,0.129,0.136,0.107,0.12,0.156,0.15,0.199,0.28,0.224,0.244,0.311,0.424,0.399,0.445,0.553,0.494,0.478,0.507,0.
462,0.587,0.641,0.655,0.724,0.674,0.63,0.649,0.659,0.566,0.595,0.523,0.493,0.373,0.329,0.374,0.279,0.288,0.228,0.317,0.325,0.356,0.395,0.45,0.496,0.549,0.4
63,0.415,0.401,0.346,0.335,0.368,0.335,0.298,0.27,0.292,0.397,0.412,0.42,0.608,0.644,0.627,0.654,0.672,0.621,0.651,0.656,0.636
```

Figure 1 Putative disorder generated by the flDPnn method for the nucleoprotein from the SARS-CoV-2 virus; DisProt ID: DP03212; UniProt ID: P0DTC9. The prediction includes four lines: 1) protein identifier; 2) sequence; 3) binary prediction (1 for disorder and 0 for order); and 4) disorder propensities. Red and green colors represent predictions that correspond to individual putative IDRs. They can be used to identify corresponding residues between lines 3 and 4.

We collected native IDRs for this protein from DisProt, version 9.3 (Quaglia et al., 2022), which were extracted from experimental data aggregated across several recent studies (Cubuk et al., 2021; Guseva et al., 2021; Savastano, Ibanez de Opakua, Rankovic, & Zweckstetter, 2020; Schiavina, Pontoriero, Uversky, Felli, & Pierattelli, 2021). This nucleoprotein has three IDRs: one at the N-terminus (positions 1 to 68; IDR1), one in the center of the chain (positions 172 to 263; IDR2) and one at the C-terminus (positions 362 to 419; IDR3). We show disorder prediction that was generated by the flDPnn method (Hu et al., 2021) directly from the nucleoprotein sequence in Figure 1. The output generated by flDPnn includes the input (protein identifier and the nucleoprotein's sequence) and the disorder prediction that consists of the binary prediction (1 for disorder and 0 for order) and the disorder propensities. Other disorder predictors provide the same results, typically in a similar format. The sequence and the two predictions are in comma-

separable format. The binary prediction is computed from the putative propensities using the threshold of 0.31, i.e., amino acids with propensities > 0.31 are predicted as disordered.

To interpret results produced by the predictors, users should first analyze the binary predictions to extract the corresponding putative IDRs, i.e., segments composed of consecutive disordered residues. Next, each predicted IDR should be assessed using the numeric propensities. Residues with high scores are more likely to be disordered and the corresponding predictions are more likely to be accurate. We suggest averaging the scores of residues in a given putative IDR to quantify the likelihood of the entire region to be correctly identified. On the other hand, low scores can be used to identify structured residues and regions. Predictions with scores close to the threshold (i.e., 0.31 for flDPnn) are arguably less accurate than the predictions with either much higher or much lower scores. We observe that binary predictions in Figure 1 are in relatively good agreement with the location of the native IDRs, i.e., disordered residues are predicted primarily at both termini and in the middle of the sequence. In particular, the two regions predicted at the termini, positions 1 to 89 and positions 408 to 419 have rather high average putative propensities of 0.56 and 0.58 (see green-colored annotations in Figure 1), respectively, and they nicely align with the native IDR1 and IDR3. Similarly, the putative disordered region in the middle of the sequence that spans positions 169 to 215 (see green-colored annotations in Figure 1) and which coincides with the native IDR2, similarly obtains a high average propensity of 0.56. To compare, some of the other predicted disordered regions, including a region between positions 97 and 100 and a region 142 to 153, have lower average propensities of 0.39 and 0.50, respectively, suggesting that these are less accurate predictions. In fact, we note that these regions do not overlap with the three native IDRs that are present in this protein. We further discuss the disorder annotations for this protein and compare them against predictions from several methods, including flDPnn, in the “Consensus-Based Disorder Predictions” section.

SELECTED COMPUTATIONAL PREDICTORS OF INTRINSIC DISORDER

The selection of a suitable disorder predictor is a rather daunting task because over 100 of these tools have been released to date (Zhao & Kurgan, 2021). One arguably compelling option to identify good predictors is to rely on results from community assessments. These assessments are organized by a community of experts where the predictions are evaluated against the ground truth on blind test datasets (i.e., data that are withheld from the authors of the predictors before the assessment) by independent assessors (i.e., assessors do not participate in the competitions). Community assessments are arguably more objective when compared to the comparative studies done by the authors of predictors. The disorder prediction has been included in several community assessments including the biannual CASP experiment between CASP4 in 2000 (Lesk, Lo Conte, & Hubbard, 2001) and CASP10 in 2012 (Monastyrskyy, Kryshtafovych, Moult, Tramontano, & Fidelis, 2014), and more recently in the Critical Assessment of Intrinsic Protein Disorder (CAID) experiment that was published in 2021 (Necci et al., 2021).

We utilize results from CAID to select a group of accurate predictors (Lang & Babu, 2021; Necci et al., 2021). These methods include the top three tools that produce the most accurate binary predictions: flDPnn (Hu et al., 2021), SPOT-Disorder2 (Hanson et al., 2019), and rawMSA (Mirabello & Wallner, 2019); and the top three methods that generate the most accurate putative propensities: flDPnn, rawMSA and ESpritz-DisProt (Walsh et al., 2012). With two methods overlapping between the two lists (flDPnn and rawMSA), which means that they provide high-quality binary and propensity predictions, altogether we identify four accurate tools. We describe these four predictors in the chronological order of their publication and discuss their

place of origin, key architectural characteristics, and several practical aspects, such as inputs, outputs, and availability to end users. The latter aspect considers whether these methods are publicly available to the end users or have to be re-implemented, and discloses the mode of their availability, which includes code and webserver. Each availability option provides certain benefits and drawbacks. The code can be integrated into other/larger bioinformatics platforms and can be applied on a larger scale of hundreds or thousands of proteins, but it has to be run on the user's own hardware and requires sometimes burdensome installation. The webservers are easier to use since predictions are run on the server side and typically do not require the installation of any software by an end user, but are harder to integrate into other platforms and more limited in scale, i.e., webservers typically constrain the input size since they might be used by multiple users and/or for other computations. Users need only a web browser and internet connection to utilize webservers and the results are delivered via the website and/or to a user-provided email.

ESpritz-DisProt (2012)

ESpritz-DisProt (Walsh et al., 2012) was created by Silvio Tosatto's lab at the University of Padua in Italy. This is a machine learning method that relies on bidirectional recursive neural networks. The predictive model consists of four such networks that are trained using different types of inputs including Atchley sequence metrics (Atchley, Zhao, Fernandes, & Druke, 2005), one-hot encoding of the input protein sequence, and multiple sequence alignment profiles generated from the sequence. The results produced by the four networks are averaged. The networks were trained on the dataset that was collected from the DisProt database.

Input: FASTA-formatted amino acid sequence. No limit on the number of input sequences.

Output: Putative binary disorder annotation and propensity scores for each amino acid.

Availability: webserver at *http://old.protein.bio.unipd.it/espritz/*; standalone code at *https://biocomputingup.it/downloads*

SPOT-Disorder2 (2019)

The SPOT-Disorder2 tool (Hanson et al., 2019) was developed by Yaoqi Zhou's group at the Griffith University in Australia. This research team has recently moved to the Shenzhen Bay Laboratory in China. SPOT-Disorder2 has evolved from the SPOT-Disorder1 tool (J. Hanson, Y. D. Yang, K. Paliwal, & Y. Q. Zhou, 2017) and applies a machine learning approach that utilizes deep neural networks. The network architecture is based on the residual convolutional network that uses squeeze-and-excitation residual inception and long short-term memory (LSTM) units. The inputs to the networks are generated from the protein sequence using several other tools, some of which are relatively time-consuming to run. These inputs include the multiple sequence alignment profiles generated using PSI-BLAST (Altschul et al., 1997) and HHblits (Remmert, Biegert, Hauser, & Soding, 2012), and sequence-based predictions of secondary structure, backbone and dihedral angles, solvent accessibility, contact number, and half-sphere exposure produced by the SPIDER2 method (Heffernan et al., 2016; Heffernan et al., 2015). This network was trained on disordered proteins extracted from DisProt and PDB.

Input: Up to 10 FASTA-formatted amino acid sequences for webserver. Sequences cannot be longer than 750 amino acids.

Output: Putative binary disorder annotation and propensity scores for each residue.

Availability: webserver at *http://zhouyq-lab.szbl.ac.cn/servers/*; standalone code at *http://zhouyq-lab.szbl.ac.cn/download/*

rawMSA (2019)

RawMSA (Mirabello & Wallner, 2019) was released by Björn Wallner's lab at the Linköping University in Sweden. This is a machine learning tool that utilizes a deep neural network. The network has a rather complex architecture that consists of an embedding layer followed by two-dimensional convolutional layer, two stacked LSTM bidirectional recurrent layers, and three stacked fully-connected layers. The sole input to this network is a multiple sequence alignment generated from the sequence using HHblits (Remmert et al., 2012). The network was trained on a dataset collected from PDB. Interestingly, the authors demonstrate empirically that a similar predictive architecture can be used to accurately predict other aspects of protein structure, such as secondary structure, solvent accessibility and inter-residue contact maps (Mirabello & Wallner, 2019).

Input: FASTA-formatted amino acid sequence.

Output: Putative binary disorder annotation and propensity scores for each amino acid.

Availability: no webserver; standalone code at *https://bitbucket.org/clami66/rawmsa*

flDPnn (2021)

The flDPnn method (Hu et al., 2021) was designed by Lukasz Kurgan's lab at the Virginia Commonwealth University in USA in collaboration with the bioinformatics groups at the Nankai University in China. This is also a machine learning approach that relies on a deep neural network. However, the network architecture is rather rudimentary, consisting of just four fully connected layers. The innovation behind flDPnn is the network input that is produced with the assistance of several tools that derive/predict a broad range of relevant structural and functional characteristics of proteins from the sequence. This include the secondary structure predicted with the single-sequence version of PSIPRED (Buchan & Jones, 2019), initial disorder prediction (which is refined and improved by flDPnn) generated with IUPred (Dosztanyi, 2018), multiple sequence alignment profiles generated using with PSI-BLAST (Altschul et al., 1997), disordered DNA and RNA binding residues predicted with DisoRDPbind (C. J. Oldfield, Peng, & Kurgan, 2020; Z. Peng & Kurgan, 2015; Z. Peng, Wang, Uversky, & Kurgan, 2017), disordered protein binding residues predicted by DisoRDPbind and fMoRFpred (Yan et al., 2016), and disordered linkers predicted by DFLpred (Meng & Kurgan, 2016). Moreover, the above information is encoded at three levels of aggregation: residue, sliding sequence window and full protein sequence, before it is passed to the neural networks. Importantly, in contrast to SPOT-Disorder2, the above tools that are used to generate inputs are specifically selected to be computationally efficient. The model was trained using data collected from DisProt. Interestingly, a variant of this predictor that applies a logistic regression model instead of the neural network, called flDPlr, is also available and produces only marginally less accurate disorder predictions (Hu et al., 2021).

Input: Up to 20 FASTA-formatted amino acid sequences for the webserver.

Output: Putative binary disorder annotation and propensity scores for each residue.

Availability: webserver at *http://biomine.cs.vcu.edu/servers/flDPnn/*; standalone code at *https://gitlab.com/sina.ghadermarzi/fldpnn* and docker container at *https://gitlab.com/sina.ghadermarzi/fldpnn_docker*

We note that the four methods rely on the same type of the machine learning algorithm, the neural networks. The three more recent methods (i.e., SPOT-Disorder2, rawMSA and flDPnn) utilize deep neural networks that are characterized by inclusion of many (i.e., more than three) layers. The main differences between these methods are the use of different network architectures and different network inputs that are derived from the protein sequence. The observation that the deep learning methods secure favorable predictive performance, which is how we selected the above methods, is supported by a recent study that empirically demonstrates that deep network-based methods statistically outperform other types of models when applied to the disorder prediction (Bi Zhao & Lukasz Kurgan, 2022).

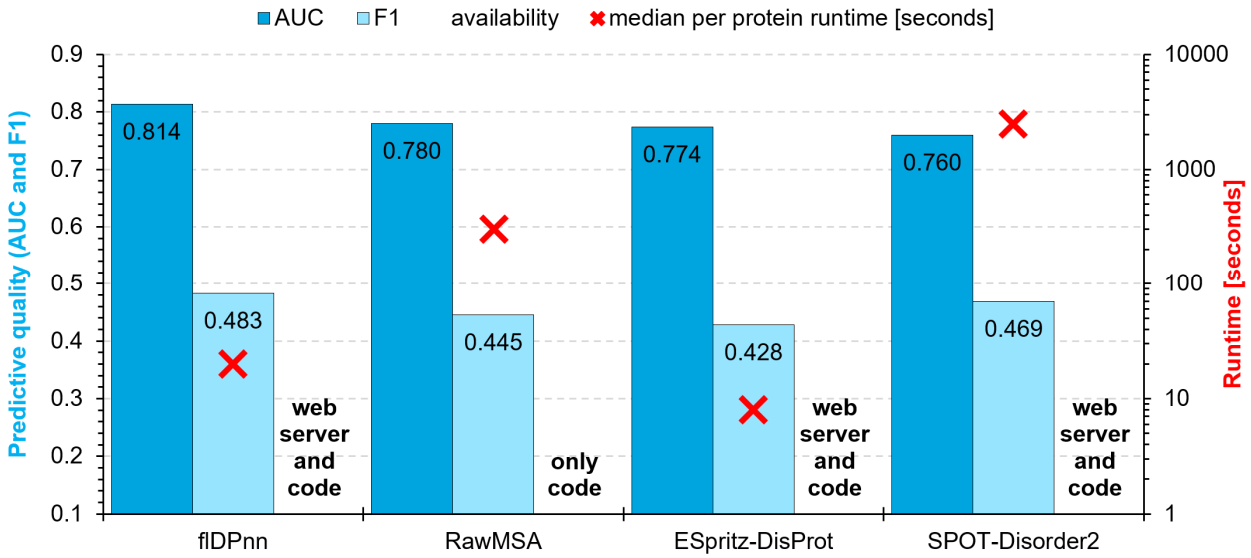

Figure 2 Comparison of predictive quality measured on the DisProt dataset, runtime and availability of the top-performing methods from the CAID experiment: flDPnn, SPOT-Disorder2, rawMSA and ESpritz-DisProt. The blue bars and the right-size axis show the AUC scores that quantify quality of the predicted propensities and F1 values that assess quality of binary predictions. The red x marker is the median per-protein runtime measured in seconds that can be quantified using the logarithmic scale shown on the right *y*-axis.

PREDICTIVE PERFORMANCE, RUNTIME AND AVAILABILITY OF ACCURATE PREDICTORS OF INTRINSIC DISORDER

Similar to other assessments of disorder predictors, the CAID experiment evaluated the quality of predictions using popular metrics including F1 to assess the binary predictions and AUC (area under the ROC curve) for the putative propensities (Necci et al., 2021).

F1, which is the harmonic mean of precision and recall, is defined as (Eqn. 1):

$$F1 = 2 * \frac{precision * recall}{precision + recall} = \frac{2 * TP}{2 * TP + FP + FN}$$

where *TP* (true positives) is the number of correctly predicted disorder residues, *FN* (false negatives) is the number of disorder residues predicted as structured, *FP* (false positives) is the

number of structured residues predicted as disordered, and *TN* (true negatives) is the number of correctly predicted structured residues. F1 ranges between 0 and 1, where larger values correspond to a better predictive performance.

AUC assesses predicted propensity scores by quantifying the area under the receiver operating characteristic (ROC) curve. ROC curve is defined as a relation between true positive rate, $TPR = TP/(TP + FN)$, and false positive rate, $FPR = FP (FP + TN)$. The curve is composed of multiple points that correspond to the *TPR* and *FPR* values computed at different thresholds imposed over the propensity scores, where the amino acids with the propensity scores above (below) the threshold are assumed to be predicted as disordered (structured). AUC values range between 0.5 (i.e., a random predictor) and 1 (i.e., perfect prediction).

CAID also quantifies runtime. This was motivated by the fact that the participating predictors were made available to the CAID organizers who in turn run them on the same computer system. They quantified the runtime per-protein and measured it in seconds. We use the results from CAID to perform side-by-side comparison of the four methods that secured the highest F1 and/or AUC values: flDPnn, SPOT-Disorder2, rawMSA, and ESpritz-DisProt. We summarize predictive quality, runtime and availability for the resulting four predictors in Figure 2.

The method that secures the highest predictive performance, quantified by both F1 and AUC metrics, is flDPnn (Figure 2). The fastest predictor among the four accurate methods is ESpritz-DisProt, which predicts a median size protein sequence in about 8 seconds. To compare, flDPnn computes the prediction for about 20 seconds, rawMSA needs about 5 minutes and SPOT-Disorder2 requires 40 minutes. The runtime differences are very substantial, with over 3 orders of magnitude change between the fastest and the slowest tools. We also highlight a recently published platform for extremely fast prediction of disorder, RIDAO (Dayhoff & Uversky, 2022). This tool predicts a single protein in about 2.5 milliseconds, which is 3 orders of magnitude faster than ESpritz-DisProt. Moreover, flDPnn, ESpritz-DisProt and SPOT-Disorder2 are conveniently available as both webservers and source code, while rawMSA does not offer the webserver option. Altogether, results from CAID suggest that the arguably best option to predict disorder is flDPnn. This analysis is in line with a commentary article for the CAID experiment where the authors conclude (Lang & Babu, 2021): "*SPOT-Disorder2 and fIDPnn, followed by RawMSA and AUCpreD, are consistently good. However, fIDPnn is at least an order of magnitude faster than its competitors, and it succeeded on all sequences, whereas SPOT-Disorder2 skipped 5% of sequences as a result of a length limitation. This might make fIDPnn the overall winner of CAID*". The skipping is due to the fact that SPOT-Disorder2 cannot predict sequences that are over 750 amino acids long. However, scenarios where predictions are needed for very large datasets of dozens of thousands or millions of proteins may require using faster methods, such as RIDAO.

DATABASES OF INTRINSIC DISORDER PREDICTIONS

Users can also employ databases that provide fast and convenient access to pre-computed predictions of intrinsic disorder, typically generated by multiple methods. These databases are particularly useful when collecting predictions for larger datasets of protein, like protein families and proteomes, and when collecting results from multiple predictors. Using predictors directly requires substantially more time and effort since predictions can be computationally costly and results from multiple methods have to be collected one at the time and may need to be re-formatted to combine them together. However, databases are limited to a specific list of proteins

that they include, whereas predictors generate putative disorder for any sequence provided by a user, including novel sequences.

There are three databases of disorder predictions (Zhao & Kurgan, 2023a):

1. MobiDB (database of protein disorder and mobility annotations) (Di Domenico, Walsh, Martin, & Tosatto, 2012; Piovesan et al., 2023) that includes 219.74 million proteins from UniProt and is available online at *https://mobidb.bio.unipd.it/*
2. D^2P^2 (**D**atabase of **D**isorder **P**rotein **P**redictions) (Oates et al., 2013) that covers 10.43 million proteins from 1,256 organisms and is available online at *https://d2p2.pro/*
3. DescribePROT (**D**atabas**e** of **s**truc**t**u**r**e and funct**i**on residue-**b**ased pr**e**dictions of **PROT**eins) (Zhao, Katuwawala, Oldfield, Dunker, et al., 2021) that has 2.26 million proteins from 273 popular/model organisms and can be accessed online at *http://biomine.cs.vcu.edu/servers/DESCRIBEPROT/*

These three databases provide access to predictions for individual proteins in two formats: a text format that can be parsed and in an interactive graphical format. They also offer convenient options to download predictions for whole proteomes. MobiDB and D^2P^2 facilitate instantaneous retrieval of results produced by 10 and 9 disorder predictors, respectively. One drawback of the MobiDB resource is that it provides only the binary predictions, with no propensity scores, while D^2P^2 and DescribePROT provide both types of prediction outputs. Furthermore, MobiDB includes experimental annotations of disorder from a large selection of relevant databases, such as DisProt, PDB, IDEAL, FuzDB (Hatos, Monzon, Tosatto, Piovesan, & Fuxreiter, 2022), MFIB (Ficho, Remenyi, Simon, & Meszaros, 2017), ELM (Kumar et al., 2020), DIBS (Schad et al., 2018), and PhaSepDB (Hou et al., 2023), while D^2P^2 is linked to the experimental disorder data from DisProt and IDEAL (Fukuchi et al., 2014). While covering the smallest number of proteins, DescribePROT delivers predictions for a wide variety of structural and functional aspects of proteins. Besides the disorder predictions, it stores predictions of solvent accessibility, secondary structure, disordered linkers, DNA binding, RNA binding, protein binding, signal peptides, and pre-computed multiple sequence alignment profiles. In total, DescribePROT provides 13.5 billion amino acid level predictions.

CONSENSUS-BASED DISORDER PREDICTIONS

How should users interpret disorder prediction results that are produced by different methods and that may disagree? We suggest a consensus approach, where the final prediction is determined by a majority of the results generated by the applied methods. The consensus binary prediction can be computed as a simple majority vote, i.e., a given amino acid is assumed disordered if most methods predict it as disordered, otherwise it is predicted as structured. If the binary prediction for a given residue is disordered (or structured), the consensus-based propensity can be computed by averaging normalized propensities from methods that predict the residue as disordered (or structured). The consensus approach incurs a higher cost since it requires running multiple tools but it typically results in a higher predictive performance when compared to using methods individually, especially if methods used in the consensus perform well individually. This is supported by a number of studies that empirically demonstrate that consensus-based predictors obtain higher predictive quality (Fan & Kurgan, 2014; Necci et al., 2017; Z. Peng & L. Kurgan, 2012). One relatively low-cost option is to collect multiple predictions from the above databases. We note that in fact both databases that offer access to multiple disorder predictions, D^2P^2 and MobiDB, generate a consensus prediction. MobiDB computes this consensus using the MobiDB-

lite algorithm (Necci et al., 2017) while D^2P^2 applies the 75% consensus approach, i.e., an amino acid is predicted as disordered if at least 75% of methods predicts it as disordered.

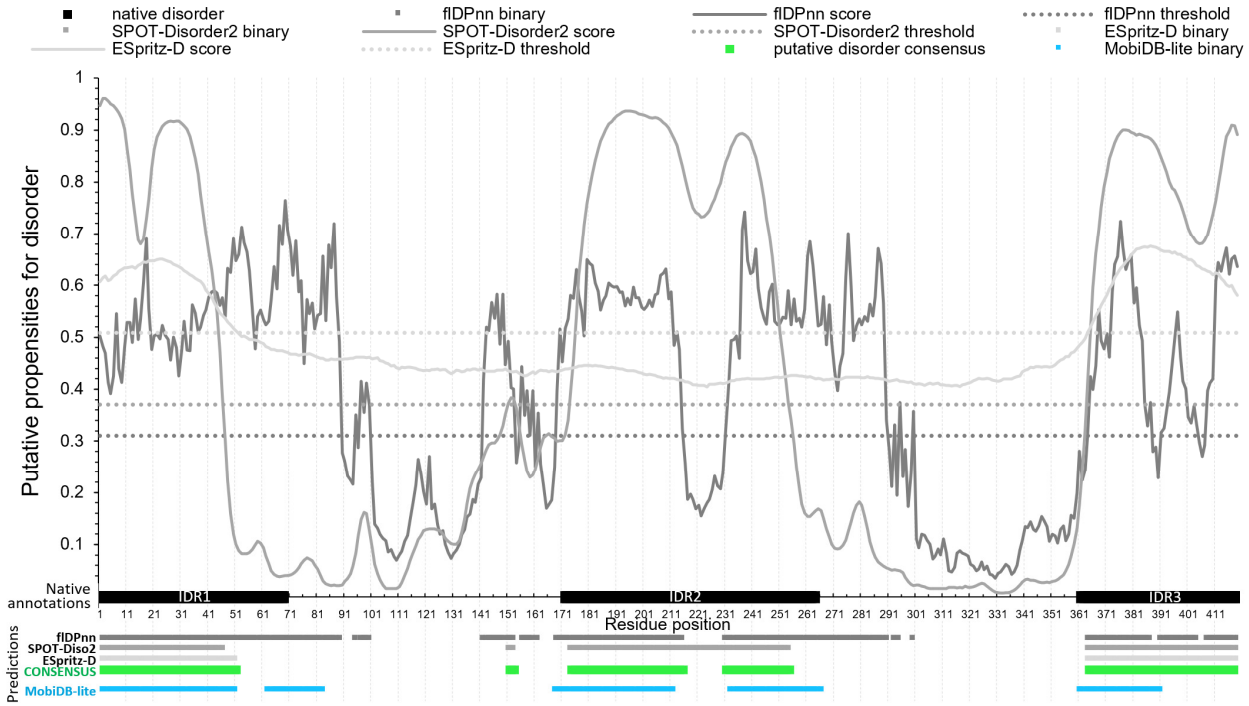

Figure 3 Visualization of native disorder annotations and disorder predictions generated by flDPnn, SPOT-Disorder2, ESpritz-DisProt, and MobiDB-lite methods for the nucleoprotein from the SARS-CoV-2 virus; UniProt ID: P0DTC9. The *x*-axis denotes the protein sequence. The black horizontal bars on the *x*-axis show the native IDRs that were collected from DisProt; DisProt ID: DP03212. Plots above the *x*-axis show the propensity scores; higher propensity values indicate higher likelihood for disorder. Propensities generated by flDPnn, SPOT-Disorder2, and ESpritz-DisProt are shown using dark grey, grey and light grey lines, respectively. The dotted horizontal lines represent the thresholds that these three tools use to generate binary predictions (0.509 for ESpritz-DisProt; 0.37 for SPOT-Disorder2; 0.31 for flDPnn), i.e., amino acids with propensities above the threshold are categorized as disordered while the remaining residues are assumed to be structured. Horizontal bars below the *x*-axis show the binary disorder predictions using the same colors as the propensity lines. The green horizontal bar represents a consensus binary prediction computed from the binary predictions of flDPnn, SPOT-Disorder2, and ESpritz-DisProt. This consensus prediction is based on a majority vote, i.e., residues are assumed disordered if most methods predict them as disordered, otherwise they are predicted as structured. The blue horizontal bar at the bottom is the consensus binary prediction generated by MobiDB-lite method from the MobiDB database. This prediction does not include the propensity scores.

Figure 3 visualizes the disorder predictions that we collected using webservers of flDPnn, SPOT-Disorder2 and ESpritz-DisProt using sequence of the nucleoprotein, which we showcase in Figure 1. We did not include predictions from rawMSA since it does not have the webserver. The top of Figure 3 shows putative propensities using solid gray lines and the corresponding thresholds using dotted horizontal lines. These thresholds are used to convert the propensities into the binary predictions (amino acids with propensities greater than threshold are predicted as disordered) that are shown at the bottom of the figure using the gray horizontal bars. They denote putative IDRs produced by the three predictors. Using the black horizontal bars on the *x*-axis of Figure 3, we annotate the three native IDRs that we collect from the DisProt database. By comparing the native and the predicted disordered regions, we find that flDPnn identifies all three

native IDRs but also overpredicts disorder near the native IDR1 and IDR2. SPOT-Disorder2 also finds the three native IDRs but slightly underpredicts sizes of IDR1 and IDR2. ESpritz-DisProt more substantially underpredicts disorder by entirely missing the native IDR2. Correspondingly, their predictive quality quantified with F1 is 0.66 for ESpritz-DisProt, 0.80 for flDPnn, and 0.85 for SPOT-Disorder2. These values and our observations suggest that the three methods produce reasonably accurate predictions from the sequence, which would allow users to identify the location of the majority of native disordered residues.

We also compute a simple majority-based consensus of the putative binary predictions generated by flDPnn, SPOT-Disorder2 and ESpritz-DisProt. This consensus is shown using the green horizontal bar at the bottom of Figure 3. This result provides a reasonable balance between the overpredictions produced by flDPnn and the underpredictions generated by the other two methods. We compare this consensus-based prediction against the consensus prediction generated by MobiDB-lite that we collected from the MobiDB database; blue horizontal bar at the bottom of Figure 3. The F1 scores for the green and blue consensuses are 0.88 and 0.80, respectively, demonstrating that both alternatives provide rather accurate predictions. The two consensus-based predictions are largely in agreement, with an exception of the C-terminus where the MobiDB's consensus misses disorder, which is why it secures a slightly lower F1 value. When compared to the three best methods based on the CAID results that we include in Figure 3, the blue MobiDB's consensus matches the predictive quality of flDPnn and is somewhat outperformed by SPOT-Disorder2 while the green majority-based consensus provides a modest improvement (i.e., F1=0.88 vs. F1=0.85 for the most accurate individual predictor). This can be explained by the fact that the green consensus relies on the three most accurate methods while the blue consensus utilizes an assortment of 10 fast and popular predictors, which might not necessarily be as accurate individually. Our observations that rely on one protein are supported by studies that compare consensus-based approaches against their input predictors using large datasets of proteins. For instance, the MobiDB-lite consensus was shown to secure F1=0.34 when compared to the disorder predictors that it uses as inputs that obtain F1 values ranging between 0.08 and 0.24 (Necci et al., 2017). More broadly, this example demonstrates how to understand and potentially combine multiple disorder predictions, conveying the richness and utility of the information that can be obtained from the disorder predictions.

EXPERIMENTAL MEANS FOR THE VALIDATION OF PREDICTED DISORDER

A detailed structural and dynamic characterization of IDPs/IDRs cannot be typically provided by a single tool. This is due to the highly heterogeneous nature of the intrinsic disorder phenomenon, where IDPs can attain highly extended conformations or to remain globally collapsed, where different parts of a protein can be affected by disorder to different degree, with some regions being more (dis)ordered than others, and where disordered structures represent conformational ensembles containing highly dynamic structures interconverting on a number of timescales. This indicates that accurate descriptions of IDPs/IDRs must rely on a multiparametric approach that have to include a wide spectrum of biophysical methods capable of providing information on the overall compactness of IDPs, their conformational stability, shape, residual secondary structure, transient long-range contacts, regions of restricted or enhanced mobility (Schramm et al., 2019; V. N. Uversky, 2015a).

Similar to the outputs of different predictors that either generate information on the overall disorder status of a whole protein molecule (i.e., disorder content predictors, such as DisCon (M. J. Mizianty et al., 2011), RAPID (Yan et al., 2013)) or that provide a per-residue

disorder score (i.e., methods described in this unit), experimental techniques also describe the whole protein or give residue-level information. There are several reviews and books that describe nearly 70 experimental techniques that can be used to characterize intrinsic disorder in proteins (Daughdrill, Pielak, Uversky, Cortese, & Dunker, 2005; Eliezer, 2009; Receveur-Brechot, Bourhis, Uversky, Canard, & Longhi, 2006; V. N. Uversky, 2015a; V.N. Uversky & A.K. Dunker, 2012a, 2012b; V. N. Uversky & A. K. Dunker, 2012; V.N. Uversky & Longhi, 2010). Detailed description of these approaches is outside the scope of this overview. Here we summarize several key techniques: X-ray crystallography, NMR, Small-Angle Scattering (SAS) of X-rays (SAXS) or Neutrons (SANS), single molecule fluorescence resonance energy transfer (smFRET), limited proteolysis, hydrogen-deuterium exchange, ion mobility mass spectrometry (IM-MS), and high-speed atomic force microscopy (HS-AFM). From the viewpoint of natural propensity of an amino acid sequence for intrinsic disorder, these techniques are non-invasive, since their application does not require the introduction of amino acid substitutions, which can affect the predisposition of a protein for intrinsic disorder.

Although X-ray crystallography is traditionally used to describe atomic-level structures of structured proteins, increased flexibility of atoms in structured regions results in high B-factor values (i.e., uncertainty), and high flexibility of atoms in disordered regions causes non-coherent X-ray scatter. As a consequence of the non-coherent X-ray scatter, the corresponding atoms become “invisible,” giving rise to missing electron density regions (Le Gall, Romero, Cortese, Uversky, & Dunker, 2007; Radivojac et al., 2004). Therefore, if a crystal structure of a protein is available, it can be used to validate predictions of disorder by looking for the presence of regions with missing electron density (remark 465) in the corresponding PDB entry. A recent study that analyzed prevalence and meaning of the regions with missing electron density shows that a missing regions represent only a weak indication of intrinsic disorder, and this uncertainty is further aggravated by the presence of ambiguous regions, where more than one structure of the same protein sequence "disagrees" in terms of the presence or absence of missing residues (DeForte & Uversky, 2016). These observations raised an important question on the nature of such ambiguous regions – are they real IDRs (i.e., represents regions with dynamic disorder was caused by continual motion in the protein region), or reflect the existence of static disorder (i.e., the presence of the multiple stable conformations or crystal packing imperfections) originating from experimental conditions and ensembles of structures, or domain wobbling that reflects cooperative movements of a structurally intact unit, which are typically facilitated by a small flexible hinge (DeForte & Uversky, 2016). This study revealed that such structural ambiguity mostly represents a natural consequence of many IDPs/IDRs crystallized under different conditions. Since it was also established that static disorder and wobbling domains are relatively rare, the authors concluded that structural ambiguity arises because many of the corresponding regions were conditionally or partially disordered (DeForte & Uversky, 2016).

NMR spectroscopy is the technique of choice for providing high-resolution, residue-level structural information on intrinsically disordered proteins. In fact, heteronuclear multidimensional NMR can generate precise structural information on IDPs/IDRs via assignment of their resonances, and they can directly measure the mobility of IDRs (Angyan & Gaspari, 2013; Bax & Clore, 2019; B. Brutscher et al., 2015; Camacho-Zarco et al., 2022; Daughdrill et al., 2005; A. K. Dunker & Oldfield, 2015; Dyson & Wright, 2021; Eliezer, 2009; Felli & Pierattelli, 2012, 2014; Gibbs, Cook, & Showalter, 2017; Grudziaz, Zawadzka-Kazimierczuk, & Kozminski, 2018; Jensen et al., 2009; Jensen, Ruigrok, & Blackledge, 2013; Jensen, Salmon, Nodet, & Blackledge, 2010; Jensen, Zweckstetter, Huang, & Blackledge, 2014; Kosol, Contreras-Martos, Cedeno, & Tompa, 2013; Kragelj, Blackledge, & Jensen, 2015; Kragelj, Ozenne, Blackledge, & Jensen, 2013; Milles, Salvi, Blackledge, & Jensen, 2018; Mittag & Forman-Kay, 2007; Murthy & Fawzi, 2020; Nodet et al., 2009; Novacek, Zidek, & Sklenar, 2014; Salmon et al., 2010; Schneider et al., 2012). Recent years evidenced a systematic increase in the number of NMR-based approaches for

the structural characterization of IDPs/IDRs allowing one to look into structures and dynamics of IDPs of increasing size and complexity (Felli & Pierattelli, 2012). This includes the possibility to completely assign the heteronuclear protein resonances by protonless NMR spectroscopy utilizing multidimensional NMR experiments based on ^{13}C direct detection (W. Bermel et al., 2005; Wolfgang Bermel, Bertini, Felli, Piccioli, & Pierattelli, 2006; Bertini, Felli, Gonnelli, Kumar, & Pierattelli, 2011; Felli & Pierattelli, 2012, 2014). Since in comparison with NMR spectra of structured proteins those of IDPs are typically very crowded, a better peak separation can be achieved using high-dimensional NMR experiments, allowing accurate analysis of the study of structure, dynamics, and interactions of IDPs (Bernhard Brutscher et al., 2015; Grudziaz et al., 2018; Kazimierczuk, Stanek, Zawadzka-Kazimierczuk, & Kozminski, 2013). Some additional NMR-based approaches suitable for structural and dynamical characterization of IDPs and IDRs include: 1) solvent paramagnetic relaxation enhancement (sPRE) experiments that provide quantitative experimental information on solvent accessibility of NMR-active nuclei that characterizes structure and dynamics of biomolecular systems (Hocking, Zangger, & Madl, 2013; Lenard, Mulder, & Madl, 2022); 2) the use of the NMR spin relaxation that delivers information-rich, site-specific data reporting on conformational fluctuations occurring throughout the molecule, thereby representing an important means for gaining atomic resolution conformational dynamics of IDPs (Abyzov et al., 2016; Salmon et al., 2010; Salvi, Abyzov, & Blackledge, 2017); 3) the use of the hyperpolarized water as universal sensitivity booster in biomolecular NMR (Hilty, Kurzbach, & Frydman, 2022; Konig et al., 2019); and 4) utilization of the recent advances in solid-state NMR (Siemer, 2020).

Another important development in the utilization of NMR for the structural characterization of IDPs is representation of the dynamic nature of IDPs in a form of conformational ensembles. In fact, generating atomic level visualization of the interconverting species that captures the conformations explored and their physico-chemical properties represents the most accurate approach for showing residual structure of IDPs, which is commonly described as transient/dynamic or expressed in terms of fractional populations (Fu & Vendruscolo, 2015; Kragelj et al., 2015; Kurzbach, Kontaxis, Coudevylle, & Konrat, 2015). One more crucial recent development in this field is in-cell NMR spectroscopy, which offers the possibility to analyze proteins and other biomolecules at the atomic resolution directly in cells (Freedberg & Selenko, 2014; Hansel, Luh, Corbeski, Trantirek, & Dotsch, 2014; Milles et al., 2018; Plitzko, Schuler, & Selenko, 2017; Sciolino, Burz, & Shekhtman, 2019; Selenko, 2019; Theillet et al., 2014).

The structure and dynamics of biomolecules (including IDPs) in solution at low resolution can be probed by SAXS and SANS (Bernado & Svergun, 2012; Cordeiro et al., 2017; Kachala, Valentini, & Svergun, 2015; Kikhney & Svergun, 2015; Receveur-Brechot & Durand, 2012). SAS provides useful information on the size and shape of individual macromolecules or their complexes, can detect structural changes upon the environmental perturbations, such as interactions with other molecules, and can also provide information on the biomolecular dynamics (Bernado & Svergun, 2012). For example, an Ensemble Optimization Method (EOM) considers the co-existence of multiple protein conformations in solution compatible with the scattering data, with the analysis of the selected ensembles providing quantitative information about structural features and flexibility (Bernado & Svergun, 2012). Furthermore, being combined with the high resolution methods of X-ray crystallography and NMR, SAXS, due to its ability to report on the three-dimensional space sampled by disordered states and thereby complement the local information provided by NMR, represents a powerful tool for the quantitative analysis of flexible systems, including IDPs (Bernado & Svergun, 2012; Cordeiro et al., 2017; Kachala et al., 2015; Kikhney & Svergun, 2015; Rodriguez-Zamora, 2020; Sibille & Bernado, 2012).

Another set of techniques uniquely suited for the analysis of the structural flexibility of highly disordered systems is given by single-molecule fluorescence spectroscopy techniques, including smFRET, which are capable of measuring conformations without ensemble averaging (Gomes & Gradinaru, 2017). When combined with computational methods and polymer physics models, smFRET can be used to infer global dimension parameters of IDPs (Gomes & Gradinaru, 2017). Furthermore, the integration of smFRET with the complementary experimental data from NMR and SAXS provides important constraints for molecular simulations and leads to a more complete structural representations of disordered proteins (Gomes & Gradinaru, 2017; Naudi-Fabra, Blackledge, & Milles, 2021).

Both limited proteolysis and hydrogen-deuterium exchange are based on the solvent accessibility of corresponding target sites. A high solvent accessibility of the potential cleavage sites makes non-folded proteins highly susceptible to proteolytic degradation in vitro (Fontana et al., 2004). Limited proteolysis can therefore be used to indirectly confirm the increased conformational flexibility of IDPs and IDRs (Fontana, de Laureto, Spolaore, & Frare, 2012), and thereby confirm the results of a disorder prediction. Similarly, structural information and detailed description of the dynamics of a protein chain can be obtained by measuring the efficiency and rates of incorporation of deuterium into a protein's backbone amide. This is achieved via monitoring hydrogen/deuterium exchange in proteins by mass spectrometry combined with the high performance liquid chromatography (Smith, Deng, & Zhang, 1997). The ability of this technique to distinguish between structured and disordered protein regions by their level of protection against hydrogen/deuterium exchange makes it suitable to detect intrinsic disorder and to validate predictions of disorder (Bobst & Kaltashov, 2012). Another useful mass spectrometry-based tool for the analysis of IDPs/IDRs is IM-MS coupled with the application of electrospray ionization (Jurneczko et al., 2012; Stuchfield & Barran, 2018). This is because this approach can examine absolute conformation(s), populations of conformation(s), and also conformational change (Jurneczko et al., 2012). The spectrum of MS-based approaches is very broad and includes hydrogen/deuterium exchange, native MS, ion-mobility MS, protein footprinting, and chemical cross-linking/MS, which are being combined together to constitute structural MS that complements high resolution structural techniques, such as NMR spectroscopy and X-ray crystallography (Faini, Stengel, & Aebersold, 2016; Sinz, 2018). Chemical cross-linking/MS is of particular interest, since it can provide distance constraints that are imposed by the chemical cross-linker (which consists of two reactive groups separated by a "molecular ruler"; i.e., a spacer of a defined length) on the protein structure and that can serve as a basis for the subsequent computational modeling to derive structural models (Kahraman et al., 2013; Sinz, 2018).

Finally, recent advances in HS-AFM provide a unique opportunity to directly visualize individual IDP molecules in dynamic motion at sub-molecular resolution without altering the dynamic structure of IDPs (Ando, 2022). Importantly, images generated by this technique can be used to estimate the number of amino acids contained in a fully disordered region (Ando, 2022).

We conclude this section by illustrating an agreement between the four computational methods introduced in this chapter and corresponding experimental data using SARS-CoV-2 nucleoprotein as an example. In a dedicated NMR study, the backbone assignment was reported for the two disordered regions of this protein, IDR1 and IDR2 that flank the NTD (Schiavina et al., 2021). This study utilized sequence-specific assignment of the resonances by combining the information available in the 2D ^{13}C-detected spectra with that provided by two 3D experiments, the (H)CBCACON and the (H)CBCANCO, and revealed that residues 1-46 and 181-248 of the analyzed IDR1-NTD-IDR2 (residues 1–248) construct are indeed intrinsically disordered (Schiavina et al., 2021). In fact, 98% of the disordered fragment IDR1 (only the first methionine is missing) and 91% of the fragment IDR2 were assigned in a sequence-specific manner (Schiavina et al., 2021). Furthermore, NMR-based analysis of secondary structure confirmed the

mostly disordered nature of both IDR1 and IDR2, with a moderate propensity of the leucine-rich region (218–232) to sample an α-helical conformation, which is in a good agreement with the bioinformatics analysis reported in Figure 3 showing a high extent of disorder for the two IDR regions and the presence of some structure in the region 216–232 (Schiavina et al., 2021).

CONCLUSIONS

The first predictor of intrinsic disorder was developed over 40 year ago (R. J. Williams, 1979). With dozens of new predictors developed in recent years (Zhao & Kurgan, 2021, 2023b), their predictive performance and availability has substantially improved (Necci et al., 2021). Modern predictors are characterized by sophisticated designs that rely on state-of-the-art machine learning algorithms, such as deep neural networks (Bi Zhao & Lukasz Kurgan, 2022), and are widely available to the users as convenient webservers and standalone code. Their predictions are accurate, with AUC values near 0.80 (Necci et al., 2021). We describe and illustrate inputs, outputs, architectures, predictive performance, and runtimes of several popular and accurate disorder predictors. We recommend accurate and fast disorder predictors. We also discuss how to proceed when combining predictions of different methods and suggest several experimental methods that can be used to validate these predictions. Furthermore, we describe several databases that provide access to the native (determined by structural studies) and putative (determined by computational predictors) annotations of disordered residues.

As these predictive methods and databases mature, research has recently shifted toward the prediction of various functions of disordered regions (Basu, Kihara, & Kurgan, 2023). These functions include protein and peptide binding (Hanson, Litfin, Paliwal, & Zhou, 2020; Katuwawala, Peng, Yang, & Kurgan, 2019; Meszaros et al., 2018; Monzon, Bonato, Necci, Tosatto, & Piovesan, 2021; Z. Peng, Li, Meng, Zhao, & Kurgan, 2023; Sharma, Sharma, Raicar, Tsunoda, & Patil, 2019); nucleic acid binding (Barik et al., 2020; Basu, Gsponer, & Kurgan, 2023; Katuwawala & Kurgan, 2020; Z. Peng & Kurgan, 2015; F. Zhang, Zhao, Shi, Li, & Kurgan, 2022); lipid binding (Dobson & Tusnady, 2021; Katuwawala, Zhao, & Kurgan, 2021); and disordered linkers (Meng & Kurgan, 2016; Z. Peng, Xing, & Kurgan, 2020). This progress is reflected in the recent CAID experiment, which for the first time, included assessment of the predictions of disordered binding regions (Necci et al., 2021).

Conflict of Interest Statement

Authors declare no conflict of interest.

Acknowledgements

This work was funded in part by the National Science Foundation (grants 2125218 and 2146027) and the Robert J. Mattauch Endowment funds to L.K.

LITERATURE CITED

Abyzov, A., Salvi, N., Schneider, R., Maurin, D., Ruigrok, R. W., Jensen, M. R., & Blackledge, M. (2016). Identification of Dynamic Modes in an Intrinsically Disordered Protein Using

Temperature-Dependent NMR Relaxation. *J Am Chem Soc, 138*(19), 6240-6251. doi:10.1021/jacs.6b02424
Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res, 25*(17), 3389-3402.
Anbo, H., Sato, M., Okoshi, A., & Fukuchi, S. (2019). Functional Segments on Intrinsically Disordered Regions in Disease-Related Proteins. *Biomolecules, 9*(3). doi:10.3390/biom9030088
Ando, T. (2022). Functional Implications of Dynamic Structures of Intrinsically Disordered Proteins Revealed by High-Speed AFM Imaging. *Biomolecules, 12*(12). doi:10.3390/biom12121876
Angyan, A. F., & Gaspari, Z. (2013). Ensemble-based interpretations of NMR structural data to describe protein internal dynamics. *Molecules, 18*(9), 10548-10567. doi:10.3390/molecules180910548
Atchley, W. R., Zhao, J., Fernandes, A. D., & Druke, T. (2005). Solving the protein sequence metric problem. *Proc Natl Acad Sci U S A, 102*(18), 6395-6400. doi:10.1073/pnas.0408677102
Babu, M. M. (2016). The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease. *Biochem Soc Trans, 44*(5), 1185-1200. doi:10.1042/BST20160172
Barik, A., Katuwawala, A., Hanson, J., Paliwal, K., Zhou, Y., & Kurgan, L. (2020). DEPICTER: Intrinsic Disorder and Disorder Function Prediction Server. *J Mol Biol, 432*(11), 3379-3387. doi:10.1016/j.jmb.2019.12.030
Basu, S., Gsponer, J., & Kurgan, L. (2023). DEPICTER2: a comprehensive webserver for intrinsic disorder and disorder function prediction. *Nucleic Acids Research*. doi:10.1093/nar/gkad330
Basu, S., Kihara, D., & Kurgan, L. (2023). Computational prediction of disordered binding regions. *Computational and Structural Biotechnology Journal, 21*, 1487-1497. doi:https://doi.org/10.1016/j.csbj.2023.02.018
Bax, A., & Clore, G. M. (2019). Protein NMR: Boundless opportunities. *J Magn Reson, 306*, 187-191. doi:10.1016/j.jmr.2019.07.037
Bermel, W., Bertini, I., Duma, L., Felli, I. C., Emsley, L., Pierattelli, R., & Vasos, P. R. (2005). Complete assignment of heteronuclear protein resonances by protonless NMR spectroscopy. *Angew Chem Int Ed Engl, 44*(20), 3089-3092. doi:10.1002/anie.200461794
Bermel, W., Bertini, I., Felli, I. C., Piccioli, M., & Pierattelli, R. (2006). 13C-detected protonlessNMR spectroscopy of proteins in solution. *Progress in Nuclear Magnetic Resonance Spectroscopy, 48*(1), 25-45.
Bernado, P., & Svergun, D. I. (2012). Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. *Molecular Biosystems, 8*(1), 151-167. doi:10.1039/c1mb05275f
Bertini, I., Felli, I. C., Gonnelli, L., Kumar, M. V. V., & Pierattelli, R. (2011). 13C direct-detection biomolecular NMR spectroscopy in living cells. *Angew Chem Int Ed Engl, 50*(10), 2339-2341. doi:10.1002/anie.201006636
Biesaga, M., Frigole-Vivas, M., & Salvatella, X. (2021). Intrinsically disordered proteins and biomolecular condensates as drug targets. *Curr Opin Chem Biol, 62*, 90-100. doi:10.1016/j.cbpa.2021.02.009
Bobst, C. E., & Kaltashov, I. A. (2012). Localizing flexible regions in proteins using hydrogen-deuterium exchange mass spectrometry. *Methods Mol Biol, 896*, 375-385. doi:10.1007/978-1-4614-3704-8_25
Bondos, S. E., Dunker, A. K., & Uversky, V. N. (2022). Intrinsically disordered proteins play diverse roles in cell signaling. *Cell Commun Signal, 20*(1), 20. doi:10.1186/s12964-022-00821-7
Brutscher, B., Felli, I. C., Gil-Caballero, S., Hosek, T., Kummerle, R., Piai, A., . . . Solyom, Z. (2015). NMR Methods for the Study of Instrinsically Disordered Proteins Structure, Dynamics, and Interactions: General Overview and Practical Guidelines. *Adv Exp Med Biol, 870*, 49-122. doi:10.1007/978-3-319-20164-1_3
Brutscher, B., Felli, I. C., Gil-Caballero, S., Hošek, T., Kümmerle, R., Piai, A., . . . Sólyom, Z. (2015). NMR methods for the study of instrinsically disordered proteins structure, dynamics, and

interactions: general overview and practical guidelines. *Intrinsically disordered proteins studied by NMR spectroscopy*, 49-122.

Buchan, D. W. A., & Jones, D. T. (2019). The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Research, 47*(W1), W402-W407. doi:10.1093/nar/gkz297

Burley, S. K., Bhikadiya, C., Bi, C., Bittrich, S., Chen, L., Crichlow, G. V., . . . Zhuravleva, M. (2021). RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Res, 49*(D1), D437-D451. doi:10.1093/nar/gkaa1038

Camacho-Zarco, A. R., Schnapka, V., Guseva, S., Abyzov, A., Adamski, W., Milles, S., . . . Blackledge, M. (2022). NMR Provides Unique Insight into the Functional Dynamics and Interactions of Intrinsically Disordered Proteins. *Chem Rev, 122*(10), 9331-9356. doi:10.1021/acs.chemrev.1c01023

Campen, A., Williams, R. M., Brown, C. J., Meng, J., Uversky, V. N., & Dunker, A. K. (2008). TOP-IDP-scale: a new amino acid scale measuring propensity for intrinsic disorder. *Protein Pept Lett, 15*(9), 956-963.

Chang, C. K., Hou, M. H., Chang, C. F., Hsiao, C. D., & Huang, T. H. (2014). The SARS coronavirus nucleocapsid protein--forms and functions. *Antiviral Res, 103*, 39-50. doi:10.1016/j.antiviral.2013.12.009

Chang, C. K., Hsu, Y. L., Chang, Y. H., Chao, F. A., Wu, M. C., Huang, Y. S., . . . Huang, T. H. (2009). Multiple nucleic acid binding sites and intrinsic disorder of severe acute respiratory syndrome coronavirus nucleocapsid protein: implications for ribonucleocapsid protein packaging. *J Virol, 83*(5), 2255-2264. doi:10.1128/JVI.02001-08

Chang, C. K., Sue, S. C., Yu, T. H., Hsieh, C. M., Tsai, C. K., Chiang, Y. C., . . . Huang, T. H. (2006). Modular organization of SARS coronavirus nucleocapsid protein. *J Biomed Sci, 13*(1), 59-72. doi:10.1007/s11373-005-9035-9

Cordeiro, T. N., Herranz-Trillo, F., Urbanek, A., Estana, A., Cortes, J., Sibille, N., & Bernado, P. (2017). Structural Characterization of Highly Flexible Proteins by Small-Angle Scattering. *Adv Exp Med Biol, 1009*, 107-129. doi:10.1007/978-981-10-6038-0_7

Cubuk, J., Alston, J. J., Incicco, J. J., Singh, S., Stuchell-Brereton, M. D., Ward, M. D., . . . Holehouse, A. S. (2021). The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. *Nature Communications, 12*(1). doi:10.1038/s41467-021-21953-3

Daughdrill, G. W., Pielak, G. J., Uversky, V. N., Cortese, M. S., & Dunker, A. K. (2005). Natively Disordered Proteins. In J. Buchner & T. Kiefhaber (Eds.), *Handbook of Protein Folding* (pp. 271-353). Weinheim, Germany: Wiley-VCH, Verlag GmbH & Co. KGaA.

Dayhoff, G. W., 2nd, & Uversky, V. N. (2022). Rapid prediction and analysis of protein intrinsic disorder. *Protein Sci, 31*(12), e4496. doi:10.1002/pro.4496

DeForte, S., & Uversky, V. N. (2016). Resolving the ambiguity: Making sense of intrinsic disorder when PDB structures disagree. *Protein Sci, 25*(3), 676-688. doi:10.1002/pro.2864

Di Domenico, T., Walsh, I., Martin, A. J. M., & Tosatto, S. C. E. (2012). MobiDB: a comprehensive database of intrinsic protein disorder annotations. *Bioinformatics, 28*(15), 2080-2081. doi:10.1093/bioinformatics/bts327

Dobson, L., & Tusnady, G. E. (2021). MemDis: Predicting Disordered Regions in Transmembrane Proteins. *Int J Mol Sci, 22*(22). doi:10.3390/ijms222212270

Dosztanyi, Z. (2018). Prediction of protein disorder based on IUPred. *Protein Sci, 27*(1), 331-340. doi:10.1002/pro.3334

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. doi:10.1093/bioinformatics/bti541

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. doi:http://dx.doi.org/10.1016/j.jmb.2005.01.071

Dunker, A. K., Babu, M. M., Barbar, E., Blackledge, M., Bondos, S. E., Dosztányi, Z., . . . Uversky, V. N. (2013). What's in a name? Why these proteins are intrinsically disordered. *Intrinsically Disordered Proteins, 1*(1), e24157. doi:10.4161/idp.24157

Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., . . . Obradovic, Z. (2001). Intrinsically disordered protein. *J Mol Graph Model, 19*(1), 26-59.

Dunker, A. K., & Oldfield, C. J. (2015). Back to the Future: Nuclear Magnetic Resonance and Bioinformatics Studies on Intrinsically Disordered Proteins. *Adv Exp Med Biol, 870*, 1-34. doi:10.1007/978-3-319-20164-1_1

Dunker, A. K., & Uversky, V. N. (2010). Drugs for 'protein clouds': targeting intrinsically disordered transcription factors. *Curr Opin Pharmacol, 10*(6), 782-788. doi:10.1016/j.coph.2010.09.005

Dyson, H. J., & Wright, P. E. (2021). NMR illuminates intrinsic disorder. *Curr Opin Struct Biol, 70*, 44-52. doi:10.1016/j.sbi.2021.03.015

Eliezer, D. (2009). Biophysical characterization of intrinsically disordered proteins. *Curr Opin Struct Biol, 19*(1), 23-30. doi:10.1016/j.sbi.2008.12.004

Erdos, G., Pajkos, M., & Dosztanyi, Z. (2021). IUPred3: prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation. *Nucleic Acids Res, 49*(W1), W297-W303. doi:10.1093/nar/gkab408

Faini, M., Stengel, F., & Aebersold, R. (2016). The Evolving Contribution of Mass Spectrometry to Integrative Structural Biology. *J Am Soc Mass Spectrom, 27*(6), 966-974. doi:10.1007/s13361-016-1382-4

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. doi:10.1080/07391102.2013.775969

Felli, I. C., & Pierattelli, R. (2012). Recent progress in NMR spectroscopy: toward the study of intrinsically disordered proteins of increasing size and complexity. *IUBMB Life, 64*(6), 473-481. doi:10.1002/iub.1045

Felli, I. C., & Pierattelli, R. (2014). Novel methods based on (13)C detection to study intrinsically disordered proteins. *J Magn Reson, 241*, 115-125. doi:10.1016/j.jmr.2013.10.020

Ficho, E., Remenyi, I., Simon, I., & Meszaros, B. (2017). MFIB: a repository of protein complexes with mutual folding induced by binding. *Bioinformatics, 33*(22), 3682-3684. doi:10.1093/bioinformatics/btx486

Fontana, A., de Laureto, P. P., Spolaore, B., & Frare, E. (2012). Identifying disordered regions in proteins by limited proteolysis. *Methods Mol Biol, 896*, 297-318. doi:10.1007/978-1-4614-3704-8_20

Fontana, A., de Laureto, P. P., Spolaore, B., Frare, E., Picotti, P., & Zambonin, M. (2004). Probing protein structure by limited proteolysis. *Acta Biochim Pol, 51*(2), 299-321. doi:035001299

Freedberg, D. I., & Selenko, P. (2014). Live cell NMR. *Annu Rev Biophys, 43*, 171-192. doi:10.1146/annurev-biophys-051013-023136

Fu, B., & Vendruscolo, M. (2015). Structure and Dynamics of Intrinsically Disordered Proteins. *Adv Exp Med Biol, 870*, 35-48. doi:10.1007/978-3-319-20164-1_2

Fukuchi, S., Amemiya, T., Sakamoto, S., Nobe, Y., Hosoda, K., Kado, Y., . . . 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. doi:10.1093/nar/gkt1010

Fuxreiter, M., Toth-Petroczy, A., Kraut, D. A., Matouschek, A., Lim, R. Y., Xue, B., . . . Uversky, V. N. (2014). Disordered proteinaceous machines. *Chem Rev, 114*(13), 6806-6843. doi:10.1021/cr4007329

Gibbs, E. B., Cook, E. C., & Showalter, S. A. (2017). Application of NMR to studies of intrinsically disordered proteins. *Arch Biochem Biophys, 628*, 57-70. doi:10.1016/j.abb.2017.05.008

Giri, R., Bhardwaj, T., Shegane, M., Gehi, B. R., Kumar, P., Gadhave, K., . . . Uversky, V. N. (2021). Understanding COVID-19 via comparative analysis of dark proteomes of SARS-CoV-2, human SARS and bat SARS-like coronaviruses. *Cell Mol Life Sci, 78*(4), 1655-1688. doi:10.1007/s00018-020-03603-x

Gomes, G. N., & Gradinaru, C. C. (2017). Insights into the conformations and dynamics of intrinsically disordered proteins using single-molecule fluorescence. *Biochim Biophys Acta Proteins Proteom, 1865*(11 Pt B), 1696-1706. doi:10.1016/j.bbapap.2017.06.008

Grudziaz, K., Zawadzka-Kazimierczuk, A., & Kozminski, W. (2018). High-dimensional NMR methods for intrinsically disordered proteins studies. *Methods, 148*, 81-87. doi:10.1016/j.ymeth.2018.04.031

Guseva, S., Perez, L. M., Camacho-Zarco, A., Bessa, L. M., Salvi, N., Malki, A., . . . Blackledge, M. (2021). (1)H, (13)C and (15)N Backbone chemical shift assignments of the n-terminal and central intrinsically disordered domains of SARS-CoV-2 nucleoprotein. *Biomol NMR Assign, 15*(2), 255-260. doi:10.1007/s12104-021-10014-x

Habchi, J., Tompa, P., Longhi, S., & Uversky, V. N. (2014). Introducing protein intrinsic disorder. *Chem Rev, 114*(13), 6561-6588. doi:10.1021/cr400514h

Hansel, R., Luh, L. M., Corbeski, I., Trantirek, L., & Dotsch, V. (2014). In-cell NMR and EPR spectroscopy of biomacromolecules. *Angew Chem Int Ed Engl, 53*(39), 10300-10314. doi:10.1002/anie.201311320

Hanson, J., Litfin, T., Paliwal, K., & Zhou, Y. (2020). Identifying molecular recognition features in intrinsically disordered regions of proteins by transfer learning. *Bioinformatics, 36*(4), 1107-1113. doi:10.1093/bioinformatics/btz691

Hanson, J., Paliwal, K. K., Litfin, T., & Zhou, Y. (2019). SPOT-Disorder2: Improved Protein Intrinsic Disorder Prediction by Ensembled Deep Learning. *Genomics Proteomics Bioinformatics, 17*(6), 645-656. doi:10.1016/j.gpb.2019.01.004

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. doi:10.1093/bioinformatics/btw678

Hanson, J., Yang, Y. D., Paliwal, K., & Zhou, Y. Q. (2017). Improving protein disorder prediction by deep bidirectional long short-term memory recurrent neural networks. *Bioinformatics, 33*(5), 685-692. doi:10.1093/bioinformatics/btw678

Hatos, A., Monzon, A. M., Tosatto, S. C. E., Piovesan, D., & Fuxreiter, M. (2022). FuzDB: a new phase in understanding fuzzy interactions. *Nucleic Acids Res, 50*(D1), D509-D517. doi:10.1093/nar/gkab1060

Heffernan, R., Dehzangi, A., Lyons, J., Paliwal, K., Sharma, A., Wang, J. H., . . . Yang, Y. D. (2016). Highly accurate sequence-based prediction of half-sphere exposures of amino acid residues in proteins. *Bioinformatics, 32*(6), 843-849. doi:10.1093/bioinformatics/btv665

Heffernan, R., Paliwal, K., Lyons, J., Dehzangi, A., Sharma, A., Wang, J., . . . Zhou, Y. (2015). Improving prediction of secondary structure, local backbone angles, and solvent accessible surface area of proteins by iterative deep learning. *Sci Rep, 5*, 11476. doi:10.1038/srep11476

Hilty, C., Kurzbach, D., & Frydman, L. (2022). Hyperpolarized water as universal sensitivity booster in biomolecular NMR. *Nat Protoc, 17*(7), 1621-1657. doi:10.1038/s41596-022-00693-8

Hocking, H. G., Zangger, K., & Madl, T. (2013). Studying the structure and dynamics of biomolecules by using soluble paramagnetic probes. *Chemphyschem, 14*(13), 3082-3094. doi:10.1002/cphc.201300219

Hou, C., Wang, X., Xie, H., Chen, T., Zhu, P., Xu, X., . . . Li, T. (2023). PhaSepDB in 2022: annotating phase separation-related proteins with droplet states, co-phase separation partners and other experimental information. *Nucleic Acids Res, 51*(D1), D460-D465. doi:10.1093/nar/gkac783

Hu, G., Katuwawala, A., Wang, K., Wu, Z., Ghadermarzi, S., Gao, J., & Kurgan, L. (2021). flDPnn: Accurate intrinsic disorder prediction with putative propensities of disorder functions. *Nat Commun, 12*(1), 4438. doi:10.1038/s41467-021-24773-7

Hu, G., Wang, K., Song, J., Uversky, V. N., & Kurgan, L. (2018). Taxonomic Landscape of the Dark Proteomes: Whole-Proteome Scale Interplay Between Structural Darkness, Intrinsic Disorder, and Crystallization Propensity. *Proteomics*, e1800243. doi:10.1002/pmic.201800243

Hu, G., Wu, Z., Uversky, V. N., & 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). doi:10.3390/ijms18122761

Hu, G., Wu, Z., Wang, K., Uversky, V. N., & Kurgan, L. (2016). Untapped Potential of Disordered Proteins in Current Druggable Human Proteome. *Curr Drug Targets, 17*(10), 1198-1205.

Ibrahim, A. Y., Khaodeuanepheng, N. P., Amarasekara, D. L., Correia, J. J., Lewis, K. A., Fitzkee, N. C., . . . Whitten, S. T. (2023). Intrinsically disordered regions that drive phase separation form a robustly distinct protein class. *J Biol Chem, 299*(1), 102801. doi:10.1016/j.jbc.2022.102801

Ishida, T., & Kinoshita, K. (2007). PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Research, 35*(suppl 2), W460-W464. doi:10.1093/nar/gkm363

Jensen, M. R., Markwick, P. R., Meier, S., Griesinger, C., Zweckstetter, M., Grzesiek, S., . . . Blackledge, M. (2009). Quantitative determination of the conformational properties of partially folded and intrinsically disordered proteins using NMR dipolar couplings. *Structure, 17*(9), 1169-1185. doi:10.1016/j.str.2009.08.001

Jensen, M. R., Ruigrok, R. W., & Blackledge, M. (2013). Describing intrinsically disordered proteins at atomic resolution by NMR. *Curr Opin Struct Biol, 23*(3), 426-435. doi:10.1016/j.sbi.2013.02.007

Jensen, M. R., Salmon, L., Nodet, G., & Blackledge, M. (2010). Defining conformational ensembles of intrinsically disordered and partially folded proteins directly from chemical shifts. *J Am Chem Soc, 132*(4), 1270-1272. doi:10.1021/ja909973n

Jensen, M. R., Zweckstetter, M., Huang, J. R., & Blackledge, M. (2014). Exploring free-energy landscapes of intrinsically disordered proteins at atomic resolution using NMR spectroscopy. *Chem Rev, 114*(13), 6632-6660. doi:10.1021/cr400688u

Jones, D. T., & Cozzetto, D. (2015). DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics, 31*(6), 857-863. doi:10.1093/bioinformatics/btu744

Jones, D. T., & Ward, J. J. (2003). Prediction of disordered regions in proteins from position specific score matrices. *Proteins: Structure, Function, and Bioinformatics, 53*(S6), 573-578. doi:10.1002/prot.10528

Jurneczko, E., Cruickshank, F., Porrini, M., Nikolova, P., Campuzano, I. D., Morris, M., & Barran, P. E. (2012). Intrinsic disorder in proteins: a challenge for (un)structural biology met by ion mobility-mass spectrometry. *Biochem Soc Trans, 40*(5), 1021-1026. doi:10.1042/BST20120125

Kachala, M., Valentini, E., & Svergun, D. I. (2015). Application of SAXS for the Structural Characterization of IDPs. *Adv Exp Med Biol, 870*, 261-289. doi:10.1007/978-3-319-20164-1_8

Kahraman, A., Herzog, F., Leitner, A., Rosenberger, G., Aebersold, R., & Malmstrom, L. (2013). Cross-link guided molecular modeling with ROSETTA. *PLoS One, 8*(9), e73411. doi:10.1371/journal.pone.0073411

Katuwawala, A., & Kurgan, L. (2020). Comparative Assessment of Intrinsic Disorder Predictions with a Focus on Protein and Nucleic Acid-Binding Proteins. *Biomolecules, 10*(12). doi:10.3390/biom10121636

Katuwawala, A., Oldfield, C. J., & Kurgan, L. (2020). Accuracy of protein-level disorder predictions. *Brief Bioinform, 21*(5), 1509-1522. doi:10.1093/bib/bbz100

Katuwawala, A., Peng, Z., Yang, J., & Kurgan, L. (2019). Computational Prediction of MoRFs, Short Disorder-to-order Transitioning Protein Binding Regions. *Comput Struct Biotechnol J, 17*, 454-462. doi:10.1016/j.csbj.2019.03.013

Katuwawala, A., Zhao, B., & Kurgan, L. (2021). DisoLipPred: Accurate prediction of disordered lipid binding residues in protein sequences with deep recurrent networks and transfer learning. *Bioinformatics*. doi:10.1093/bioinformatics/btab640

Kazimierczuk, K., Stanek, J., Zawadzka-Kazimierczuk, A., & Kozminski, W. (2013). High-dimensional NMR spectra for structural studies of biomolecules. *Chemphyschem, 14*(13), 3015-3025. doi:10.1002/cphc.201300277

Kikhney, A. G., & Svergun, D. I. (2015). A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett, 589*(19 Pt A), 2570-2577. doi:10.1016/j.febslet.2015.08.027

Konig, A., Scholzel, D., Uluca, B., Viennet, T., Akbey, U., & Heise, H. (2019). Hyperpolarized MAS NMR of unfolded and misfolded proteins. *Solid State Nucl Magn Reson, 98*, 1-11. doi:10.1016/j.ssnmr.2018.12.003

Kosol, S., Contreras-Martos, S., Cedeno, C., & Tompa, P. (2013). Structural characterization of intrinsically disordered proteins by NMR spectroscopy. *Molecules, 18*(9), 10802-10828. doi:10.3390/molecules180910802

Kozlowski, L. P., & Bujnicki, J. M. (2012). MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins. *BMC Bioinformatics, 13*(1), 1-11. doi:10.1186/1471-2105-13-111

Kragelj, J., Blackledge, M., & Jensen, M. R. (2015). Ensemble Calculation for Intrinsically Disordered Proteins Using NMR Parameters. *Adv Exp Med Biol, 870*, 123-147. doi:10.1007/978-3-319-20164-1_4

Kragelj, J., Ozenne, V., Blackledge, M., & Jensen, M. R. (2013). Conformational propensities of intrinsically disordered proteins from NMR chemical shifts. *Chemphyschem, 14*(13), 3034-3045. doi:10.1002/cphc.201300387

Kulkarni, P., & Uversky, V. N. (2018). Intrinsically Disordered Proteins: The Dark Horse of the Dark Proteome. *Proteomics, 18*(21-22). doi:ARTN 1800061

10.1002/pmic.201800061

Kulkarni, P., & Uversky, V. N. (2019). Intrinsically Disordered Proteins in Chronic Diseases. *Biomolecules, 9*(4). doi:10.3390/biom9040147

Kumar, M., Gouw, M., Michael, S., Samano-Sanchez, H., Pancsa, R., Glavina, J., . . . Gibson, T. J. (2020). ELM-the eukaryotic linear motif resource in 2020. *Nucleic Acids Res, 48*(D1), D296-D306. doi:10.1093/nar/gkz1030

Kurgan, L. (2022). Resources for computational prediction of intrinsic disorder in proteins. *Methods, 204*, 132-141. doi:10.1016/j.ymeth.2022.03.018

Kurzbach, D., Kontaxis, G., Coudevylle, N., & Konrat, R. (2015). NMR Spectroscopic Studies of the Conformational Ensembles of Intrinsically Disordered Proteins. *Adv Exp Med Biol, 870*, 149-185. doi:10.1007/978-3-319-20164-1_5

Lang, B., & Babu, M. M. (2021). A community effort to bring structure to disorder. *Nature Methods, 18*(5), 454-455. doi:10.1038/s41592-021-01123-5

Le Gall, T., Romero, P. R., Cortese, M. S., Uversky, V. N., & Dunker, A. K. (2007). Intrinsic disorder in the Protein Data Bank. *J Biomol Struct Dyn, 24*(4), 325-342. doi:10.1080/07391102.2007.10507123

Lenard, A. J., Mulder, F. A. A., & Madl, T. (2022). Solvent paramagnetic relaxation enhancement as a versatile method for studying structure and dynamics of biomolecular systems. *Prog Nucl Magn Reson Spectrosc, 132-133*, 113-139. doi:10.1016/j.pnmrs.2022.09.001

Lesk, A. M., Lo Conte, L., & Hubbard, T. J. P. (2001). Assessment of novel fold targets in CASP4: Predictions of three-dimensional structures, secondary structures, and interresidue contacts. *Proteins-Structure Function and Bioinformatics*, 98-118.

Linding, R., Jensen, L. J., Diella, F., Bork, P., Gibson, T. J., & Russell, R. B. (2003). Protein Disorder Prediction: Implications for Structural Proteomics. *Structure, 11*(11), 1453-1459. doi:http://dx.doi.org/10.1016/j.str.2003.10.002

Linding, R., Russell, R. B., Neduva, V., & Gibson, T. J. (2003). GlobPlot: exploring protein sequences for globularity and disorder. *Nucleic Acids Research, 31*(13), 3701-3708. doi:10.1093/nar/gkg519

Liu, J., Perumal, N. B., Oldfield, C. J., Su, E. W., Uversky, V. N., & Dunker, A. K. (2006). Intrinsic disorder in transcription factors. *Biochemistry, 45*(22), 6873-6888. doi:10.1021/bi0602718

Liu, J., & Rost, B. (2003). NORSp: predictions of long regions without regular secondary structure. *Nucleic Acids Research, 31*(13), 3833-3835. doi:10.1093/nar/gkg515

Liu, Y., Wang, X., & Liu, B. (2019). A comprehensive review and comparison of existing computational methods for intrinsically disordered protein and region prediction. *Brief Bioinform, 20*(1), 330-346. doi:10.1093/bib/bbx126

McBride, R., van Zyl, M., & Fielding, B. C. (2014). The coronavirus nucleocapsid is a multifunctional protein. *Viruses, 6*(8), 2991-3018. doi:10.3390/v6082991

McGuffin, L. J., Atkins, J. D., Salehe, B. R., Shuid, A. N., & Roche, D. B. (2015). IntFOLD: an integrated server for modelling protein structures and functions from amino acid sequences. *Nucleic Acids Research, 43*(W1), W169-W173. doi:10.1093/nar/gkv236

Meng, F., & Kurgan, L. (2016). DFLpred: High-throughput prediction of disordered flexible linker regions in protein sequences. *Bioinformatics, 32*(12), i341-i350. doi:10.1093/bioinformatics/btw280

Meng, F., Uversky, V., & Kurgan, L. (2017a). Computational Prediction of Intrinsic Disorder in Proteins. *Curr Protoc Protein Sci, 88*, 2 16 11-12 16 14. doi:10.1002/cpps.28

Meng, F., Uversky, V. N., & Kurgan, L. (2017b). Comprehensive review of methods for prediction of intrinsic disorder and its molecular functions. *Cell Mol Life Sci, 74*(17), 3069-3090. doi:10.1007/s00018-017-2555-4

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. doi:10.1093/nar/gky384

Metallo, S. J. (2010). Intrinsically disordered proteins are potential drug targets. *Curr Opin Chem Biol, 14*(4), 481-488. doi:10.1016/j.cbpa.2010.06.169

Milles, S., Salvi, N., Blackledge, M., & Jensen, M. R. (2018). Characterization of intrinsically disordered proteins and their dynamic complexes: From in vitro to cell-like environments. *Prog Nucl Magn Reson Spectrosc, 109*, 79-100. doi:10.1016/j.pnmrs.2018.07.001

Mirabello, C., & Wallner, B. (2019). rawMSA: End-to-end Deep Learning using raw Multiple Sequence Alignments. *PLoS One, 14*(8), e0220182. doi:10.1371/journal.pone.0220182

Mitrea, D. M., & Kriwacki, R. W. (2013). Regulated unfolding of proteins in signaling. *FEBS Lett, 587*(8), 1081-1088. doi:10.1016/j.febslet.2013.02.024

Mittag, T., & Forman-Kay, J. D. (2007). Atomic-level characterization of disordered protein ensembles. *Curr Opin Struct Biol, 17*(1), 3-14. doi:10.1016/j.sbi.2007.01.009

Mizianty, M. J., Peng, Z., & Kurgan, L. (2013). MFDp2: Accurate predictor of disorder in proteins by fusion of disorder probabilities, content and profiles. *Intrinsically Disord Proteins, 1*(1), e24428. doi:10.4161/idp.24428

Mizianty, M. J., Stach, W., Chen, K., Kedarisetti, K. D., Disfani, F. M., & Kurgan, L. (2010). Improved sequence-based prediction of disordered regions with multilayer fusion of multiple information sources. *Bioinformatics, 26*(18), i489-i496. doi:10.1093/bioinformatics/btq373

Mizianty, M. J., Uversky, V., & Kurgan, L. (2014). Prediction of intrinsic disorder in proteins using MFDp2. *Methods Mol Biol, 1137*, 147-162. doi:10.1007/978-1-4939-0366-5_11

Mizianty, M. J., Zhang, T., Xue, B., Zhou, Y., Dunker, A. K., Uversky, V. N., & Kurgan, L. (2011). In-silico prediction of disorder content using hybrid sequence representation. *BMC Bioinformatics, 12*, 245. doi:10.1186/1471-2105-12-245

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. doi:10.1002/prot.24391

Monzon, A. M., Bonato, P., Necci, M., Tosatto, S. C. E., & Piovesan, D. (2021). FLIPPER: Predicting and Characterizing Linear Interacting Peptides in the Protein Data Bank. *J Mol Biol, 433*(9), 166900. doi:10.1016/j.jmb.2021.166900

Monzon, A. M., Necci, M., Quaglia, F., Walsh, I., Zanotti, G., Piovesan, D., & Tosatto, S. C. E. (2020). Experimentally Determined Long Intrinsically Disordered Protein Regions Are Now Abundant in the Protein Data Bank. *Int J Mol Sci, 21*(12). doi:10.3390/ijms21124496

Murthy, A. C., & Fawzi, N. L. (2020). The (un)structural biology of biomolecular liquid-liquid phase separation using NMR spectroscopy. *J Biol Chem, 295*(8), 2375-2384. doi:10.1074/jbc.REV119.009847

Naudi-Fabra, S., Blackledge, M., & Milles, S. (2021). Synergies of Single Molecule Fluorescence and NMR for the Study of Intrinsically Disordered Proteins. *Biomolecules, 12*(1). doi:10.3390/biom12010027

Necci, M., Piovesan, D., Dosztanyi, Z., Tompa, P., & Tosatto, S. C. E. (2018). A comprehensive assessment of long intrinsic protein disorder from the DisProt database. *Bioinformatics, 34*(3), 445-452. doi:10.1093/bioinformatics/btx590

Necci, M., Piovesan, D., Dosztanyi, Z., & Tosatto, S. C. E. (2017). MobiDB-lite: fast and highly specific consensus prediction of intrinsic disorder in proteins. *Bioinformatics, 33*(9), 1402-1404. doi:10.1093/bioinformatics/btx015

Necci, M., Piovesan, D., Predictors, C., DisProt, C., & Tosatto, S. C. E. (2021). Critical assessment of protein intrinsic disorder prediction. *Nature Methods, 18*(5), 472-481. doi:10.1038/s41592-021-01117-3

Nodet, G., Salmon, L., Ozenne, V., Meier, S., Jensen, M. R., & Blackledge, M. (2009). Quantitative description of backbone conformational sampling of unfolded proteins at amino acid resolution from NMR residual dipolar couplings. *J Am Chem Soc, 131*(49), 17908-17918. doi:10.1021/ja9069024

Novacek, J., Zidek, L., & Sklenar, V. (2014). Toward optimal-resolution NMR of intrinsically disordered proteins. *J Magn Reson, 241*, 41-52. doi:10.1016/j.jmr.2013.12.008

Oates, M. E., Romero, P., Ishida, T., Ghalwash, M., Mizianty, M. J., Xue, B., . . . Gough, J. (2013). D(2)P(2): database of disordered protein predictions. *Nucleic Acids Res, 41*(Database issue), D508-516. doi:10.1093/nar/gks1226

Oldfield, C. J., Fan, X., Wang, C., Dunker, A. K., & Kurgan, L. (2020). Computational Prediction of Intrinsic Disorder in Protein Sequences with the disCoP Meta-predictor. *Methods Mol Biol, 2141*, 21-35. doi:10.1007/978-1-0716-0524-0_2

Oldfield, C. J., Peng, Z., & Kurgan, L. (2020). Disordered RNA-Binding Region Prediction with DisoRDPbind. *Methods Mol Biol, 2106*, 225-239. doi:10.1007/978-1-0716-0231-7_14

Oldfield, C. J., Uversky, V. N., Dunker, A. K., & Kurgan, L. (2019). Introduction to intrinsically disordered proteins and regions. In N. Salvi (Ed.), *Intrinsically Disordered Proteins* (pp. 1-34): Academic Press.

Peng, Z., & Kurgan, L. (2012). On the complementarity of the consensus-based disorder prediction. *Pac Symp Biocomput*, 176-187.

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. doi:10.1093/nar/gkv585

Peng, Z., Li, Z., Meng, Q., Zhao, B., & Kurgan, L. (2023). CLIP: accurate prediction of disordered linear interacting peptides from protein sequences using co-evolutionary information. *Brief Bioinform, 24*(1). doi:10.1093/bib/bbac502

Peng, Z., Oldfield, C. J., Xue, B., Mizianty, M. J., Dunker, A. K., Kurgan, L., & Uversky, V. N. (2014). A creature with a hundred waggly tails: intrinsically disordered proteins in the ribosome. *Cell Mol Life Sci, 71*(8), 1477-1504. doi:10.1007/s00018-013-1446-6

Peng, Z., Wang, C., Uversky, V. N., & Kurgan, L. (2017). Prediction of Disordered RNA, DNA, and Protein Binding Regions Using DisoRDPbind. *Methods Mol Biol, 1484*, 187-203. doi:10.1007/978-1-4939-6406-2_14

Peng, Z., Xing, Q., & Kurgan, L. (2020). APOD: accurate sequence-based predictor of disordered flexible linkers. *Bioinformatics, 36*(Supplement_2), i754-i761. doi:10.1093/bioinformatics/btaa808

Peng, Z., Yan, J., Fan, X., Mizianty, M. J., Xue, B., Wang, K., . . . Kurgan, L. (2015). Exceptionally abundant exceptions: comprehensive characterization of intrinsic disorder in all domains of life. *Cell Mol Life Sci, 72*(1), 137-151. doi:10.1007/s00018-014-1661-9

Peng, Z. L., & Kurgan, L. (2012). Comprehensive comparative assessment of in-silico predictors of disordered regions. *Curr Protein Pept Sci, 13*(1), 6-18.

Peng, Z. L., Mizianty, M. J., Xue, B., Kurgan, L., & Uversky, V. N. (2012). More than just tails: intrinsic disorder in histone proteins. *Molecular Biosystems, 8*(7), 1886-1901. doi:10.1039/c2mb25102g

Piovesan, D., Del Conte, A., Clementel, D., Monzon, A. M., Bevilacqua, M., Aspromonte, M. C., . . . Tosatto, S. C. E. (2023). MobiDB: 10 years of intrinsically disordered proteins. *Nucleic Acids Res, 51*(D1), D438-D444. doi:10.1093/nar/gkac1065

Piovesan, D., Tabaro, F., Micetic, I., Necci, M., Quaglia, F., Oldfield, C. J., . . . Tosatto, S. C. (2017). DisProt 7.0: a major update of the database of disordered proteins. *Nucleic Acids Res, 45*(D1), D219-D227. doi:10.1093/nar/gkw1056

Plitzko, J. M., Schuler, B., & Selenko, P. (2017). Structural Biology outside the box-inside the cell. *Curr Opin Struct Biol, 46*, 110-121. doi:10.1016/j.sbi.2017.06.007

Quaglia, F., Meszaros, B., Salladini, E., Hatos, A., Pancsa, R., Chemes, L. B., . . . Piovesan, D. (2022). DisProt in 2022: improved quality and accessibility of protein intrinsic disorder annotation. *Nucleic Acids Res, 50*(D1), D480-D487. doi:10.1093/nar/gkab1082

Radivojac, P., Obradovic, Z., Smith, D. K., Zhu, G., Vucetic, S., Brown, C. J., . . . Dunker, A. K. (2004). Protein flexibility and intrinsic disorder. *Protein Sci, 13*(1), 71-80. doi:10.1110/ps.03128904

Receveur-Brechot, V., Bourhis, J. M., Uversky, V. N., Canard, B., & Longhi, S. (2006). Assessing protein disorder and induced folding. *Proteins, 62*(1), 24-45. doi:10.1002/prot.20750

Receveur-Brechot, V., & Durand, D. (2012). How random are intrinsically disordered proteins? A small angle scattering perspective. *Curr Protein Pept Sci, 13*(1), 55-75. doi:10.2174/138920312799277901

Remmert, M., Biegert, A., Hauser, A., & Soding, J. (2012). HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nature Methods, 9*(2), 173-175. doi:10.1038/nmeth.1818

Rodriguez-Zamora, P. (2020). Conjugation of NMR and SAXS for flexible and multidomain protein structure determination: From sample preparation to model refinement. *Prog Biophys Mol Biol, 150*, 140-144. doi:10.1016/j.pbiomolbio.2019.08.009

Saikatendu, K. S., Joseph, J. S., Subramanian, V., Neuman, B. W., Buchmeier, M. J., Stevens, R. C., & Kuhn, P. (2007). Ribonucleocapsid formation of severe acute respiratory syndrome coronavirus through molecular action of the N-terminal domain of N protein. *J Virol, 81*(8), 3913-3921. doi:10.1128/JVI.02236-06

Salmon, L., Nodet, G., Ozenne, V., Yin, G., Jensen, M. R., Zweckstetter, M., & Blackledge, M. (2010). NMR characterization of long-range order in intrinsically disordered proteins. *J Am Chem Soc, 132*(24), 8407-8418. doi:10.1021/ja101645g

Salvi, N., Abyzov, A., & Blackledge, M. (2017). Atomic resolution conformational dynamics of intrinsically disordered proteins from NMR spin relaxation. *Prog Nucl Magn Reson Spectrosc, 102-103*, 43-60. doi:10.1016/j.pnmrs.2017.06.001

Savastano, A., Ibanez de Opakua, A., Rankovic, M., & Zweckstetter, M. (2020). Nucleocapsid protein of SARS-CoV-2 phase separates into RNA-rich polymerase-containing condensates. *Nat Commun, 11*(1), 6041. doi:10.1038/s41467-020-19843-1

Schad, E., Ficho, E., Pancsa, R., Simon, I., Dosztanyi, Z., & Meszaros, B. (2018). DIBS: a repository of disordered binding sites mediating interactions with ordered proteins. *Bioinformatics, 34*(3), 535-537. doi:10.1093/bioinformatics/btx640

Schiavina, M., Pontoriero, L., Uversky, V. N., Felli, I. C., & Pierattelli, R. (2021). The highly flexible disordered regions of the SARS-CoV-2 nucleocapsid N protein within the 1-248 residue construct: sequence-specific resonance assignments through NMR. *Biomol NMR Assign, 15*(1), 219-227. doi:10.1007/s12104-021-10009-8

Schneider, R., Huang, J. R., Yao, M., Communie, G., Ozenne, V., Mollica, L., . . . Blackledge, M. (2012). Towards a robust description of intrinsic protein disorder using nuclear magnetic resonance spectroscopy. *Molecular Biosystems, 8*(1), 58-68. doi:10.1039/c1mb05291h

Schramm, A., Bignon, C., Brocca, S., Grandori, R., Santambrogio, C., & Longhi, S. (2019). An arsenal of methods for the experimental characterization of intrinsically disordered proteins - How to choose and combine them? *Arch Biochem Biophys, 676*, 108055. doi:10.1016/j.abb.2019.07.020

Sciolino, N., Burz, D. S., & Shekhtman, A. (2019). In-Cell NMR Spectroscopy of Intrinsically Disordered Proteins. *Proteomics, 19*(6), e1800055. doi:10.1002/pmic.201800055

Selenko, P. (2019). Quo Vadis Biomolecular NMR Spectroscopy? *Int J Mol Sci, 20*(6). doi:10.3390/ijms20061278

Sharma, R., Sharma, A., Raicar, G., Tsunoda, T., & Patil, A. (2019). OPAL+: Length-Specific MoRF Prediction in Intrinsically Disordered Protein Sequences. *Proteomics, 19*(6), e1800058. doi:10.1002/pmic.201800058

Sibille, N., & Bernado, P. (2012). Structural characterization of intrinsically disordered proteins by the combined use of NMR and SAXS. *Biochem Soc Trans, 40*(5), 955-962. doi:10.1042/BST20120149

Sickmeier, M., Hamilton, J. A., LeGall, T., Vacic, V., Cortese, M. S., Tantos, A., . . . Dunker, A. K. (2007). DisProt: the Database of Disordered Proteins. *Nucleic Acids Res, 35*(Database issue), D786-793. doi:10.1093/nar/gkl893

Siemer, A. B. (2020). Advances in studying protein disorder with solid-state NMR. *Solid State Nucl Magn Reson, 106*, 101643. doi:10.1016/j.ssnmr.2020.101643

Sinz, A. (2018). Cross-Linking/Mass Spectrometry for Studying Protein Structures and Protein-Protein Interactions: Where Are We Now and Where Should We Go from Here? *Angew Chem Int Ed Engl, 57*(22), 6390-6396. doi:10.1002/anie.201709559

Smith, D. L., Deng, Y., & Zhang, Z. (1997). Probing the non-covalent structure of proteins by amide hydrogen exchange and mass spectrometry. *J Mass Spectrom, 32*(2), 135-146. doi:10.1002/(SICI)1096-9888(199702)32:2<135::AID-JMS486>3.0.CO;2-M

Staby, L., O'Shea, C., Willemoes, M., Theisen, F., Kragelund, B. B., & Skriver, K. (2017). Eukaryotic transcription factors: paradigms of protein intrinsic disorder. *Biochem J, 474*(15), 2509-2532. doi:10.1042/BCJ20160631

Stuchfield, D., & Barran, P. (2018). Unique insights to intrinsically disordered proteins provided by ion mobility mass spectrometry. *Curr Opin Chem Biol, 42*, 177-185. doi:10.1016/j.cbpa.2018.01.007

Theillet, F. X., Binolfi, A., Frembgen-Kesner, T., Hingorani, K., Sarkar, M., Kyne, C., . . . Selenko, P. (2014). Physicochemical properties of cells and their effects on intrinsically disordered proteins (IDPs). *Chem Rev, 114*(13), 6661-6714. doi:10.1021/cr400695p

Theillet, F. X., Kalmar, L., Tompa, P., Han, K. H., Selenko, P., Dunker, A. K., . . . Uversky, V. N. (2013). The alphabet of intrinsic disorder: I. Act like a Pro: On the abundance and roles of proline residues in intrinsically disordered proteins. *Intrinsically Disord Proteins, 1*(1), e24360. doi:10.4161/idp.24360

Toth-Petroczy, A., Oldfield, C. J., Simon, I., Takagi, Y., Dunker, A. K., Uversky, V. N., & Fuxreiter, M. (2008). Malleable machines in transcription regulation: the mediator complex. *PLoS Comput Biol, 4*(12), e1000243. doi:10.1371/journal.pcbi.1000243

Tsafou, K., Tiwari, P. B., Forman-Kay, J. D., Metallo, S. J., & Toretsky, J. A. (2018). Targeting Intrinsically Disordered Transcription Factors: Changing the Paradigm. *J Mol Biol, 430*(16), 2321-2341. doi:10.1016/j.jmb.2018.04.008

UniProt, C. (2023). UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res, 51*(D1), D523-D531. doi:10.1093/nar/gkac1052

Uversky, V. N. (2012). Intrinsically disordered proteins and novel strategies for drug discovery. *Expert Opin Drug Discov, 7*(6), 475-488. doi:10.1517/17460441.2012.686489

Uversky, V. N. (2013). The alphabet of intrinsic disorder: II. Various roles of glutamic acid in ordered and intrinsically disordered proteins. *Intrinsically Disord Proteins, 1*(1), e24684. doi:10.4161/idp.24684
Uversky, V. N. (2015a). Biophysical Methods to Investigate Intrinsically Disordered Proteins: Avoiding an "Elephant and Blind Men" Situation. *Adv Exp Med Biol, 870*, 215-260. doi:10.1007/978-3-319-20164-1_7
Uversky, V. N. (2015b). The intrinsic disorder alphabet. III. Dual personality of serine. *Intrinsically Disord Proteins, 3*(1), e1027032. doi:10.1080/21690707.2015.1027032
Uversky, V. N. (2015c). The multifaceted roles of intrinsic disorder in protein complexes. *FEBS Lett, 589*(19 Pt A), 2498-2506. doi:10.1016/j.febslet.2015.06.004
Uversky, V. N. (2017). Intrinsically disordered proteins in overcrowded milieu: Membrane-less organelles, phase separation, and intrinsic disorder. *Curr Opin Struct Biol, 44*, 18-30. doi:10.1016/j.sbi.2016.10.015
Uversky, V. N. (2018). Bringing Darkness to Light: Intrinsic Disorder as a Means to Dig into the Dark Proteome. *Proteomics, 18*(21-22), e1800352. doi:10.1002/pmic.201800352
Uversky, V. N., Dave, V., Iakoucheva, L. M., Malaney, P., Metallo, S. J., Pathak, R. R., & Joerger, A. C. (2014). Pathological unfoldomics of uncontrolled chaos: intrinsically disordered proteins and human diseases. *Chem Rev, 114*(13), 6844-6879. doi:10.1021/cr400713r
Uversky, V. N., & Dunker, A. K. (2010). Understanding protein non-folding. *Biochim Biophys Acta, 1804*(6), 1231-1264. doi:10.1016/j.bbapap.2010.01.017
Uversky, V. N., & Dunker, A. K. (2012a). *Intrinsically Disordered Protein Analysis: Volume I. Methods and Experimental Tools* (V. N. Uversky & A. K. Dunker Eds. Vol. 895). Totowa, NJ, USA: Humana Press.
Uversky, V. N., & Dunker, A. K. (2012b). *Intrinsically Disordered Protein Analysis: Volume II. Methods and Experimental Tools*. Totowa, NJ, USA: Humana Press.
Uversky, V. N., & Dunker, A. K. (2012). Multiparametric analysis of intrinsically disordered proteins: looking at intrinsic disorder through compound eyes. *Anal Chem, 84*(5), 2096-2104. doi:10.1021/ac203096k
Uversky, V. N., Gillespie, J. R., & Fink, A. L. (2000). Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins, 41*(3), 415-427.
Uversky, V. N., & Longhi, S. (2010). *Instrumental Analysis of Intrinsically Disordered Proteins: Assessing Structure and Conformation* (V. N. Uversky & S. Longhi Eds.). New Jersey, USA: John Wiley & Sons.
Uversky, V. N., Oldfield, C. J., & Dunker, A. K. (2005). Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J Mol Recognit, 18*(5), 343-384. doi:10.1002/jmr.747
Uversky, V. N., Oldfield, C. J., & Dunker, A. K. (2008). Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu Rev Biophys, 37*, 215-246. doi:10.1146/annurev.biophys.37.032807.125924
Vacic, V., Oldfield, C. J., Mohan, A., Radivojac, P., Cortese, M. S., Uversky, V. N., & Dunker, A. K. (2007). Characterization of molecular recognition features, MoRFs, and their binding partners. *J Proteome Res, 6*(6), 2351-2366. doi:10.1021/pr0701411
van der Lee, R., Buljan, M., Lang, B., Weatheritt, R. J., Daughdrill, G. W., Dunker, A. K., . . . Babu, M. M. (2014). Classification of Intrinsically Disordered Regions and Proteins. *Chemical Reviews, 114*(13), 6589-6631. doi:10.1021/cr400525m
Varadi, M., Zsolyomi, F., Guharoy, M., & Tompa, P. (2015). Functional Advantages of Conserved Intrinsic Disorder in RNA-Binding Proteins. *PLoS One, 10*(10), e0139731. doi:10.1371/journal.pone.0139731
Vucetic, S., Obradovic, Z., Vacic, V., Radivojac, P., Peng, K., Iakoucheva, L. M., . . . Dunker, A. K. (2005). DisProt: a database of protein disorder. *Bioinformatics, 21*(1), 137-140. doi:10.1093/bioinformatics/bth476

Walsh, I., Giollo, M., Di Domenico, T., Ferrari, C., Zimmermann, O., & Tosatto, S. C. (2015). Comprehensive large-scale assessment of intrinsic protein disorder. *Bioinformatics, 31*(2), 201-208. doi:10.1093/bioinformatics/btu625

Walsh, I., Martin, A. J., Di Domenico, T., & Tosatto, S. C. (2012). ESpritz: accurate and fast prediction of protein disorder. *Bioinformatics, 28*(4), 503-509. doi:10.1093/bioinformatics/btr682

Walsh, I., Martin, A. J., Di Domenico, T., Vullo, A., Pollastri, G., & Tosatto, S. C. (2011). CSpritz: accurate prediction of protein disorder segments with annotation for homology, secondary structure and linear motifs. *Nucleic Acids Res, 39*(Web Server issue), W190-196. doi:10.1093/nar/gkr411

Wang, C., Uversky, V. N., & 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. doi:10.1002/pmic.201500177

Wang, S., Ma, J., & Xu, J. (2016). AUCpreD: proteome-level protein disorder prediction by AUC-maximized deep convolutional neural fields. *Bioinformatics, 32*(17), i672-i679. doi:10.1093/bioinformatics/btw446

Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F., & Jones, D. T. (2004). Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol, 337*(3), 635-645. doi:10.1016/j.jmb.2004.02.002

Williams, R. J. (1979). The conformation properties of proteins in solution. *Biol Rev Camb Philos Soc, 54*(4), 389-437. doi:10.1111/j.1469-185x.1979.tb00843.x

Williams, R. M., Obradovi, Z., Mathura, V., Braun, W., Garner, E. C., Young, J., . . . Dunker, A. K. (2001). The protein non-folding problem: amino acid determinants of intrinsic order and disorder. *Pac Symp Biocomput*, 89-100. doi:10.1142/9789814447362_0010

Xie, H., Vucetic, S., Iakoucheva, L. M., Oldfield, C. J., Dunker, A. K., Uversky, V. N., & 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. doi:10.1021/pr060392u

Xue, B., Dunbrack, R. L., Williams, R. W., Dunker, A. K., & Uversky, V. N. (2010). PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochim Biophys Acta, 1804*(4), 996-1010. doi:10.1016/j.bbapap.2010.01.011

Xue, B., Dunker, A. K., & Uversky, V. N. (2012). Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. *J Biomol Struct Dyn, 30*(2), 137-149. doi:10.1080/07391102.2012.675145

Yan, J., Cheng, J., Kurgan, L., & Uversky, V. N. (2020). Structural and functional analysis of "non-smelly" proteins. *Cell Mol Life Sci, 77*(12), 2423-2440. doi:10.1007/s00018-019-03292-1

Yan, J., Dunker, A. K., Uversky, V. N., & Kurgan, L. (2016). Molecular recognition features (MoRFs) in three domains of life. *Molecular Biosystems, 12*(3), 697-710. doi:10.1039/c5mb00640f

Yan, J., Mizianty, M. J., Filipow, P. L., Uversky, V. N., & Kurgan, L. (2013). RAPID: fast and accurate sequence-based prediction of intrinsic disorder content on proteomic scale. *Biochim Biophys Acta, 1834*(8), 1671-1680. doi:10.1016/j.bbapap.2013.05.022

Zhang, F., Zhao, B., Shi, W., Li, M., & Kurgan, L. (2022). DeepDISOBind: accurate prediction of RNA-, DNA- and protein-binding intrinsically disordered residues with deep multi-task learning. *Brief Bioinform, 23*(1). doi:10.1093/bib/bbab521

Zhang, T., Faraggi, E., Xue, B., Dunker, A. K., Uversky, V. N., & 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. doi:10.1080/073911012010525022

Zhao, B., Katuwawala, A., Oldfield, C. J., Dunker, A. K., Faraggi, E., Gsponer, J., . . . Kurgan, L. (2021). DescribePROT: database of amino acid-level protein structure and function predictions. *Nucleic Acids Res, 49*(D1), D298-D308. doi:10.1093/nar/gkaa931

Zhao, B., Katuwawala, A., Oldfield, C. J., Hu, G., Wu, Z., Uversky, V. N., & Kurgan, L. (2021). Intrinsic Disorder in Human RNA-Binding Proteins. *J Mol Biol, 433*(21), 167229. doi:10.1016/j.jmb.2021.167229

Zhao, B., & Kurgan, L. (2021). Surveying over 100 predictors of intrinsic disorder in proteins. *Expert Rev Proteomics, 18*(12), 1019-1029. doi:10.1080/14789450.2021.2018304
Zhao, B., & Kurgan, L. (2022a). Compositional Bias of Intrinsically Disordered Proteins and Regions and Their Predictions. *Biomolecules, 12*(7). doi:10.3390/biom12070888
Zhao, B., & Kurgan, L. (2022b). Deep learning in prediction of intrinsic disorder in proteins. *Comput Struct Biotechnol J, 20*, 1286-1294. doi:10.1016/j.csbj.2022.03.003
Zhao, B., & Kurgan, L. (2022). Deep learning in prediction of intrinsic disorder in proteins. *Computational and Structural Biotechnology Journal, 20*, 1286-1294. doi:https://doi.org/10.1016/j.csbj.2022.03.003
Zhao, B., & Kurgan, L. (2023a). Databases of Protein Structure and Function Predictions at the Amino Acid Level. In *Machine Learning in Bioinformatics of Protein Sequences* (pp. 329-353).
Zhao, B., & Kurgan, L. (2023b). Machine Learning for Intrinsic Disorder Prediction. In *Machine Learning in Bioinformatics of Protein Sequences* (pp. 205-236).

KEY REFERENCES

Habchi, Tompa, Longhi & Uversky, 2014. See above.

Provides a comprehensive introduction to intrinsic disorder in proteins.

Hanson, Paliwal, Litfin, & Zhou, 2019. See above.

Describes SPOT-Disorder2, one of the most accurate methods for the prediction of intrinsic disorder.

Hu, Katuwawala, Wang, Wu, Ghadermarzi, Gao, & Kurgan, 2021. See above.

Describes flDPnn, winner of the CAID experiment and one of the most accurate disorder predictors.

Necci, Piovesan, Predictors, DisProt Curators, Tosatto, 2021. See above.

Describes the CAID experiment that performs empirical evaluation of disorder predictors.

Piovesan, Del Conte, Clementel, Monzon, Bevilacqua, Aspromonte, … Tosatto, 2023. See above.

Describes the MobiDB database that provides access to experimental and putative annotations of intrinsic disorder.

Zhao, & Kurgan, 2021. See above.

The most comprehensive to date survey of disorder predictors.

INTERNET RESOURCES

https://d2p2.pro/

D^2P^2 database

http://biomine.cs.vcu.edu/servers/DESCRIBEPROT/

DescribePROT database

http://old.protein.bio.unipd.it/espritz/

ESpritz-DisProt's webserver

http://biomine.cs.vcu.edu/servers/flDPnn

flDPnn's webserver

https://mobidb.bio.unipd.it/

MobiDB database

http://zhouyq-lab.szbl.ac.cn/servers/

SPOT-Disorder2 webserver