DEPICTER: Intrinsic Disorder and Disorder Function Prediction Server

Amita Barik[1,2,†], Akila Katuwawala[1,†], Jack Hanson[3], Kuldip Paliwal[3], Yaoqi Zhou[4,5] and Lukasz Kurgan[1]

*1 - **Department of Computer Science,** Virginia Commonwealth University, Richmond, VA, 23284, USA*
*2 - **Department of Biotechnology,** National Institute of Technology, Durgapur, India*
*3 - **Signal Processing Laboratory,** Griffith University, Brisbane, QLD, 4122, Australia*
*4 - **School of Information and Communication Technology,** Griffith University, Gold Coast, QLD, 4222, Australia*
*5 - **Institute for Glycomics,** Griffith University, Gold Coast, QLD, 4222, Australia*

***Correspondence to Lukasz Kurgan:** Fax: +804 828-2771. lkurgan@vcu.edu*
https://doi.org/10.1016/j.jmb.2019.12.030
Edited by Michael Sternberg

Abstract

Computational predictions of the intrinsic disorder and its functions are instrumental to facilitate annotation for the millions of unannotated proteins. However, access to these predictors is fragmented and requires substantial effort to find them and to collect and combine their results. The DEPICTER (DisorderEd PredictIon CenTER) server provides first-of-its-kind centralized access to 10 popular disorder and disorder function predictions that cover protein and nucleic acids binding, linkers, and moonlighting regions. It automates the prediction process, runs user-selected methods on the server side, visualizes the results, and outputs all predictions in a consistent and easy-to-parse format. DEPICTER also includes two accurate consensus predictors of disorder and disordered protein binding. Empirical tests on an independent (low similarity) benchmark dataset reveal that the computational tools included in DEPICTER generate accurate predictions that are significantly better than the results secured using sequence alignment. The DEPICTER server is freely available at http://biomine.cs.vcu.edu/servers/DEPICTER/.

Introduction

Intrinsic disordered proteins (IDPs) and intrinsically disordered protein regions (IDRs) lack stable tertiary structure and form dynamic conformational ensembles under physiological conditions [1–3]. Recent computational studies estimate that they are highly abundant in nature, with up to 17% of eukaryotic proteins (depending on an organism) that are entirely disordered [4] and between 30 and 50% that have at least one long IDR (≥30 consecutive residues) [5,6]. IDPs and IDRs are instrumental for a wide range of cellular functions that include signaling and molecular recognition [7,8], translation [9–11], regulation of transcription [12], cell death processes [13–15], innate immune response [16], viral life cycle [17–19], and many others. They are implicated in the dark proteome [20,21], often found to contribute to human diseases [22,23], and are being considered as attractive new class of targets for drug discovery [24,25]. However, experimental annotations of IDRs and their functions are limited to only about 1600 proteins that are stored in the DisProt database [26,27]. This gave rise to the development of a large collection of over 70 computational methods that predict IDRs and IDPs from the protein sequences [3,28–33]. Recent empirical studies have shown that some of these methods provide highly accurate predictions [34–37]. Moreover, two dozen computational tools that predict several functions of IDRs were published and released over the last decade [32,38,39]. These methods address sequence-based prediction of molecular partners that interact with IDRs, including proteins, DNA and RNAs, and a selected set of cellular functions, such as flexible linkers and

moonlighting regions. These computational predictors can be used to accurately and in cost- and runtime-efficient way predict and functionally annotate IDPs and IDRs for the millions of proteins that lack these annotations.

The predictions produced by these methods can be collected either by the means of their webservers/implementations provided and supported by the authors and by using popular and large databases of precomputed predictions: D^2P^2 [40] and MobiDB [41]. Both databases offer access to results generated by a large pool of disorder predictors for millions of sequences proteins. More specifically, D^2P^2 covers 9 disorder predictions for about 10.5 million proteins, while MobiDB provides 10 disorder predictions for the contents of the October 2017 version of the UniProt repository [42], which includes around 90 million proteins. While these two resources provide unrivaled coverage of the disorder predictions, they offer a rather limited selection of the disorder function predictions that includes only the protein-binding regions predicted by the ANCHOR method [43]. Moreover, they are constrained to the set of proteins that they currently include. This means that the users must use the webservers or implementations of the available predictors for the huge number of proteins that are not currently included in these databases. More specifically, the current version of UniProt includes over 171 million proteins compared to 90 million in MobiDB. Collection of these predictions is rather difficult since it demands finding their websites/implementations, running the predictions using multiple different interfaces that require reformatting the input protein data, and assembling the results that are provided in a wide range of formats. The aforementioned issues can be alleviated by the development of a predictive resource that provides integrated access to a comprehensive set of disorder and disorder function predictors. While there are no such resources for the disorder prediction, they are already available for the prediction of various aspects of protein structure, including PSIPRED workbench [44], SCRATCH [45], PredictProtein [46], and MULTICOM [47].

To this end, we provide first-of-its-kind webserver for the sequence-based prediction of intrinsic disorder and disorder functions. The DEPICTER (DisorderEd PredictIon CenTER) server integrates predictions of intrinsic disorder and several disorder functions using several popular and runtime-efficient methods. The server automates the entire process of prediction across all these tools, visualizes the results, and outputs an easy-to-parse text file that provides all predictions in a consistent format. DEPICTER is freely available at http://biomine.cs.vcu.edu/servers/DEPICTER/.

Materials and Methods

Selection of predictors for inclusion into the DEPICTER webserver

There are over 70 disorder predictors and another 25 predictors of disorder function [28,29,31–33,38]. It would be infeasible and unnecessary to develop a platform that provides access to all these tools. We select a collection of fast (needing short runtime to make predictions), recently published and empirically shown to provide accurate predictions tools that predict disorder and that provide a comprehensive coverage of the currently predicted disorder functions. In total, the DEPICTER server includes 10 predictive tools: SPOT-Disorder-Single [48], two versions of IUPred2 [49]: $IUPred2_{long}$ and $IUPred2_{short}$, DFLpred [50], DMRpred [51], $DisoRDPbind_{RNA}$ [52,53], $DisoRDPbind_{DNA}$ [52,53], fMoRFpred [54], $DisoRDPbind_{protein}$ [52,53], and ANCHOR2 [49]. Each of these methods was empirically shown to provide competitive levels of predictive performance, outperforming or at least matching the performance of other currently available approaches that can be used to make the same predictions [48–51,53,54]. Further discussion of the selection process is available in the Supplement.

Interface and architecture of the DEPICTER server

The server is available at http://biomine.cs.vcu.edu/servers/DEPICTER/. Users submits a FASTA-formatted sequence of the input protein using the interface shown in Fig. 1A. The main page of the server offers a brief tutorial that explains how to use the interface. We encourage the users to provide an email address where the server will send an email notification with links to the prediction results upon completion of the prediction. Users have the option to select a subset of predictors to run—by default the server runs all available predictors. Once the sequence is submitted by clicking the "RUN" button, the browser is redirected to a status page that provides the current position of this request in the server queue. We utilize the first-come-first-serve queue where each user is limited to submit up to five concurrent requests. This limit is imposed to ensure a fair access across users. The status page is automatically redirected to the results page when predictions are completed. Closing the status webpage prevents the users for advancing to the results page; however, links to the predictions will still be sent through email. The entire prediction takes less than 1 min for an average-size protein sequence. The front end of the server is implemented in HTML and Javascript, while the back end relies on PHP, Java, Python, and MySQL database.

The programs required to produce predictions are run automatically by scripts on the server side. The workflow of the DEPICTER server is summarized in Fig. 2. The 10 predictors that are included in the DEPICTER server (left side of Fig. 2) make six distinct color-coded types of predictions of disordered regions (by SPOT-Disorder-Single, $IUPred2_{long}$, and $IUPred2_{short}$ methods), protein-binding IDRs, which include MoRFs ($DisoRDPbind_{protein}$,

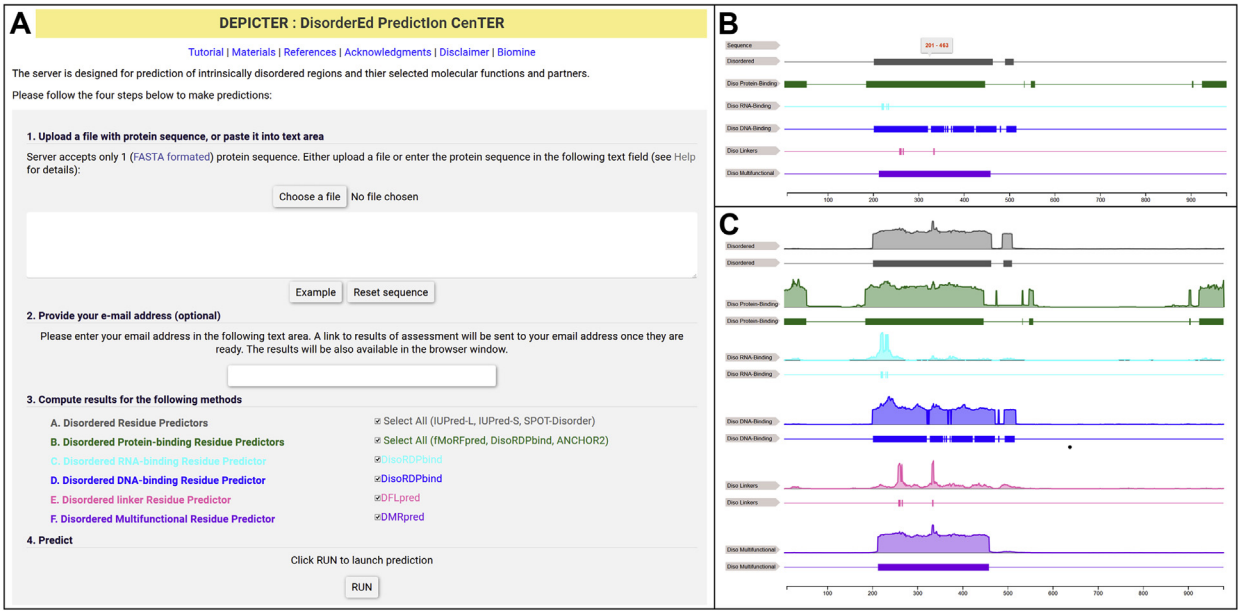

Fig. 1. Web interface of the DEPICTER server (panel A) and predictions produced by the DEPICTER server for the silent information regulator Sir3p (DisProt id: DP00533; UniProt id: P06701). Panel B shows the short prediction profile that includes the binary predictions of the disordered regions (in gray), disordered protein-binding regions (green), RNA-binding regions (light blue), DNA-binding regions (dark blue), linkers (pink), and multifunctional regions (violet). Panel C shows the complete prediction profile that includes the color-coded binary predictions (horizontal lines) and the corresponding real-valued propensities (located above the binary prediction lines). The numeric line at the bottom shows positions in the protein chain.

ANCHOR2, and fMoRFpred), DNA-binding IDRs ($DisoRDPbind_{DNA}$), RNA-binding IDRs ($DisoRDPbind_{RNA}$), disordered linkers (DLFpred), and moonlighting IDRs (DMRpred). Each predictor generates two types of outputs for each residue in the input protein sequence: real-valued propensity that quantifies likelihood that a given residue is disordered or carries out a given function, and a binary value where 1 means that a given residue is disordered (or carries out a given function) and 0 denotes that it is structured (or not associated with the function). We designed consensus predictors for the prediction of disordered regions and protein-binding IDRs for which DEPICTER includes multiple methods. A consensus predictor combines results produced by multiple predictive tools to generate a new prediction with the premise that the new result has higher accuracy compared to each of the input predictions. The inclusion of the consensus is motivated by two factors. First, predictions generated by multiple methods could be conflicting with each other, which would confuse the end users. The consensus offers a single prediction that resolves these conflicts. Second, previously developed consensus-based predictors of the intrinsic disorder were empirically shown to provide improvements in the predictive quality [55–57]. This means that inclusion of the consensus is likely to strengthen the quality of the outputs produced by the DEPICTER server. Moreover, we use the accurate disorder consensus prediction to improve the capability of the disorder function predictors to differentiate functional IDRs from structured protein sequences that carry out similar functions. Namely, we multiply the propensities generated by the fMoRFpred, DLFpred, DMRpred DisoRDPbind, and ANCHOR2 by the propensities produced by the disorder prediction consensus. This ensures that the resulting values are going to be high for the disordered regions predicted by the aforementioned tools and low for the structured protein sequences. The consensus predictor of the protein-binding IDRs uses these adjusted propensities as the inputs. After the predictions are completed, the users are presented with the *short prediction profile* that is produced by the server (right side of Fig. 2). This profile incorporates six putative binary annotations of the disordered residues (based on the consensus), disordered protein-binding residues (based on the consensus), disordered protein–DNA and protein–RNA binding residues, linker residues, and multifunctional disordered residues. This profile provides a concise and complete overview of the location and function of the putative disordered regions in the input protein chain. Users are also provided with the *complete prediction profile* that includes the complete set of 12 propensities and the corresponding 12 binary predictions produced by the 10 individual predictors and the two consensuses. The predictions are color-coded and visualized on the results page using the BioJS program [58]. The server also provides an option to download the results in a parsable comma-separated text file. We archive user-generated predictions for at least 1 month. They can be accessed directly via a unique link sent by email.

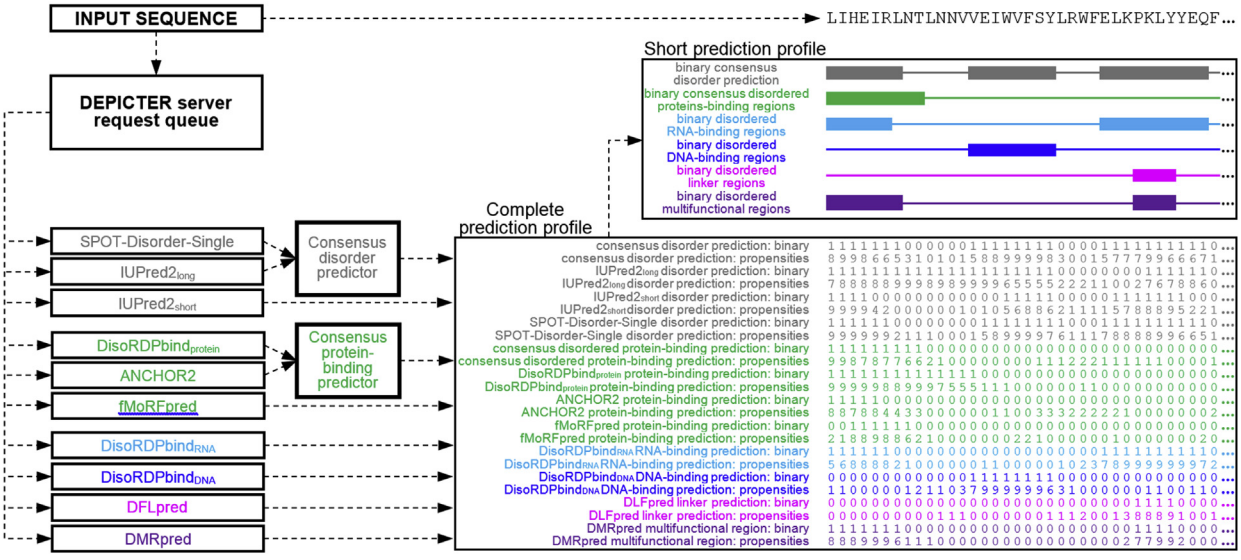

Fig. 2. The workflow of the predictions in the DEPICTER server.

Design and evaluation setup

We use training and test datasets to empirically design the two consensus methods and to quantify the predictive performance of the predictions generated by the DEPICTER server. Supplementary Table S1 summarizes the contents of the two datasets. The functionally annotated test dataset is available at http://biomine.cs.vcu.edu/servers/DEPICTER/. Details describing the process of collection and annotations of these datasets are provided in the Supplement.

We quantify the predictive quality with several popular metrics that were used in the most recent Critical Assessment of protein Structure Prediction (CASP) experiments that included disorder predictions: CASP9 [59] and CASP10 [60] and in several recent empirical assessments [37,61,62]. They include Matthews Correlation Coefficient (MCC), sensitivity, specificity, area under the Receiver Operating Curve (AUC_{ROC}), and area under the Precision Recall Curve (AUC_{PRC}). We also sample the test dataset to equalize the rates in functional and disordered residues to allow for direct comparison of the predictive performance across different predictive targets. Definitions of these metrics and details of the sampling process are given in the Supplement.

Consensus predictors for the disordered and the disordered protein-binding residues

We design two consensus predictors that combine results produced by the multiple methods that predict disorder and disorder protein binding that are included in DEPICTER. The aim of the consensuses is to provide improved predictive performance when compared to their input predictions and resolve potential conflicts between multiple individual predictions. The design process, the corresponding empirical results, and comparison of predictions between the consensus method and the corresponding input predictors are included in the Supplement.

Results

Assessment of predictive quality

Table 1 summarizes the quality of the results produced by the predictors included in the DEPICTER server on the test dataset. This dataset shares low (<30%) similarity to the proteins that were used to develop these predictors. We compare the results generated by the server against a sequence alignment-based prediction. Details concerning the calculation of the alignment-based predictor and the experimental setup are described in the Supplement. Table 1 reveals that each of the 10 predictors included in the DEPICTER server and the two consensuses provide accurate results. The AUC_{ROC} values span between 0.79 ($IUPred2_{short}$) and 0.85 (consensus method) for the disorder prediction and between 0.74 (fMoRFpred) and 0.87 (consensus method) for the prediction of the protein-binding regions. The predictions of the RNA binding, DNA binding, linker, and multifunctional regions secure AUC_{ROC} values equal 0.80, 0.83, 0.71, and 0.83, respectively. The corresponding ROC and precision-recall curves are shown in the Supplementary Figure S2. We note that these results are comparable to the previously published benchmarks. More specifically, $IUPred_{short}$'s, $IUPred_{long}$'s, and SPOT-Disorder-Single's AUC_{ROC} values were reported to range (depending on the test dataset used) between 0.720 and 0.830 (we report 0.791), between 0.706 and 0.838 (we report 0.804), and between 0.792 and 0.905 (we report 0.845), respectively [48]. For the prediction of the protein-binding regions, fMORFpred and ANCHOR2 were

Table 1. Predictive performance on the test dataset for the prediction of disordered residues, disordered protein-, DNA-, and RNA-binding residues, disordered linker residues, and disordered moonlighting (multifunctional) residues.

Target of prediction	Predictor	AUC_{ROC}	AUC_{PRC}	Sensitivity at specificity = 0.9	Sensitivity	Specificity	MCC
Disordered residues	$IUPred2_{long}$	0.804*	0.525*	0.504	0.623	0.85	0.446#
	$IUPred2_{short}$	0.791*	0.485*	0.424	0.665	0.81	0.418#
	SPOT-Disorder-Single	0.845*	0.585*	0.559	0.716	0.83	0.487#
	Sequence alignment	*N/A*	*N/A*	*N/A*	*0.138*	*0.97*	*0.203*
	Consensus predictor	**0.853**	**0.611**	**0.575**	**0.712**	**0.85**	**0.507#**
Disordered protein-binding/MoRF residues	fMORFpred	0.738*	0.429*	0.238	0.615	0.76	0.323#
	$DisoRDPbind_{protein}$	0.863*	0.552*	0.577	0.760	0.84	0.515#
	ANCHOR2	0.854*	0.510*	0.510	0.771	0.81	0.509#
	Sequence alignment	*N/A*	*N/A*	*N/A*	*0.150*	*0.99*	*0.312*
	Consensus predictor	**0.872**	**0.637**	**0.561**	**0.771**	**0.82**	**0.520#**
Disordered RNA-binding residues	*Sequence alignment*	*N/A*	*N/A*	*N/A*	*0.008*	*1.00*	*0.078*
	$DisoRDPbind_{RNA}$	**0.799**	**0.555**	**0.543**	**0.470**	**0.93**	**0.455#**
Disordered DNA-binding residues	*Sequence alignment*	*N/A*	*N/A*	*N/A*	*0.000*	*1.00*	*0.000*
	$DisoRDPbind_{DNA}$	**0.831**	**0.511**	**0.430**	**0.805**	**0.77**	**0.487#**
Disordered linker residues	*Sequence alignment*	*N/A*	*N/A*	*N/A*	*0.000*	*1.00*	*0.000*
	DFLpred	**0.711**	**0.361**	**0.254**	**0.554**	**0.78**	**0.297#**
Disordered multifunctional residues	*Sequence alignment*	*N/A*	*N/A*	*N/A*	*0.000*	*1.00*	*0.000*
	DMRpred	**0.833**	**0.455**	**0.502**	**0.924**	**0.67**	**0.473#**

The consensus predictors use the best-performing model that relies on the extreme gradient boosting tree (see Supplementary Table S4). Results produced by the methods included in the DEPICTER webserver are compared against alignment-based predictions. The binary predictions were generated from the propensity scores using a threshold such that residues with propensities > threshold are predicted with the target label while the remaining residues are predicted not to have the label. The binary predictive performance measures (sensitivity, specificity, and MCC) are based on the thresholds that were optimized to maximize the MCC value. We also provide the value of sensitivity for the threshold that corresponds to a predefined specificity = 0.9; these sensitivities can be directly compared between different predictors. N/A means that the corresponding score could not be computed since the alignment-based predictions produce only the binary results, i.e., a given residue is aligned to a training residue with the target annotation or it lacks such alignment. The alignment-based predictions could not be used to produce specificity = 0.9 due to the low number of target predictions that they produce. Statistical significance of differences in the AUC_{ROC} and AUC_{PRC} values for IUPred2, SPOT-Disorder-Single, fMoRFpred, $DisoRDPbind_{protein,}$ and ANCHOR2 against the corresponding consensus predictors was assessed using *t*-test (for normal measurement) or Wilcoxon rank test (otherwise); normality was tested with the Anderson–Darling test at 0.05 significance. * denotes that the performance of a given predictor is significantly worse than the corresponding consensus (p-value < 0.001). Using the same tests,# denotes that MCC of the alignment-based prediction is significantly worse than the MCC of a given predictor (p-value < 0.001).

shown to secure AUC_{ROC} of 0.671 (we report 0.738) [54] and 0.865 (we report 0.854) [49], respectively. Finally, the previously published AUC_{ROC} for DLFpred equals 0.715 (we report 0.711) [50] and for DMRpred it equals 0.856 (we report 0.833) [51]. Table 1 shows that the sensitivity values of these predictions range between 0.24 (for fMoRFpred) and 0.58 (for the disorder consensus) when the specificity is fixed at 0.9. This means that between 24% and 58% of the native disordered/functional residues are predicted correctly when the false-positive rate (rate of the nonfunctional/ordered residues predicted as functional/disordered) is at 10%. The MCCs for the prediction of disorder (consensus method), protein binding (consensus method), RNA binding, DNA binding, linker, and multifunctional regions are 0.51, 0.52, 0.46, 0.49, 0.30, and 0.47 respectively. This means that binary predictions are correlated with the native annotations. To compare, the corresponding MCCs obtained via the alignment-based approach are much lower and equal to 0.20, 0.31, 0.08, 0, 0, and 0, respectively. The right-most columns show that the alignment-based results are significantly worse than the predictions generated by the server (p-value < 0.001). Moreover, we also compare the predictive quality of the two consensus methods that we designed for this server with the predictions generated by the corresponding individual predictors. Consistent with the results on the training dataset, the consensuses generate consistently higher values of AUC_{ROC}, $AUC_{PRC,}$ and MCC when compared with the best individual predictors for the same type of prediction. These differences are modest in magnitude and statistically significant (p-value < 0.001), meaning that the modest improvements are robust across different proteins. Moreover, Supplementary Figure S2 shows that both consensuses offer comparable values of precision for high values of recall and substantially higher precision for lower values of recall, when compared with the best individual disorder and protein-binding predictors. This means that the consensuses provide much better predictions for the residues for which they generate high scores. This figure also shows that the two consensuses provide the end users with a high precision >0.8 when correctly predicting about 18% of the native disordered residues and 22% of the native disordered protein-binding residues.

To summarize, the 10 predictors included in the DEPICTER server were previously shown to outperform or at least match the predictive performance of other currently available approaches that make the same predictions [48–51,53,54]. Here, we empirically demonstrate that the two consensuses that we designed for the prediction of the disorder and the disordered protein-binding regions also provide high-quality results. Moreover, our assessment confirms that all 12 methods included in DEPICTER provide accurate predictions, even for the protein sequences that share low similarity with the proteins that were used to develop the predictors included in this server. This means that DEPICTER is capable of providing high-quality predictions for the proteins for which alignment cannot provide accurate results.

Case study

We demonstrate results produced by the DEPICTER server using one of the test proteins, silent information regulator Sir3p (DisProt id: DP00533; UniProt id: P06701). The silent information regulator proteins in the budding yeast include Sir1p, Sir2p, Sir3p, and Sir4p. They are responsible for silencing of chromatin. Sir3p participates in the initiation, propagation, and maintenance of the silenced chromatin [63] and functions as chromatin architectural protein being involved in the compaction of chromatin fiber [64,65]. Sir3p is a multidomain protein that has a long IDR located between structured N- and C-terminal regions. The N-terminus includes structured BAH domain (positions 1–215) [63]. The C-terminus comprises of a long structured region (positions 550–980) that features binding sites for RAP1p [66], histones H3 and H4 [67], and RAD7p [68]. The middle region of Sir3p (positions 216–549) is intrinsically disordered [63]. This IDR interacts with RAP1p [66], RAD7p [68], and Sir4p coiled-coil domain [69]. This implies that this disordered region is multifunctional and interacts with DNA and proteins.

Predictions generated by DEPICTER for the Sir3p protein are shown in panels B and C in Fig. 1. Fig. 1B shows the short prediction profile that includes binary predictions of the disordered regions (in gray), disordered protein-binding regions (green), RNA-binding regions (light blue), DNA-binding regions (dark blue), linkers (pink), and multifunctional regions (violet). The server predicts two IDRs (positions 201–463, and 490–509). These predictions are in good agreement with the location of the native IDR and with the disorder predictions that are available in the MobiDB resource [41]. More importantly, DEPICTER finds a long segment of DNA-binding residues (blue line; positions 201–515) and several protein-binding regions (positions 1–53, 184–446, and 924–978). The putative protein-binding regions (green line) at both termini should be dismissed since they disagree with the prediction of disorder (gray line). The location of the central DNA-binding and protein-binding regions coincides with the disorder predictions and is in line with the native annotations for this IDR [66,68,69]. Moreover, DEPICTER also predicts a long multifunctional IDR (violet line; positions 212–458), providing further support for the aforementioned predictions of interactions with the two distinct partner types. Fig. 1C shows the full prediction profile that includes the color-coded binary predictions and real-valued propensities that quantify likelihood of a given target annotation (disorder and disorder function). The propensity traces reveal that the predicted DNA-binding, protein-binding and multifunctional regions are associated with high likelihood values. Altogether, DEPICTER suggests that Si3p has a multifunctional IDR that interacts with DNA and proteins, and the location of this prediction is in agreement with the experimental data.

Summary

The access to the current methods that predict disorder and disorder functions is currently fragmented and requires substantial amount of effort. Users must find and visit multiple websites that require inputs in different formats, collect their results, and reformat and combine these results. The exception are the two comprehensive databases of disorder predictions, D^2P^2 and MobiDB. However, they offer limited scope of the function predictions and are constrained to a subset of proteins that they already include, with no facilities to provide results for the millions of other proteins. The DEPICTER server addresses the need for a centralized resource that offers access to a comprehensive set of disorder and disorder function predictions. It automates the prediction process using a single access point, visualizes the results, and outputs predictions in a consistent and easy-to-parse format. One limitation of our server is that it processes one sequence at the time. We do not allow batch predictions of multiple proteins due to a high computational cost of running the multiple predictors.

DEPICTER integrates predictions of intrinsic disorder generated by three popular tools (SPOT-Disorder-Single, $IUPred2_{short}$, and $IUPred2_{long}$) and prediction of disorder function produced by seven methods ($DisoRDPbind_{RNA}$, ANCHOR2, fMoRFpred, $DisoRDPbind_{DNA}$, $DisoRDPbind_{protein}$, DLFpred, and DMRpred). We also design, implement and test two consensuses to resolve potential conflicts in the results generated by the three predictors of disorder and the three predictors of protein binding that are included in DEPICTER. We test the 12 predictors (including the two

consensuses) on a benchmark dataset that was constructed to share low similarity with the training datasets of the included predictors. This test reveals that the predictions produced by DEPICTER are accurate; AUC_{ROC} values range between 0.71 and 0.87 and MCCs range between 0.3 and 0.52, depending on the target of the prediction. We demonstrate that these results are significantly better than the predictions obtained with a sequence alignment-based solution. Moreover, our tests also suggest that the two new consensuses provide modest and statistically significant improvements in the predictive performance when compared to the corresponding three input disorder predictors and the three protein-binding predictors.

The DEPICTER server is freely available at http://biomine.cs.vcu.edu/servers/DEPICTER/.

Acknowledgments

This research was supported in part by the United States National Science Foundation (grant 1617369) and the Robert J. Mattauch Endowment funds to LK, and by the Australian Research Council (DP180102060) to YZ and KP.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmb.2019.12.030.

Received 7 October 2019;
Received in revised form 7 December 2019;
Accepted 15 December 2019
Available online 21 December 2019

Keywords:
webserver;
prediction center;
intrinsically disordered proteins;
intrinsically disordered region;
protein–nucleic acids interactions

[†]These authors contributed equally.

References

[1] J. Habchi, P. Tompa, S. Longhi, V.N. Uversky, Introducing protein intrinsic disorder, Chem. Rev. 114 (2014) 6561–6588.

[2] P. Lieutaud, F. Ferron, A.V. Uversky, L. Kurgan, V.N. Uversky, S. Longhi, 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 (2016), e1259708.

[3] C.J. Oldfield, V.N. Uversky, A.K. Dunker, L. Kurgan, Chapter 1 - introduction to intrinsically disordered proteins and regions, in: N. Salvi (Ed.), Intrinsically Disordered Proteins, Academic Press, 2019, pp. 1–34.

[4] A.K. Dunker, Z. Obradovic, P. Romero, E.C. Garner, C.J. Brown, Intrinsic protein disorder in complete genomes, Genome Inf Serv Workshop Genome Inf 11 (2000) 161–171.

[5] B. Xue, A.K. Dunker, V.N. Uversky, Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life, J. Biomol. Struct. Dyn. 30 (2012) 137–149.

[6] Z. Peng, J. Yan, X. Fan, M.J. Mizianty, B. Xue, K. Wang, et al., Exceptionally abundant exceptions: comprehensive characterization of intrinsic disorder in all domains of life, Cell. Mol. Life Sci. 72 (2015) 137–151.

[7] H.J. Dyson, P.E. Wright, Intrinsically unstructured proteins and their functions, Nat. Rev. Mol. Cell Biol. 6 (2005) 197–208.

[8] V.N. Uversky, C.J. Oldfield, A.K. Dunker, Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling, J. Mol. Recognit. 18 (2005) 343–384.

[9] Z. Peng, C.J. Oldfield, B. Xue, M.J. Mizianty, A.K. Dunker, L. Kurgan, et al., A creature with a hundred waggly tails: intrinsically disordered proteins in the ribosome, Cell. Mol. Life Sci. 71 (2014) 1477–1504.

[10] Z. Peng, M.J. Mizianty, B. Xue, L. Kurgan, V.N. Uversky, More than just tails: intrinsic disorder in histone proteins, Mol. Biosyst. 8 (2012) 1886–1901.

[11] C. Wang, V.N. Uversky, L. Kurgan, Disordered nucleiome: abundance of intrinsic disorder in the DNA- and RNA-binding proteins in 1121 species from Eukaryota, Bacteria and Archaea, Proteomics 16 (2016) 1486–1498.

[12] J. Liu, N.B. Perumal, C.J. Oldfield, E.W. Su, V.N. Uversky, A.K. Dunker, Intrinsic disorder in transcription factors, Biochemistry 45 (2006) 6873–6888.

[13] I. Na, F. Meng, L. Kurgan, V.N. Uversky, Autophagy-related intrinsically disordered proteins in intra-nuclear compartments, Mol. Biosyst. 12 (2016) 2798–2817.

[14] Z. Peng, B. Xue, L. Kurgan, V.N. Uversky, Resilience of death: intrinsic disorder in proteins involved in the programmed cell death, Cell Death Differ. 20 (2013) 1257–1267.

[15] A.V. Uversky, B. Xue, Z. Peng, L. Kurgan, V.N. Uversky, On the intrinsic disorder status of the major players in programmed cell death pathways, F1000Res 2 (2013) 190.

[16] B. Xue, V.N. Uversky, Intrinsic disorder in proteins involved in the innate antiviral immunity: another flexible side of a molecular arms race, J. Mol. Biol. 426 (2014) 1322–1350.

[17] B. Xue, M.J. Mizianty, L. Kurgan, V.N. Uversky, Protein intrinsic disorder as a flexible armor and a weapon of HIV-1, Cell. Mol. Life Sci. 69 (2012) 1211–1259.

[18] X. Fan, B. Xue, P.T. Dolan, D.J. LaCount, L. Kurgan, V.N. Uversky, The intrinsic disorder status of the human hepatitis C virus proteome, Mol. Biosyst. 10 (2014) 1345–1363.

[19] F. Meng, R.A. Badierah, H.A. Almehdar, E.M. Redwan, L. Kurgan, V.N. Uversky, Unstructural biology of the Dengue virus proteins, FEBS J. 282 (2015) 3368–3394.

[20] G. Hu, K. Wang, J. Song, V.N. Uversky, L. Kurgan, Taxonomic landscape of the dark proteomes: whole-proteome scale interplay between structural darkness,

intrinsic disorder, and crystallization propensity, Proteomics (2018), e1800243.
[21] A. Bhowmick, D.H. Brookes, S.R. Yost, H.J. Dyson, J.D. Forman-Kay, D. Gunter, et al., Finding our way in the dark proteome, J. Am. Chem. Soc. 138 (2016) 9730–9742.
[22] V.N. Uversky, V. Dave, L.M. Iakoucheva, P. Malaney, S.J. Metallo, R.R. Pathak, et al., Pathological unfoldomics of uncontrolled chaos: intrinsically disordered proteins and human diseases, Chem. Rev. 114 (2014) 6844–6879.
[23] V.N. Uversky, C.J. Oldfield, A.K. Dunker, Intrinsically disordered proteins in human diseases: introducing the D2 concept, Annu. Rev. Biophys. 37 (2008) 215–246.
[24] G. Hu, Z. Wu, K. Wang, V.N. Uversky, L. Kurgan, Untapped potential of disordered proteins in current druggable human proteome, Curr. Drug Targets 17 (2016) 1198–1205.
[25] S. Ambadipudi, M. Zweckstetter, Targeting intrinsically disordered proteins in rational drug discovery, Expert Opin. Drug Discov. (2015) 1–13.
[26] D. Piovesan, F. Tabaro, I. Micetic, M. Necci, F. Quaglia, C.J. Oldfield, et al., DisProt 7.0: a major update of the database of disordered proteins, Nucleic Acids Res. D1 (2016) D219–D227.
[27] M. Sickmeier, J.A. Hamilton, T. LeGall, V. Vacic, M.S. Cortese, A. Tantos, et al., DisProt: the database of disordered proteins, Nucleic Acids Res. 35 (2007) D786–D793.
[28] B. He, K. Wang, Y. Liu, B. Xue, V.N. Uversky, A.K. Dunker, Predicting intrinsic disorder in proteins: an overview, Cell Res. 19 (2009) 929–949.
[29] Z. Dosztányi, B. Mészáros, I. Simon, Bioinformatical approaches to characterize intrinsically disordered/unstructured proteins, Briefings Bioinf. 11 (2010) 225–243.
[30] M. Pentony, J. Ward, D. Jones, Computational resources for the prediction and analysis of native disorder in proteins, in: S.J. Hubbard, A.R. Jones (Eds.), Proteome Bioinformatics, Humana Press, 2010, pp. 369–393.
[31] X. Deng, J. Eickholt, J. Cheng, A comprehensive overview of computational protein disorder prediction methods, Mol. Biosyst. 8 (2012) 114–121.
[32] F. Meng, V.N. Uversky, L. Kurgan, Comprehensive review of methods for prediction of intrinsic disorder and its molecular functions, Cell. Mol. Life Sci. 74 (2017) 3069–3090.
[33] F. Meng, V. Uversky, L. Kurgan, Computational prediction of intrinsic disorder in proteins, Curr. Protein Pept. Sci. 88 (2017), 2 16 1-2 4.
[34] M. Necci, D. Piovesan, Z. Dosztanyi, P. Tompa, S.C.E. Tosatto, A comprehensive assessment of long intrinsic protein disorder from the DisProt database, Bioinformatics 34 (3) (2017) 445–452.
[35] I. Walsh, M. Giollo, T. Di Domenico, C. Ferrari, O. Zimmermann, S.C. Tosatto, Comprehensive large-scale assessment of intrinsic protein disorder, Bioinformatics 31 (2015) 201–208.
[36] B. Monastyrskyy, A. Kryshtafovych, J. Moult, A. Tramontano, K. Fidelis, Assessment of protein disorder region predictions in CASP10, Proteins 82 (Suppl 2) (2014) 127–137.
[37] Z.L. Peng, L. Kurgan, Comprehensive comparative assessment of in-silico predictors of disordered regions, Curr. Protein Pept. Sci. 13 (2012) 6–18.
[38] A. Katuwawala, S. Ghadermarzi, L. Kurgan, Chapter Nine - computational prediction of functions of intrinsically disordered regions, in: V.N. Uversky (Ed.), Progress in Molecular Biology and Translational Science, Academic Press, 2019, pp. 341–369.
[39] A. Katuwawala, Z. Peng, J. Yang, L. Kurgan, Computational prediction of MoRFs, short disorder-to-order transitioning protein binding regions, Comput. Struct. Biotechnol. J. 17 (2019) 454–462.
[40] M.E. Oates, P. Romero, T. Ishida, M. Ghalwash, M.J. Mizianty, B. Xue, et al., D(2)P(2): database of disordered protein predictions, Nucleic Acids Res. 41 (2013) D508–D516.
[41] D. Piovesan, F. Tabaro, L. Paladin, M. Necci, I. Micetic, C. Camilloni, et al., MobiDB 3.0: more annotations for intrinsic disorder, conformational diversity and interactions in proteins, Nucleic Acids Res. 46 (2018) D471–D476.
[42] C. The UniProt, UniProt: the universal protein knowledgebase, Nucleic Acids Res. 45 (2017) D158–D169.
[43] Z. Dosztanyi, B. Meszaros, I. Simon, ANCHOR: web server for predicting protein binding regions in disordered proteins, Bioinformatics 25 (2009) 2745–2746.
[44] D.W.A. Buchan, D.T. Jones, The PSIPRED protein analysis workbench: 20 years on, Nucleic Acids Res. 47 (2019) W402–W407.
[45] J. Cheng, A.Z. Randall, M.J. Sweredoski, P. Baldi, SCRATCH: a protein structure and structural feature prediction server, Nucleic Acids Res. 33 (2005) W72–W76.
[46] G. Yachdav, E. Kloppmann, L. Kajan, M. Hecht, T. Goldberg, T. Hamp, et al., PredictProtein—an open resource for online prediction of protein structural and functional features, Nucleic Acids Res. 42 (2014) W337–W343.
[47] J. Cheng, J. Li, Z. Wang, J. Eickholt, X. Deng, The MULTICOM toolbox for protein structure prediction, BMC Bioinf. 13 (2012) 65.
[48] J. Hanson, K.K. Paliwal, Y. Zhou, Accurate single-sequence prediction of protein intrinsic disorder by an ensemble of deep recurrent and convolutional architectures, J. Chem. Inf. Model. 58 (11) (2018) 2369–2376.
[49] B. Meszaros, G. Erdos, Z. Dosztanyi, IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding, Nucleic Acids Res. 46 (2018) W329–W337.
[50] F. Meng, L. Kurgan, DFLpred: high-throughput prediction of disordered flexible linker regions in protein sequences, Bioinformatics 32 (2016) i341–i350.
[51] F. Meng, L. Kurgan, High-throughput prediction of disordered moonlighting regions in protein sequences, Proteins 10 (2018) 1097–1110.
[52] Z. Peng, C. Wang, V.N. Uversky, L. Kurgan, Prediction of disordered RNA, DNA, and protein binding regions using DisoRDPbind, Methods Mol. Biol. 1484 (2017) 187–203.
[53] Z. Peng, L. Kurgan, High-throughput prediction of RNA, DNA and protein binding regions mediated by intrinsic disorder, Nucleic Acids Res. 43 (2015) e121.
[54] J. Yan, A.K. Dunker, V.N. Uversky, L. Kurgan, Molecular recognition features (MoRFs) in three domains of life, Mol. Biosyst. 12 (2016) 697–710.
[55] M. Necci, D. Piovesan, Z. Dosztanyi, S.C.E. Tosatto, MobiDB-lite: fast and highly specific consensus prediction of intrinsic disorder in proteins, Bioinformatics 33 (2017) 1402–1404.
[56] X. Fan, L. Kurgan, Accurate prediction of disorder in protein chains with a comprehensive and empirically designed consensus, J. Biomol. Struct. Dyn. 32 (2014) 448–464.
[57] Z. Peng, L. Kurgan, On the complementarity of the consensus-based disorder prediction, Pac Symp Biocomput (2012) 176–187.

[58] M. Corpas, R. Jimenez, S.J. Carbon, A. Garcia, L. Garcia, T. Goldberg, et al., BioJS: an open source standard for biological visualisation - its status in 2014, F1000Res 3 (2014) 55.
[59] B. Monastyrskyy, K. Fidelis, J. Moult, A. Tramontano, A. Kryshtafovych, Evaluation of disorder predictions in CASP9, Proteins 79 (Suppl 10) (2011) 107–118.
[60] B. Monastyrskyy, A. Kryshtafovych, J. Moult, A. Tramontano, K. Fidelis, Assessment of protein disorder region predictions in CASP10, Proteins 82 (2014) 127–137.
[61] X. Deng, J. Eickholt, J. Cheng, A comprehensive overview of computational protein disorder prediction methods, Mol. Biosyst. 8 (2012) 114–121.
[62] I. Walsh, M. Giollo, T. Di Domenico, C. Ferrari, O. Zimmermann, S.C.E. Tosatto, Comprehensive large-scale assessment of intrinsic protein disorder, Bioinformatics 31 (2015) 201–208.
[63] S.J. McBryant, C. Krause, J.C. Hansen, Domain organization and quaternary structure of the Saccharomyces cerevisiae silent information regulator 3 protein, Sir3p. Biochem 45 (2006) 15941–15948.
[64] P.T. Georgel, M.A. Palacios DeBeer, G. Pietz, C.A. Fox, J.C. Hansen, Sir3-dependent assembly of supramolecular chromatin structures in vitro, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 8584–8589.
[65] S.J. McBryant, V.H. Adams, J.C. Hansen, Chromatin architectural proteins, Chromosome Res. 14 (2006) 39–51.
[66] C. Liu, A.J. Lustig, Genetic analysis of Rap1p/Sir3p interactions in telomeric and HML silencing in Saccharomyces cerevisiae, Genetics 143 (1996) 81–93.
[67] L.M. Johnson, P.S. Kayne, E.S. Kahn, M. Grunstein, Genetic-evidence for an interaction between Sir3 and histone-H4 in the repression of the silent mating loci in saccharomyces-cerevisiae, Proc Natl Acad Sci USA 87 (1990) 6286–6290.
[68] D.W. Paetkau, J.A. Riese, W.S. MacMorran, R.A. Woods, R.D. Gietz, Interaction of the yeast RAD7 and SIR3 proteins: implications for DNA repair and chromatin structure, Genes Dev. 8 (1994) 2035–2045.
[69] J.F. Chang, B.E. Hall, J.C. Tanny, D. Moazed, D. Filman, T. Ellenberger, Structure of the coiled-coil dimerization motif of Sir4 and its interaction with Sir3, Structure 11 (2003) 637–649.