Resources for computational prediction of intrinsic disorder in proteins

Lukasz Kurgan1*

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

*Corresponding author: Lukasz Kurgan (lkurgan@vcu.edu)

Abstract

With over 40 years of research, intrinsic disorder prediction field has developed over 100 computational predictors. This review offers a holistic perspective of this field by highlighting accurate and popular disorder predictors and introducing a wide range of practical resources that support collection, interpretation and application of disorder predictions. These resources include meta webservers that expedite collection of multiple disorder predictions, large databases of pre-computed disorder predictions that ease collection of predictions particularly for large datasets of proteins, and modern quality assessment tools. The latter methods facilitate identification of accurate predictions in a specific protein sequence, reducing uncertainty associated to the use of the putative disorder. Altogether, we review eleven predictors, four meta webservers, three databases and two quality assessment tools, all of which are conveniently available online. We also offer a perspective on future developments of the disorder prediction and the quality assessment tools. The availability of this comprehensive toolbox of useful resources should stimulate further growth in the application of the disorder predictions across many areas including rational drug design, systems medicine, structural bioinformatics and structural genomics.

Keywords: intrinsic disorder; intrinsically disordered proteins; machine learning; webserver; prediction; protein function.

1 Introduction

Intrinsic disorder is manifested by presence of regions in protein sequences that are absent of a well-defined equilibrium structure under physiological conditions [1-4]. Intrinsically disordered proteins (IDPs) include one or more intrinsically disordered regions (IDRs) interspersed with structured regions, and in some cases, they are composed of one sequence-wide IDR. Several bioinformatics studies suggest that IDPs are relatively common in all kingdoms of life and viruses [5-9]. Some studies predict that about 1/3 of eukaryotic proteins have long IDRs, i.e., sequence regions composed of 30 or more consecutive disordered residues [5, 6, 10]. IDPs are instrumental for numerous cellular functions including molecular assembly and recognition, cell cycle regulation, signaling, transcription and translation, to name a few [11-23]. They are found across several membrane-bound cellular compartments [24, 25] and are important for the formation of membraneless compartments via the liquid-liquid phase separation [26-28]. They are also central to the formation of dark proteomes, defined as proteins that are not amenable to commonly used methods of experimental structure determination, such as X-ray crystallography [29-32].

Several databases, such as DisProt [33-37], PDB [38], IDEAL [39], DIBS [40], FuzDB [41] and MFIB [42], provide access to experimental data on IDPs. However, these resources cover a small sliver of the protein sequence space. More specifically, version 9.01 of DisProt includes about 2,400 IDPs [37] and a recent study identified approximately 25,000 IDPs in PDB [43], compared to over 220 million proteins in the RefSeq database [44]. This large and growing annotation gap motivates development of accurate computational methods that predict disordered residues and regions in a given protein sequence. The underlying ability to produce accurate predictions stems from the observation that disorder is an
inherent/intrinsic property of the input amino acid sequences [4]. Moreover, the development and applications of the bioinformatics methods, including the disorder predictors, was shown to stimulate rapid acceleration in the research on IDPs and IDR [45].

A recent study finds that over 100 disorder predictors were developed to date [46]. Numerous surveys summarize these predictors [43, 46-56]. These articles offer a historical perspective, describe and classify selected collections of disorder predictors and, in some cases, compare their predictive quality. The comparative studies offer invaluable advice on how to select the most accurate methods [43, 48, 53, 54, 57-63]. However, these surveys and comparative studies miss the opportunity to introduce other important resources that facilitate use and applications of disorder predictors. Nowadays, users have the options to conveniently collect pre-computed disorder predictions, generate multiple disorder predictions with one request and utilize disorder quality assessment tools. We provide a holistic perspective of the disorder prediction field. We summarize a selected collection of popular and/or accurate disorder predictors, identify several webservers that offer multiple disorder predictions, introduce and contrast large databases of putative disorder, and discuss how to obtains useful insights into correctness of the disorder predictions using modern disorder quality assessment tools.

2 Computational predictors of intrinsic disorder

Prior surveys identify six disorder predictors that were released between 1979, when the first method was published [64], and 2002 [46, 52, 55]. The development efforts have accelerated after 2002, which coincides with the inclusion of the disorder prediction assessment into the Critical Assessment of protein Structure Prediction (CASP) experiment in 2002 [62]. Since then, on average new five methods are developed annually [46], including at least nine predictors that were published since 2020: DisoMine [65], ODiNPred [66], IDP-Seq2Seq [67], flDPnn [68], flDPnlr [68], IUPred3 [69], RFPR-IDP [70], MetaPredict [71], and DeepCLD [72]. Recent survey finds that majority of recent disorder predictors rely on deep neural network models [46], which partly stems from the success of deep learners in recent disorder prediction assessments [48, 56, 57]. Here, we focus on providing practical advice for the end users by identifying a selection of popular and/or most accurate disorder predictors.

About a dozen large-scale comparative assessments of the intrinsic disorder predictors were carried out so far [43, 48, 53, 54, 57-63]. They include several community assessments where disorder predictors are tested on blind datasets by independent assessors. These datasets are withheld from the authors of the predictors before the assessment (i.e., authors are blind to the content of these datasets) and the assessors do not participate in the competitions, arguably making these evaluations more objective when compared to the comparative studies done by the authors of predictors. The disorder prediction was included in the biannual CASP experiment between CASP4 (2002) and CASP10 (2012) [58-63], and it was discontinued after CASP10. Another community assessment, Critical Assessment of Intrinsic Protein Disorder (CAID), was published in 2021 [57]. We focus on the two most recent community assessments, CASP10 and CAID to identify a pool of well-performing disorder predictors. We include the two assessments since they utilize benchmark data from two complementary sources, PDB depositions in CASP10 and DisProt records in CAID.

To quote conclusions from the CASP10 assessment: “Four prediction groups – Prdos-CNF, DISOPRED3, biomine_dr Mixed, and biomine_dr pdb c – perform better than the others according to the majority of the evaluation measures.” [61]. The corresponding top-three disorder predictors are PrDOS [73], DISOPRED3 [74], and MFDp [75-77], given that biomine_dr Mixed, and biomine_dr pdb c predictions cover two versions of MFDp where the better performing version is implemented by the MFDp tool. Similarly, the results from CAID are perhaps best summarized by the quote from the accompanying commentary [78]: “SPOT-Disorder2 and flDPnn, followed by RawMSA and AUCpreD, are consistently good. However, flDPnn 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, but since fIDPnn and RawMSA are not yet published, SPOT-Disorder2 and AUCpreD are the best options available to researchers today.” The fIDPnn and RawMSA methods were published in the meantime and thus the four methods, which include AUCpreD [79], rawMSA [80], SPOT-Disorder2 [81], and fIDPnn [68], are available to the end users.

We note that the CAID benchmark was also recently used to investigate potential impact of results produced by AlphaFold2 [82], which has disrupted the protein structure prediction area, on the disorder prediction field. Two recently released manuscripts use the CAID dataset to assess whether the AlphaFold2’s predictions can rival current disorder predictors [83, 84]. They conclude that while AlphaFold2’s results can be used to identify disordered regions, they are outperformed by the modern disorder predictors that include the best methods from the CAID experiment. This suggests that disorder and structure prediction areas require different solutions and they effectively complement each other.

We supplement the above list of the seven accurate disorder predictors, three from CASP10 and four from CAID, with a few popular methods. We quantify popularity with the citations collected using Google Scholar in January 2022. We normalize the total number of citations by the time since the corresponding articles were published, measured in months. The four highly cited methods include (in the order of their citation rates): IUPred [69, 85-87], DISOPRED [74, 88, 89], PONDR VSL2 [90, 91], and DisEMBL [92] (Table 1). We use the recent CAID results to provide context of the differences in the predictive performance between the accurate and the popular predictors. Using the results for the DisProt dataset from CAID [57], the F1 metric values of the accurate predictors that were included in this experiment are 0.48 (fIDPnn), 0.47 (SPOT-Disorder2), 0.45 (RawMSA), 0.43 (AUCpreD) and 0.39 (DISOPRED3). To compare, the corresponding F1 scores for the popular methods are 0.42 (IUPred), 0.41 (PONDR VSL2) and 0.36 (DisEMBL). While these results show that certain methods produce accurate predictions of disorder, they are not necessarily equally good when applied in the context of related predictions. A recent report analyzes predictive performance of ten disorder predictors applied to identify protein-binding and nucleic acid-binding regions [48]. This study finds that the predictions from SPOT-Disorder are the best for the identification of nucleic acid-binding while the ESpritz method [93] produces disorder that works well for the prediction of protein-binding.

We summarize the 11 most accurate and/or popular disorder predictors in Table 1. We list the citation data, types of predictive models that are used by these methods, and websites where these methods can be accessed. The most cited methods are published in mid 2000s and they rely on traditional machine learning models, such as shallow feed-forward neural networks (DISOPRED and DisEMBL) or support vector machines (PONDR VSL2 and MFDp), and scoring functions that approximate physical principles governing protein folding (IUPred). In contrast, the recent CAID-winning predictors utilize more sophisticated deep learning models. We note that these deep learners utilize several different neural network architectures including the deep feed-forward topology (fIDPnn), recurrent topology (SPOT-Disorder2), and a hybrid architecture that combines convolutional and recurrent elements (RawMSA and AUCpred). Importantly, the 11 selected disorder predictors are freely available to the end users via the websites shown in Table 1. Most of these methods, including DISOPRED, DisEMBL, IUPred, SPOT-Disorder2 and fIDPnn are provided in two complementary ways: as a standalone code that can be installed and executed on the user’s hardware and a webserver that conveniently runs on the server-side without the need for installation. Four predictors, PONDR VSL2, PrDOS, MFDp and AUCpred are available as webservers, while rawMSA is offered solely as the standalone code. Altogether, we conclude that users have access to a variety of readily available, popular and accurate options.
Table 1. Popular and accurate predictors of intrinsic disorder. The methods are sorted in the chronological order of their publication. The citations were collected from Google Scholar as of January 2022. Total citations number is calculated as the sum over all listed publications. The per month citations number is the total number of citations divided by the number of months since the first publication. Predictive models used include SVM (support vector machine), SFFNN (shallow feed-forward neural network), and DNN (deep neural network).

<table>
<thead>
<tr>
<th>Predictor</th>
<th>First published</th>
<th>Ref</th>
<th>Reason for inclusion</th>
<th>Total citations (per month)</th>
<th>Predictive model used</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>DISOPRED</td>
<td>Oct. 2003</td>
<td>[74, 88, 89]</td>
<td>High citation rate</td>
<td>1,593 (7.3)</td>
<td>SFFNN</td>
<td><a href="http://bioinf.cs.ucl.ac.uk/psipred/">http://bioinf.cs.ucl.ac.uk/psipred/</a></td>
</tr>
<tr>
<td>DisEMBL</td>
<td>Nov. 2003</td>
<td>[92]</td>
<td>High citation rate</td>
<td>1,377 (6.3)</td>
<td>SFFNN</td>
<td><a href="http://dis.embl.de/">http://dis.embl.de/</a></td>
</tr>
<tr>
<td>PONDR VSL2</td>
<td>Sept. 2005</td>
<td>[90, 91]</td>
<td>High citation rate</td>
<td>1,357 (6.9)</td>
<td>SVM</td>
<td><a href="http://www.pondr.com/">http://www.pondr.com/</a></td>
</tr>
<tr>
<td>PrDOS</td>
<td>July 2007</td>
<td>[73]</td>
<td>High CASP10 rank</td>
<td>681 (3.9)</td>
<td>SVM</td>
<td><a href="https://prdos.hgc.jp/cgi-bin/top.cgi">https://prdos.hgc.jp/cgi-bin/top.cgi</a></td>
</tr>
<tr>
<td>MFDp</td>
<td>Sept. 2010</td>
<td>[75-77]</td>
<td>High CASP10 rank</td>
<td>293 (2.2)</td>
<td>SVM</td>
<td><a href="http://biomine.cs.vcu.edu/servers/MFDp/">http://biomine.cs.vcu.edu/servers/MFDp/</a></td>
</tr>
<tr>
<td>DISOPRED3</td>
<td>March 2015</td>
<td>[74]</td>
<td>High CASP10 rank</td>
<td>573 (7.0)</td>
<td>Ensemble of SVM, SFFNN and nearest neighbor</td>
<td><a href="http://bioinf.cs.ucl.ac.uk/psipred/">http://bioinf.cs.ucl.ac.uk/psipred/</a></td>
</tr>
<tr>
<td>AUCpred</td>
<td>Sept. 2016</td>
<td>[79]</td>
<td>High CAID rank</td>
<td>66 (1.0)</td>
<td>DNN (hybrid convolutional and conditional random field)</td>
<td><a href="http://raptorx.uchicago.edu/StructurePropertyPred/predict/">http://raptorx.uchicago.edu/StructurePropertyPred/predict/</a></td>
</tr>
<tr>
<td>rawMSA</td>
<td>Aug. 2019</td>
<td>[80]</td>
<td>High CAID rank</td>
<td>23 (0.8)</td>
<td>DNN (hybrid convolutional and recurrent)</td>
<td><a href="https://bitbucket.org/clami66/rawmsa/src/master/">https://bitbucket.org/clami66/rawmsa/src/master/</a></td>
</tr>
<tr>
<td>flDPnn</td>
<td>July 2021</td>
<td>[68]</td>
<td>High CAID rank</td>
<td>8 (1.3)</td>
<td>DNN (feed-forward)</td>
<td><a href="http://biomine.cs.vcu.edu/servers/flDPnn/">http://biomine.cs.vcu.edu/servers/flDPnn/</a></td>
</tr>
</tbody>
</table>
Figure 1. Experimental and predicted disorder annotations for the 50S ribosomal protein L4 (UniProtKB id: P60723).
Panel A shows experimental annotations of an IDR (positions 41 to 103) collected from DisProt (DisProt id: DP00600) (https://www.disprot.org/). Panel B shows putative disorder produced by flDPnn (http://biomine.cs.vcu.edu/servers/flDPnn/), where black plot gives the putative disorder propensity and horizontal black bar corresponds to putative IDR derived from propensity values using the cutoff denoted with the dotted line. Panel C is the putative disorder generated by IUPred3(https://iupred3.elte.hu/), where red plot shows the putative disorder propensity and horizontal black line is the cutoff used to identify putative IDRs. Panel D shows disorder predictions generated by multiple color-coded methods that are included in the MetaDisorder webserver (http://iimcb.genesilico.pl/metadisorder/), where the darker blue background identifies the threshold used to derive IDRs. Panel E provides predictions of disorder that are available in MobiDB (https://mobidb.bio.unipd.it/), where the horizontal orange bars denote predicted IDRs. Panel F gives putative disorder generated by a consensus of SPOT-DISORDER-Single [94], DISOPRED3 [74] and IUPred-short [86]
(black and gray horizontal bar line) that is accompanied by the quality assessment (QA) scores produced by QUARTERPlus (http://biomine.cs.vcu.edu/servers/QUARTERplus). The QA scores (color-coded horizontal bar) quantify predictive quality of the consensus disorder prediction, i.e., amino acids shown in green and yellow colors are more likely to be accurately predicted compared to the residues colored in orange or red.

We illustrate results that these methods generate based on an example that predicts disorder for the 50S ribosomal protein L4 (UniProtKB id: P60723). This 201-residues long protein facilitates early stages of the ribosome assembly and acts as a transcriptional repressor [95, 96]. Experimental annotations of disorder, which we collect from DisProt (DisProt id: DP00600) [37], reveal that the L4 protein harbors a long IDR (positions 41 to 103) that is engaged during the ribosome assembly process [96, 97]; see Figure 1A. Using data from Table 1, we show predictions generated by the most recent method, flDPnn (Figure 1B), and the most cited predictor, IUPred (Figure 1C). The disorder predictions are composed of two parts: numerical propensity scores that quantify likelihood for disorder for each amino acid in a given protein sequences, and binary scores that annotate each residue as disordered vs. ordered. The binary predictions typically form regions (IDRs) and are computed from the putative propensities such that amino acids with propensities greater than a threshold are designated as disordered. The thresholds used by flDPnn and IUPred are shown using black horizontal dotted line (Figure 1B) and black horizontal solid line (Figure 1C), respectively. We find that flDPnn correctly identifies the presence and location of the IDR in this protein (Figure 1B). Similarly, IUPred also generates relatively high propensities for residues that coincide with the location of the experimental IDR (Figure 1C).

Table 2. Meta webservers that provide at least four disorder predictions generated by different methods. The webservers are sorted in the chronological order of their publication. We use bold font to highlight the popular and/or accurate methods from Table 1.

<table>
<thead>
<tr>
<th>Webserver</th>
<th>First published</th>
<th>Ref</th>
<th>Disorder predictors included</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>MetaPrDOS</td>
<td>June 2008</td>
<td>[98]</td>
<td>PrDOS [73], DISOPRED2 [5], DisEMBL [92], PONDR VSL2 [90], DISpro [99], IUPred [85, 86], POODLE- S [100], DISoClust [101], and metaPrDOS [98]</td>
<td><a href="https://prdos.hgc.jp/meta/about.html">https://prdos.hgc.jp/meta/about.html</a></td>
</tr>
<tr>
<td>MetaDisorder</td>
<td>May 2012</td>
<td>[102]</td>
<td>DisEMBL [92], DISOPRED2 [5], DISpro [99, 103], GlobPlot [104], iPDA [105], IUPred [85, 86], Pdisorder, POODLE-S [100], POODLE-L [106], PrDOS [73], Spritz [107], and RONN [108]</td>
<td><a href="http://iimcb.genesilico.pl/metadisorder/">http://iimcb.genesilico.pl/metadisorder/</a></td>
</tr>
<tr>
<td>DisMeta</td>
<td>Oct. 2013</td>
<td>[109]</td>
<td>DisEMBL [92], DISOPRED2 [5], DISpro [99], FoldIndex [110], GlobPlot [104], IUPred [85, 86], RONN [108], and PONDR VSL2 [90]</td>
<td><a href="http://montelionelab.chem.rpi.edu/dismeta/">http://montelionelab.chem.rpi.edu/dismeta/</a></td>
</tr>
</tbody>
</table>

3 Meta webservers for disorder prediction

We examine recent surveys [43, 46, 48, 50-52, 54, 56] and manually screen results of a PubMed search using the “prediction[Title] AND intrinsic AND disorder” query to identify online resources that provide access to multiple disorder predictions. We require that these resources generate and output results of at least four different intrinsic disorder predictors. The key benefit of having access to multiple results is that they can be assessed for convergence in order to boost confidence in the resulting disorder prediction. The latter is motivated by a few studies that empirically demonstrate that the use of a consensus-based disorder prediction typically leads to an improved predictive performance [43, 111-114]. We found four meta webservers, which we summarize in Table 2. They include MetaDisorder [102], MetaPrDOS [98], DisMeta [109] and MFDp [75-77].
that generate 12, 9, 8 and 4 disorder predictions, respectively. Availability via the webserver is especially valuable for users who perform predictions in an *ad hoc* manner and less computer savvy users. Users simply input their protein sequence(s) via a web browser using the URL listed in Table 2, and collect the resulting predictions that are delivered via the browser window and/or via email. The prediction process that covers running multiple methods is automated and executed on the server side. This means that users do not need to use their own hardware or install software to collect the results, which makes these resources very convenient. Table 2 reveals that these meta webservers generate predictions for the four popular methods (DISOPRED, DisEMBL, IUPred and PONDR VSL2) and two of the top-performers from CASP10 (PrDOS and MFDp). However, since these webservers were released and published several years ago, they do not include the newer accurate predictors, such as DISOPRED3, AUCpred, rawMSA, SPOT-Disorder2 and flDPnn.

Figure 1D shows predictions for the L4 protein produced by MetaDisorder, which is the resource that covers the largest number of disorder predictors. Propensity scores output by different predictors are color-coded and they reveal that majority of the methods converge on predicting residues in the region spanning between positions 40 and 80 as disordered. This prediction overlaps with the location of the experimentally-found IDR (Figure 1A). While some methods predict disorder at the sequence termini, it is clear that these methods are in minority, which suggests that this prediction should be disregarded (Figure 1D). This illustrates how the consensus analysis can be used to guide the process of identifying IDRs.

4 Databases of intrinsic disorder predictions

Collection of the disorder predictions could be difficult and time consuming, especially when assembling results of multiple methods that are not available via one of the meta webservers. In that case, users must navigate multiple websites and/or install multiple software, deliver inputs (i.e., sequences, identifiers and/or emails) in multiple formats, and parse and standardize the different formats of outputs that disorder predictors use. Another substantial drawback is that the runtime required to make predictions could be substantial, as much as several minutes for one protein [57]. This is particularly challenging when collecting predictions for large protein datasets, such as big protein families or proteomes. We note that studies that perform proteome-wide analysis of disorder are done regularly [5-7, 24, 29, 115-120].

A suitable alternative to the direct use the disorder predictors is to collect pre-computed disorder predictions from one of the three available databases: Database of Disorder Protein Predictions (D\(^2\)P\(^2\)) [121], MobiDB [122-125], and DescribePROT [126]. These resources provide access to disorder prediction for large collections of proteins ranging from 1.37 million in DescribePROT, 10.43 million proteins in D\(^2\)P\(^2\), to 219.74 million proteins in MobiDB (Table 3). The key feature of these databases is the instantaneous retrieval of the pre-computed predictions generated by multiple methods. Users do not have to wait for the computation of predictions and do not need to assemble the various predictions together. Moreover, this reduces wasteful duplication in the use of the predictors that are often tasked to make many predictions for the same protein when the same sequence is submitted by different users. The three databases facilitate collection of predictions for individual proteins, which are provided in a parsable text format and in an interactive graphical format (Figure 1E). They also provide options to conveniently download predictions for whole proteomes. For example, DescribePROT offers access to the raw predictions (propensities and binary values) and protein-level summaries that include disorder content at the whole proteome level (http://biomine.cs.vcu.edu/servers/DESCRIBEPROT/download.html). MobiDB and D\(^2\)P\(^2\) databases deliver predictions from 8 and 6 disorder predictors (Table 3), respectively, and also produce a consensus result. MobiDB computes the consensus using the MobiDB-lite algorithm [113] and D\(^2\)P\(^2\) uses the 75% consensus approach, i.e., an amino acid is predicted as disorder if at least 75% of methods predicts it as disordered. Figure 1E shows disorder predictions for the L4 protein that we collect from MobiDB, which covers the largest number of disorder predictors. The first two lines give the binary predictions and the corresponding propensities generated by MobiDB-lite. The subsequent lines provide binary predictions from the individual disorder predictors. We note a good agreement between these predictions, in particular the consensus-based result, and the corresponding experimental data from Figure 1A.
Table 3. Databases of intrinsic disorder predictions. The databases are sorted in the chronological order of their publication. We use bold font to highlight the popular and/or accurate methods from Table 1.

<table>
<thead>
<tr>
<th>Database</th>
<th>First published</th>
<th>Ref</th>
<th>Size [millions proteins]</th>
<th>Disorder predictors included</th>
<th>Other predictions included</th>
<th>URL</th>
</tr>
</thead>
</table>
| MobiDB version 4.1 | Aug. 2012       | [124] | 219.74                   | DisEMBL [92], DynaMine [127], ESpritz [93], GlobPlot [104], IUPred2A [87], JRONN [108], MobiDB-lite [113], POND VRSL2 [90, 91] | Disordered protein-binding by ANCHOR [128]  
Secondary structure by FeSS [129]  
Low complexity regions by SEG [130]  
Domains by Gene3D [131] | https://mobidb.bio.unipd.it/ |
| D²p² version 1.0 | Jan. 2013       | [121] | 10.43                     | POND VRSL2 [90, 91], PrDOS [98], PV2 [133], ESpritz [93], IUPred [86] | Disordered protein-binding residues by ANCHOR [128]  
Domains by SUPERFAMILY [134] | https://d2p2.pro/ |
Secondary structure by PSIPRED [136-138]  
Disordered DNA- and RNA-binding by DisoRDPbind [139-141]  
Disordered protein-binding by DisoRDPbind [139-141] and MoRFChibi [142]  
Structured protein-binding by SCRIBER [143, 144]  
Structured RNA-binding and DNA-binding by DRNApred [145]  
Disordered DNA-binding by Disordered linkers by DFLpred [146]  
We also point to the key advantages and drawbacks of these resources. MobiDB covers by far the largest number of proteins. It is also cross-linked and includes experimental data from ten external sources: CoDNaS [149], DIBS [40], DisProt [33], ELM [150], FuzDB [41], IDEAL [39], MFIB [42], PDBe [151], PhasePro [152], and UniProt [153]. On the other hand, MobiDB is almost exclusively focused on the intrinsic disorder. D²P² offers arguably the most comprehensive annotations of protein domains and posttranslational modification sites that are collected using SUPERFAMILY [134] and PhosphoSitePlus [154], respectively. However, it is also heavily focused on the intrinsic disorder and is no longer maintained. It was last updated in 2013.

DescribePROT provides access to the predictions for a diverse set of structural and functional characteristics of proteins, including intrinsic disorder. The other characteristics consist of alignment profiles and putative solvent accessibility, disordered linkers, secondary structure, signal peptides, and protein-binding, DNA-binding and RNA-binding residues. Predictors used to derive these annotations are listed in Table 3. Altogether, release 1.4 of DescribePROT provides access to over 7.8 billion amino acid-level predictions. However, DescribePROT covers the smallest number of proteins and only one disorder predictor.

Finally, while undoubtedly these are very helpful resources, they do not offer a holistic solution. In particular, users who would like to collect result outside of the protein sets covered in these databases, e.g., for a novel protein sequence, have to rely on the disorder predictors and meta webservers

5 Methods for quality assessment of disorder predictions

The comparative assessments of disorder predictors, including CASP10 and CAID, report results on datasets composed of dozens or hundreds of proteins. This quantifies an overall, dataset-wide predictive performance of these methods. However, these studies may not offer an accurate guidance when predicting specific proteins. A recent study shows that predictive performance of disorder predictors varies widely between proteins, where some sequences are predicted very accurately while other predictions are hardly better than random [56]. Other works find that the predictive quality varies between amino acids in the same sequences and that the outputs generated by the predictors do not allow to accurately identify these differences [155, 156]. To this end, one of the more recent advances in the disorder prediction field is the development and release of methods that provide quality assessment (QA) scores [157]: QUARTER [156, 158] and its upgraded version QUARTERplus [159]. These QA scores quantify confidence in the disorder predictions at an amino acid level [156]. In other words, these scores facilitate identification of amino acids in a given protein chain for which the disorder predictions are more likely to be accurate.

QUARTERplus webserver (http://biomine.cs.vcu.edu/servers/QUARTERplus/) produces disorder predictions and the associated QA scores for several disorder predictors including DISOPRED3 [74], IUPred [85, 86], PONDR VSL2 [90, 91], DisEMBL [92], GlobPlot [104], and SPOT-Disorder-Single [94]. Moreover, it uses a deep convolutional neural network to make accurate, consensus-based disorder predictions using outputs from SPOT-Disorder-Single, IUPred and DISOPRED3 that are accompanied by the QA scores. These QA scores allow users to conveniently pinpoint which disorder predictions are more trustworthy. We illustrate this in Figure 1F where we present the consensus disorder prediction produced by QUARTERplus (pink line and black and gray horizontal bar) for the L4 protein. This prediction is coupled with the color-coded QA scores where green and yellow denote residues that are more likely to be accurately predicted compared to the predictions colored in orange or red. The usefulness of the QA scores becomes apparent when comparing the disorder predictions from QUARTERplus (Figure 1F) and the corresponding experimental annotations (Figure 1A). While QUARTERplus predicts a short IDR at the N-terminus, the associated QA scores are orange, which suggests that this is a false prediction, in agreement with the native annotations. The long experimental IDR located between positions 41 and 103 is underpredicted by QUARTERplus around positions 80 to 103. However, the corresponding QA scores are colored red and yellow, which correctly suggests low quality of these disorder predictions. At the same time, the QA scores for the positions between 103 and the C-terminus are green and yellow, which indicate that the related disorder predictions are likely accurate. This agrees with the experimental annotations for this part of the sequence (Figure 1A). This example demonstrates how the QA scores can be used to identify correct vs. incorrect disorder predictions for a given protein sequence. Altogether,
we conclude that QUARTERplus a is a convenient tool which provides QA scores that empower users to identify more trustworthy predictions for several popular disorder predictors.

6 Summary and outlook

Intrinsic disorder prediction field produced over 100 predictors [46]. We highlight 11 disorder predictors selected based on the results of two most recent community assessments, CASP10 [61] and CAID [57], and citation data. This collection represents arguably the most accurate and most popular methods at this time. More importantly, we introduce and discuss a wide range of other practical resources including meta webservers, databases and disorder QA tools. These resources facilitate convenient collection, interpretation and application of disorder predictions.

The meta webservers, which include MetaDisorder [102], MetaPrDOS [98], DisMeta [109] and MFDp [75-77], expedite collection of multiple disorder predictions, which can be effectively used to perform a consensus-based analysis. The users have an option to obtain pre-computed disorder predictions from three large databases, such as MobiDB [125], D²P² [121], and DescribePROT [126]. This is particularly valuable when targeting analysis of disorder for large datasets of proteins and when analyzing disorder in the context of other structural and functional features, such as protein domains (available in D²P² and MobiDB), posttranslational modification sites (D²P²), secondary structure (MobiDB and DescribePROT), protein-protein interactions (MobiDB and DescribePROT), and protein-nucleic acid interactions, linkers and solvent accessibility (DescribePROT).

Moreover, we discuss QUARTERplus [159], a modern deep learning-based tool that produces QA scores for several popular disorder predictors. The QA scores facilitate identification of accurate predictions in a specific protein sequence, reducing uncertainty associated with the use of putative disorder annotations. Availability of this comprehensive toolbox of practical resources will fuel further growth of the computational intrinsic disorder field, ensuring that the disorder predictions will continue to make impact across many application areas, such as drug design [160-163], systems medicine [164], structural bioinformatics [143, 165-169], and structural genomics [29, 32, 92]. In the context of the rational drug design, the disorder predictions are used to identify currently underrepresented disordered proteins as drug targets [160-162, 170] and to facilitate the development of new computational tools for modelling protein-drug interactions [171-174]. As another example, the putative annotations of intrinsic disorder facilitate target selection in structural genomics by avoiding disordered proteins when selecting targets to be solved using the commonly applied X-ray crystallography [32, 92].

We find that the most accurate disorder predictors according to CAID rely exclusively on the deep neural networks. More broadly, a recent analysis of the deep learning-based disorder predictors reveals that these methods rely on a wide range of network types and that they outperform other types of predictive models [175]. The currently used deep network topologies include feed-forward (flDPnn [68]), convolutional (DeepCNF-D [176] and AUcPred [79]), recurrent (SPOT-Disorder [177], IDP-Seq2Seq [67], MetaPredict [71] and DeepCLD [72]) and convolutional/recurrent hybrids (SPOT-Disorder-Single, rawMSA [80], SPOT-Disorder2 [81] and RFPR-IDP [70]). Some of the recently released methods that predict specific types of functional disordered regions also use deep networks. Examples include SPOT-MoRF [178], MoRFPred_en [179] and en_DCNNMoRF [180] that predict protein-binding regions; DeepDISObind [181] that predicts protein and nucleic acid binding regions; and DisoLipPred [182] that finds putative disordered lipid binding regions. One recent advancement in deep networks that is yet to penetrate into the mainstream of the disorder predictions are the transformer networks. These architectures arguably improve over the convolutional and recurrent designs by applying attention mechanisms and positional embeddings. The transformer networks were already used with success in several related problems including prediction of tertiary protein structure [183], protein-protein interactions [184] and protein-compound interactions [185].

Moreover, authors of the recent flDPnn method conclude that predictive quality can be improved by extending inputs of the deep networks to cover additional functional and structural characteristics that are derived from the protein sequence [68]. We believe that further advances in the disorder prediction could stem from hybrid designs that combine multiple network topologies and tailor their predictive inputs. These hybrid designs should aim to mimic the underlying diversity of disorder types/flavors [186, 187]. For example, IDRs are classified into native coils, native pre-molten globules and native molten globules, which stems from differences in their
conformational space [188]. IDRs also vary in length and localization in the sequence including typically short regions that are located at the termini [189] vs. long regions that can cover the whole length of the sequence [190, 191]. Simultaneous tuning of the topologies and inputs should offer sufficient amount of flexibility to accurately model different flavors of disorder. Hybridizing these models together will provide a more complete approach to cover the multiplicity of disorder and consequently should lead to further improvement in the predictive performance.

Another interesting future direction would be to expand the current scope of the QA tools to cover more disorder predictors. The current version of QUARTERPlus works with six disorder predictors (DISOPRED3, IUPred, PONDR VSL2, DisEMBL, GlobPlot, and SPOT-Disorder-Single) that exclude the top-performers from CAID, such as AUCpred, rawMSA, SPOT-Disorder2, and flDPnn. Adding QA scores to the results produced by the latter set of methods would make them more practical, further boosting confidence in the results that they produce.

**Declaration of interest**

The authors declare no conflicts of interest.

**Funding**

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

**References**


190. Nielsen, J.T. and F.A. Mulder, There is Diversity in Disorder-"In all Chaos there is a Cosmos, in all Disorder a Secret Order". Front Mol Biosci, 2016. 3: p. 4.