Prediction of disordered RNA, DNA, and protein binding regions using DisoRDPbind

Zhenling Peng[1], Chen Wang[2], Vladimir N. Uversky[3,4,5,6] and Lukasz Kurgan[7]*

[1]Center for Applied Mathematics, Tianjin University, Tianjin, 300072, China

[2]Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Alberta, T6G 2V4, Canada

[3]Department of Molecular Medicine, Morsani College of Medicine, University of South Florida, Tampa, FL, 33612, USA

[4]USF Health Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, Tampa, FL, 33612, USA

[5]Institute for Biological Instrumentation, Russian Academy of Sciences, Pushchino, Moscow Region, 142292, Russian Federation

[6]Department of Biology, Faculty of Science, King Abdulaziz University, P.O. Box 80203, Jeddah, 21589, Kingdom of Saudi Arabia

[7]Department of Computer Science, Virginia Commonwealth University, Richmond, VA 23284, USA

*Corresponding author: lkurgan@vcu.edu.

Summary
Intrinsically disordered proteins and regions (IDPs and IDRs) are involved in a wide range of cellular functions and they often facilitate interactions with RNAs, DNAs, and proteins. Although many computational methods can predict IDPs and IDRs in protein sequences, only a few methods predict their functions and these functions primarily concern protein-binding. We describe how to use the first computational method DisoRDPbind for high-throughput prediction of multiple functions of disordered regions. Our method predicts the RNA-, DNA-, and protein-binding residues located in IDRs in the input protein sequences. DisoRDPbind provides accurate predictions and is sufficiently fast to make predictions for full genomes. Our method is implemented as a user-friendly webserver that is freely available at http://biomine.ece.ualberta.ca/DisoRDPbind/. We overview our predictor, discuss how to run the webserver, and show how to interpret the corresponding results. We also demonstrate the utility of our method based on two case studies, human BRCA1 protein that binds various proteins and DNA, and yeast 60S ribosomal protein L4 that interacts with proteins and RNA.

Key words
Intrinsic disorder; prediction; protein-protein interactions; protein-DNA interactions; protein-RNA interactions; DisoRDPbind.

1 Introduction
Intrinsically disordered proteins and regions (IDPs and IDRs) lack a stable 3-dimensional structure under physiological conditions in-vitro and form an ensemble of structural conformations [1-3]. They participate in a wide range of cellular functions and are common in nature, particularly in eukaryotic species [3-6]. Many computational methods are available for the prediction of intrinsic disorder from protein sequences [7-13]. These predictors were used to estimate the amount of disorder in various species and domains of life and to characterize cellular functions of disorder [5,14-20]. IDPs and IDRs were shown to be significantly involved in the protein-protein, protein-DNA and protein-RNA interactions [5,21-25,18,20,26,27]; for convenience, here we utilize the terms disordered RNA-, DNA-, and protein-binding to denote the RNA-, DNA-, and protein-binding located in IDRs. Prediction of residues that bind proteins, RNAs and DNAs has attracted strong research interest in the last decade [28-33]. However, these predictions address interactions annotated from crystal structures, which means that they are primarily focused on the structured (ordered) regions.

A number of studies that predict functions of IDPs and IDRs were also recently discussed [34] . A prediction of over one hundred Gene Ontology (GO) annotations associated with IDPs and IDRs was carried out by Khan *et al*. [35]. Moreover, several methods were developed for the prediction of disordered protein binding regions including alpha-MoRF-Pred [36], ANCHOR [37], MoRFpred [38], PepBindPred [39], MFSPSSMpred [40], DISOPRED3 [41], MoRFCHiBi [42], and fMoRFpred [43]. This implies that functions of IDRs and IDPs are predictable from protein sequences. Availability of hundreds of regions annotated as disordered RNA-, DNA-, and protein-binding in the DisProt database [44] and the lack of methods that predict disordered RNA- and DNA-binding motivated the development of a new predictor DisoRDPbind [45]. This is the first method that predicts multiple functions mediated by IDPs and IDRs. DisoRDPbind obtains favourable predictive performance for these three types of disordered binding regions. It is also very fast and can be applied to predict full genomes in a matter of hours using its convenient webserver (the largest human genome can be predicted in about 2 days) [45]. The DisoRDPbind's webserver outputs three propensity scores for each input residue that quantify the likelihood for this residue to be involved in the disordered RNA-, DNA-, and protein-binding. We overview architecture of our method and provide details on how to use the webserver and how to

interpret the results. Finally, we use two case studies that involve analysis of RNA-, DNA- and protein-binding proteins to illustrate how our method can be used to suggest localization of disordered RNA-, DNA- and protein-binding regions in protein sequences.

2 Materials and Method

2.1 Datasets

We extracted 315, 114 and 36 proteins from the DisProt database [44] to develop three datasets: TRAINING, TEST114 and TEST36, respectively. Each dataset includes disordered regions that were annotated to bind RNAs, DNAs and proteins. The TRAINING dataset was used for empirical design of DisoRDPbind while the other two datasets were used to assess its predictive performance and compare it against other methods. Proteins in TEST114 were collected to share < 30% sequence similarity with proteins in TRAINING to allow for assessment of predictive performance on dissimilar proteins. The second test dataset, TEST36, includes new depositions to DisProt as compared with the proteins from TRAINING. These three datasets are available at http://biomine.ece.ualberta.ca/DisoRDPbind/. Ref. [45] provides further details.

2.2 Architecture

Recently, we developed the DisoRDPbind method to predict the disordered RNA-, DNA-, and protein-binding residues in the input protein sequences. Our method is based on a runtime efficient four-layer design; see Figure 1. First, we represent the input protein using several structural and functional properties. Second, these properties are used to represent each residue in the input protein chain using a vector of numeric descriptors/features. Third, these features are inputted into a predictive model. Fourth, the outputs of the predictive model are merged with an alignment-based prediction to derive the final result. Following we provide a more detailed explanation.

In layer 1, we represent each input protein sequence based on its amino acid (AA) composition, its sequence complexity generated by the SEG algorithm [46], intrinsic disorder predicted by IUPred (including IUPred L and IUPred S) [47], secondary structure predicted by PSIPRED [48], and 17 physiochemical properties of amino acids (AAs) including hydrophobicity, net charge, and free energy. In layer 2, we use this information to compute a vector of features for each residue in the input protein chain and each predicted function. We utilized sliding window with different window size (WS) to obtain the numerical features for different binding events where WS = 55, 21, and 33 for disordered RNA-, DNA-, and protein-binding, respectively. We quantified the abovementioned putative sequence structural and functional characteristics, such as disorder, secondary structure, hydrophobicity, etc. in a window centered over the predicted residue by computing their averages and content values. These values represent local (in the sequence) bias that contributes towards the prediction of the residue in the middle of the window. Since the many features considered in this layer are redundant and/or irrelevant to the predicted functions, we performed an empirical feature selection for each function using the TRAINING dataset. Consequently, we selected a small sets of 11, 7 and 7 features for the prediction of disordered RNA-, DNA-, and protein-binding, respectively. In layer 3, for each residue in the input protein sequence we pass a given selected set of features into a logistic regression model for the corresponding binding event. This means that three regression-based models are used to find the putative disordered RNA-, DNA-, and protein-binding residues. We picked this type of model based on its popularity, short runtime and the ability to output a real-valued propensity. In the last layer 4, we merge the regression-based predictions with functional annotations found through sequence similarity to generate the final predictions. We utilize BLAST [49] to align the input sequence against a database of functionally annotated proteins (TRAINING dataset) and then we transfer the functional annotations for each input residue that was aligned to a functional residue in a sufficiently similar annotated protein.

The final predictions include three propensity scores for each residue in the input sequence that quantify its likelihood to be disordered RNA-, DNA-, and protein-binding residues, respectively; higher values of propensity correspond to a higher likelihood of binding. DisoRDPbind also provides a binary prediction for each function by using a threshold on a given putative propensity score; residues with propensities higher than the threshold are predicted as binding and the other residues are predicted as non-binding (see section 2.6).

2.3 Predictive quality and runtime

The predictive quality of our method was assessed in the original manuscript [45]. DisoRDPbind was shown to secure the area under the ROC curve (AUC) values between 0.62 and 0.72, depending on the benchmark dataset used (TEST114 and TEST36) and the disorder function that was assessed. The TP-rate (fraction of correctly predicted binding residues) of DisoRDPbind computed at the FP-rate (fraction of incorrectly predicted non-binding residues) of 0.1 is 0.27, 0.25, and 0.24 for the prediction of the disordered DNA-, protein-, and RNA-binding residues, respectively. These are reasonable levels of TP-rate and AUC values, which were shown to be higher than the corresponding values of the closest alternatives (predictors of disordered protein-binding residues and predictors of ordered DNA- and RNA-binding residues) [45]. Interestingly, predictions from DisoRDPbind complement predictions from the predictors of structured DNA- and RNA-binding residues (they are characterized by low correlation < 0.3), while as expected they are similar to the outputs of methods that predict disordered protein-binding residues (correlation > 0.5 with ANCHOR) [45]. Overall, these observations demonstrate that our method is relatively accurate and complements the other available methods.

Using a modern desktop (Intel i7-950 CPU at 3.06GHz with 24GB or RAM), the runtime of DisoRDPbind for a single protein is between 0.3 seconds and 1 minute, depending on the chain length, and is characterized by a quadratic increase with the chain size [45]. An average size protein with about 200 residues is predicted in 1 second (see **Note 1**). This includes the combined runtime of the prediction of the three binding events. To put that into perspective, the runtime to predict the entire human proteome is just over 40 hours on the abovementioned desktop computer, which means that DisoRDPbind can be used to predict full genomes.

2.4 Webserver

The webserver of DisoRDPbind was designed to be user-friendly and is freely available at http://biomine.ece.ualberta.ca/DisoRDPbind/. The end user only needs a modern web browser (Firefox, IE, and Chrome were tested) and internet connection to use the webserver.

The main (start) page of the webserver is for the submission of the user's query. It includes a text field where up to 5000 input protein sequences in FASTA format can be pasted and another text field for the e-mail of the user. For convenience, the server also provides an option to submit the input proteins in a FASTA-formatted file. The e-mail is required and is used to send notification when the predictions are completed. The notification provides a link to a summary page that explains the format of the outputs and the formatted text file with the predictions.

The DisoRDPbind method uses other programs to generate its inputs. Specifically, our method generates the disorder profiles utilizing IUPred [47], predicts the secondary structure with the fast version of PSIPRED (without using PSI-BLAST) [48], obtains the information about low complexity regions with the SEG algorithm [46], and transfers the functional annotations based on the alignment generated by BLAST [49]. These methods are used in a fully automated manner by the scripts that implement the webserver. Once the user provides the sequences and the e-mail and hits the "Run DisoRDPbind!" button, the results are generated without further interaction with the webserver.

2.5 Running DisoRDPbind

Three easy steps should be followed to use the DisoRDPbind webserver (the step numbers are highlighted in red in Figure 2):

1. Copy and paste protein sequences formatted in the FASTA format into text field or upload FASTA-formatted file (an "Example" button may be used to see properly formatted example inputs) (see **Notes 2** and **3**).
2. Provide e-mail address (required, see **Note 4**). The notification e-mail, including the hyperlinks to the results page and the downloadable outputs, will send to the user once the predictions are completed.
3. Click "Run DisoRDPbind!" button to start the predictions.

Note that the webserver generates predictions of RNA-, DNA-and protein-binding at the same time for each input protein sequence. Once the “Run DisoRDPbind!” button is clicked, the user’s web browser is redirected to another page that shows the current status of the prediction. The user’s query is added to a queue of predictions on the biomine server (this server also implements a few other methods) and the position in the queue is shown and updated. The prediction is executed when the query reaches the first position in the queue. After the prediction is completed the user’s web browser is automatically redirected to the page with the results and the notification e-mail with a link to this page is sent (see **Notes 4** and **5**). The prediction is completed and e-mail is sent even in the case when the user closes the web browser before the completion of the prediction.

2.6 Results generated by DisoRDPbind

This webpage with the results includes a hyperlink to the downloadable text file (red number 1 in Figure 3) and the description of the format of this file (red number 2 in Figure 3). The text file, named DisoRDPbind.pred, is provided for download to the end user. This file includes the prediction of disordered RNA-, DNA-, and protein-binding residues for all submitted protein sequences. For each of the three types of binding we provide a binary prediction (1 for putative binding residues and 0 for putative non-binding residues) and a real-valued propensity (higher values indicates higher likelihood for binding) for each input residue. The results are organized in eight lines per protein where six lines provide prediction for the entire input sequence and two lines lists the residues from the input sequence and its name:

- The first line lists the protein name (as provided in the user’s input)
- The second line is the protein sequence where each letter identifies a residue and where a lower (upper) case indicates the residue was predicted to interact (not to interact) with RNA, DNA, or protein. This is based on the binary prediction across the three types of binding.
- The third/fifth/seventh line provides the putative binary prediction of the RNA-binding/DNA-binding/protein-binding residues (see **Note 6**)
- The fourth/sixth/eighth line provides the putative propensity for the RNA-binding/DNA-binding/protein-binding for each input residue. The values of the propensities are separated by commas, they range between 0 (lowest propensity) and 1 (highest propensity), and they are provided with the precision of 3 digits after the decimal point.

The notification e-mail includes the hyperlinks to the page with the results (red number 1 in Figure 4) and to the downloadable outputs (red number 2 in Figure 4). The first hyperlink leads the user directly to the "DisoRDPbind Results Page" (Figure 3). We also provide a unique job ID at the top of the e-mail. This ID can be used to trace a given prediction query. In case if the user encounters problems then (s)he should simply reply to the e-mail with a description of what is wrong making sure that the job ID is included.

3 Case studies

3.1 Case 1. BRCA1

BRCA1 is the breast cancer type 1 susceptibility protein which is known to play a number of important roles in controlling the development of breast cancer. The *BRCA1* gene expression is dependent on the cell cycle, and the G1–S and the G2–M transition checkpoints are controlled by the BRCA1 protein [50]. However, major functions of BRCA1 are related to the repair of chromosomal damage and to the error-free repair of DNA double-strand breaks [51]. In the norm, BRCA1 is involved in repair of the damaged DNA, or, if the DNA damage cannot be repaired, it initiates the cell destruction. The mutation-induced decrease or loss of the BRCA1 functions results in the accumulation of the damaged DNA, increasing the probability of the development of breast cancer [51]. Of the 1863 amino acids of BRCA1, only ~20% terminally located residues are involved in the formation of structured domains (residues 1-169 and 1646-1863 are folded into the RNG and tandem BRCT domains, respectively), whereas a long central region (residues 170-1645) is mostly disordered, acting as a scaffold that determines the exceptional binding promiscuity of BRCA1 [52]. Among known interacting partners of the central region of BRCA1 are several proteins involved in regulation of various biological processes. They include c-Myc, which is a proto-oncogene that is implicated in tumorigenesis, embryonic development and apoptosis, which binds to BRCA1 at residues 173–303 and 433–511 [53]; retinoblastoma protein (pRB) that is a tumor suppressor protein dysfunctional in several tumors interacts with residues 304–394 of BRCA1 [54]; p53, which is known to acts as the guardian of the genome and a tumor suppressor [55] and binds BRCA1 at residues 224–500 [56]; Rad50, which forms a complex with Mre11 and p95/nibrin that acts in meiotic recombination, homologous recombination, non-homologous end joining, the DNA damage response, and telomere maintenance [57], and that binds BRCA1 at residues 341–748 [58]; Rad51, which is a member of a protein family that mediates DNA strand–exchange functions related to normal recombination [59], and which interacts with the residues 758–1064 of BRCA1 [60]; FANCA, a member of the proteins related to Fanconi anemia that form a nuclear complex [61], which binds BRCA1 at residues 740–1083 [62]; whereas JunB, a transcription factor involved in regulation of the gene activity following the primary growth factor response interacts with BRCA1 at residues 1343–1440 [63]. Finally, residues 452–1079 of human BRCA1 are known to interact with DNA [64].

Results of the DisoRDPbind analysis of human BRCA1 (UniProt ID: P38398) are shown in Figure 5A. We observe that the putative propensities clearly illustrate that this protein has a number of identifiable disordered protein- and DNA-binding sites that are located in the intrinsically disordered region of this protein. The entire long central region is predicted as protein binding, which is in agreement with the annotated native binding sites that are discussed in the previous paragraph. The known DNA-binding region, which is shown as a blue horizontal line at the bottom of Figure 5A also lines up with the higher values of the predicted propensities for the DNA binding; we note that the DisoRDPbind webserver predicts residues with the propensities for DNA binding ≥ 0.245 as DNA binding (see **Note 6**), and such residues are fairly abundant in the native DNA-binding region.

3.2 Case 2. Yeast 60S ribosomal protein L4

Every living cell contains ribosomes, which are ancient ribonucleoprotein complexes serving as molecular machines for protein biosynthesis. Ribosomes are large (with the molecular mass of at least 2.5 MDa) macromolecular complexes composed of one or more ribosomal RNA molecules and a variety of proteins. Being the major force in the cellular protein production, these highly specialized machines have two major components known as the small and the large ribosomal subunits. These components have different roles in protein biosynthesis, with the small ribosomal subunit being responsible for "reading" the mRNA and with the large ribosomal subunit catalyzing the peptide bond formation.

Although overall function and organization of ribosomes is similar between different organisms, prokaryotic and eukaryotic ribosomes have significant differences. For example, in prokaryotes, ribosomes are composed of ~65% of rRNA and 35% of ribosomal proteins, whereas in eukaryotic ribosomes, the rRNA : protein ratio is close to 1. Furthermore, in prokaryotic ribosomes, small (30S) subunit includes 16S rRNA and 21 ribosomal proteins, whereas large (50S) subunit contains 5S and 23S rRNA molecules and 31 proteins [65]. In the 80S eukaryotic ribosome, the small 40S subunit contains 18S rRNA and 33 proteins, and the large 60S subunit is composed of 3 rRNA molecules (5S, 28S, and 5.8S) and 46 proteins [66]. Proteins derived from the small and large ribosomal subunits are named S1, S2, S3... and L1, L2, L3..., respectively. Their high conservation during evolution suggests that they have critical roles in ribosome biogenesis or functions of the mature ribosome. The ribosomal proteins are known to be enriched in intrinsic disorder [18] which is why they are relevant for our case study.

Since ribosomal proteins are abundant in every cell, and since they can interact with nucleic acids and other proteins, these RNA-binding proteins are known to be recruited to carry out many extra-ribosomal or auxiliary functions; i.e., they serve as moonlighting proteins [67-70]. It has been pointed out that the ribosomal proteins might have over 30 extra-ribosomal functions including regulation of the gene-specific control of transcription, transcript-specific translational control, and surveillance of ribosome synthesis and they could be involved in induction of cell-cycle arrest or apoptosis and in regulation of normal development and cancer [67,69,70]. One of the characteristic examples of such moonlighting ribosomal proteins is given by the protein L4. The L4 protein is annotated to have 24% of disordered residues in the MobiDB database [71] which are localized in the several regions including residues 1-11, 52-91, 189-196, 300-313, and at the C-terminus starting at the residue 341. This is also in agreement with the D^2P^2 database [72] that lists residues 1-12, 72-81, 193-194, 306-311, and 347-351 as disordered. This protein is known to both inhibit [73] and attenuate [74] the translation of the S10 operon, which, in *E.coli*, encodes eleven different ribosomal proteins, one of which is L4 itself [75]. Also, L4 can bind to RNase E (which is a part of the degradosome that plays an important role in mRNA turnover as well as in the processing and decay of non-coding RNAs), modulate activity of this crucial nuclease and thereby regulate mRNA composition in response to stress [76]. Curiously, eukaryotic L4 seems to be also engaged in the extra-ribosomal functions. In fact, recently it has been pointed out that this protein plays an important role in the ribosome biogenesis, since the deletion of the universally conserved internal loop of yeast L4 resulted in severe impairment of the growth and reduction of the levels of large ribosomal subunits [77], and since the eukaryote-specific acidic C-terminal extension (residues 265-362) is involved in several distinct interactions with the 60S surface needed for the hierarchical ribosome assembly [78]. Therefore, the internal loop (~60 residues) is known to be involved in interaction with the chaperone Acl4 involved in the assembly of the mature ribosome and later binds to the cognate nascent rRNA site [78]. In fact, in mature ribosome, the aforementioned loop (residues 46-111) protrudes from the globular folded core of L4 and deeply projects into the 25S rRNA core, lining a peptide exit tunnel of the mature ribosome [78].

Figure 5B represents results of the DisoRDPbind-based analysis of the interactions of yeast 60S ribosomal protein L4 (UniProt ID: P10664) with RNA (red lines) and proteins (green lines). We also annotate the native RNA-binding regions (red horizontal line at the bottom of Figure 5B) which were collected from the protein-ligand binding database BioLiP [79]; they are in agreement with the discussion in the above paragraph. We observe that the two large peaks in the putative propensities for disordered RNA binding align with the localization of the native RNA-binding regions; the DisoRDPbind webserver predicts residues with the propensities for RNA binding ≥ 0.151 as RNA binding (see **Note 6**). Also, both MobiDB and D^2P^2 suggest that these regions are intrinsically disordered. Although the predicted propensities for the protein-binding are below a cut-off value (the DisoRDPbind webserver

predicts residues with the propensities for protein binding ≥ 0.799 as protein binding; see **Note 6**), the N-terminus is predicted with relatively high values that suggest potential for disordered protein-binding.

Overall, we conclude that both case studies demonstrate that our webserver generates predictions that provide useful clues to find native disordered DNA-, RNA- and protein-binding regions.

4 Notes

1. The runtime in milliseconds for a given sequence with n residues can be estimated using the following formula, time = $0.0077*n^2 + 0.9028*n + 301.06$. This formula was estimated based on empirical data discussed in [45]. Given n = 200, time = 789.6 [milliseconds] = 0.79 [seconds]. Given n = 1000, time = 8903.7 [milliseconds] = 8.9 [second]. This formula can be used to estimate a total runtime for a large set of proteins since predictions on the webserver are run serially.
2. Server accepts between 1 and 5000 protein sequences. The user must submit their sequence(s) in FASTA format to guarantee they will receive the correct response from DisoRDPbind webserver. This format is described at https://en.wikipedia.org/wiki/FASTA_format
3. Due to a limitation of one of the methods that is used to generate DisoRDPbind sequence features (i.e., secondary structure profile predicted by PSIPRED), the webserver cannot process very long (>10000 residues) protein chains.
4. Although DisoRDPbind can predict an average size protein with about 200 residues within 1 second, it may take hours to process the prediction for thousands of (up to 5000) protein sequences. Keeping the web browser window open this long could be prohibitive. Therefore, we require the user to provide an e-mail address where (s)he will be notified when the results are available and how to access these results.
5. User should store the link to the results for future reference. We store the results of the prediction for at least 3 months under the provided link. Although the same link that is shown in the web browser window is sent via e-mail, we advise users to copy the link from the web browser. This is in case if an invalid e-mail address was entered and thus no e-mail will reach the user.
6. The binary prediction is generated from the predicted propensity scores using a threshold, i.e., residues with the propensity higher than the threshold are assigned with the binary value 1 and the remaining residues are assigned with 0. These thresholds equal 0.245, 0.151, and 0.799 for the predictions of the disordered DNA-, RNA- and protein-binding, respectively. They correspond to the FP-rate (fraction of incorrectly predicted non-binding residues) of 0.1 that was estimated using the TRAINING dataset. This means that the user should expect that among the predicted binding residues there are about 10% of the non-binding residues.

5 Acknowledgement

This work was supported by the National Natural Science Foundation of China (Grant No. 11501407) and the China National 863 High-Tech Program (Grant No. 2015AA020101) to Z.P. and by the Discovery grant from the Natural Sciences and Engineering Research Council of Canada to L.K. that was used to fund C.W.

6 References

1. 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
2. Guharoy M, Pauwels K, Tompa P (2015) SnapShot: Intrinsic Structural Disorder. Cell 161 (5):1230-1230 e1231. doi:10.1016/j.cell.2015.05.024
3. Habchi J, Tompa P, Longhi S, Uversky VN (2014) Introducing protein intrinsic disorder. Chem Rev 114 (13):6561-6588. doi:10.1021/cr400514h
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. Chem Rev 114 (13):6589-6631. doi:10.1021/cr400525m
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. doi:10.1007/s00018-014-1661-9
6. 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. doi:10.1016/j.bbapap.2013.05.022
7. Atkins JD, Boateng SY, Sorensen T, McGuffin LJ (2015) Disorder Prediction Methods, Their Applicability to Different Protein Targets and Their Usefulness for Guiding Experimental Studies. Int J Mol Sci 16 (8):19040-19054. doi:10.3390/ijms160819040
8. Bhowmick P, Guharoy M, Tompa P (2015) Bioinformatics Approaches for Predicting Disordered Protein Motifs. Adv Exp Med Biol 870:291-318. doi:10.1007/978-3-319-20164-1_9
9. Deng X, Eickholt J, Cheng J (2012) A comprehensive overview of computational protein disorder prediction methods. Mol Biosyst 8 (1):114-121. doi:10.1039/c1mb05207a
10. Dosztanyi Z, Meszaros B, Simon I (2010) Bioinformatical approaches to characterize intrinsically disordered/unstructured proteins. Brief Bioinform 11 (2):225-243. doi:10.1093/bib/bbp061
bbp061 [pii]
11. 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. doi:10.1038/cr.2009.87
12. Monastyrskyy B, Kryshtafovych A, Moult J, Tramontano A, Fidelis K (2014) Assessment of protein disorder region predictions in CASP10. Proteins 82 Suppl 2:127-137. doi:10.1002/prot.24391
13. Peng ZL, Kurgan L (2012) Comprehensive comparative assessment of in-silico predictors of disordered regions. Curr Protein Pept Sci 13 (1):6-18
14. Galea CA, High AA, Obenauer JC, Mishra A, Park CG, Punta M, Schllessinger A, Ma J, Rost B, Slaughter CA, Kriwacki RW (2009) Large-Scale Analysis of Thermostable, Mammalian Proteins Provides Insights into the Intrinsically Disordered Proteome. Journal of Proteome Research 8 (1):211-226. doi:10.1021/pr800308v
15. Tompa P, Dosztanyi Z, Simon I (2006) Prevalent structural disorder in E-coli and S-cerevisiae proteomes. Journal of Proteome Research 5 (8):1996-2000. doi:Doi 10.1021/Pr0600881
16. 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. Journal of Molecular Biology 337 (3):635-645. doi:10.1016/j.jmb.2004.02.002
S0022283604001482 [pii]
17. 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. doi:10.1080/07391102.2012.675145

18. 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. doi:10.1007/s00018-013-1446-6
19. 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. doi:10.1038/cdd.2013.65
20. 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. doi:10.1039/c2mb25102g
21. Chen JW, Romero P, Uversky VN, Dunker AK (2006) Conservation of intrinsic disorder in protein domains and families: II. Functions of conserved disorder. Journal of Proteome Research 5 (4):888-898. doi:Doi 10.1021/Pr060049p
22. Cumberworth A, Lamour G, Babu MM, Gsponer J (2013) Promiscuity as a functional trait: intrinsically disordered regions as central players of interactomes. Biochemical Journal 454:361-369. doi:Doi 10.1042/Bj20130545
23. Dyson HJ (2012) Roles of intrinsic disorder in protein-nucleic acid interactions. Molecular Biosystems 8 (1):97-104. doi:Doi 10.1039/C1mb05258f
24. Fuxreiter M, Toth-Petroczy A, Kraut DA, Matouschek AT, Lim RYH, Xue B, Kurgan L, Uversky VN (2014) Disordered Proteinaceous Machines. Chem Rev 114 (13):6806-6843. doi:Doi 10.1021/Cr4007329
25. Haynes C, Oldfield CJ, Ji F, Klitgord N, Cusick ME, Radivojac P, Uversky VN, Vidal M, Iakoucheva LM (2006) Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. PLoS Comput Biol 2 (8):890-901. doi:ARTN e100
DOI 10.1371/journal.pcbi.0020100
26. Tompa P, Csermely P (2004) The role of structural disorder in the function of RNA and protein chaperones. Faseb Journal 18 (11):1169-1175. doi:DOI 10.1096/fj.04-1584rev
27. Wu Z, Hu G, Yang J, Peng Z, Uversky VN, Kurgan L (2015) In various protein complexes, disordered protomers have large per-residue surface areas and area of protein-, DNA- and RNA-binding interfaces. FEBS Lett 589 (19 Pt A):2561-2569. doi:10.1016/j.febslet.2015.08.014
28. Shen J, Zhang J, Luo X, Zhu W, Yu K, Chen K, Li Y, Jiang H (2007) Predicting protein-protein interactions based only on sequences information. Proc Natl Acad Sci U S A 104 (11):4337-4341. doi:0607879104 [pii]
10.1073/pnas.0607879104
29. Zhang QC, Petrey D, Deng L, Qiang L, Sin Y, Thu CA, Bisikirska B, Lefebvre C, Accili D, Hunter T, Maniatis T, Califano A, Honig B (2013) Structure-based prediction of protein-protein interactions on a genome-wide scale (vol 490, pg 556, 2012). Nature 495 (7439):127-127. doi:Doi 10.1038/Nature11977
30. Puton T, Kozlowski L, Tuszynska I, Rother K, Bujnicki JM (2012) Computational methods for prediction of protein-RNA interactions. J Struct Biol 179 (3):261-268. doi:10.1016/j.jsb.2011.10.001
31. Zhao HY, Yang YD, Zhou YQ (2013) Prediction of RNA binding proteins comes of age from low resolution to high resolution. Mol Biosyst 9 (10):2417-2425. doi:Doi 10.1039/C3mb70167k
32. Kauffman C, Karypis G (2012) Computational tools for protein-DNA interactions. Wiley Interdisciplinary Reviews-Data Mining and Knowledge Discovery 2 (1):14-28. doi:Doi 10.1002/Widm.48
33. Gromiha MM, Nagarajan R (2013) Computational Approaches for Predicting the Binding Sites and Understanding the Recognition Mechanism of Protein-DNA Complexes. Protein-Nucleic Acids Interactions 91:65-99. doi:Doi 10.1016/B978-0-12-411637-5.00003-2

34. Varadi M, Vranken W, Guharoy M, Tompa P (2015) Computational approaches for inferring the functions of intrinsically disordered proteins. Front Mol Biosci 2:45. doi:10.3389/fmolb.2015.00045
35. Sharma A, Dehzangi A, Lyons J, Imoto S, Miyano S, Nakai K, Patil A (2014) Evaluation of Sequence Features from Intrinsically Disordered Regions for the Estimation of Protein Function. PLoS One 9 (2). doi:ARTN e89890
DOI 10.1371/journal.pone.0089890
36. Cheng Y, Oldfield CJ, Meng J, Romero P, Uversky VN, Dunker AK (2007) Mining alpha-helix-forming molecular recognition features with cross species sequence alignments. Biochemistry 46 (47):13468-13477. doi:10.1021/bi7012273
37. Meszaros B, Simon I, Dosztanyi Z (2009) Prediction of protein binding regions in disordered proteins. PLoS Comput Biol 5 (5):e1000376. doi:10.1371/journal.pcbi.1000376
38. 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. doi:10.1093/bioinformatics/bts209
39. Khan W, Duffy F, Pollastri G, Shields DC, Mooney C (2013) Predicting Binding within Disordered Protein Regions to Structurally Characterised Peptide-Binding Domains. PLoS One 8 (9). doi:ARTN e72838
DOI 10.1371/journal.pone.0072838
40. Fang C, Noguchi T, Tominaga D, Yamana H (2013) MFSPSSMpred: identifying short disorder-to-order binding regions in disordered proteins based on contextual local evolutionary conservation. BMC Bioinformatics 14:300. doi:10.1186/1471-2105-14-300
41. Jones DT, Cozzetto D (2014) DISOPRED3: precise disordered region predictions with annotated protein-binding activity. Bioinformatics. doi:10.1093/bioinformatics/btu744
42. Malhis N, Gsponer J (2015) Computational identification of MoRFs in protein sequences. Bioinformatics 31 (11):1738-1744. doi:10.1093/bioinformatics/btv060
43. Yan J, Dunker AK, Uversky VN, Kurgan L (2015) Molecular Recognition Features (MoRFs) in three domains of life. Mol Biosyst in revision
44. Sickmeier M, Hamilton JA, LeGall T, Vacic V, Cortese MS, Tantos A, Szabo B, Tompa P, Chen J, Uversky VN, Obradovic Z, Dunker AK (2007) DisProt: the Database of Disordered Proteins. Nucleic Acids Res 35 (Database issue):D786-793. doi:10.1093/nar/gkl893
45. Peng Z, Kurgan L (2015) High-throughput prediction of RNA, DNA and protein binding regions mediated by intrinsic disorder. Nucleic Acids Res. doi:gkv585 [pii]
10.1093/nar/gkv585
46. Wootton JC, Federhen S (1993) Statistics of Local Complexity in Amino-Acid-Sequences and Sequence Databases. Comput Chem 17 (2):149-163. doi:Doi 10.1016/0097-8485(93)85006-X
47. 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. doi:10.1093/bioinformatics/bti541
48. McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. Bioinformatics 16 (4):404-405
49. Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25 (17):3389-3402. doi:DOI 10.1093/nar/25.17.3389
50. Vaughn JP, Davis PL, Jarboe MD, Huper G, Evans AC, Wiseman RW, Berchuck A, Iglehart JD, Futreal PA, Marks JR (1996) BRCA1 expression is induced before DNA

synthesis in both normal and tumor-derived breast cells. Cell Growth Differ 7 (6):711-715
51. Friedenson B (2007) The BRCA1/2 pathway prevents hematologic cancers in addition to breast and ovarian cancers. BMC Cancer 7:152. doi:1471-2407-7-152 [pii]
10.1186/1471-2407-7-152
52. Mark WY, Liao JC, Lu Y, Ayed A, Laister R, Szymczyna B, Chakrabartty A, Arrowsmith CH (2005) Characterization of segments from the central region of BRCA1: an intrinsically disordered scaffold for multiple protein-protein and protein-DNA interactions? J Mol Biol 345 (2):275-287. doi:10.1016/j.jmb.2004.10.045
53. Wang Q, Zhang H, Kajino K, Greene MI (1998) BRCA1 binds c-Myc and inhibits its transcriptional and transforming activity in cells. Oncogene 17 (15):1939-1948. doi:10.1038/sj.onc.1202403
54. Aprelikova ON, Fang BS, Meissner EG, Cotter S, Campbell M, Kuthiala A, Bessho M, Jensen RA, Liu ET (1999) BRCA1-associated growth arrest is RB-dependent. Proc Natl Acad Sci U S A 96 (21):11866-11871
55. Lane DP (1992) p53, guardian of the genome. Nature 358 (6381):15-16
56. Zhang H, Somasundaram K, Peng Y, Tian H, Bi D, Weber BL, El-Deiry WS (1998) BRCA1 physically associates with p53 and stimulates its transcriptional activity. Oncogene 16 (13):1713-1721. doi:10.1038/sj.onc.1201932
57. Haber JE (1998) The many interfaces of Mre11. Cell 95 (5):583-586. doi:S0092-8674(00)81626-8 [pii]
58. Zhong Q, Chen CF, Li S, Chen Y, Wang CC, Xiao J, Chen PL, Sharp ZD, Lee WH (1999) Association of BRCA1 with the hRad50-hMre11-p95 complex and the DNA damage response. Science 285 (5428):747-750. doi:7719 [pii]
59. Baumann P, Benson FE, West SC (1996) Human Rad51 protein promotes ATP-dependent homologous pairing and strand transfer reactions in vitro. Cell 87 (4):757-766. doi:S0092-8674(00)81394-X [pii]
60. Scully R, Chen J, Plug A, Xiao Y, Weaver D, Feunteun J, Ashley T, Livingston DM (1997) Association of BRCA1 with Rad51 in mitotic and meiotic cells. Cell 88 (2):265-275. doi:S0092-8674(00)81847-4 [pii]
61. Garcia-Higuera I, Kuang Y, Naf D, Wasik J, D'Andrea AD (1999) Fanconi anemia proteins FANCA, FANCC, and FANCG/XRCC9 interact in a functional nuclear complex. Mol Cell Biol 19 (7):4866-4873
62. Folias A, Matkovic M, Bruun D, Reid S, Hejna J, Grompe M, D'Andrea A, Moses R (2002) BRCA1 interacts directly with the Fanconi anemia protein FANCA. Hum Mol Genet 11 (21):2591-2597
63. Hu YF, Li R (2002) JunB potentiates function of BRCA1 activation domain 1 (AD1) through a coiled-coil-mediated interaction. Genes Dev 16 (12):1509-1517. doi:10.1101/gad.995502
64. Paull TT, Cortez D, Bowers B, Elledge SJ, Gellert M (2001) Direct DNA binding by Brca1. Proc Natl Acad Sci U S A 98 (11):6086-6091. doi:10.1073/pnas.111125998
111125998 [pii]
65. Cate JH, Yusupov MM, Yusupova GZ, Earnest TN, Noller HF (1999) X-ray crystal structures of 70S ribosome functional complexes. Science 285 (5436):2095-2104. doi:7861 [pii]
66. Ben-Shem A, Garreau de Loubresse N, Melnikov S, Jenner L, Yusupova G, Yusupov M (2011) The structure of the eukaryotic ribosome at 3.0 A resolution. Science 334 (6062):1524-1529. doi:science.1212642 [pii]
10.1126/science.1212642
67. Wool IG (1996) Extraribosomal functions of ribosomal proteins. Trends Biochem Sci 21 (5):164-165. doi:S0968-0004(96)20011-8 [pii]
68. Weisberg RA (2008) Transcription by moonlight: structural basis of an extraribosomal activity of ribosomal protein

S10. Mol Cell 32 (6):747-748. doi:S1097-2765(08)00851-4 [pii]
10.1016/j.molcel.2008.12.010
69. Lindstrom MS (2009) Emerging functions of ribosomal proteins in gene-specific transcription and translation. Biochem Biophys Res Commun 379 (2):167-170. doi:S0006-291X(08)02492-3 [pii]
10.1016/j.bbrc.2008.12.083
70. Warner JR, McIntosh KB (2009) How common are extraribosomal functions of ribosomal proteins? Mol Cell 34 (1):3-11. doi:S1097-2765(09)00177-4 [pii]
10.1016/j.molcel.2009.03.006
71. 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. doi:10.1093/nar/gku982
72. 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. doi:10.1093/nar/gks1226
73. Gaal T, Bartlett MS, Ross W, Turnbough CL, Jr., Gourse RL (1997) Transcription regulation by initiating NTP concentration: rRNA synthesis in bacteria. Science 278 (5346):2092-2097
74. Zengel JM, Lindahl L (1994) Diverse mechanisms for regulating ribosomal protein synthesis in Escherichia coli. Prog Nucleic Acid Res Mol Biol 47:331-370
75. Mikhaylina AO, Kostareva OS, Sarskikh AV, Fedorov RV, Piendl W, Garber MB, Tishchenko SV (2014) Investigation of the regulatory function of archaeal ribosomal protein L4. Biochemistry (Mosc) 79 (1):69-76. doi:BCM79010087 [pii]
10.1134/S0006297914010106
76. Singh D, Chang SJ, Lin PH, Averina OV, Kaberdin VR, Lin-Chao S (2009) Regulation of ribonuclease E activity by the L4 ribosomal protein of Escherichia coli. Proc Natl Acad Sci U S A 106 (3):864-869. doi:0810205106 [pii]
10.1073/pnas.0810205106
77. Gamalinda M, Woolford JL, Jr. (2014) Deletion of L4 domains reveals insights into the importance of ribosomal protein extensions in eukaryotic ribosome assembly. RNA 20 (11):1725-1731. doi:rna.046649.114 [pii]
10.1261/rna.046649.114
78. Stelter P, Huber FM, Kunze R, Flemming D, Hoelz A, Hurt E (2015) Coordinated Ribosomal L4 Protein Assembly into the Pre-Ribosome Is Regulated by Its Eukaryote-Specific Extension. Mol Cell 58 (5):854-862. doi:S1097-2765(15)00220-8 [pii]
10.1016/j.molcel.2015.03.029
79. Yang J, Roy A, Zhang Y (2013) BioLiP: a semi-manually curated database for biologically relevant ligand-protein interactions. Nucleic Acids Res 41 (Database issue):D1096-1103. doi:10.1093/nar/gks966

Figure Captions

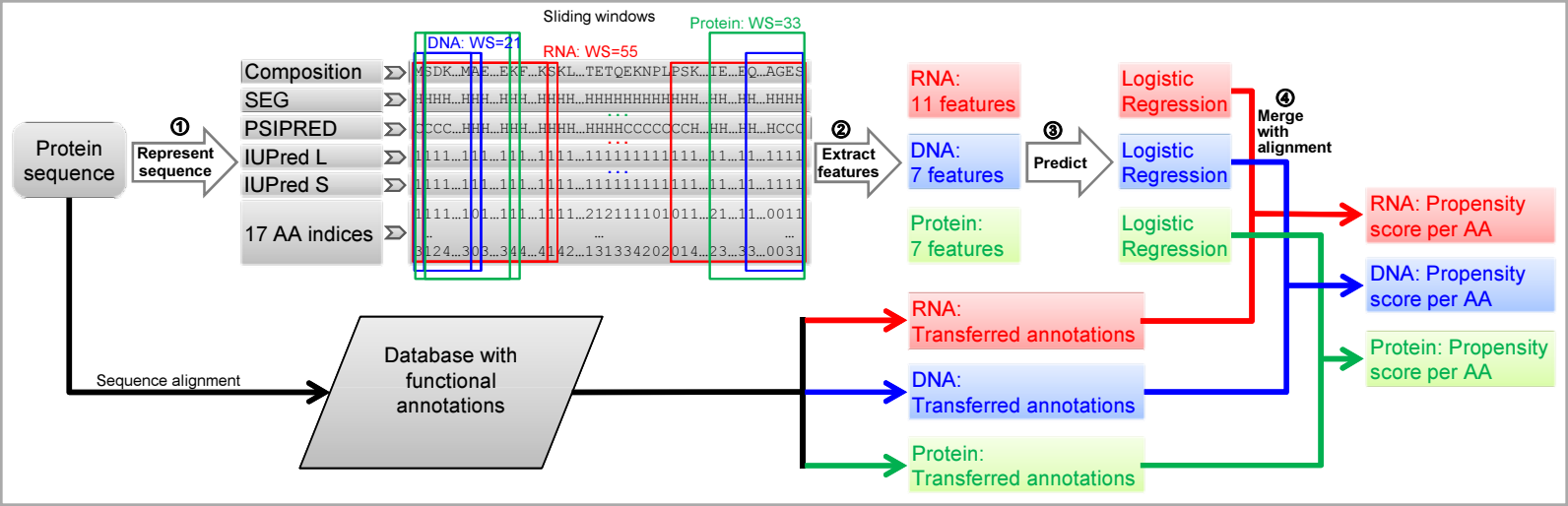
Figure 1. The architecture of the DisoRDPbind method. The four layers are denoted by the corresponding numbers shown inside circles. We use term “composition” to denote the amino acid composition. The SEG algorithm is used to generate the sequence complexity and PSIPRED and IUPred L(S) are utilized to predict the profiles of secondary structure and intrinsic disorder, respectively. The “17 AA indices” denote the physiochemical properties of amino acids (AAs) including their hydrophobicity, net charge, and free energy.

Figure 2. Screenshot of DisoRDPbind input form on the main webserver page. The red numbers annotate the three steps that must be followed to run the predictions.

Figure 3. Screenshot of page with the results generated by DisoRDPbind. The red numbers indicate the two main parts of this page.

Figure 4. Screenshot of the notification e-mail. The red numbers indicate the two main parts of this e-mail.

Figure 5. Predictions generated by DisoRDPbind for human BRCA1 protein (UniProt ID: P38398) (panel A) and yeast 60S ribosomal protein L4 (UniProt ID: P10664) (panel B). The putative propensities for DNA binding in panel A and RNA binding in panel B are shown using blue and red line, respectively; and putative propensities for protein binding are shown using green lines. The native annotations are shown using horizontal lines that lie on the *x*-axis.


Sliding windows
DNA: WS=21
RNA: WS=55
Protein: WS=33
Protein sequence
① Represent sequence
Composition
SEG
PSIPRED
IUPred L
IUPred S
17 AA indices
② Extract features
RNA: 11 features
DNA: 7 features
Protein: 7 features
③ Predict
Logistic Regression
Logistic Regression
Logistic Regression
④ Merge with alignment
RNA: Propensity score per AA
DNA: Propensity score per AA
Protein: Propensity score per AA
Sequence alignment
Database with functional annotations
RNA: Transferred annotations
DNA: Transferred annotations
Protein: Transferred annotations

Please follow the three steps below to make predictions:

1. Upload a file with protein sequences, or paste them into text area

Server accepts up to 5000 (**FASTA** FORMATED) protein sequences.
Either upload a file or enter each protein in a new line in the following text field (see **HELP** for details):

1

1

Example | Reset sequence(s)

2. Provide your e-mail address (required): 2

Please provide your e-mail address to be notified when results are ready.

3. Predict: Run DisoRDPbind! 3

DisoRDPbind results page

Results for **DisoRDPbind** webserver.

Use this link to download the results as a text file: **DisoRDPbind.pred** 1

Format of Results 2

Prediction for each protein is given in 8 lines

line 1: >protein name
line 2: protein sequence - 1-letter encoded protein sequence, where the lower (upper) case indicates the residue was predicted to interact (not to interact) with RNA/DNA/protein
line 3: RNA-binding residues - 1 represents the putative disordered RNA-binding residues; 0 otherwise
line 4: RNA-binding propensity scores separated by comma
line 5: DNA-binding residues - 1 represents the putative disordered DNA-binding residues; 0 otherwise
line 6: DNA-binding propensity scores separated by comma
line 7: protein-binding residues - 1 represents the putative disordered protein-binding residues; 0 otherwise
line 8: protein-binding propensity scores separated by comma
Note: The propensity score, which indicates the likelihood of a residue for the RNA-, DNA-, and/or protein-binding located in a disordered region, is predicted for each residue.

Visit biomine lab web page

http://biomine.ece.ualberta.ca

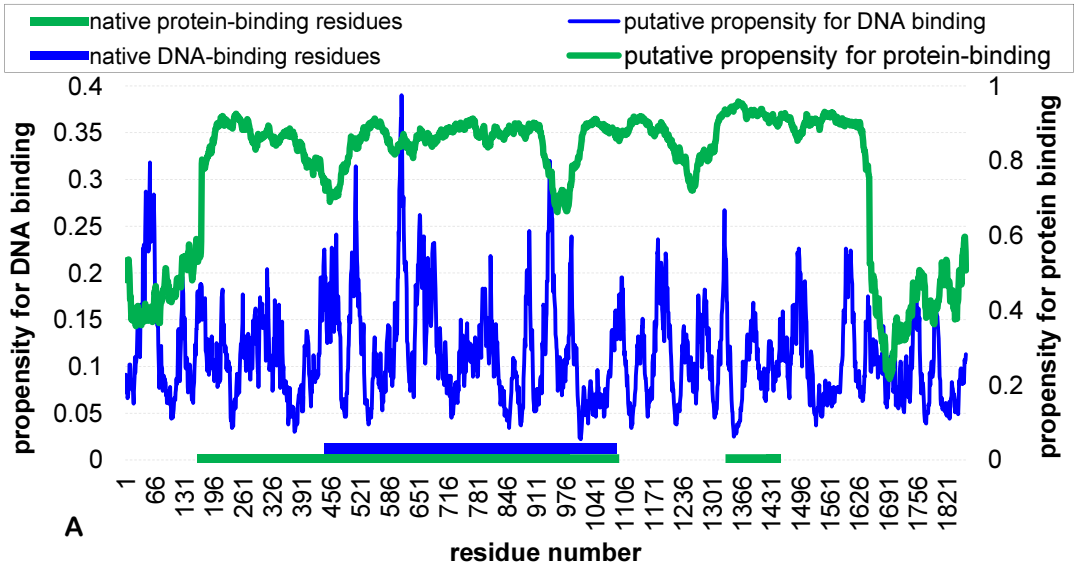
Predictions for DisoRDPbind job id: 20151005020757 are ready.

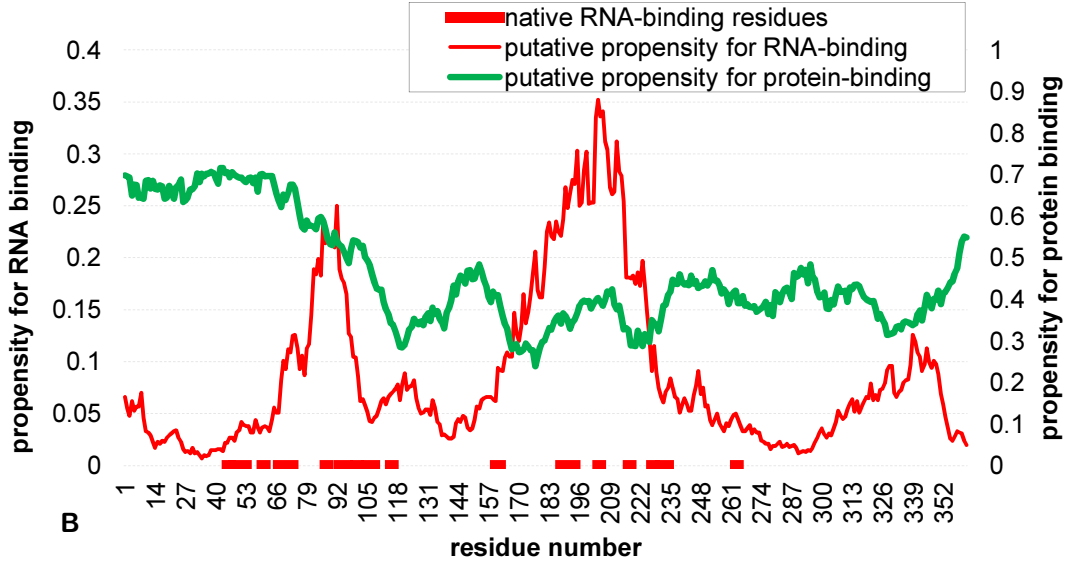
You can find the results for this job at:
http://biomine-ws.ece.ualberta.ca/webresults/DisoRDPbind/20151005020757/results.html 1
The text file can be found here:
http://biomine-ws.ece.ualberta.ca/webresults/DisoRDPbind/20151005020757/DisoRDPbind.pred 2

Upon the usage the users are requested to use the following citations:
Peng Z., Kurgan L. High-throughput prediction of RNA, DNA and protein binding regions mediated by intrinsic disorder. Nucleic Acids Res. doi:10.1093/nar/gkv585, 2015.

The webserver can be found here:
http://biomine.ece.ualberta.ca/DisoRDPbind/

Thank you for using our webserver,
Biomine group


native protein-binding residues
native DNA-binding residues
putative propensity for DNA binding
putative propensity for protein-binding
propensity for DNA binding
propensity for protein binding
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
1
0.8
0.6
0.4
0.2
0
1 66 131 196 261 326 391 456 521 586 651 716 781 846 911 976 1041 1106 1171 1236 1301 1366 1431 1496 1561 1626 1691 1756 1821
A
residue number


native RNA-binding residues
putative propensity for RNA-binding
putative propensity for protein-binding
propensity for RNA binding
propensity for protein binding
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1 14 27 40 53 66 79 92 105 118 131 144 157 170 183 196 209 222 235 248 261 274 287 300 313 326 339 352
B
residue number