Prediction of intrinsic disorder with quality assessment using QUARTER

Zhonghua Wu[1], Gang Hu[1], Christopher J. Oldfield[2,*] and Lukasz Kurgan[2,*]

[1]School of Mathematical Sciences and LPMC, Nankai University, Tianjin, 300071, PR China
[2]Department of Computer Science, Virginia Commonwealth University, Richmond, VA 23284, USA
*Corresponding authors: lkurgan@vcu.edu and cjoldfield@vcu.edu

Summary
Intrinsically disordered regions (IDRs) are estimated to be highly abundant in nature. While only several thousand proteins are annotated with experimentally derived IDRs, computational methods can be used to predict IDRs for the millions of currently uncharacterized protein chains. Several dozen disorder predictors were developed over the last few decades. While some of these methods provide accurate predictions, unavoidably they also make some mistakes. Consequently, one of the challenges facing users of these methods is how to decide which predictions can be trusted and which are likely incorrect. This practical problem can be solved using quality assessment (QA) scores that predict correctness of the underlying (disorder) predictions at a residue level. We motivate and describe a first-of-its-kind toolbox of QA methods, QUARTER (QUality Assessment for pRotein inTrinsic disordEr pRedictions), which provides the scores for a diverse set of ten disorder predictors. QUARTER is available to the end users as a free and convenient webserver at http://biomine.cs.vcu.edu/servers/QUARTER/. We briefly describe the predictive architecture of QUARTER and provide detailed instructions on how to use the webserver. We also explain how to interpret results produced by QUARTER with the help of a case study.

Keywords
Intrinsic disorder; intrinsically disordered regions; prediction; quality assessment; QUARTER.

1 Introduction
Intrinsically disordered regions (IDRs) in protein sequences lack stable tertiary structure under physiological conditions and instead they form dynamic conformational ensembles [1-4]. Several large-scale computational estimates reveal that proteins with IDRs are highly abundant in nature [5-10] and that these regions are functionally important [11-31]. The disordered nature of these regions is encoded in their sequences; IDRs often feature high net charge and low hydrophobicity, when compared to structured protein regions. IDRs are typically depleted in aromatic residues, large hydrophobic amino acids, and valine [32]. These marked differences between the sequences of IDRs and structured regions have motivated the development of accurate computational tools for the prediction of disorder.

Over 50 disorder predictors have already been developed. A near complete list of these methods can be assembled from several recent surveys and comparative studies [33-36,2,37-39]. Arguably the most popular methods (sorted by the number of Google Scholar citations as of Feb 21, 2019) include DisEMBL [40] (1115 citations), IUPred [41] (1561 citations), DISOPRED [42] (617 citations), VSL2 [43] (588 citations), PrDOS [44] (402 citations), ESpritz [45] (195 citations), MFDp [46] (131 citations), and SPINE-D [47] (117 citations). Several recent comparative analyses show that some disorder predictors offer high predictive quality. For instance, the disorder assessment in the CASP10 experiment (the latest CASP that

has assessed these predictions) reveals that the top three methods achieve areas under the ROC curves (AUCs) equal 0.907 (PrDOS), 0.897 (DISOPRED3), and 0.890 (MFDp) [48].

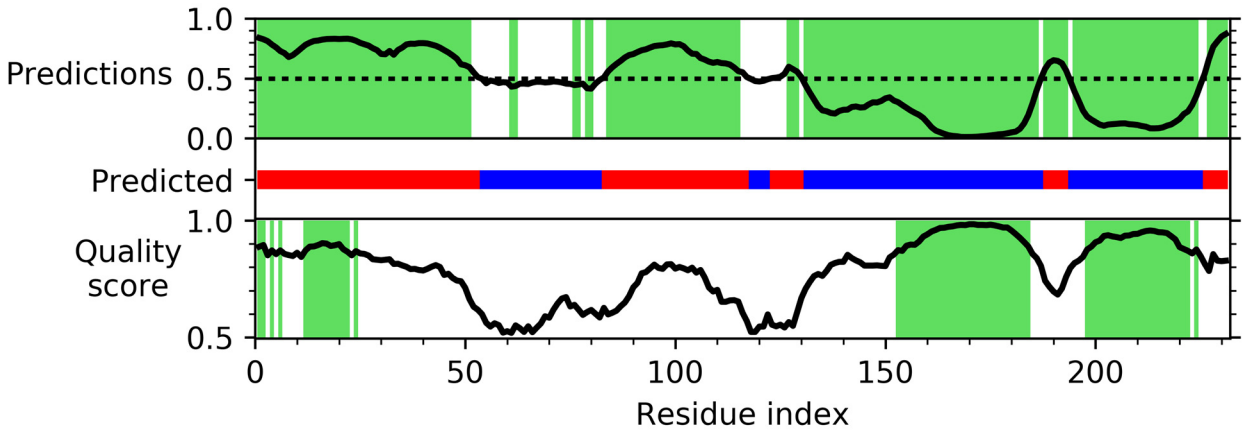

Figure 1. Prediction of intrinsic disorder and the associated quality assessment (QA) scores for the PsaF protein (UniProt ID: P12355). The top panel of the figure shows the putative propensities for disorder (black line) and the binary predictions track, i.e., disordered (in red) vs. ordered (in blue) residues, which were generated by the VSL2B method. The bottom panel of the figure shows the corresponding QA scores (black line) produced with the QUARTER method. Green shading denotes regions where the predictions are assumed to be correct according to the putative propensities for disorder (at the top) and according to the QA scores (at the bottom).

Disorder predictors generate two types of outputs for each amino acid in the input protein sequence: a real-value propensity for disordered conformation and/or a binary category (disordered vs. ordered). High values of propensity suggest that the corresponding residues are likely disordered while low values suggest that they are ordered. The binary category is usually generated using a predictor-specific threshold, where residues with propensities greater than the threshold are predicted as disordered while the remaining residues are predicted as ordered. Figure 1 shows an example prediction from the VSL2B method [43] for the PsaF protein from spinach (UniProt ID: P12355). This relatively obscure example is used to illustrate prediction for a protein with disorder status that is most likely unknown to the reader. The black line in the top panel of Figure 1 represents the real-value propensities, the dotted horizontal line denotes the threshold, and the color-coded horizontal track right below shows binary predictions (red for disordered residues vs. blue for ordered) generated by VSL2B. After computing the predictions, it is left to the user to decide whether and which of these predictions can be trusted. Assuming all predictions are correct would be unreasonable since none of the predictors is or should be expected to be 100% accurate. The VSL2B predictor was previously benchmarked to have 81.6% accuracy [43], which means that on average 18.4% of its lowest quality predictions should be discarded. The residues with presumably the lowest predictive quality are those with propensities closest to the binary prediction threshold, i.e., residues with higher propensities (lower propensities) are more likely to be correctly predicted as disordered (ordered). The green shading of the VSL2B predictions in Figure 1 shows the location of the corresponding set of 81.6% correctly predicted residues. This correct/green part of the prediction suggests that the PsaF protein has IDRs at both termini, a long IDR in the middle of the chain (positions 83 to 117), and a short IDR at positions 187 to 194. The remaining short putative IDR at positions 124 to 130 is likely inaccurately predicted. There are at least two problems with this interpretation of predictions. First, the 81.6% accuracy was measured on a benchmark dataset and it does not imply that this level of accuracy applies to each individual prediction. In fact, results for individual proteins vary widely, from highly accurate predictions to cases where majority of predictions are wrong. Second, the propensity scores are not guaranteed to be accurate indicators of the quality of predictions, as shown in a recent investigation that considered a collection of 10 representative disorder

predictors [49]. This study has revealed that for 9 out of 10 of these methods (the exception being VSL2B) the propensity scores provide useful information to select accurate predictions only for the natively ordered residues, while being virtually unusable for the natively disordered residues. Altogether, we argue that selection of the subset of accurate predictions is a challenging task which is influenced by the choice of a particular predictor and a particular protein sequence.

One solution to this problem is to generate quality assessment (QA) scores together with the disorder predictions. QA scores quantify correctness (confidence) of the disorder predictions at a residue level to reveal which predictions are more likely to be correct. Correctly predicted native disordered and structured residues should have high values of the QA scores, while residues that are incorrectly predicted should have low QA scores. The QA score predictions must be optimized for specific predictors of disorder since these methods use different types of disorder annotations, were designed using different training datasets and use different predictive architectures [35,36,2,37]. While prediction of QA scores has been pursued for over a decade for predictions of protein structure [50-54], so far only one method, called QUARTER (QUality Assessment for pRotein inTrinsic disordEr pRedictions), was developed to generate these scores for predictions of intrinsic disorder [55]. The bottom panel of Figure 1 shows QUARTER's QA scores for VSL2B's predictions. The black line shows the QA scores and the green shading corresponds to the regions where VSL2B's predictions are likely correct. They reveal that the N-terminus is intrinsically disordered while the region between positions 153 and 224 is ordered. The QA scores also suggest that the central part of this protein where the QA scores are low (positions 50 to130) is likely incorrectly predicted.

QUARTER was empirically shown to provide accurate and individually optimized QA scores for ten popular disorder predictors [55]. Empirical tests on a large test dataset show that QUARTER's QA scores are accurate and significantly better than the propensity scores generated by the disorder predictors, particularly for the native disordered residues [55]. Consequently, these QA scores avoid the two pitfalls of disorder propensities: (1) they are tailored to individual proteins and (2) they work equally well for both native disordered and native ordered residues. We note that at the end of this chapter, QA scores and disorder propensities are compared in the context of native annotations of disorder for the PsaF protein from Figure 1. Overall, QUARTER's QA scores provide a useful context for the underlying disorder predictions, guiding the user towards a subset of high quality predictions that are calibrated for specific protein sequences and predictors.

This chapter describes the predictive architecture of the QUARTER tool, explains how to use QUERTER's webserver and clarifies how to interpret results produced by this webserver. It concludes with a case study that focuses on the PsaF protein.

2 Materials

2.1 Disorder Predictors Supported by QUARTER

QUARTER supports ten disorder predictors that include three version of the ESpritz method that were designed to predict disorder annotated using X-ray crystallography ($Espritz_{X\text{-}ray}$), NMR ($Espritz_{NMR}$), and the DisProt database ($Espritz_{DisProt}$) [56]; two versions of IUPred that predict short ($IUPred_{short}$) and long ($IUPred_{long}$) disordered regions [57]; two versions of the DisEMBL method that predict disordered regions defined as hot loops ($DisEMBL_{HotLoops}$) and based on remark 465 from Protein Data Bank ($DisEMBL_{remark465}$) [58], GlobPlot [59], RONN [60], and VSL2B [43]. The selection of the ten predictors was motivated by several factors: 1) they are sufficiently computationally efficient to perform genome-scale predictions, i.e., their runtime is under 1 min for an average size protein sequence; 2) they are incorporated into the two popular databases of predicted disorder: MobiDB [61-63] and D^2P^2 [64]; 3)

they provide relatively accurate disorder predictions [65,66]; and they are freely accessible online, see Table 1.

Table 1. Availability for the ten disorder predictors that are supported by the QUARTER method.

Disorder predictor	URL
$DisEMBL_{HotLoops}$ $DisEMBL_{remark456}$ GlobPlot	http://dis.embl.de/
$Espritz_{DisProt}$ $Espritz_{NMR}$ $Espritz_{X\text{-}ray}$	http://protein.bio.unipd.it/espritz/
$IUPred_{long}$ $IUPred_{short}$	https://iupred2a.elte.hu/
RONN	https://www.strubi.ox.ac.uk/RONN
VSL2B	http://www.dabi.temple.edu/disprot/predictorVSL2.php

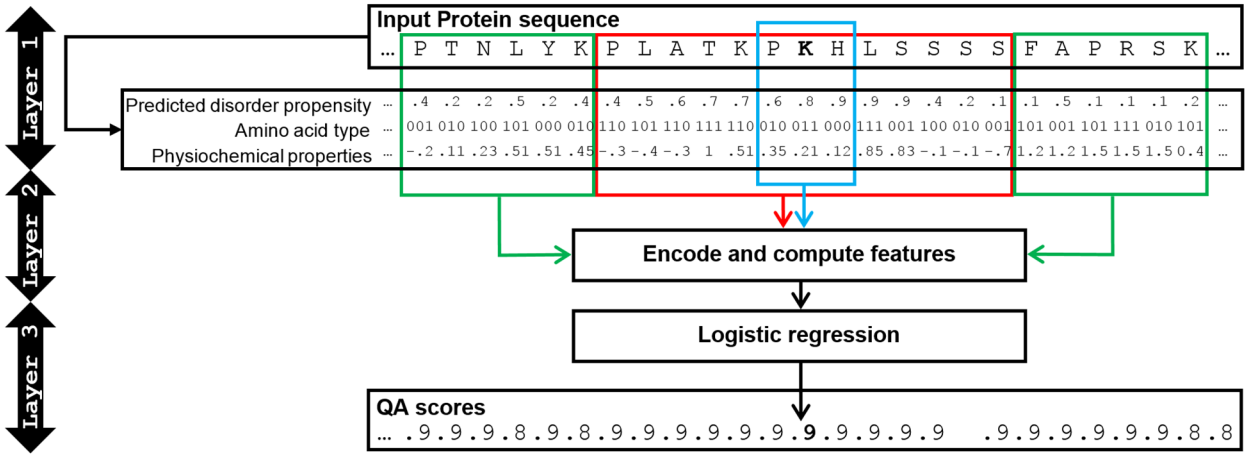

Figure 2. Architecture of the QUARTER method.

2.2 Architecture of the QUARTER Method

QUARTER uses a three-step procedure to generate QA scores for each of the ten considered disorder predictors. The three steps correspond to the following three architectural layers (Figure 2):

1. *Layer 1: Sequence profile*. The first layer uses the input protein sequence to produce sequence profile that includes sequence-derived information useful for the QA prediction. The profile integrates the disorder predictions, amino acid type (AA type) encoded in binary, and several selected physicochemical properties of residues including putative solvent accessibility, hydrophobicity, flexibility, net charge, propensity for disorder and sequence complexity (see **Note 1**).
2. *Layer 2: Feature encoding*. The second layer converts the sequence profile into a fixed number of custom-designed numerical features that combine information across multiple elements of the profile and across multiple residues. Three types of sliding windows are used to encode profile information across residues (Fig. 1): window of 3 adjacent residues (in blue); window of 13 neighboring residues (in red); and window of background residues (in green). QUARTER is optimized for specific disorder predictors by empirically selecting a set of features that maximizes predictive performance for the resulting putative QA scores on a training dataset (see **Note 2**).

3. *Layer 3: Predictive model*. The third layer inputs the disorder predictor-specific features into a logistic regression model that outputs the QA scores. The use of this model type is motivated by several factors. First, logistic regression generates real values in [0, 1] interval that intuitively correspond to the QA scores. Second, this model is computationally efficient which consequently speeds up generation of the QA score. Third, the logistic regression-based models have been used to make numerous related types of predictions including predictions of disorder [10,67], disordered protein and nucleic acids binding [68], disordered linkers [69], protease cleavage sites [70] and phosphorylation sites [71].

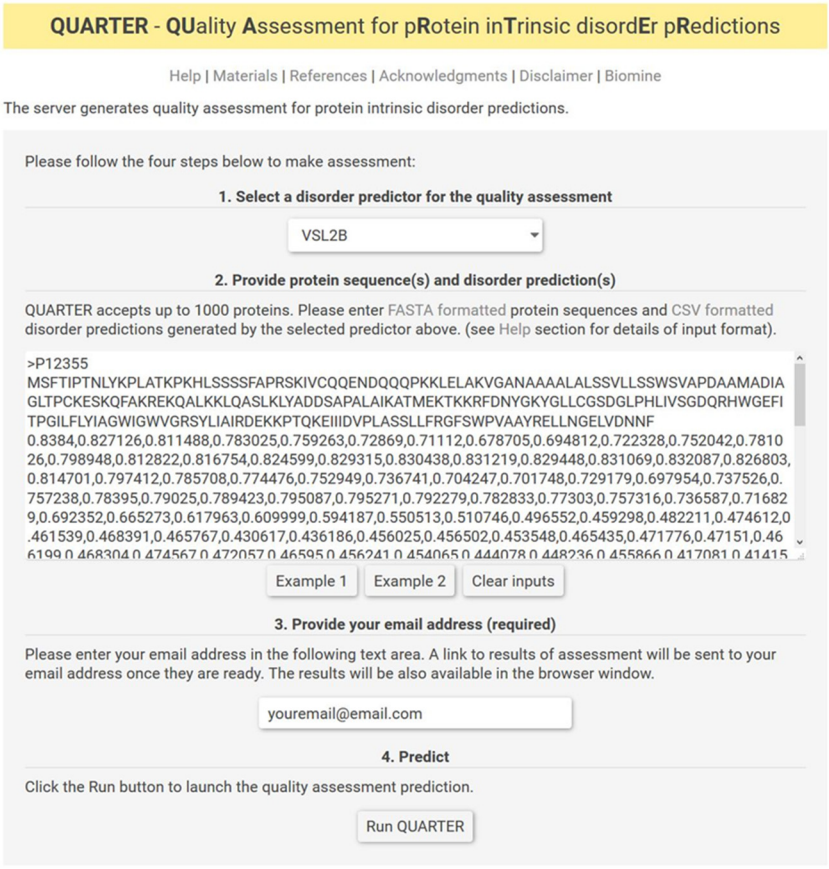

Figure 3. Submission page for the QUARTER webserver. The webpage is setup to make predictions of the QA scores for the VSL2B's disorder predictor and the PsaF protein.

3 Methods

3.1 Running the QUARTER Webserver

The QUARTER predictor is available to the end users as a convenient and free webserver at http://biomine.cs.vcu.edu/servers/QUARTER/. The webserver calculates the QA scores for the ten popular disorder predictors listed in Table 1. The computations are done on the server side and the user only needs a modern web browser (Internet Explorer, Firefox, Opera, or Chrome) and internet connection to make predictions. After arriving at the QUARTER webserver page, four easy steps are required to request the prediction of the disorder QA scores (Figure 3):

1. Select a disorder predictor using a pre-set drop box. The QA scores will be computed for the selected predictor (see **Note 3**).

2. Insert the FASTA-formatted protein sequence(s) and disorder propensities for the selected disorder predictor into the white text box. If your input has more than one protein then these proteins should be placed in consecutive lines (see **Notes 4** and **5**). Figure 2 shows an example input for the PsaF protein (see **Note 6**).
3. Enter your email address. This is the address where a unique web links to the results will be sent (see **Note 7**).
4. Click "Run QUARTER". This submits the sequences to the webserver for the prediction of the QA scores.

Once the job is submitted, the browser is redirected to the QUARTER processing page that provides information about the current position of this submission in the *biomine* server queue (see **Notes 8** and **9**). This page is automatically updated to indicate when the prediction reaches the top of the queue and when it is being processed (see **Note 10**). The webpage automatically redirects to the page with the results when the prediction is completed. This page includes a direct link to the result, which is the same link that is communicated to the user-provided email. The email with the link to the results is sent even in the event when the processing or results page is closed or when the web browser is shut down in the middle of the prediction process.

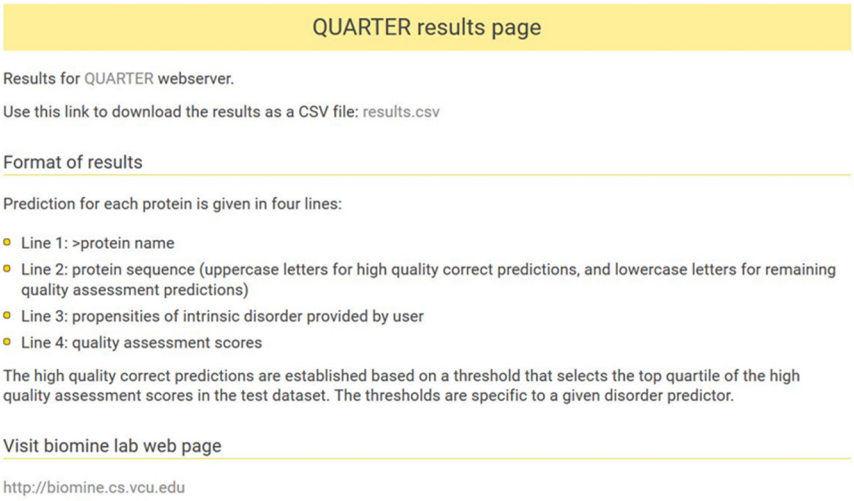

Figure 4. Webpage that summarizes results obtained from the QUARTER webserver.

3.2 Results generated by the QUARTER Webserver

Figure 4 shows the webpage that summarizes the results generated by the QUARTER webserver. The QA score results can be downloaded by clicking "Download CSV file with the results" link. The same link is sent to the user's email address. The link leads to a text file that provides the predicted QA scores for each submitted sequence using four lines (see **Note 11**):

1. Protein name that corresponds to the annotation header from the FASTA-formatted input.
2. The input sequence where the residues predicted as correct predictions are capitalized.
3. Comma delimited disorder prediction scores.
4. Comma delimited predicted disorder QA scores.

The residues identified in the second line of the output text file as correctly predicted are annotated by processing the predicted QA scores using a threshold. The residues with high values of the QA scores that are above the threshold are assumed to be correctly predicted. The value of the threshold is set to

balance the fraction of the residues that are predicted to be correct disorder predictions and the corresponding false positive rate, i.e., fraction of residues that are incorrectly predicted by a given disorder predictor but identified by QUARTER as correct predictions. Figure 5 shows this relation for the 10 predictors that are included in the QUARTER webserver. As expected, the false positive rate increases as the coverage by the correct predictions goes up. The best coverage is secured for the $Espritz_{X\text{-}ray}$ predictor while the worst is for the GlobPlot predictor. The threshold values for the ten disorder predictors are selected to result in a low, 10%, false positive rate. Figure 5 demonstrates that this rate corresponds to the coverage that ranges between 22% (for GlobPlot) and 58% (for $Espritz_{X\text{-}ray}$). Precise, numerical values of thresholds for several selected false positive rates are shown in Table 2. These values are useful for the users who would like to process the predicted QA scores to annotate the correct predictions at different levels of false positive rates and the corresponding coverage values. For instance, a user who would like to annotate correct predictions at the 1% false positive rate for the VSL2B predictor should use threshold = 0.960.

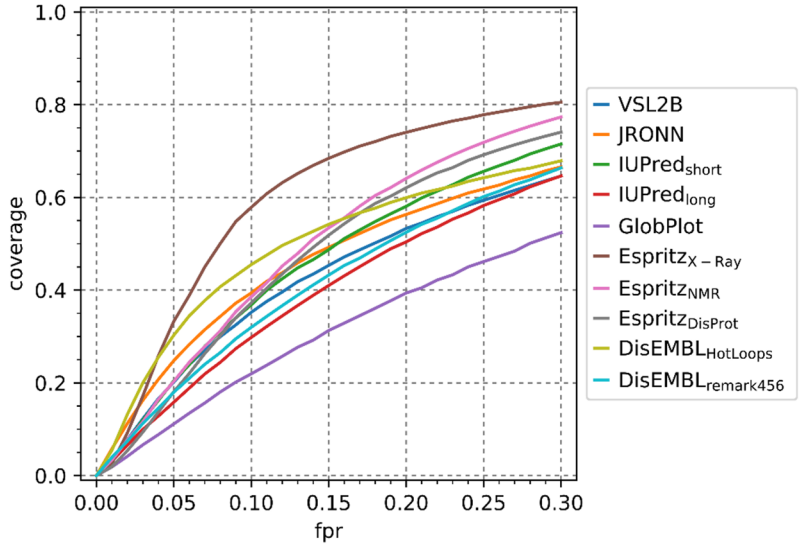

Figure 5. Relation between false positive rate (fpr) and coverage (fraction of residues predicted as correct predictions) for the quality assessment (QA) scores generated with the QUARTER method. These data were computed using the benchmark dataset from [55]. Each color-coded curve corresponds to the QA scores generated for a different disorder predictor.

Table 2. Threshold values that should be used to attain specific false positive rates for the prediction of QA scores for the ten disorder predictors covered by the QUARTER webserver.

	False positive rates									
Predictors	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.1**	**0.15**	**0.2**	**0.25**	**0.3**
VSL2B	0.960	0.946	0.933	0.921	0.911	0.865	0.829	0.797	0.77	0.745
RONN	0.887	0.863	0.845	0.83	0.817	0.771	0.739	0.715	0.696	0.679
$IUPred_{short}$	0.929	0.92	0.913	0.907	0.901	0.875	0.85	0.824	0.797	0.771
$IUPred_{long}$	0.89	0.879	0.871	0.865	0.859	0.831	0.806	0.782	0.759	0.738
GlobPlot	0.785	0.766	0.753	0.744	0.736	0.708	0.689	0.674	0.662	0.651
$Espritz_{X\text{-}ray}$	0.838	0.826	0.819	0.813	0.808	0.783	0.753	0.725	0.701	0.682
$Espritz_{NMR}$	0.872	0.86	0.853	0.846	0.841	0.819	0.797	0.774	0.749	0.724
$Espritz_{DisProt}$	0.691	0.659	0.644	0.633	0.624	0.593	0.568	0.544	0.519	0.494
$DisEMBL_{HotLoops}$	0.868	0.832	0.806	0.786	0.769	0.713	0.679	0.656	0.639	0.625
$DisEMBL_{remark465}$	0.934	0.924	0.915	0.908	0.901	0.873	0.848	0.824	0.8	0.777

The notification that is sent to the end user-provided email address is shown in Figure 6. It provides direct links to the webpage from Figure 4 and to the text file with the results. This email can be used to access results at a later time (see **Note 12**).

Predictions for QUARTER job id: 20190203022026 are ready.

Upon the usage the users are requested to use the following citation(s):

Hu G, Wu Z, Oldfield C, Wang C, Kurgan LA. Quality assessment for the putative intrinsic disorder in proteins. Submitted

You can find the results for this job at: http://biomine.cs.vcu.edu/webresults/QUARTER/20190203022026/results.html

The CSV file can be found here: http://biomine.cs.vcu.edu/webresults/QUARTER/20190203022026/results.csv

The webserver can be found here: http://biomine.cs.vcu.edu/servers/QUARTER/

Thank you for using our webserver,
Biomine group

Figure 6. Notification email generated by the QUARTER webserver.

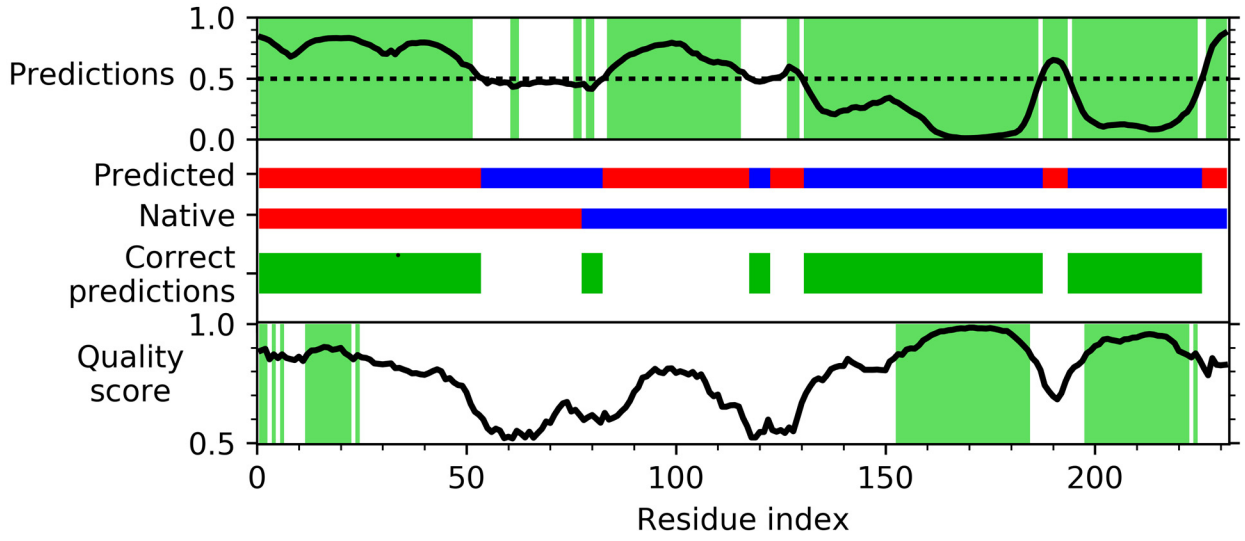

Figure 7. Quality assessment (QA) of the disorder predictions for the PsaF protein (UniProt ID: P12355). The top of the figure shows the putative propensities for disorder (black line) and the binary predictions track, i.e., disordered (in red) vs. ordered (in blue) residues, which were produced by the VSL2B method. The "native" annotation track visualizes the native annotation of intrinsic disorder collected from DisProt (DisProt ID: DP00990), where red and blue denote disordered and structured regions. The green regions in the "correct predictions" line indicate regions of agreement between native and predicted disorder. The bottom of the figure shows the QA scores (black line) computed by the QUARTER method. Green shading denotes regions where the predictions are assumed to be correct according to the putative propensities for disorder (at the top) and according to the QA scores (at the bottom).

4 Case Study

The PsaF protein is a component of the Photosystem I (PSI) complex. It facilitates electron transfer to the PSI from electron donors – plastocyanin and cytochrome c_6 [72] – where absence of PsaF from PSI drastically reduces electron transfer [73]. PsaF's mechanism of action is through direct interaction with electron donors through a N-terminal region, which is helically amphipathic [74]. This interaction region falls within a larger IDR at the N-terminus (Fig. 7, middle panel, "native" annotation track), as determined from the structure of the PSI complex [75]. Molecular recognition is a common function for IDRs [4]. Frequently, as is the case for PsaF, short recognition regions are located within longer IDRs; these short regions have been called molecular recognition features [76-78] and are predicted to be common in nature [79,80]. The N-terminal interaction region of PsaF recruits electron donors and/or activates them for electron transfer.

Figure 7 illustrates the QA predictions from the QUARTER webserver and compares them to the putative disorder propensities for the PsaF protein in the context of the native annotation of the disorder. Intrinsic disorder predictions by VSL2B [43] for PsaF identify several potential IDRs throughout the protein (Fig. 7, middle panel, "predicted" track). These predicted IDRs correspond to prediction scores (Fig. 7, top panel, black line) greater than 0.5, and prediction scores less than 0.5 correspond to the predictions of order. As discussed in the introduction, 81.6% of residues in PsaF are assumed to be predicted correctly by VSL2B, where the most likely correct predictions correspond to the most extreme putative disorder propensities generated by VSL2B. In the case of the 231 residues long PsaF protein, this approach identifies 188 confidently predicted residues (Fig. 7, top panel, green shaded regions). However, comparing predicted ordered and disordered regions with known ordered and disordered regions shows that only 152 residues are predicted correctly (Fig. 7, middle panel, "correct" predictions track). Confident VSL2B predictions overlap with these correct predictions, but incorrectly predicted disordered regions in the middle and C-terminus of PsaF are spuriously indicated to be confident predictions.

QUARTER predicted QA scores for the VSL2B predictions of the PsaF protein (Fig. 7, bottom panel, black line) contrast with putative disorder propensity-based assessment. Adjusted to the false positive rate of 10% at a threshold of a quality score of 0.865 (Table 2), gives 74 residues predicted to be classified correctly by VSL2B for PsaF (Fig. 7, bottom panel, green shaded regions). These residues include the known disordered region at the N-terminus, as well as the C-terminal portion of the ordered region. All of these 74 residues are in fact correctly predicted by VSL2B. Conversely, incorrectly predicted IDRs, including a large IDR in the center of the sequence and two smaller regions at the C-terminus, are not predicted to be correctly identified by VSL2B. Overall, this example demonstrates effectiveness of QUARTER in finding high quality predictions that are calibrated for a specific protein and a specific disorder predictor.

5 Notes

1. The putative solvent accessibility is predicted with the ASAquick method [81]. The sequence complexity is computed using the SEG algorithm [82]. The hydrophobicity is estimated using the Kyte and Doolittle index [83]. The flexibility, which is expressed with the B-factors, is computed by the method described in [84]. Lastly, propensity for intrinsic disorder is estimated using the TOPIDP scale [32].
2. The training and test datasets that were used to optimize and benchmark the QUARTER method, respectively, are available on the "Materials" section of the webpage of the QUARTER webserver at http://biomine.cs.vcu.edu/servers/QUARTER/.
3. Links to the websites of the ten disorder predictors that are covered by QUARTER are given in the "Help" section of the QUARTER webserver page. This is useful to collect their prediction that must be entered in the second step of the prediction process.
4. The webserver accepts up to 1,000 protein sequences for a single run. Information for each input protein must be placed in three consecutive lines: line 1) protein identifier and/or name; line 2) protein sequence using one-letter amino acid encoding; and line 3) comma-separated disorder predictions.
5. The inputs must satisfy several requirements. First, the protein sequence must not contain any spaces or incorrect characters, i.e., only letters that denote amino acids are allowed. Second, the sequences must be longer than 20 residues. This is required by the ASAquick method [81] that is run in the background to collect putative solvent accessibility. Third, the number of putative disorder propensities must be equal to the number of residues in the input protein

chain. The webserver returns a descriptive error message in case if any of these requirements are not met.

6. The three buttons underneath the input box provide two correctly formatted sample inputs, one for the Espritz-DisProt predictor and another for the VSL2B predictor, and ability to clear the input box.
7. The users are required to provide an email address where notification of completed prediction and a private URL to the results are sent. The email is required since this is the most reliable way to inform the user how to locate the results. While the results also appear in the browser window, closing this window or shutting down the web browser would effectively prevent the users from having access to the results.
8. The biomine server services several other predictors including (in alphabetical order) CONNECTOR [85], CRYSTALP2 [86], Cypred [87], DFLpred [69], DisCon [88], disCoP [89], DisoRDPbind [90,68], DMRpred [91], DRNApred [92], fDETECT [93,94], fMoRFpred [80], hybridNAP [95], ILbind [96], MFDp [97], MFDp2 [98,99], MoRFpred [78,100], NsitePred [101], PPCpred [102], RAPID [103], SSCon [104] and SLIDER [10].
9. The biomine webserver utilizes the first-come-first-serve queue. However, the number of concurrent submissions across all predictors listed in note 10 that are coming from the same source is limited to three. Users that submit too many times receive a message that informs them to resubmit after one of their pending submissions is completed. This limit intends to equalize access to the webserver across different users.
10. Prediction of a single protein by the QUARTER webserver takes less than 1 second.
11. A sample text file with the results produced by the QUARTER webserver for a user's query that includes one protein follows:

>P12355
MSfTiPtnlykPLATKPKHLSSsSfaprskivcqqendqqqpkklelakvganaaaalalssvllsswsvapdaamadiagltpckeskqfakrekqalkklqaslklyaddsapalaikat
mektkkrfdnygkygllcgsdglphlivsgDQRHWGEFITPGILFLYIAGWIGWVGRSYLIAirdekkptqkeiiIDVPLASSLLFRGFSWPVAAYRELLnGelvdnnf
0.838,0.827,0.811,0.783,0.759,0.729,0.711,0.679,0.695,0.722,0.752,0.781,0.799,0.813,0.817,0.825,0.829,0.830,0.831,0.829,0.831,0.
832,0.827,0.815,0.797,0.786,0.774,0.753,0.737,0.704,0.702,0.729,0.698,0.738,0.757,0.784,0.790,0.789,0.795,0.795,0.792,0.783,0.7
73,0.757,0.737,0.717,0.692,0.665,0.618,0.610,0.594,0.551,0.511,0.497,0.459,0.482,0.475,0.462,0.468,0.466,0.431,0.436,0.456,0.45
7,0.454,0.465,0.472,0.472,0.466,0.468,0.475,0.472,0.466,0.456,0.454,0.444,0.448,0.456,0.417,0.414,0.456,0.486,0.508,0.558,0.593,
0.621,0.647,0.675,0.695,0.703,0.720,0.735,0.743,0.758,0.771,0.775,0.778,0.784,0.793,0.784,0.786,0.769,0.733,0.707,0.705,0.670,0.
657,0.640,0.630,0.637,0.628,0.625,0.616,0.591,0.561,0.552,0.512,0.495,0.478,0.474,0.480,0.493,0.502,0.506,0.514,0.539,0.598,0.5
77,0.556,0.510,0.442,0.381,0.324,0.268,0.232,0.225,0.210,0.206,0.227,0.237,0.239,0.260,0.265,0.256,0.259,0.282,0.299,0.299,0.31
1,0.333,0.342,0.314,0.301,0.277,0.254,0.231,0.212,0.195,0.169,0.131,0.090,0.063,0.049,0.036,0.024,0.019,0.016,0.010,0.009,0.010,
0.011,0.012,0.016,0.018,0.022,0.027,0.032,0.034,0.041,0.048,0.055,0.077,0.124,0.193,0.291,0.402,0.499,0.581,0.632,0.652,0.646,0.
626,0.566,0.479,0.390,0.312,0.231,0.180,0.155,0.135,0.114,0.104,0.114,0.117,0.120,0.123,0.124,0.117,0.114,0.110,0.102,0.096,0.0
82,0.082,0.083,0.090,0.104,0.115,0.136,0.161,0.202,0.224,0.277,0.354,0.450,0.542,0.664,0.765,0.810,0.851,0.872
0.888,0.895,0.851,0.872,0.856,0.872,0.857,0.852,0.848,0.862,0.844,0.872,0.888,0.889,0.895,0.903,0.901,0.891,0.894,0.899,0.877,0.
866,0.851,0.869,0.859,0.857,0.850,0.834,0.832,0.830,0.832,0.834,0.814,0.815,0.808,0.802,0.792,0.793,0.788,0.786,0.793,0.802,0.8
09,0.801,0.767,0.772,0.754,0.743,0.740,0.712,0.664,0.631,0.618,0.599,0.562,0.546,0.560,0.552,0.520,0.526,0.519,0.551,0.539,0.52
3,0.547,0.522,0.541,0.559,0.590,0.584,0.618,0.645,0.666,0.672,0.632,0.639,0.618,0.596,0.608,0.617,0.601,0.585,0.626,0.598,0.605,
0.614,0.633,0.649,0.676,0.714,0.733,0.774,0.779,0.795,0.811,0.797,0.793,0.812,0.813,0.794,0.801,0.795,0.787,0.798,0.777,0.788,0.
756,0.713,0.696,0.703,0.652,0.652,0.658,0.659,0.653,0.605,0.570,0.523,0.523,0.546,0.546,0.598,0.554,0.547,0.554,0.542,0.565,0.5
49,0.608,0.668,0.708,0.737,0.762,0.771,0.763,0.784,0.810,0.818,0.823,0.823,0.853,0.839,0.829,0.821,0.807,0.807,0.807,0.808,0.80
6,0.805,0.842,0.854,0.873,0.870,0.892,0.897,0.896,0.910,0.931,0.941,0.953,0.960,0.968,0.969,0.973,0.975,0.979,0.981,0.980,0.984,
0.984,0.981,0.981,0.980,0.982,0.979,0.975,0.970,0.970,0.957,0.942,0.926,0.908,0.889,0.857,0.839,0.806,0.742,0.717,0.696,0.683,0.
709,0.759,0.797,0.818,0.828,0.844,0.876,0.888,0.907,0.914,0.934,0.937,0.932,0.929,0.925,0.935,0.936,0.942,0.943,0.950,0.955,0.9
57,0.955,0.950,0.949,0.946,0.933,0.919,0.886,0.878,0.870,0.859,0.876,0.848,0.814,0.784,0.856,0.828,0.826,0.828

12. The predictions are kept on the webserver for the period of at least three months. They can be accessed via the direct link send in the return email.

Acknowledgement

This research was supported in part by the National Science Foundation grant 1617369 and the Robert J. Mattauch Endowment funds to L.K.

References

1. Habchi J, Tompa P, Longhi S, Uversky VN (2014) Introducing protein intrinsic disorder. Chem Rev 114 (13):6561-6588.
2. Lieutaud P, Ferron F, Uversky AV, Kurgan L, Uversky VN, Longhi S (2016) How disordered is my protein and what is its disorder for? A guide through the "dark side" of the protein universe. Intrinsically Disord Proteins 4 (1):e1259708.
3. A. Keith Dunker MMB, Elisar Barbar, Martin Blackledge, Sarah E. Bondos, Zsuzsanna Dosztányi, H. Jane Dyson, Julie Forman-Kay, Monika Fuxreiter, Jörg Gsponer, Kyou-Hoon Han, David T. Jones, Sonia Longhi, Steven J. Metallo, Ken Nishikawa, Ruth Nussinov, Zoran Obradovic, Rohit V. Pappu, Burkhard Rost, Philipp Selenko, Vinod Subramaniam, Joel L. Sussman, Peter Tompa & Vladimir N Uversky (2013) What's in a name? Why these proteins are intrinsically disordered. Intrinsically Disordered Proteins 1 (1):e24157
4. van der Lee R, Buljan M, Lang B, Weatheritt RJ, Daughdrill GW, Dunker AK, Fuxreiter M, Gough J, Gsponer J, Jones DT, Kim PM, Kriwacki RW, Oldfield CJ, Pappu RV, Tompa P, Uversky VN, Wright PE, Babu MM (2014) Classification of Intrinsically Disordered Regions and Proteins. Chemical Reviews 114 (13):6589-6631.
5. Peng Z, Yan J, Fan X, Mizianty MJ, Xue B, Wang K, Hu G, Uversky VN, Kurgan L (2015) Exceptionally abundant exceptions: comprehensive characterization of intrinsic disorder in all domains of life. Cell Mol Life Sci 72 (1):137-151.
6. Xue B, Dunker AK, Uversky VN (2012) Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. J Biomol Struct Dyn 30 (2):137-149.
7. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J Mol Biol 337 (3):635-645.
8. Hu G, Wang K, Song J, Uversky VN, Kurgan L (2018) Taxonomic Landscape of the Dark Proteomes: Whole-Proteome Scale Interplay Between Structural Darkness, Intrinsic Disorder, and Crystallization Propensity. Proteomics:e1800243.
9. Dunker AK, Obradovic Z, Romero P, Garner EC, Brown CJ (2000) Intrinsic protein disorder in complete genomes. Genome Inform Ser Workshop Genome Inform 11:161-171.
10. Peng Z, Mizianty MJ, Kurgan L (2014) Genome-scale prediction of proteins with long intrinsically disordered regions. Proteins 82 (1):145-158.
11. Wang C, Uversky VN, Kurgan L (2016) Disordered nucleiome: Abundance of intrinsic disorder in the DNA- and RNA-binding proteins in 1121 species from Eukaryota, Bacteria and Archaea. Proteomics 16 (10):1486-1498.
12. Romero PR, Zaidi S, Fang YY, Uversky VN, Radivojac P, Oldfield CJ, Cortese MS, Sickmeier M, LeGall T, Obradovic Z, Dunker AK (2006) Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. Proc Natl Acad Sci U S A 103 (22):8390-8395.
13. Hu G, Wu Z, Uversky VN, Kurgan L (2017) Functional Analysis of Human Hub Proteins and Their Interactors Involved in the Intrinsic Disorder-Enriched Interactions. Int J Mol Sci 18 (12).
14. Na I, Meng F, Kurgan L, Uversky VN (2016) Autophagy-related intrinsically disordered proteins in intra-nuclear compartments. Mol Biosyst 12 (9):2798-2817.
15. Meng F, Na I, Kurgan L, Uversky VN (2016) Compartmentalization and Functionality of Nuclear Disorder: Intrinsic Disorder and Protein-Protein Interactions in Intra-Nuclear Compartments. Int J Mol Sci 17 (1).

16. Xue B, Blocquel D, Habchi J, Uversky AV, Kurgan L, Uversky VN, Longhi S (2014) Structural disorder in viral proteins. Chem Rev 114 (13):6880-6911.
17. Peng Z, Oldfield CJ, Xue B, Mizianty MJ, Dunker AK, Kurgan L, Uversky VN (2014) A creature with a hundred waggly tails: intrinsically disordered proteins in the ribosome. Cell Mol Life Sci 71 (8):1477-1504.
18. Fuxreiter M, Toth-Petroczy A, Kraut DA, Matouschek A, Lim RY, Xue B, Kurgan L, Uversky VN (2014) Disordered proteinaceous machines. Chem Rev 114 (13):6806-6843.
19. Fan X, Xue B, Dolan PT, LaCount DJ, Kurgan L, Uversky VN (2014) The intrinsic disorder status of the human hepatitis C virus proteome. Mol Biosyst 10 (6):1345-1363.
20. Peng Z, Xue B, Kurgan L, Uversky VN (2013) Resilience of death: intrinsic disorder in proteins involved in the programmed cell death. Cell Death Differ 20 (9):1257-1267.
21. Xue B, Mizianty MJ, Kurgan L, Uversky VN (2012) Protein intrinsic disorder as a flexible armor and a weapon of HIV-1. Cell Mol Life Sci 69 (8):1211-1259.
22. Peng Z, Mizianty MJ, Xue B, Kurgan L, Uversky VN (2012) More than just tails: intrinsic disorder in histone proteins. Mol Biosyst 8 (7):1886-1901.
23. Buljan M, Chalancon G, Dunker AK, Bateman A, Balaji S, Fuxreiter M, Babu MM (2013) Alternative splicing of intrinsically disordered regions and rewiring of protein interactions. Current opinion in structural biology 23 (3):443-450.
24. Korneta I, Bujnicki JM (2012) Intrinsic disorder in the human spliceosomal proteome. PLoS Comput Biol 8 (8):e1002641.
25. Dyson HJ (2012) Roles of intrinsic disorder in protein-nucleic acid interactions. Mol Biosyst 8 (1):97-104.
26. Dunker AK, Silman I, Uversky VN, Sussman JL (2008) Function and structure of inherently disordered proteins. Curr Opin Struct Biol 18 (6):756-764.
27. Tompa P, Fuxreiter M, Oldfield CJ, Simon I, Dunker AK, Uversky VN (2009) Close encounters of the third kind: disordered domains and the interactions of proteins. Bioessays 31 (3):328-335.
28. Varadi M, Zsolyomi F, Guharoy M, Tompa P (2015) Functional Advantages of Conserved Intrinsic Disorder in RNA-Binding Proteins. PLoS One 10 (10):e0139731.
29. Dosztanyi Z, Chen J, Dunker AK, Simon I, Tompa P (2006) Disorder and sequence repeats in hub proteins and their implications for network evolution. Journal of Proteome Research 5 (11):2985-2995.
30. Pancsa R, Tompa P (2016) Coding Regions of Intrinsic Disorder Accommodate Parallel Functions. Trends Biochem Sci 41 (11):898-906.
31. Tantos A, Kalmar L, Tompa P (2015) The role of structural disorder in cell cycle regulation, related clinical proteomics, disease development and drug targeting. Expert Rev Proteomics 12 (3):221-233.
32. Campen A, Williams RM, Brown CJ, Meng J, Uversky VN, Dunker AK (2008) TOP-IDP-scale: a new amino acid scale measuring propensity for intrinsic disorder. Protein Pept Lett 15 (9):956-963.
33. Peng ZL, Kurgan L (2012) Comprehensive comparative assessment of in-silico predictors of disordered regions. Curr Protein Pept Sci 13 (1):6-18.
34. Walsh I, Giollo M, Di Domenico T, Ferrari C, Zimmermann O, Tosatto SCE (2015) Comprehensive large-scale assessment of intrinsic protein disorder. Bioinformatics 31 (2):201-208.
35. Meng F, Uversky VN, Kurgan L (2017) Comprehensive review of methods for prediction of intrinsic disorder and its molecular functions. Cell Mol Life Sci 74 (17):3069-3090.
36. Meng F, Uversky V, Kurgan L (2017) Computational Prediction of Intrinsic Disorder in Proteins. Curr Protoc Protein Sci 88:2 16 11-12 16 14.
37. He B, Wang K, Liu Y, Xue B, Uversky VN, Dunker AK (2009) Predicting intrinsic disorder in proteins: an overview. Cell Res 19 (8):929-949.

38. Uversky VN, Radivojac P, Iakoucheva LM, Obradovic Z, Dunker AK (2007) Prediction of intrinsic disorder and its use in functional proteomics. Methods Mol Biol 408:69-92.
39. Necci M, Piovesan D, Dosztanyi Z, Tompa P, Tosatto SCE (2017) A comprehensive assessment of long intrinsic protein disorder from the DisProt database. Bioinformatics.
40. Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ, Russell RB (2003) Protein Disorder Prediction: Implications for Structural Proteomics. Structure (London, England : 1993) 11 (11):1453-1459.
41. Dosztányi Z, Csizmok V, Tompa P, Simon I (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics 21 (16):3433-3434.
42. Ward JJ, McGuffin LJ, Bryson K, Buxton BF, Jones DT (2004) The DISOPRED server for the prediction of protein disorder. Bioinformatics 20 (13):2138-2139.
43. Peng K, Radivojac P, Vucetic S, Dunker AK, Obradovic Z (2006) Length-dependent prediction of protein intrinsic disorder. BMC bioinformatics 7 (1):208.
44. Ishida T, Kinoshita K (2007) PrDOS: prediction of disordered protein regions from amino acid sequence. Nucleic acids research 35 (suppl 2):W460-W464.
45. Walsh I, Martin AJM, Di Domenico T, Tosatto SCE (2012) ESpritz: accurate and fast prediction of protein disorder. Bioinformatics 28 (4):503-509.
46. Mizianty MJ, Stach W, Chen K, Kedarisetti KD, Disfani FM, Kurgan L (2010) Improved sequence-based prediction of disordered regions with multilayer fusion of multiple information sources. Bioinformatics 26 (18):i489-i496.
47. Zhang T, Faraggi E, Xue B, Dunker AK, Uversky VN, Zhou Y (2012) SPINE-D: Accurate Prediction of Short and Long Disordered Regions by a Single Neural-Network Based Method. Journal of biomolecular structure & dynamics 29 (4):799-813.
48. Monastyrskyy B, Kryshtafovych A, Moult J, Tramontano A, Fidelis K (2014) Assessment of protein disorder region predictions in CASP10. Proteins 82 Suppl 2:127-137.
49. Wu Z, Hu G, Wang K, Kurgan L (2017) Exploratory Analysis of Quality Assessment of Putative Intrinsic Disorder in Proteins. 6th International Conference on Artificial Intelligence and Soft Computing, vol LNAI 10245. Zakopane, Poland.
50. Kihara D, Chen H, Yang YD (2009) Quality assessment of protein structure models. Curr Protein Pept Sci 10 (3):216-228.
51. Skwark MJ, Elofsson A (2013) PconsD: ultra rapid, accurate model quality assessment for protein structure prediction. Bioinformatics 29 (14):1817-1818.
52. McGuffin LJ, Buenavista MT, Roche DB (2013) The ModFOLD4 server for the quality assessment of 3D protein models. Nucleic Acids Res 41 (Web Server issue):W368-372.
53. Cao R, Bhattacharya D, Adhikari B, Li J, Cheng J (2016) Massive integration of diverse protein quality assessment methods to improve template based modeling in CASP11. Proteins 84 Suppl 1:247-259.
54. Cao R, Adhikari B, Bhattacharya D, Sun M, Hou J, Cheng J (2017) QAcon: single model quality assessment using protein structural and contact information with machine learning techniques. Bioinformatics 33 (4):586-588.
55. Hu G, Wu Z, Oldfield C, Wang C, Kurgan L (2018) Quality Assessment for the Putative Intrinsic Disorder in Proteins. Bioinformatics.
56. Walsh I, Martin AJ, Di Domenico T, Tosatto SC (2012) ESpritz: accurate and fast prediction of protein disorder. Bioinformatics 28 (4):503-509.
57. Dosztanyi Z, Csizmok V, Tompa P, Simon I (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics 21 (16):3433-3434.
58. Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ, Russell RB (2003) Protein disorder prediction: implications for structural proteomics. Structure 11 (11):1453-1459.

59. Linding R, Russell RB, Neduva V, Gibson TJ (2003) GlobPlot: Exploring protein sequences for globularity and disorder. Nucleic Acids Res 31 (13):3701-3708.
60. Yang ZR, Thomson R, McNeil P, Esnouf RM (2005) RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. Bioinformatics 21 (16):3369-3376.
61. Piovesan D, Tabaro F, Paladin L, Necci M, Micetic I, Camilloni C, Davey N, Dosztanyi Z, Meszaros B, Monzon AM, Parisi G, Schad E, Sormanni P, Tompa P, Vendruscolo M, Vranken WF, Tosatto SCE (2018) MobiDB 3.0: more annotations for intrinsic disorder, conformational diversity and interactions in proteins. Nucleic Acids Res 46 (D1):D471-D476.
62. Di Domenico T, Walsh I, Martin AJM, Tosatto SCE (2012) MobiDB: a comprehensive database of intrinsic protein disorder annotations. Bioinformatics 28 (15):2080-2081.
63. Potenza E, Di Domenico T, Walsh I, Tosatto SC (2015) MobiDB 2.0: an improved database of intrinsically disordered and mobile proteins. Nucleic Acids Res 43 (Database issue):D315-320.
64. Oates ME, Romero P, Ishida T, Ghalwash M, Mizianty MJ, Xue B, Dosztanyi Z, Uversky VN, Obradovic Z, Kurgan L, Dunker AK, Gough J (2013) D(2)P(2): database of disordered protein predictions. Nucleic Acids Res 41 (Database issue):D508-516.
65. Walsh I, Giollo M, Di Domenico T, Ferrari C, Zimmermann O, Tosatto SC (2015) Comprehensive large-scale assessment of intrinsic protein disorder. Bioinformatics 31 (2):201-208.
66. Necci M, Piovesan D, Dosztanyi Z, Tosatto SCE (2017) MobiDB-lite: fast and highly specific consensus prediction of intrinsic disorder in proteins. Bioinformatics 33 (9):1402-1404.
67. Obradovic Z, Peng K, Vucetic S, Radivojac P, Dunker AK (2005) Exploiting heterogeneous sequence properties improves prediction of protein disorder. Proteins 61 Suppl 7:176-182.
68. Peng Z, Wang C, Uversky VN, Kurgan L (2017) Prediction of Disordered RNA, DNA, and Protein Binding Regions Using DisoRDPbind. Methods Mol Biol 1484:187-203.
69. Meng F, Kurgan L (2016) DFLpred: High-throughput prediction of disordered flexible linker regions in protein sequences. Bioinformatics 32 (12):i341-i350.
70. Song J, Li F, Leier A, Marquez-Lago TT, Akutsu T, Haffari G, Chou KC, Webb GI, Pike RN, Hancock J (2018) PROSPERous: high-throughput prediction of substrate cleavage sites for 90 proteases with improved accuracy. Bioinformatics 34 (4):684-687.
71. Li F, Li C, Marquez-Lago TT, Leier A, Akutsu T, Purcell AW, Smith AI, Lithgow T, Daly RJ, Song J, Chou KC (2018) Quokka: a comprehensive tool for rapid and accurate prediction of kinase family-specific phosphorylation sites in the human proteome. Bioinformatics.
72. Hippler M, Drepper F, Farah J, Rochaix JD (1997) Fast electron transfer from cytochrome c6 and plastocyanin to photosystem I of Chlamydomonas reinhardtii requires PsaF. Biochemistry 36 (21):6343-6349.
73. Farah J, Rappaport F, Choquet Y, Joliot P, Rochaix JD (1995) Isolation of a psaF-deficient mutant of Chlamydomonas reinhardtii: efficient interaction of plastocyanin with the photosystem I reaction center is mediated by the PsaF subunit. The EMBO journal 14 (20):4976-4984.
74. Hippler M, Drepper F, Haehnel W, Rochaix JD (1998) The N-terminal domain of PsaF: precise recognition site for binding and fast electron transfer from cytochrome c6 and plastocyanin to photosystem I of Chlamydomonas reinhardtii. Proceedings of the National Academy of Sciences of the United States of America 95 (13):7339-7344.
75. Amunts A, Toporik H, Borovikova A, Nelson N (2010) Structure determination and improved model of plant photosystem I. The Journal of biological chemistry 285 (5):3478-3486.
76. Oldfield CJ, Cheng Y, Cortese MS, Romero P, Uversky VN, Dunker AK (2005) Coupled folding and binding with alpha-helix-forming molecular recognition elements. Biochemistry 44 (37):12454-12470.

77. Disfani FM, Hsu W-L, Mizianty MJ, Oldfield CJ, Xue B, Dunker AK, Uversky VN, Kurgan L (2012) MoRFpred, a computational tool for sequence-based prediction and characterization of short disorder-to-order transitioning binding regions in proteins. Bioinformatics 28 (12):i75-i83.
78. Oldfield CJ, Uversky VN, Kurgan L (2018) Predicting Functions of Disordered Proteins with MoRFpred. Methods Mol Biol 1851.
79. Mohan A, Oldfield CJ, Radivojac P, Vacic V, Cortese MS, Dunker AK, Uversky VN (2006) Analysis of molecular recognition features (MoRFs). J Mol Biol 362 (5):1043-1059.
80. Yan J, Dunker AK, Uversky VN, Kurgan L (2016) Molecular recognition features (MoRFs) in three domains of life. Mol Biosyst 12 (3):697-710.
81. Faraggi E, Zhou YQ, Kloczkowski A (2014) Accurate single-sequence prediction of solvent accessible surface area using local and global features. Proteins 82 (11):3170-3176.
82. Wootton JC, Federhen S (1993) Statistics of Local Complexity in Amino-Acid-Sequences and Sequence Databases. Comput Chem 17 (2):149-163.
83. Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157 (1):105-132.
84. Vihinen M, Torkkila E, Riikonen P (1994) Accuracy of protein flexibility predictions. Proteins 19 (2):141-149.
85. Wang C, Kurgan L (2018) Review and comparative assessment of similarity-based methods for prediction of drug-protein interactions in the druggable human proteome. Brief Bioinform.
86. Kurgan L, Razib AA, Aghakhani S, Dick S, Mizianty M, Jahandideh S (2009) CRYSTALP2: sequence-based protein crystallization propensity prediction. Bmc Structural Biology 9.
87. Kedarisetti P, Mizianty MJ, Kaas Q, Craik DJ, Kurgan L (2014) Prediction and characterization of cyclic proteins from sequences in three domains of life. Biochim Biophys Acta 1844 (1 Pt B):181-190.
88. Mizianty MJ, Zhang T, Xue B, Zhou Y, Dunker AK, Uversky VN, Kurgan L (2011) In-silico prediction of disorder content using hybrid sequence representation. BMC Bioinformatics 12:245.
89. Fan X, Kurgan L (2014) Accurate prediction of disorder in protein chains with a comprehensive and empirically designed consensus. J Biomol Struct Dyn 32 (3):448-464.
90. Peng Z, Kurgan L (2015) High-throughput prediction of RNA, DNA and protein binding regions mediated by intrinsic disorder. Nucleic Acids Res 43 (18):e121.
91. Meng F, Kurgan L (2018) High-throughput prediction of disordered moonlighting regions in protein sequences. Proteins.
92. Yan J, Kurgan L (2017) DRNApred, fast sequence-based method that accurately predicts and discriminates DNA- and RNA-binding residues. Nucleic Acids Res 45 (10):e84.
93. Meng F, Wang C, Kurgan L (2018) fDETECT webserver: fast predictor of propensity for protein production, purification, and crystallization. BMC Bioinformatics 18 (1):580.
94. Mizianty MJ, Fan X, Yan J, Chalmers E, Woloschuk C, Joachimiak A, Kurgan L (2014) Covering complete proteomes with X-ray structures: a current snapshot. Acta Crystallogr D Biol Crystallogr 70 (Pt 11):2781-2793.
95. Zhang J, Ma Z, Kurgan L (2017) Comprehensive review and empirical analysis of hallmarks of DNA-, RNA- and protein-binding residues in protein chains. Brief Bioinform.
96. Hu G, Gao J, Wang K, Mizianty MJ, Ruan J, Kurgan L (2012) Finding protein targets for small biologically relevant ligands across fold space using inverse ligand binding predictions. Structure 20 (11):1815-1822.
97. Mizianty MJ, Stach W, Chen K, Kedarisetti KD, Disfani FM, Kurgan L (2010) Improved sequence-based prediction of disordered regions with multilayer fusion of multiple information sources. Bioinformatics 26 (18):i489-496.
98. Mizianty MJ, Uversky V, Kurgan L (2014) Prediction of intrinsic disorder in proteins using MFDp2. Methods Mol Biol 1137:147-162.

99. Mizianty MJ, Peng ZL, Kurgan L (2013) MFDp2: Accurate predictor of disorder in proteins by fusion of disorder probabilities, content and profiles. Intrinsically Disordered Proteins 1 (1):e24428.
100. Disfani FM, Hsu WL, Mizianty MJ, Oldfield CJ, Xue B, Dunker AK, Uversky VN, Kurgan L (2012) MoRFpred, a computational tool for sequence-based prediction and characterization of short disorder-to-order transitioning binding regions in proteins. Bioinformatics 28 (12):i75-83.
101. Chen K, Mizianty MJ, Kurgan L (2012) Prediction and analysis of nucleotide-binding residues using sequence and sequence-derived structural descriptors. Bioinformatics 28 (3):331-341.
102. Mizianty MJ, Kurgan L (2011) Sequence-based prediction of protein crystallization, purification and production propensity. Bioinformatics 27 (13):i24-33.
103. Yan J, Mizianty MJ, Filipow PL, Uversky VN, Kurgan L (2013) RAPID: fast and accurate sequence-based prediction of intrinsic disorder content on proteomic scale. Biochim Biophys Acta 1834 (8):1671-1680.
104. Yan J, Marcus M, Kurgan L (2014) Comprehensively designed consensus of standalone secondary structure predictors improves Q3 by over 3%. J Biomol Struct Dyn 32 (1):36-51.