Published online 4 May 2023 *Nucleic Acids Research, 2023, Vol. 51, Web Server issue* ***W141–W147***
https://doi.org/10.1093/nar/gkad330

DEPICTER2: a comprehensive webserver for intrinsic disorder and disorder function prediction

Sushmita Basu[1], Jörg Gsponer[2] and Lukasz Kurgan[1,*]

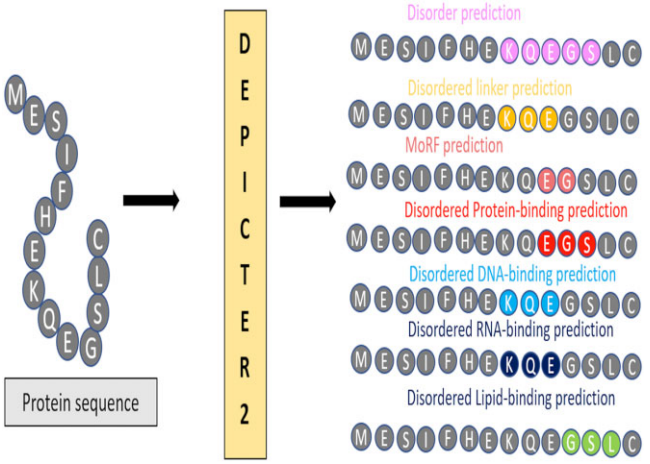
[1]Department of Computer Science, Virginia Commonwealth University, Richmond, VA 23284, USA and [2]Michael Smith Laboratories, University of British Columbia, Vancouver, British Columbia, Canada

Received March 15, 2023; Revised April 12, 2023; Editorial Decision April 17, 2023; Accepted April 18, 2023

ABSTRACT

Intrinsic disorder in proteins is relatively abundant in nature and essential for a broad spectrum of cellular functions. While disorder can be accurately predicted from protein sequences, as it was empirically demonstrated in recent community-organized assessments, it is rather challenging to collect and compile a comprehensive prediction that covers multiple disorder functions. To this end, we introduce the DEPICTER2 (DisorderEd PredictIon CenTER) webserver that offers convenient access to a curated collection of fast and accurate disorder and disorder function predictors. This server includes a state-of-the-art disorder predictor, flDPnn, and five modern methods that cover all currently predictable disorder functions: disordered linkers and protein, peptide, DNA, RNA and lipid binding. DEPICTER2 allows selection of any combination of the six methods, batch predictions of up to 25 proteins per request and provides interactive visualization of the resulting predictions. The webserver is freely available at http://biomine.cs.vcu.edu/servers/DEPICTER2/

GRAPHICAL ABSTRACT



INTRODUCTION

Intrinsically disordered proteins (IDPs) have one or more intrinsically disordered regions (IDRs) that lack stable tertiary structure under physiological conditions (1–3). Bioinformatics studies estimate that IDPs and IDRs are relatively common in nature, with 30 to 50% of eukaryotic proteins, depending on the organism, that have at least one long IDR with 30 or more consecutive disordered amino acids (4,5). IDPs are involved in a variety of cellular functions (6–15), are located across several cellular compartments (16), contribute to human diseases (17,18), and are considered to be promising drug targets (19,20). However, only several hundred IDRs that are included in the DisProt database have experimental annotations of their functions (21,22). Availability of these annotations and the fact that IDRs have intrinsic compositional bias that makes them predictable from sequence (23,24) motivate development of computational methods that predict disorder from the protein sequences. There are over 100 disorder predictors (25,26) and over three dozen predictors of disorder functions (27–29). Most of them rely on machine learning models that are generated using training datasets composed of the experimentally annotated IDRs (30,31). The function predictors

*To whom correspondence should be addressed. Tel: +1 804 827 3986; Email: lkurgan@vcu.edu

address prediction of IDRs that interact with specific types of molecular partners, such as proteins, peptides, DNA, RNA and lipids, as well as disordered linker regions. Predictive performance of these tools was evaluated in a number of comparative assessments including the community-driven Critical Assessment of techniques for protein Structure Prediction (CASP) experiments between CASP5 and CASP10 (32,33) and more recently the Critical Assessment of Intrinsic disorder (CAID) experiment (34). The CAID's results and subsequent follow-up studies reveal that modern disorder predictors, particularly those that rely on deep neural networks, produce accurate results (31,34,35). Example deep learning-based tools are flDPnn (36), SPOT-Disorder2 (37), RawMSA (38), AUCpreD (39), IDP-Seq2Seq (40), DeepIDP-2L (41) and DeepCLD (42).

Computational methods offer an accurate and cost-efficient way to predict and functionally annotate IDPs and IDRs for the millions of protein sequences that lack annotations. Predictions can be obtained with webservers and implementations that are provided and supported by the authors and by using popular and large databases of pre-computed disorder predictions: D^2P^2 (43) and MobiDB (44). While these databases conveniently provide predictions for millions of proteins, they offer a rather narrow selection of the disorder function predictions that covers only protein and peptide binding. They are also limited to the proteins that they currently include. Collecting predictions using webservers and/or code is rather difficult. This requires identifying suitable methods that cover disorder prediction and desired disorder function predictions, installing code if this option was selected, converting between multiple input/output formats, working with multiple interfaces, and assembling different predictions. There is a prototype solution that solves this problem by integrating disorder and disorder function predictions, the DEPICTER (DisorderEd PredictIon CenTER) webserver (45). DEPICTER incorporates prediction of disorder using SPOT-Disorder-Single (46) and IUPred2 (47), disordered linkers with DFLpred (48), nucleic acid binding with DisoRDPbind (49,50), and protein and peptide binding with ANCHOR2 (47) and fMoRFpred (51). However, this resource utilizes a selection of methods that are now outperformed by more recent solutions (SPOT-Disorder-Single, IUPred2 and fMoRFpred), predicts only one sequence at the time, and omits disorder functions for which methods were developed recently. To this end, we provide a new and significantly improved DEPICTER2 resource. DEPICTER2 provides access to a comprehensive selection of fast tools that include state-of-the-art disorder predictor, flDPnn (36), and five methods that cover the currently predictable disorder functions: disordered linkers (DFLpred (48)); protein and peptide binding IDRs (ANCHOR2 (47)); MoRFs (51), which are short protein-binding segments that are typically located in IDRs and that undergo disorder-to-order transitions upon binding ($MoRF_{CHiBi\_Light}$ (52)); DNA and RNA binding IDRs (DisoRDPbind (49,50,53)); and lipid-binding IDRs (DisoLipPred (54)). The DEPICTER2 webserver allows for batch predictions of up to 25 proteins, automates the entire prediction process, provides an interactive visualization of the results, and delivers results in a consistent format across the six tools using easy to parse files in multiple format (comma-separable, xml and json). DEPICTER2 is freely available at http://biomine.cs.vcu.edu/servers/DEPICTER2/.

MATERIALS AND METHODS

Predictive performance and selection of methods included in DEPICTER2

With nearly 150 disorder and disorder function predictors (25,27), it would be impractical to provide access to all these tools. Thus, DEPICTER2 covers a curated collection of six predictors, where each method targets prediction of a different aspect of intrinsic disorder. We select fast, recently published and empirically shown to provide accurate predictions tools that include a predictor of disorder and five tools that comprehensively cover the five currently predicted disorder functions. These predictors generate two results for each residue in the input sequence: real-values propensities and binary scores (disorder vs. structure; function vs. no function). Correspondingly, we quantify predictive accuracy with two popular metrics: area under receiver operating characteristic curve (ROC-AUC) that evaluates propensities and F1 for the binary predictions.

A post-CAID commentary that analyzed CAID results concludes that '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' (35). More precisely, ROC-AUC and F1 values are 0.814 and 0.48 for flDPnn and 0.760 and 0.47 for SPOT-Disorder2, respectively (34). Consequently, DEPICTER2 applies flDPnn, the fastest among the most accurate disorder predictors in the CAID experiment, to generate the disorder predictions. To compare, the ROC-AUC and F1 in CAID for the two methods that were used in DEPICTER are 0.757 and 0.43 (SPOT-Disorder-Single) and 0.740 and 0.42 (IUPred2), respectively (34).

The current disorder function predictors address predictions of disordered linkers and IDRs that interact with several types of molecular partners: proteins, peptides, DNA, RNA and lipids (27,28). DEPICTER2 includes one predictor for each of these functions, selected based on its favorable predictive performance from the CAID experiment if multiple methods are available. In fact, CAID is the first community-driven effort that evaluates predictions of binding IDRs. The top three predictors in this category are ANCHOR2 with ROC-AUC = 0.742 and F1 = 0.22, DisoRDPbind with ROC-AUC = 0.729 and F1 = 0.21, and $MoRF_{CHiBi\_Light}$ with AUC = 0.720 and F1 = 0.21 (34). We include these three tools in DEPICTER2. They predict disordered residues that interact with proteins and peptides (ANCHOR2), RNA and DNA (DisoRDPbind), as well as MoRF regions ($MoRF_{CHiBi\_Light}$). MoRFs are short regions that are embedded in longer IDRs that undergo disorder-to-order transition when interacting with proteins and peptides (55,56). To compare, the MoRF predictor that is included in the original DEPICTER, fMoRFpred, obtains ROC-AUC = 0.55 and F1 = 0.07 in the CAID experiment. Moreover, we re-use the predictor of disordered linkers, DFLpred, from DEPICTER. This tool secures

ROC-AUC = 0.715 on a low-similarity test dataset in the original publication (linker prediction was not included in CAID) (48). Finally, DEPICTER2 incorporates DisoLipPred, the sole predictor of the disordered lipid-binding residues that was released after CAID experiment was completed. DisoLipPred obtains ROC-AUC = 0.781 and F1 = 0.15 on a low-similarity test dataset, outperforming other indirect ways to predict this functional type of disorder (54). Altogether, the six selected methods (flDPnn, ANCHOR2, DisoRDPbind, $MoRF_{CHiBi\_Light}$, DFLpred and DisoLipPred) are relatively accurate and most of them, except for DisoLipPred, are also optimized for speed. Their predictions can be completed in approximately 15, 30 and 80 seconds for sequences with length of 100, 300 and 1000 amino acids, respectively.

RESULTS

Architecture

Figure 1 summarizes workflow of the DEPICTER2 webserver. We use the input sequence (step 1) to generate a comprehensive profile with several third-party methods (step 2). The profile quantifies sequence-derived information that is useful for the disorder and disorder function prediction including sequence conservation, putative secondary structure, solvent accessibility and other characteristics. This profile is shared between the six predictors that are run individually. Each predictor relies on its own feature engineering procedure which converts a specific part of the profile into inputs that are utilized by the predictive model (step 3); details are described in their publications (36,47,48,50,52,54). We input the features into the corresponding predictive models (step 4) that produce predictions separately for each of the six predictors (step 5). This is a rather complex architecture, with nine third-party programs, several vastly different feature engineering procedures, including the most sophisticated one for flDPnn that generates features at amino acid, sequence window and whole-chain levels, and seven diverse predictive models. These models range from relatively simple regressions (DFLpred and DisoRDPbind), scoring functions (ANCHOR2) and Bayesian models ($MoRF_{CHiBi\_Light}$), to more sophisticated deep feedforward and recurrent neural networks (flDPnn and DisoLipPred). The results include color-coded binary predictions (graphically represented as horizontal bars) and the corresponding real-valued propensities (Figure 1). The webserver is available at http://biomine.cs.vcu.edu/servers/DEPICTER2/. We focus on convenience. The programs and models are run automatically by scripts on the server side. Users do not need to install any additional software beside a web browser. The front end is implemented in HTML and JavaScript while the back end is based on PHP, Java, Python and MySQL database. We offer a simple to navigate input interface and parsable text file and graphical outputs.

Inputs and interface

We discuss the inputs and outputs of DEPICTER2 using results for an example human protein, Ataxin-3 (Disprot ID: DP00576; Uniprot ID: P54252). Ataxin-3 is a deubiquitinating enzyme that cleaves ubiquitin from proteins just before they are degraded. Ataxin-3 has a disordered C-terminal domain (positions 174–361) (57,58) which hosts multiple ubiquitin interacting motifs (UIM) that are essential for the deubiquitination (59,60).

Figure 2A shows the interface of DEPICTER2. A user is asked to provide either the FASTA-formatted protein sequence(s) or the UniProt accession(s), and (optionally) an email address. Figure 2A shows an example submission with the sequence of Ataxin-3. We recommend providing the email since this is where links to the results are sent upon completion of the prediction process; otherwise, users must ensure that the browser window is open and active as the prediction progresses. The input interface that allows selection of any combination of the six methods and by default the faster five predictors are selected (see 'Runtime' section). We support batch predictions for up to 25 proteins when fast tools are selected and limit the input to two proteins when the slow DisoLipPred is included. After selecting the methods, predictions are launched by clicking the 'RUN' button. The browser redirects to the status page that shows the current position in the server queue. To provide fair access to users, first-come-first-serve queue is applied with a limit of five concurrent requests per user. We also limit the time allocated to each submission to about 15 minutes, which is why we constrain the number of input proteins to 25. Once the predictions are completed, the status page redirects to the results page.

Outputs

The results page provides links to the graphical output of each input sequence and to download raw formatted outputs for each selected method that are available in several easy to parse formats that include comma-separable text, json and xml. The files include explanations of the included data, which comprise of raw propensity scores, propensity scores that are normalized to a unit interval using the min-max normalization, and binary predictions. We store these results on the server for at least 3 months. The graphical format is color-coded and interactive with zoom, selection, image download, pan, and callouts features. The interactive color-coded panels (Figures 1 and 2B) are grouped into three parts: (i) putative disorder (in pink); (ii) putative linkers (in yellow) and (iii) putative disordered binding regions (MoRF in light red, protein-binding in dark red, DNA-binding in blue, RNA-binding in light blue and lipid-binding in green). Each panel displays protein-level data at the top, which includes percentage of predicted residues and number of predicted regions (length $\geq$ 4 residues). Residue-level predictions are displayed as propensity scores plotted in a line graph. The binarized labels are shown above as horizontal bars. Threshold values that are used to derive binary predictions (residues with propensities > threshold are classified as disordered/functional) are marked as dashed horizontal line on the line-graphs. The threshold values were established by the authors of the methods and they are typically calibrated to ensure near native rate of the predicted disordered/functional residues (36,47,48,50,52,54). The range of residues in the predicted region and their

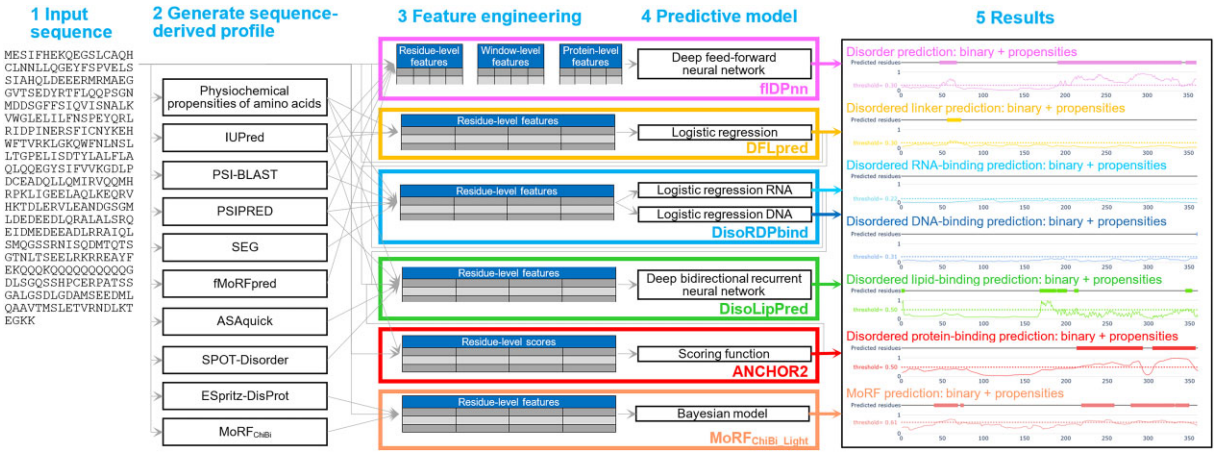


Figure 1. Workflow of the DEPICTER2 webserver.

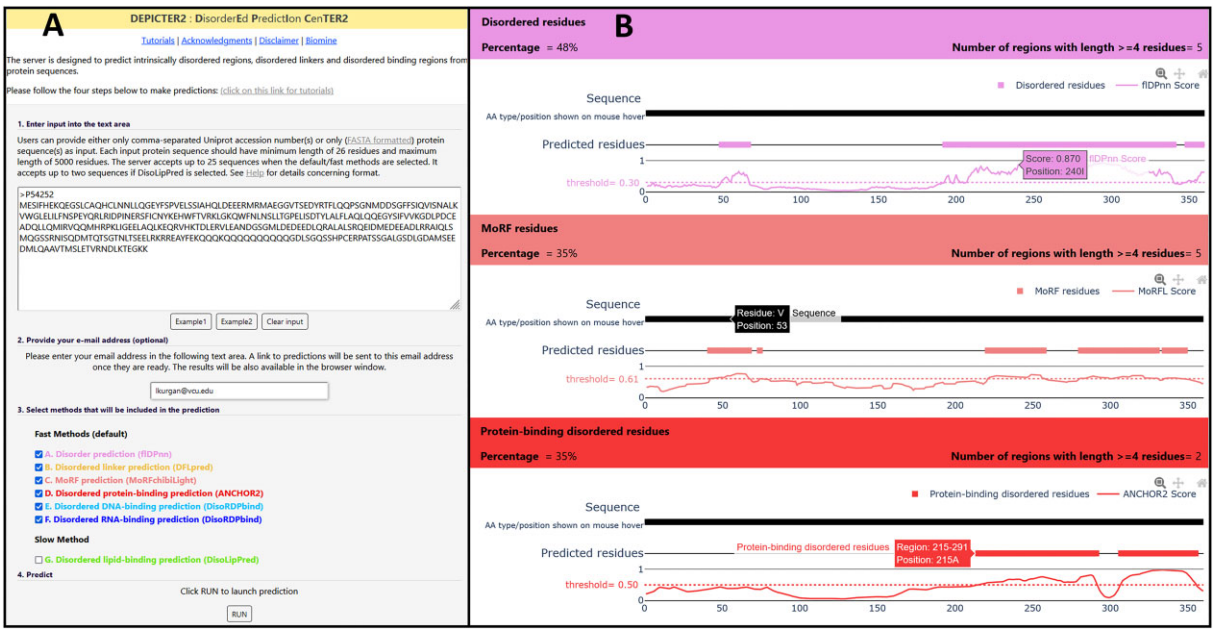


Figure 2. Web interface of the DEPICTER2 server (panel **A**) and prediction results (panel **B**) for the human Ataxin-3 protein (Disprot ID: DP00576; Uniprot ID: P54252). Panel B shows the interactive color-coded panels for predictions of disordered residues (pink), MoRF residues (light-red) and protein-binding residues (red).

underlying propensity scores can be viewed on the mouse hover. Each panel allows zooming into a section of the plot, panning axes on both sides, resetting axes to original view, and downloading it as an image in the PNG and SVG formats.

We explain how to read the outputs using the predictions for Ataxin-3 (Figure 2B). Our webserver predicts that Ataxin-3 has 48% of disordered residues (top of the pink panel in Figure 2B), which comes close to the native disorder content of 52.1% reported in the reference database DisProt (Disprot ID: DP00576). DEPICTER2 predicts four IDRs at the C-terminus (positions 193–196, 199–202, 207–340 and 349–361; pink panel in Figure 2B). These regions coincide with the position of the native IDR (positions 174–361) (57,58). DEPICTER2 also predicts a putative IDR at positions 49–66; however, this region has lower values of the underlying predicted propensity scores when compared to the regions at the C-terminus. More broadly, putative IDRs (binary predictions) that are associated with higher propensities are more likely to correspond to correct predictions. The webserver also predicts two protein-binding regions (positions 215–291 and 307–355; dark red panel at the bottom of Figure 2B) that are in good agreement with the protein-binding UIM domains of Ataxin-3 (58,61). MoRF predictions (light-red panel in Figure 2B) include five regions, three of which coincide with the protein-binding regions of Ataxin-3, while the two short regions near the N-terminus are likely spurious predictions.

Runtime

We include a runtime analysis for the six methods in DEPICTER2 in the context of the size of the input

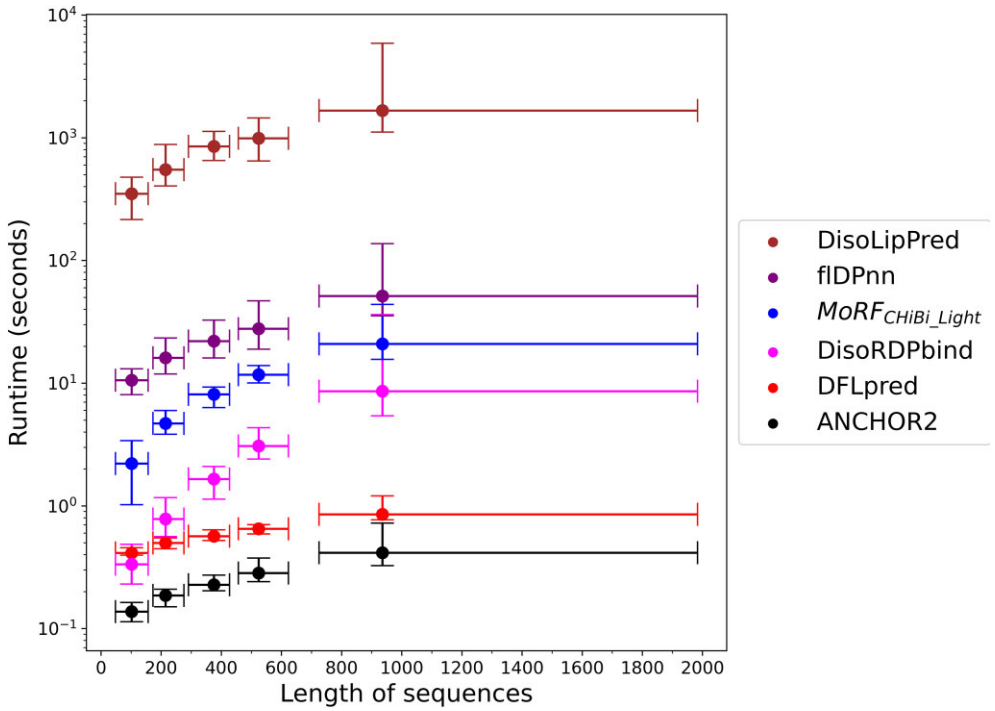


Figure 3. Comparison of runtime for the six methods included in DEPICTER2. We measure runtime on 100 randomly selected proteins from the DisProt dataset from the CAID experiment. Sequences are sorted in the ascending order by their length and divided into 5 equally sized subsets. The y-axis reports median runtime for each protein subset measured in seconds using base 10 logarithmic scale. The x-axis shows the corresponding median sequence length. The error bars along both axes denote the 5th and 95th percentiles of the values for a given protein set.

proteins. We collect and compare their runtime using randomly selected 100 proteins from the DisProt dataset that was used in CAID (34) utilizing the same hardware and operating system: Linux OS (Ubuntu v14.04.5) with 48 64-bit Intel processors and 128 GM RAM. To accommodate for performance variation due to a background workload, we measure the runtime three times for each predictor with a break in between each run, and record average of the three replicates. To study the effect of sequence length on runtime we sort the sequences in the ascending order by their length into five equally sized bins. Figure 3 plots the median runtime measured in seconds (y-axis in base 10 logarithmic scale) against the median sequence length (x-axis) for each bin. Runtime of DisoLipPred is considerably higher than that of the other five methods, by about 3 orders of magnitude compared to ANCHOR2 and DFLpred and 2 orders of magnitude compared to $MoRF_{CHiBi\_Light}$, DisoRDPbind and flDPnn. Consequently, we categorize DisoLipPred as a slow method and limit the webserver input for that method to two proteins. The five fast predictors take <30 s to produce result for an average length protein, with ANCHOR2 and DFLpred completing predictions in under 1 s. Figure 3 also reveals that the runtime increases for longer sequences. However, the degree of the increase varies between tools. DFLpred and ANCHOR2 are the least affected by sequence length as their runtime increases 2 times between the shortest and longest sequence bins, compared to the DisoRDPbind that suffers the worst increase by 25 times.

SUMMARY

Despite the availability of nearly 150 intrinsic disorder and disorder function predictors, convenient options to obtain high-quality predictions that comprehensively cover disorder and a broad selection of its functions are lacking. DEPICTER2 webserver substantially extends its prototype DEPICTER and offers a one-stop solution that includes prediction of intrinsic disorder by the accurate and fast flDPnn along with five state-of-the-art methods that deliver complete coverage of the currently available disorder function predictions: disordered linkers, MoRFs, and disordered protein-, RNA-, DNA- and lipid-interacting regions. Ability to predict interacting regions will facilitate downstream efforts to utilize this knowledge for other applications, such as drug design. Recent works point to an untapped value of utilizing IDPs as drug targets (19,62), for instance in the context of host-pathogen interactions and formation of protein assemblies and biomolecular condensates (63,64). This will require the development of novel disorder-specific scoring functions, following similar efforts for the structured interactions (65), and access to a curated collection of annotations of IDP-drug interactions, with the latter suffering a limited size.

DEPICTER2 automatically runs the six methods on the server-backend without the need to install any software. It provides an easy to navigate input interface that supports selection of any combination of methods and batch submission. The webserver generates predictions in two ways, as a consistently formatted and easy to parse text files and

color-coded graphical interface with interactive features that include residue-level and protein-level results. In a nutshell, DEPICTER2 is an accurate and fast platform that provides a holistic approach for disorder and disorder function predictions. The DEPICTER2 webserver is freely available at http://biomine.cs.vcu.edu/servers/DEPICTER2/. We are committed to maintaining this resource in the long term and plan to update it periodically to incorporate newer versions of the predictors that it covers and to extend the scope by inclusion of additional functions that will become predictable in the future. Moreover, users interested in predictions for large collections of proteins should consider the DescribePROT database (66) at http://biomine.cs.vcu.edu/servers/DESCRIBEPROT/. DescribePROT provides access to pre-computed predictions from several methods included in DEPICTER2, such as DFLpred, DisoRDPbind and MoRFchibi, for 2.3 million proteins from 273 complete proteomes of popular/model organisms. We plan to incorporate predictions of the other three methods into this resource in a near future.

DATA AVAILABILITY

DEPICTER2 is freely available at http://biomine.cs.vcu.edu/servers/DEPICTER2/.

FUNDING

National Science Foundation [DBI2146027, IIS2125218]; Robert J. Mattauch Endowment funds (to L.K.). Funding for open access charge: NSF.
Conflict of interest statement. None declared.

REFERENCES

1. Habchi,J., Tompa,P., Longhi,S. and Uversky,V.N. (2014) Introducing protein intrinsic disorder. *Chem. Rev.*, **114**, 6561–6588.
2. Lieutaud,P., Ferron,F., Uversky,A.V., Kurgan,L., Uversky,V.N. and Longhi,S. (2016) How disordered is my protein and what is its disorder for? A guide through the “dark side” of the protein universe. *Intrinsically Disord. Proteins*, **4**, e1259708.
3. Oldfield,C.J., Uversky,V.N., Dunker,A.K. and Kurgan,L. (2019) Introduction to intrinsically disordered proteins and regions. In: Salvi,N. (ed). *Intrinsically Disordered Proteins*. Academic Press, pp. 1–34.
4. Xue,B., Dunker,A.K. and 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**, 137–149.
5. Peng,Z., Yan,J., Fan,X., Mizianty,M.J., Xue,B., Wang,K., Hu,G., Uversky,V.N. and Kurgan,L. (2015) Exceptionally abundant exceptions: comprehensive characterization of intrinsic disorder in all domains of life. *Cell. Mol. Life Sci.*, **72**, 137–151.
6. Peng,Z., Oldfield,C.J., Xue,B., Mizianty,M.J., Dunker,A.K., Kurgan,L. and Uversky,V.N. (2014) A creature with a hundred waggly tails: intrinsically disordered proteins in the ribosome. *Cell. Mol. Life Sci.*, **71**, 1477–1504.
7. Peng,Z., Mizianty,M.J., Xue,B., Kurgan,L. and Uversky,V.N. (2012) More than just tails: intrinsic disorder in histone proteins. *Mol. Biosyst.*, **8**, 1886–1901.
8. Peng,Z., Xue,B., Kurgan,L. and Uversky,V.N. (2013) Resilience of death: intrinsic disorder in proteins involved in the programmed cell death. *Cell Death Differ.*, **20**, 1257–1267.
9. Xue,B. and Uversky,V.N. (2014) Intrinsic disorder in proteins involved in the innate antiviral immunity: another flexible side of a molecular arms race. *J. Mol. Biol.*, **426**, 1322–1350.
10. Ibrahim,A.Y., Khaodeuanepheng,N.P., Amarasekara,D.L., Correia,J.J., Lewis,K.A., Fitzkee,N.C., Hough,L.E. and Whitten,S.T. (2023) Intrinsically disordered regions that drive phase separation form a robustly distinct protein class. *J. Biol. Chem.*, **299**, 102801.
11. Zhao,B., Katuwawala,A., Oldfield,C.J., Hu,G., Wu,Z., Uversky,V.N. and Kurgan,L. (2021) Intrinsic Disorder in Human RNA-Binding Proteins. *J. Mol. Biol.*, **433**, 167229.
12. Zhou,J.H., Zhao,S.W. and Dunker,A.K. (2018) Intrinsically Disordered Proteins Link Alternative Splicing and Post-translational Modifications to Complex Cell Signaling and Regulation. *J. Mol. Biol.*, **430**, 2342–2359.
13. Staby,L., O’Shea,C., Willemoes,M., Theisen,F., Kragelund,B.B. and Skriver,K. (2017) Eukaryotic transcription factors: paradigms of protein intrinsic disorder. *Biochem. J.*, **474**, 2509–2532.
14. Kjaergaard,M. and Kragelund,B.B. (2017) Functions of intrinsic disorder in transmembrane proteins. *Cell. Mol. Life Sci.*, **74**, 3205–3224.
15. 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.
16. Zhao,B., Katuwawala,A., Uversky,V.N. and Kurgan,L. (2021) IDPology of the living cell: intrinsic disorder in the subcellular compartments of the human cell. *Cell. Mol. Life Sci.*, **78**, 2371–2385.
17. Uversky,V.N., Dave,V., Iakoucheva,L.M., Malaney,P., Metallo,S.J., Pathak,R.R. and Joerger,A.C. (2014) Pathological unfoldomics of uncontrolled chaos: intrinsically disordered proteins and human diseases. *Chem. Rev.*, **114**, 6844–6879.
18. Kulkarni,P. and Uversky,V.N. (2019) Intrinsically disordered proteins in chronic diseases. *Biomolecules*, **9**, 147.
19. Hu,G., Wu,Z., Wang,K., Uversky,V.N. and Kurgan,L. (2016) Untapped potential of disordered proteins in current druggable human proteome. *Curr. Drug Targets*, **17**, 1198–1205.
20. Ambadipudi,S. and Zweckstetter,M. (2016) Targeting intrinsically disordered proteins in rational drug discovery. *Expert Opin Drug Discov*, **11**, 65–77.
21. Sickmeier,M., Hamilton,J.A., LeGall,T., Vacic,V., Cortese,M.S., Tantos,A., Szabo,B., Tompa,P., Chen,J., Uversky,V.N. *et al.* (2007) DisProt: the Database of Disordered Proteins. *Nucleic Acids Res.*, **35**, D786–D793.
22. Quaglia,F., Meszaros,B., Salladini,E., Hatos,A., Pancsa,R., Chemes,L.B., Pajkos,M., Lazar,T., Pena-Diaz,S., Santos,J. *et al.* (2022) DisProt in 2022: improved quality and accessibility of protein intrinsic disorder annotation. *Nucleic Acids Res.*, **50**, D480–D487.
23. Campen,A., Williams,R.M., Brown,C.J., Meng,J., Uversky,V.N. and Dunker,A.K. (2008) TOP-IDP-scale: a new amino acid scale measuring propensity for intrinsic disorder. *Protein Pept. Lett.*, **15**, 956–963.
24. Zhao,B. and Kurgan,L. (2022) Compositional bias of intrinsically disordered proteins and regions and their predictions. *Biomolecules*, **12**, 888 .
25. Zhao,B. and Kurgan,L. (2021) Surveying over 100 predictors of intrinsic disorder in proteins. *Expert Rev Proteomics*, **18**, 1019–1029.
26. Liu,Y., Wang,X. and Liu,B. (2019) A comprehensive review and comparison of existing computational methods for intrinsically disordered protein and region prediction. *Brief Bioinform*, **20**, 330–346.
27. Basu,S., Kihara,D. and Kurgan,L. (2023) Computational prediction of disordered binding regions. *Comput. Struct. Biotechnol. J.*, **21**, 1487–1497.
28. Katuwawala,A., Peng,Z.L., Yang,J.Y. and Kurgan,L. (2019) Computational prediction of MoRFs, short disorder-to-order transitioning protein binding regions. *Comput. Struct. Biotech.*, **17**, 454–462.
29. Katuwawala,A., Ghadermarzi,S. and Kurgan,L. (2019) Computational prediction of functions of intrinsically disordered regions. *Prog. Mol. Biol. Transl. Sci.*, **166**, 341–369.
30. Zhao,B. and Kurgan,L. (2023) Machine learning for intrinsic disorder prediction. *Machine Learning in Bioinformatics of Protein Sequences*. pp. 205–236.
31. Zhao,B. and Kurgan,L. (2022) Deep learning in prediction of intrinsic disorder in proteins. *Comput. Struct. Biotechnol. J.*, **20**, 1286–1294.
32. Monastyrskyy,B., Kryshtafovych,A., Moult,J., Tramontano,A. and Fidelis,K. (2014) Assessment of protein disorder region predictions in CASP10. *Proteins*, **82**, 127–137.

33. Melamud,E. and Moult,J. (2003) Evaluation of disorder predictions in CASP5. *Proteins*, **53**, 561–565.
34. Necci,M., Piovesan,D., Predictors,C., DisProt,C. and Tosatto,S.C.E. (2021) Critical assessment of protein intrinsic disorder prediction. *Nat. Methods*, **18**, 472–481.
35. Lang,B. and Babu,M.M. (2021) A community effort to bring structure to disorder. *Nat. Methods*, **18**, 454–455.
36. Hu,G., Katuwawala,A., Wang,K., Wu,Z., Ghadermarzi,S., Gao,J. and Kurgan,L. (2021) flDPnn: accurate intrinsic disorder prediction with putative propensities of disorder functions. *Nat. Commun.*, **12**, 4438.
37. Hanson,J., Paliwal,K.K., Litfin,T. and Zhou,Y. (2019) SPOT-Disorder2: improved protein intrinsic disorder prediction by ensembled deep learning. *Genomics Proteomics Bioinformatics*, **17**, 645–656.
38. Mirabello,C. and Wallner,B. (2019) rawMSA: end-to-end deep learning using raw multiple sequence alignments. *PLoS One*, **14**, e0220182.
39. Wang,S., Ma,J.Z. and Xu,J.B. (2016) AUCpreD: proteome-level protein disorder prediction by AUC-maximized deep convolutional neural fields. *Bioinformatics*, **32**, 672–679.
40. Tang,Y.J., Pang,Y.H. and Liu,B. (2021) IDP-Seq2Seq: identification of intrinsically disordered regions based on sequence to sequence learning. *Bioinformatics*, **36**, 5177–5186.
41. Tang,Y.J., Pang,Y.H. and Liu,B. (2022) DeepIDP-2L: protein intrinsically disordered region prediction by combining convolutional attention network and hierarchical attention network. *Bioinformatics*, **38**, 1252–1260.
42. Fang,M., He,Y., Du,Z. and Uversky,V.N. (2022) DeepCLD: an Efficient Sequence-Based Predictor of Intrinsically Disordered Proteins. *IEEE/ACM Trans. Comput. Biol. Bioinform.*, **19**, 3154–3159.
43. Oates,M.E., Romero,P., Ishida,T., Ghalwash,M., Mizianty,M.J., Xue,B., Dosztanyi,Z., Uversky,V.N., Obradovic,Z., Kurgan,L. *et al.* (2013) D(2)P(2): database of disordered protein predictions. *Nucleic Acids Res.*, **41**, D508–D516.
44. Piovesan,D., Del Conte,A., Clementel,D., Monzon,A.M., Bevilacqua,M., Aspromonte,M.C., Iserte,J.A., Orti,F.E., Marino-Buslje,C. and Tosatto,S.C.E. (2023) MobiDB: 10 years of intrinsically disordered proteins. *Nucleic Acids Res.*, **51**, D438–D444.
45. Barik,A., Katuwawala,A., Hanson,J., Paliwal,K., Zhou,Y. and Kurgan,L. (2020) DEPICTER: intrinsic Disorder and Disorder Function Prediction Server. *J. Mol. Biol.*, **432**, 3379–3387.
46. Hanson,J., Paliwal,K.K. and Zhou,Y. (2018) Accurate single-sequence prediction of protein intrinsic disorder by an ensemble of deep recurrent and convolutional architectures. *J. Chem. Inf. Model.*, **58**, 2369–2376.
47. Meszaros,B., Erdos,G. and Dosztanyi,Z. (2018) IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.*, **46**, W329–W337.
48. Meng,F. and Kurgan,L. (2016) DFLpred: high-throughput prediction of disordered flexible linker regions in protein sequences. *Bioinformatics*, **32**, i341–i350.
49. Peng,Z., Wang,C., Uversky,V.N. and Kurgan,L. (2017) Prediction of disordered RNA, DNA, and protein binding regions using DisoRDPbind. *Methods Mol. Biol.*, **1484**, 187–203.
50. Peng,Z. and Kurgan,L. (2015) High-throughput prediction of RNA, DNA and protein binding regions mediated by intrinsic disorder. *Nucleic Acids Res.*, **43**, e121.
51. Yan,J., Dunker,A.K., Uversky,V.N. and Kurgan,L. (2016) Molecular recognition features (MoRFs) in three domains of life. *Mol. Biosyst.*, **12**, 697–710.
52. Malhis,N., Jacobson,M. and Gsponer,J. (2016) MoRFchibi SYSTEM: software tools for the identification of MoRFs in protein sequences. *Nucleic Acids Res.*, **44**, W488–W493.
53. Oldfield,C.J., Peng,Z. and Kurgan,L. (2020) Disordered RNA-binding region prediction with DisoRDPbind. *Methods Mol. Biol.*, **2106**, 225–239.
54. Katuwawala,A., Zhao,B. and Kurgan,L. (2021) DisoLipPred: accurate prediction of disordered lipid binding residues in protein sequences with deep recurrent networks and transfer learning. *Bioinformatics*, **38**, 115–124.
55. Yan,J., Dunker,A.K., Uversky,V.N. and Kurgan,L. (2016) Molecular recognition features (MoRFs) in three domains of life. *Mol. Biosyst.*, **12**, 697–710.
56. Vacic,V., Oldfield,C.J., Mohan,A., Radivojac,P., Cortese,M.S., Uversky,V.N. and Dunker,A.K. (2007) Characterization of molecular recognition features, MoRFs, and their binding partners. *J. Proteome Res.*, **6**, 2351–2366.
57. Sicorello,A., Kelly,G., Oregioni,A., Nováček,J., Sklenář,V. and Pastore,A. (2018) The structural properties in solution of the intrinsically mixed folded protein Ataxin-3. *Biophys. J.*, **115**, 59–71.
58. Masino,L., Musi,V., Menon,R.P., Fusi,P., Kelly,G., Frenkiel,T.A., Trottier,Y. and Pastore,A. (2003) Domain architecture of the polyglutamine protein ataxin-3: a globular domain followed by a flexible tail. *FEBS Lett.*, **549**, 21–25.
59. Burnett,B., Li,F. and Pittman,R.N. (2003) The polyglutamine neurodegenerative protein ataxin-3 binds polyubiquitylated proteins and has ubiquitin protease activity. *Hum. Mol. Genet.*, **12**, 3195–3205.
60. Donaldson,K.M., Li,W., Ching,K.A., Batalov,S., Tsai,C.-C. and Joazeiro,C.A.P. (2003) Ubiquitin-mediated sequestration of normal cellular proteins into polyglutamine aggregates. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 8892–8897.
61. Bai,J.J., Safadi,S.S., Mercier,P., Barber,K.R. and Shaw,G.S. (2013) Ataxin-3 is a multivalent ligand for the parkin Ubl domain. *Biochemistry*, **52**, 7369–7376.
62. Hosoya,Y. and Ohkanda,J. (2021) Intrinsically disordered proteins as regulators of transient biological processes and as untapped drug targets. *Molecules*, **26**, 2118.
63. Biesaga,M., Frigole-Vivas,M. and Salvatella,X. (2021) Intrinsically disordered proteins and biomolecular condensates as drug targets. *Curr. Opin. Chem. Biol.*, **62**, 90–100.
64. Blundell,T.L., Gupta,M.N. and Hasnain,S.E. (2020) Intrinsic disorder in proteins: relevance to protein assemblies, drug design and host-pathogen interactions. *Prog. Biophys. Mol. Biol.*, **156**, 34–42.
65. Li,H.J., Sze,K.H., Lu,G. and Ballester,P.J. (2021) Machine-learning scoring functions for structure-based virtual screening. *Wires Comput Mol Sci*, **11**, e1478.
66. Zhao,B., Katuwawala,A., Oldfield,C.J., Dunker,A.K., Faraggi,E., Gsponer,J., Kloczkowski,A., Malhis,N., Mirdita,M., Obradovic,Z. *et al.* (2021) DescribePROT: database of amino acid-level protein structure and function predictions. *Nucleic Acids Res.*, **49**, D298–D308.