



METHOD

SPOT-Disorder2: Improved Protein Intrinsic Disorder Prediction by Ensembled Deep Learning

Jack Hanson^{1,a}, Kuldip K. Paliwal^{1,b}, Thomas Litfin^{2,c}, Yaoqi Zhou^{2,3,*},d

¹ Signal Processing Laboratory, Griffith University, Brisbane 4111, Australia

² School of Information and Communication Technology, Griffith University, Gold Coast 4222, Australia

³ Institute for Glycomics, Griffith University, Gold Coast 4222, Australia

Received 26 October 2018; revised 18 January 2019; accepted 15 February 2019

Available online xxxx

Handled by Ziding Zhang

KEYWORDS

Intrinsic disorder;
Molecular recognition feature;
Machine learning;
Deep learning;
Protein structure

Abstract Intrinsically disordered or unstructured proteins (or regions in proteins) have been found to be important in a wide range of biological functions and implicated in many diseases. Due to the high cost and low efficiency of experimental determination of **intrinsic disorder** and the exponential increase of unannotated protein sequences, developing complementary computational prediction methods has been an active area of research for several decades. Here, we employed an ensemble of deep Squeeze-and-Excitation residual inception and long short-term memory (LSTM) networks for predicting protein intrinsic disorder with input from evolutionary information and predicted one-dimensional structural properties. The method, called SPOT-Disorder2, offers substantial and consistent improvement not only over our previous technique based on LSTM networks alone, but also over other state-of-the-art techniques in three independent tests with different ratios of disordered to ordered amino acid residues, and for sequences with either rich or limited evolutionary information. More importantly, semi-disordered regions predicted in SPOT-Disorder2 are more accurate in identifying **molecular recognition features** (MoRFs) than methods directly designed for MoRFs prediction. SPOT-Disorder2 is available as a web server and as a standalone program at <https://sparks-lab.org/server/spot-disorder2/>.

Introduction

Intrinsic disorder in proteins is the lack of tendency of a protein to fold into a well-defined, rigid structure. These dynamic protein structures can be experimentally observed as their backbone angles vary over time due to their innate flexibility [1]. The discovery of intrinsically disordered proteins (IDPs) or intrinsically disordered regions (IDRs) in proteins challenged the dogmatic structure–function paradigm, forcing a

* Corresponding author.

E-mail: yaoqi.zhou@griffith.edu.au (Zhou Y).

^a ORCID: 0000-0001-6956-6748.

^b ORCID: 0000-0002-3553-3662.

^c ORCID: 0000-0002-4863-3865.

^d ORCID: 0000-0002-9958-5699.

Peer review under responsibility of Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China.

<https://doi.org/10.1016/j.gpb.2019.01.004>

1672-0229 © 2019 The Authors. Published by Elsevier B.V. and Science Press on behalf of Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Please cite this article as: J. Hanson, K. K. Paliwal, T. Litfin et al., SPOT-Disorder2: Improved Protein Intrinsic Disorder Prediction by Ensembled Deep Learning, *Genomics Proteomics Bioinformatics*, <https://doi.org/10.1016/j.gpb.2019.01.004>

new perspective where protein rigidity is no longer a foregone conclusion [2].

IDPs are able to fulfil a wide range of niche, yet biologically crucial functional roles despite their lack of a rigid structure, due to their ability to transition between a set of transient, interconverting structural states [3]. Advantaged by their disordered flexibility [4], IDPs play essential roles in signaling, assembling, and regulatory functions [5], and are implicated in numerous human diseases, such as cancer, amyloidoses, cardiovascular disease, neurodegenerative diseases, and various genetic diseases [6]. A recent study on the amino acid (AA) residue-wise coverage of disorder has estimated that 19.6% of AA residues in eukaryotic organisms and 9.6% of AA residues in viral organisms are disordered [7]. This prevalence is vindicated by the fact that naturally-occurring proteins, particularly those in eukaryotes and viruses [7–9], tend to be more disordered than random sequences [10]. Thus, determining the identity and locations of IDPs and IDRs is fundamental to understanding and addressing the cause and effect of these unstructured states [11].

Due to the extensive monetary and time cost of experimental procedures, such as nuclear magnetic resonance (NMR), X-ray crystallography, and circular dichroism (CD) [12,13], many computational methods have been designed to bridge the growing gap between unannotated and annotated protein structures and/or their intrinsic disorder. Early work in protein disorder prediction was often based on small machine learning models [14], such as neural networks and support vector machines (SVMs). Other computational methods calculated disorder through the analysis of AA propensities and other sequence properties [15]. As more data and powerful tools became available, deep learning and recurrent architectures have taken the forefront, in methods such as SPINE-D [16], ESpritz [17], AUCpreD [18], SPOT-Disorder [19], and NetSurfP-2.0 [20]. A recent review by Liu et al. [21] placed SPOT-Disorder and AUCpreD as the two top-performing predictors for protein disorder prediction.

SPOT-Disorder [19], previously introduced by us, employed long short-term memory (LSTM) cells [22] in a bidirectional recurrent neural network (BRNN) [23] for protein disorder prediction. Since the publication of SPOT-Disorder, the single LSTM-BRNN topology for deep learning has been enhanced by utilizing an ensemble set of hybrid models consisting of both LSTM-BRNNs and residual convolutional neural networks (residual CNNs, called ResNets) [24] for protein contact map prediction [25], protein ω angle prediction [26], and protein secondary structure prediction [27]. This architec-

ture of network ensembles is advantageous because it can congregate and propagate both short-distance (ResNets) and long-distance (LSTM-BRNN) interactions throughout the protein sequence. Furthermore, the residual connections in these models alleviate the issues brought about by the vanishing gradient problem and allow for much deeper models (in the case of CNNs) and more effective gradient flow.

The effectiveness of ResNets and their various derivatives is displayed by their high performance in recent image classification competitions (ImageNet) [28]. Two such derivatives yet to be applied in bioinformatics are residual-inception networks [28] and Squeeze-and-Excitation networks [29]. Inception networks (v4) expand on the basic ResNets by increasing the number of paths available for data to be passed through. As such, the identity mapping function provided by the residual connection has a deeper level of abstraction due to the independent data paths. Squeeze-and-Excitation networks are another effective modification to ResNets that compresses the passing information into an excitation signal. This excitation signal can control the specific values added to the residual connection through the convolutional paths, behaving similarly to the learned gates of an LSTM cell. These models are currently cutting edge in image and speech processing tasks.

In this work, we examine models incorporating inception paths, residual connections, and Squeeze-and-Excitation networks (IncReSeNet) for their usefulness in disorder prediction. We find that the ensemble of different deep learning models leads to a stable and superior performance in four independent test sets with different ratios of ordered to disordered AA residues.

Methods

Neural network

The neural network topology employed in SPOT-Disorder2 consists of various models sequentially combining IncReSeNet, LSTM, and fully-connected (FC) topographical segments. Several models have been individually trained and then combined as an ensemble by averaging the disorder prediction output from each model. The hyperparameters of each individual method are outlined in Table 1.

The IncReSeNet segments follow the order of operations in the pre-activation ResNets architecture [24], with a multi-path inception-style architecture similar to Inception V4 [28]. As shown in the flow diagram (Figure 1), each block has three

Table 1 Architecture of five network models in the ensemble

Model	First segment	N_{LSTM}	N_{CNN}	K_{CNN}	N_{FC}	No. of blocks		
						LSTM	CNN	FC
0	RNN	250	60	5	250	2	10	1
1	RNN	250	60	7	500	2	10	1
2	RNN	250	60	9	250	2	10	1
3	CNN	250	60	9	250	2	10	2
4	RNN	250	60	9	250	1	10	1

Note: The order of layers in the model is presented in the column for the first segment, with LSTM and IncReSeNet blocks being the first segment for RNN and CNN, respectively (*i.e.*, at the input layers). N_{LSTM} , N_{CNN} , and N_{FC} refer to the number of hidden units in each LSTM cell, convolutional filters in each CNN layer, and nodes in each FC block, respectively. K_{CNN} refer to the kernel size of CNNs. LSTM, long short-term memory; RNN, recurrent neural network; CNN, convolutional neural network; FC, fully-connected layer; IncReSeNet, model incorporating inception paths, residual connections, and Squeeze-and-Excitation networks.

branches, including the residual connection and two convolution paths with 3 and 1 convolution operation, respectively. Each convolution operation is performed with a one-dimensional (1D) kernel with its size denoted as K_{CNN} , except for the first convolution in each branch, which has a kernel size of 1. These two paths are then concatenated and passed into a convolution of kernel size 1 and its depth denoted as N_{CNN} . The input to every convolution is normalized by the batch normalization technique [30], and is then activated by the exponential linear unit (ELU) activation function [31]. As each residual connection is preactivated, at the conclusion of all of the IncReSeNet layers, the output is both normalized and activated. Dropout of 25% is applied internally in some of the InReSeNet convolutions to avoid overtraining [32]. As

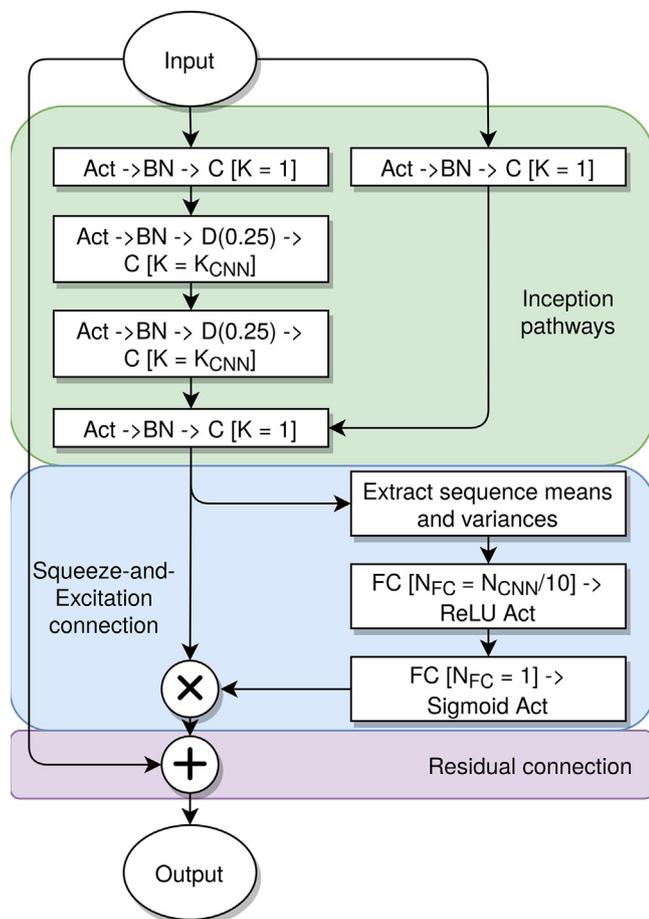


Figure 1 IncReSeNet blocks

This plot shows the data pathways from the input (top) to the output (bottom) of each IncReSeNet block. The Squeeze-and-Excitation (blue) section takes the output of the inception paths (green) and uses this information to control how much of itself is output from this block onto the residual pathway (purple). This is repeated for each sequential IncReSeNet block. The network-dependent parameters are detailed in Table 1. IncReSeNet, model incorporating inception paths, residual connections, and Squeeze-and-Excitation networks; BN, batch normalization; Act, activation; C, 1D convolution with kernel width K_{CNN} ; D(0.25), dropout of 25%; FC, fully-connected layer; K, parameter denoting layer kernel size; CNN, convolutional neural network; N_{FC} , number of neurons in FC; N_{CNN} , number of nodes in each convolutional layer; ReLU, rectified linear unit.

shown in Figure 1, dropout is applied after batch normalization (to not affect the moving average and variance measurements), but before the convolution operations (to not to affect the residual connection).

The Squeeze-and-Excitation segments in the residual blocks consist of two FC layers applied directly before the residual connection is applied. The means and variances across the protein for the outputs of the prior convolution layer are calculated to provide $2 \times N_{CNN}$ values per protein. These values are then passed through two FC layers with $N_{CNN}/10$ outputs and a single output, respectively, and an ReLU and sigmoid activation. The outputs of the second FC layer are then used as a makeshift logic gate to select which values from the final convolution of the block will be added to the residual connection.

The LSTM layers follow a similar format to our previous experiments [19,27]. Each LSTM block consists of one bidirectional LSTM layer with a memory cell size annotated as N_{LSTM} in both directions, resulting in $N_{LSTM} \times 2$ output values. Dropout of 50% is applied to the output of the LSTM blocks. Each FC layer's size is denoted as N_{FC} , is activated by a rectified linear unit (ReLU) [33] and regularized by dropout. No dropout is employed for the output layer, which employs a sigmoid activation to convert the singular output into a probability of the AA residue being disordered.

The use of an ensemble predictor minimizes the effect of generalization errors between models [34]. A large corpus of models with varying hyperparameters are trained and their performance is analyzed on a validation set. These hyperparameters are swept through in a grid search and include the layout of the network, the number of nodes in each layer (one parameter each for LSTM, IncReSeNet, and FC layers), and the number of layers for each layer type. The five top-performing models with hyperparameters listed in Table 1 are chosen from this validation period and used in the final ensemble for SPOT-Disorder2. Selecting more models did not contribute to an increase in accuracy (data not shown).

SPOT-Disorder2 has been trained using the inbuilt Adam optimizer [35] in TensorFlow v1.10 [36], on an NVIDIA TITAN X GPU. A typical IncReSeNet model takes 40 s/epoch over our whole training set, whereas an LSTM network takes 3 min/epoch.

Input features

SPOT-Disorder2 employed a similar set of features to SPOT-Disorder. Besides the same evolutionary content consisting of the position-specific substitution matrix (PSSM) profile from PSI-BLAST [37], SPOT-Disorder2 also includes the hidden Markov model (HMM) profile from HHblits [38]. The PSSM profile is generated by 3 iterations of PSI-BLAST against the UniRef90 sequence database (UniProt release 2018_03), and consists of 20 substitution values of each position for each AA residue type. The HMM profile consists of 30 values generated by using HHblits v3.0.3 with the UniProt sequence profile database from Oct 2017 [39]. These 30 values themselves consist of 20 AA substitution probabilities, 10 transition frequencies, and the number of effective homologous sequences of a given protein (Neff). In addition, we utilized the predicted structural properties from SPOT-1D [27], a significant update from SPOT-Disorder which utilized SPIDER2

[40,41]. The features from SPOT-1D consist of 11 secondary structure probabilities (both three- and eight-state predicted secondary structure elements), 4 sine and 4 cosine θ , τ , ϕ , and ψ backbone angles, 1 relative solvent-accessible surface area (ASA), 1 contact number (CN), and 2 half-sphere exposure (HSE) values based on the carbon- α atoms.

These feature groups are concatenated to form 73 input features for each protein residue. Features of each residue are standardized to have zero mean and unit variance before being input in the network by the means and standard deviations of the training data.

Datasets

The datasets used in these experiments, as shown in **Table 2**, were obtained from our previous disorder prediction publications [16,19]. To summarize, we obtained 4229 non-redundant, high-resolution protein sequences from the Protein Data Bank (PDB) and Database of Protein Disorder (DisProt) [42]. These include 4157 X-ray crystallography structures (deposited to the PDB prior to August 05, 2003) and 72 fully-disordered proteins from DisProt v5.0. These chains were randomly split into a training set (Training) of 2700 chains, a validation set (Validation) of 300 chains, and a testing set (Test) of 1229 chains. Sequence similarity among these proteins is <25% according to BLASTClust [37]. As SPOT-1D has not been trained for proteins of length >700 AA residues, we remove all proteins of length >700 from all datasets. This reduces our training, validation, and test sets to 2615, 293, and 1185 proteins, respectively. For convenience, we will label this test set as Test1185.

We also obtained three independent test datasets (SL250, Mobi9414, and DisProt228) for a fair comparison against other methods. These datasets were the subsets from the established SL477 [16], MobiDB [43], and DisProt Complement [44] sets, respectively, after removing long proteins (>700 residues) and homologous proteins in our training dataset (25% sequence identity cutoff with BLASTClust). The proteins in DisProt228 are newly-annotated proteins that are deposited in the DisProt database v7.0 [45]. The proteins in SL477 with unknown residue types were also removed. The annotations in Mobi9414 (*i.e.*, from MobiDB) contain direct labels from the DisProt database, inferred labels from the PDB, and predicted labels from a large ensemble of disorder predictors such as ESpritz [17]. Predicted labels in MobiDB are not utilized due to their potential inaccuracy. Residues listed as ‘conflicting’ labels in MobiDB are omitted for performance analysis. Because some predictors employed MobiDB as a part of their training set, we also made a reduced subset of the Mobi9414, called Mobi4730 for independent testing for all methods compared. Because not all training sets are available for all meth-

ods, Mobi4730 was obtained by clustering Mobi9414 against the largest disorder training dataset for NetSurfP-2.0 at a sequence similarity of 25% by BLASTClust.

Performance evaluation

Analyzing the performance of a disorder predictor is difficult due to the innate class imbalance present between disordered and ordered AA residues. As such, several skew-independent metrics are used to gauge the overall classification accuracy of the predictor. They include sensitivity (the fraction of predicted positives in all true positives), precision (the fraction of true positives in predicted positives), specificity (the fraction of true negatives in predicted negatives), the weighted score Sw (Sw = sensitivity + specificity - 1), the area under the receiver operating characteristic (ROC) curve (AUC_{ROC}), and the area under the precision-recall curve (AUC_{PR}). The difference between two AUC_{ROC} values can be qualified as statistically significant according to a *P* value from a bivariate statistical test [46], where a smaller *P* value indicates a higher likelihood of the difference being significant. AUC_{PR} emphasizes the performance on positive labels, which is particularly informative when the fraction of positive labels is low, as the case of protein disorder [47].

In addition, we obtain the Matthew’s correlation coefficient (MCC) between the predicted and true labels with $MCC = (TP \times TN + FN \times FP) / \sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}$, where TP, TN, FP, and FN indicate true positive, true negative, false positive, and true negative, respectively.

These metrics all have a maximum value of 1, and as such the highest performing predictor can be taken as the one that provides the overall highest metrics across our testing datasets.

Method comparison

We compare SPOT-Disorder2 to several high-performing protein disorder predictors. These include the local versions of DISOPRED2 [48] and DISOPRED3 [49] (<http://bioinfadmin.cs.ucl.ac.uk/downloads/DISOPRED/>), MobiDB-lite [50], AUCpred [18] (https://github.com/realbigws/RaptorX_Property_Fast), s2D [51] (<http://www-mvsoftware.ch.cam.ac.uk/index.php/s2D>), SPOT-Disorder [19], SPOT-Disorder-Single (denoted as SPOT-Disorder-S for brevity) [52], and SPINE-D [16] (<http://sparks-lab.org/>). We also used the webserver of NetSurfP-2.0 [20] (<http://www.cbs.dtu.dk/services/NetSurfP-2.0/>). Additionally, different versions of ESpritz method [17] downloaded were (<http://protein.bio.unipd.it/download/>), which were based on either single-sequence (seq) or sequence profile (prof) information. These ESpritz methods were trained based on structural information obtained from DisProt data-

Table 2 Order and disorder propensity in each of the datasets

Dataset	No. of proteins	No. of ordered residues	No. of disordered residues	Percentage of disorder
Training	2615	542,532	59,743	9.92%
Validation	293	57,470	5765	9.12%
Test1185	1185	246,616	26,515	9.71%
SL250	250	32,261	21,173	39.62%
Mobi9414	9414	1,932,536	127,362	6.18%
DisProt228	228	30,772	18,811	37.94%

base, or PDB as determined by NMR or X-ray crystallography, which were termed as ESpritz-D, ESpritz-N, and ESpritz-X, respectively.

Application to prediction of molecular SPOT-Disorder2 were obtained by Necci recognition motifs

In order to predict molecular recognition features (MoRFs), we define two thresholds as upper and lower bounds to classify the outputs of SPOT-Disorder2 as MoRFs. We also smooth the outputs of SPOT-Disorder2 to prevent the prediction of short MoRFs regions of < 3 residues, since MoRFs typically are longer based on our analysis. To do this, we apply a sliding window of size w_L to the predicted labels (y_M) of SPOT-Disorder2, and apply the following function $\hat{y}_m(i) = 1$ if $\sum y_m(j) > w_L$, and 0, if otherwise.

Results and discussion

Importance of ensembled learning and features for disorder prediction

One novel aspect of SPOT-Disorder2 is the use of an ensemble of IncReSeNet, LSTM, and FC network topologies, rather than a single LSTM topology in the previous version (SPOT-Disorder). Thus, it is necessary to examine if additional network models lead to an improvement of SPOT-Disorder2 over SPOT-Disorder. As shown in Table 3, there is a clear significant improvement across all test datasets based on four different measures (AUC_{ROC} , AUC_{PR} , MCC, and Sw). For example, MCC values are improved by 7%, 8%, 7%, and 8% for the four independent test sets of Test1185, SL250, Mobi9414, and DisProt228, respectively. Improvement on AUC_{PR} is less consistent with 7%, 2%, 13%, and 8% improvement for the four test sets, respectively, because AUC_{PR} is very sensitive to precision at low sensitivity.

To demonstrate the effectiveness of using an ensemble over using a single model for intrinsic disorder prediction, we compared the performance of the single component models to that of the ensemble using the Mobi9414 dataset. As shown in Table S1, the use of ensembled learning enables more accurate final output when compared to the Model 2, the highest-performing component model on this dataset. However, Model 4, rather than Model 2, is the highest performing component for Test1185 (MCC of 0.599 and 0.593 for Models 4 and 2, respectively, against 0.607 for SPOT-Disorder2). This variation in model ranking for different test sets indicates the

effectiveness of ensembling in increasing the robustness of the culminating model.

To examine the contribution of each feature type to the performance of SPOT-Disorder2, we analyzed the performance of Model 0 for the Mobi9414 dataset. The features have been separated into groups provided by the following programs: PSI-BLAST (PSSM), HHblits (HHblits), and SPOT-1D (SPOT-1D). As shown in Table S2, PSSM is the most critical feature for maximizing both AUC metrics, while SPOT-1D is the most critical for enhancing the single threshold metrics MCC and Sw. HHblits, on the other hand, does not seem to have a significant contribution to the performance of the model, probably because the HHblits profile has already been used in the input pipeline through SPOT-1D. The difference between AUC_{ROC} for the full Model 0 and HHblits-omitted model is insignificant ($P < 0.15$). We also analyzed the performance of Model 0 with the removal of the LSTM layers. The performance between the modified and original Model 0 is comparable for AUC_{ROC} , but is significantly worse in terms of AUC_{PR} and MCC, indicating that the combination of LSTM and IncReSeNet layers adds significant performance gains to the ensemble as a whole.

Improved disorder prediction over existing techniques

We further compared the prediction performance of SPOT-Disorder2 with that of 26 other predictors using the newest annotated proteins in DisProt (DisProt228) [44]. The results of all methods except SPOT-Disorder, NetSurfP-2.0, and SPOT-Disorder2 were obtained by Necci et al. [44]. JRONN [67], IUPred optimized for short and long disorder [IUPred (short)] and [IUPred (long)] [15], and PONDR-VSL [14] are not discussed here because of lower performance except the second-best shown below. However, predictions for two proteins were missing from these data, so the comparisons in this section are based on a 226-protein subset of DisProt228. As shown in Table 4, SPOT-Disorder2 improves over the second-best ESpritz-X (prof) by 2% in AUC_{ROC} , 4% in AUC_{PR} , 5% in MCC, and 5% in Sw. The precision-recall curves of the top 10 predictors according to AUC_{PR} are shown in Figure 2. The curve for SPOT-Disorder2 is above all other curves at all sensitivity values tested, except that its performance is slightly worse than that of IUPred (long) at sensitivity < 0.15 , or ESpritz-X (prof) at sensitivity between 0.4 and 0.6. It should be noted that ESpritz-X (prof) has very poor precision at extremely low sensitivity (or near the highest possible threshold that separate disordered residues from ordered residues), suggesting that false positives exist even for the highest

Table 3 Performance of the SPOT-Disorder2 and SPOT-Disorder on four independent test datasets

Dataset	SPOT-Disorder2				SPOT-Disorder			
	AUC_{ROC}	AUC_{PR}	MCC	Sw	AUC_{ROC}	AUC_{PR}	MCC	Sw
Validation	0.938	0.725	0.621	0.739	–	–	–	–
Test1185	0.914	0.698	0.607	0.676	0.894	0.65	0.567	0.477
SL250	0.901	0.889	0.679	0.625	0.893	0.875	0.629	0.567
Mobi9414	0.943	0.71	0.642	0.744	0.924	0.628	0.598	0.567
DisProt228	0.81	0.722	0.5	0.452	0.793	0.668	0.463	0.465

Note: For SPOT-Disorder2, both MCC and Sw are obtained using thresholds that maximize MCC and Sw on the Validation set, (thresholds are 0.370 for MCC and 0.070 for Sw, respectively). AUC_{ROC} , area under the receiver operating characteristic curve; AUC_{PR} , area under the precision-recall curve; MCC, Matthew's correlation coefficient; Sw, weighted score.

Table 4 Performance of various disorder prediction methods on a 226-chains subset of the DisProt228 dataset

Method group	Method name	AUC _{ROC}	AUC _{PR}	MCC	Sw
Single-sequence-based	GlobPlot [64]	0.028	0.237	0.169	0.106
	DisEMBL (Hot-Loops) [66]	0.655	0.522	0.209	0.213
	DisEMBL (465)	0.718	0.589	0.318	0.280
	IUPred (long)	0.764	0.685	0.418	0.410
	IUPred (short)	0.764	0.664	0.405	0.386
	ESpritz-N (seq)	0.751	0.644	0.391	0.395
	ESpritz-D (seq)	0.756	0.636	0.375	0.357
	ESpritz-X (seq)	0.770	0.653	0.419	0.404
	SPOT-Disorder-S	0.79	0.702	0.44	0.418
Sequence-profile-based	s2D	0.727	0.625	0.267	0.272
	AUCpreD	0.748	0.312 [#]	0.434	0.436
	JRONN	0.753	0.654	0.379	0.388
	ESpritz-D (prof)	0.759	0.645	0.379	0.364
	MFDp2 [68]	0.768	0.594 [#]	0.371	0.375
	DISOPRED	0.771	0.608	0.406	0.387
	MobiDB-lite	0.772	0.596 [#]	0.422	0.360
	NetSurfP-2.0	0.774	0.615	0.421	0.400
	ESpritz-N (prof)	0.776	0.674	0.432	0.440
	MFDp [69]	0.776	0.603 [#]	0.357	0.366
	SPINE-D	0.786	0.684	0.423	0.436
	SPOT-Disorder	0.792	0.666	0.462	0.465
	ESpritz-X (prof)	0.796	0.69	0.476	0.481
Metapredictor	MetaDisorder [65]	0.591	0.113 [#]	0.445	0.453
	MetaDisorder-3D	0.609	0.230 [#]	0.208	0.175
	MetaDisorder-md	0.632	0.354 [#]	0.454	0.455
	MetaDisorder-md2	0.717	0.611	0.454	0.462
	PONDR-VSL [70]	0.779	0.666	0.397	0.408
This work (sequence-profile-based)	SPOT-Disorder2	0.809	0.716	0.499	0.507

Note: Performance of NetSurfP-2.0, SPOT-Disorder, SPOT-Disorder-S, and SPOT-Disorder2 was obtained from this work, whereas performance of other methods was reported previously [44]. Two proteins were missing in the DisProt228 dataset [44], so the results here are calculated from the remaining 226 chains. MCC and Sw values for SPOT-Disorder2 were obtained using the disorder probability thresholds that maximize MCC and Sw on the Validation dataset. AUC_{PR} labeled with # is unreliable because the sensitivity (recall) does not cover the whole range from 0 to 1 for the respective methods. seq and prof indicate single sequence-based and sequence profile-based, respectively. See above-mentioned references [64–66, 68–70] for further information.

confidence scores when using ESpritz-X (prof). The difference between AUC_{ROC} from SPOT-Disorder2 and that from ESpritz-X (prof) is statistically significant ($P < 1 \times 10^{-5}$, bivariate statistical test).

DisProt provides experimental evidence for the labels of about 50% of the residues in the dataset [44]. The remaining ‘undefined’ residues are labeled by DisProt as ordered by default, which would likely introduce some mis-classification

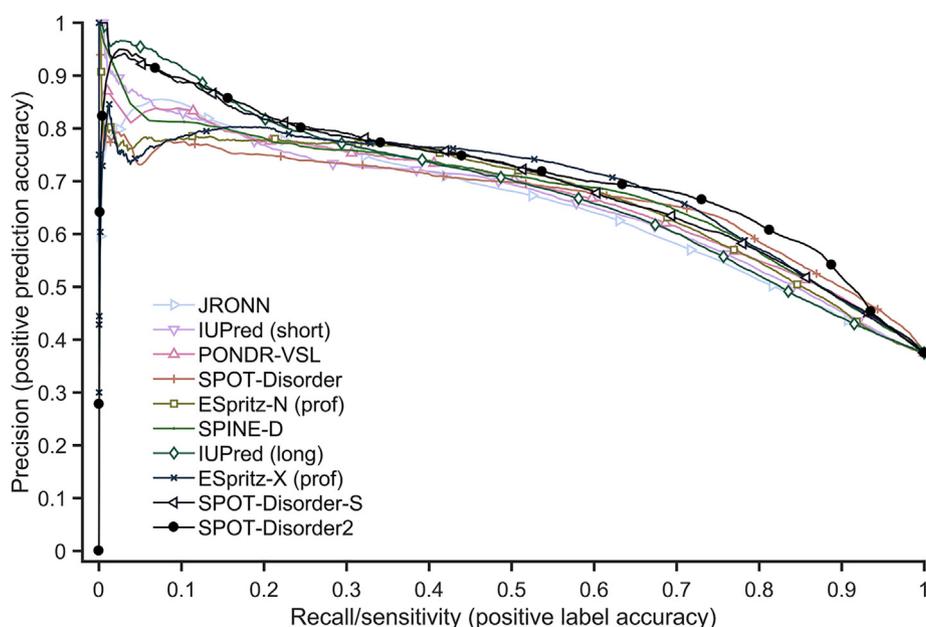


Figure 2 Precision–recall curves of the top 10 predictors for the DisProt228 dataset

The precision–recall curves were plotted by varying the threshold for defining disordered residues. ESpritz-N (prof) and ESpritz-X (prof) indicate profile-based ESpritz methods trained based on structural information obtained from PDB as determined by NMR or X-ray crystallography, respectively. SPOT-Disorder-S stands for SPOT-Disorder-Single.

Table 5 Performance of various disorder prediction methods on the Mobi4730 dataset

Method	AUC _{ROC}	AUC _{PR}	MCC	Sw
s2D	0.761	0.215	0.234	0.409
ESpritz-D (prof)	0.762	0.226	0.274	0.366
MobiDB-lite	0.811	0.434 [#]	0.45	0.449
DISOPRED2	0.859	0.536	0.394	0.577
ESpritz-N (prof)	0.864	0.567	0.299	0.524
SPOT-Disorder-S	0.878	0.567	0.51	0.394
ESpritz-X (prof)	0.893	0.608	0.439	0.635
SPINE-D	0.904	0.644	0.469	0.661
DISOPRED3	0.912	0.641	0.601	0.531
SPOT-Disorder	0.913	0.638	0.595	0.562
AUCpreD	0.917	0.297 [#]	0.603	0.611
NetSurfP-2.0	0.926	0.716	0.632	0.511
SPOT-Disorder2	0.933	0.723	0.648	0.715

Note: MCC and Sw values for SPOT-Disorder2 were obtained using the disorder probability thresholds that maximize MCC and Sw on the Validation dataset. AUC_{PR} labeled with # is unreliable because the sensitivity (recall) does not cover the whole range from 0 to 1 for the respective methods.

of disordered residues. The PR curve is particularly affected by label error due to the increased susceptibility to false positive predictions. We speculate that this label error may account for the (0,0) point of the SPOT-Disorder2 PR curve as well as the poor performance of several methods at low precision (ESpritz-X, MetaDisorder-md2, etc.). For example, the first 33 residues for actin-related protein 8 (UniProt: Q9H981; DisProt: DP00873) are amongst the highest confidence disorder prediction hits by SPOT-Disorder2. Despite being labeled as ordered by DisProt, there is no experimental evidence to support this labeling as these residues are missing from the solved X-ray structure [45,53]. However, we opt not to remove ambiguous residues from the dataset as they do not change

the performance ranking of the methods compared. Furthermore, SPOT-Disorder2 shows consistent improvement in terms of other metrics that are more robust to potential label noise, as well as in other datasets where undefined residues have been excluded (e.g., MobiDB).

We further employed other independent test datasets to compare our methods with other top performing methods for DisProt228 that are available to us as either a local implementation or online server. The performance of other methods for Mobi4730 after excluding training datasets is shown in Table 5. ESpritz-X (prof), the second-best predictor for DisProt228, performs significantly worse than SPOT-Disorder2 for Mobi4730, with a 19% difference in AUC_{PR} and 47% dif-

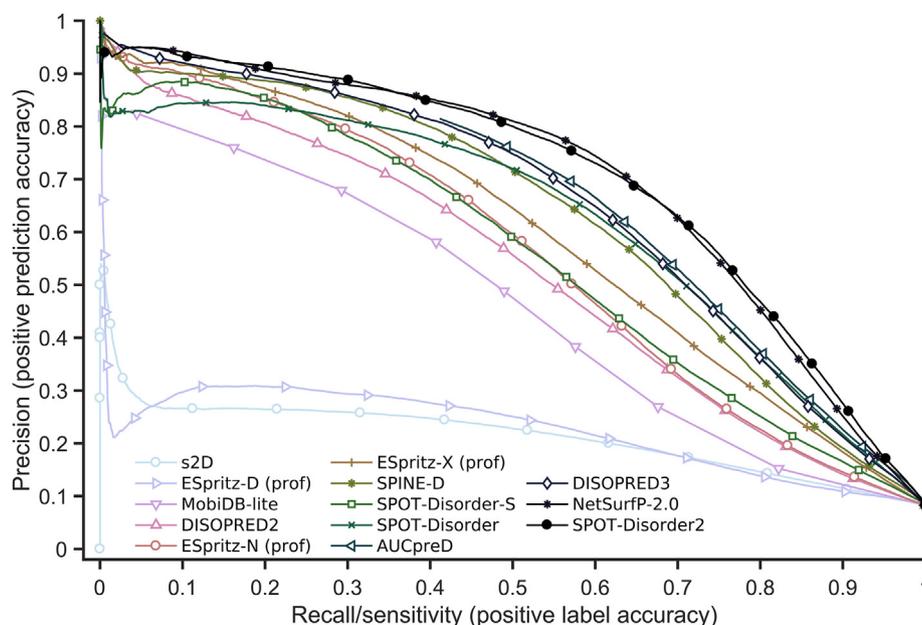


Figure 3 Precision–recall curves of 13 predictors for the Mobi4730 dataset

The methods compared are s2D, ESpritz-D (prof), MobiDB-lite, DISOPRED2, ESpritz-N (prof), ESpritz-X (prof), SPINE-D, SPOT-Disorder-S, SPOT-Disorder, AUCpreD, DISOPRED3, NetSurfP-2.0, and SPOT-Disorder2.

ference in MCC. The second best for Mobi4730 is NetSurfP-2.0. SPOT-Disorder2 achieved a 1% increase in AUC_{PR} and a 2.5% increase in MCC over NetSurfP-2.0 for Mobi4730, while the corresponding improvements are 17% in AUC_{PR} and 18% in MCC for DisProt228, respectively. PR curves for all the methods tested are shown in Figure 3. SPOT-Disorder2 has only a slight edge over NetSurfP-2.0, but both are significantly better than other methods examined. It is noted that AUCpreD is optimized for AUC_{ROC} , but performs poorly in terms of AUC_{PR} . Low AUC_{PR} values result from the inability of methods, such as AUCpreD and MobiDB-lite, to resolve high-confidence true and false positives for this dataset. For example, the PR curve of AUCpreD ends at roughly a sensitivity of 0.4 and a precision of 0.83 because AUCpreD predicts a high number of false positives even when the predicted disorder probability is 1. Calculating AUC without complete coverage of sensitivity from 0 to 1 makes the AUC_{PR} value somewhat arbitrary. To stress the inapplicability of this metric to AUCpreD (and others), we have included a note in Tables 4, 5, S3, and S4 for the predictors whose sensitivity values do not reach close to 0 and therefore having significantly disadvantaged AUC_{PR} scores. Nevertheless, the AUC_{ROC} of SPOT-Disorder2 is still significantly better than that of the nearest competitor, NetSurfP-2.0 ($P < 1 \times 10^{-7}$, bivariate statistical test).

To further demonstrate the stability of the performance of SPOT-Disorder2, we repeated the performance comparison of the aforementioned methods for the SL250 dataset. As shown in Table S3 and Figure 4, SPOT-Disorder2 continues to be the best performer with SPOT-Disorder being the second best. The PR curve of SPOT-Disorder2 is clearly above the curves of all other predictors for this dataset, including SPOT-Disorder and the two second-best methods for the two datasets tested previously, NetSurfP-2.0 and ESpritz-X (prof). The difference in AUC_{ROC} is significant between SPOT-Disorder2 and the near-

est predictor AUCpreD ($P < 1 \times 10^{-7}$, bivariate statistical test), as well as between the SPOT-Disorder2 and SPOT-Disorder ($P < 1 \times 10^{-3}$, bivariate statistical test) according to a bivariate statistical test.

Application of SPOT-Disorder2 to long proteins

Analysis on the UniProtKB/Swiss-Prot database (as of Dec 2018) [54] has shown that more than 91% of proteins consist of <700 AA residues, indicating that SPOT-Disorder2 is applicable to the vast majority of available sequences. However, it is also important to see how SPOT-Disorder2 performs for longer proteins representative of the remaining 9% that are not covered. Note that the size of 700 AA residues is not a hard limit in the software, but the size which was found to maximize the memory usage of GPU on our workstation.

The size limitation is mainly due to the use of SPOT-1D in the input of SPOT-Disorder2 input, which relies on the contact map prediction tool SPOT-Contact [25]. The computational memory necessary for using SPOT-1D with extremely long sequences becomes far too high for a typical user's workstation. To test the utility of SPOT-Disorder2 for long proteins, we replaced SPOT-1D by the secondary structure prediction tool SPIDER3 [55]. We generated the disorder profiles of SPOT-Disorder2 for 31 proteins that were initially omitted from the DisProt complement set from Necci et al. [44] in DisProt228, using the outputs of the secondary structure prediction tool SPIDER3 [55] in place of SPOT-1D (one protein consisting of >18,000 AA residues is still omitted). As SPIDER3 does not predict for 8-state secondary structure, we merely assign the 3-state probability predictions of SPIDER3 to the C, H, and E states for the 8-state predictions (and 0 for the S, T, I, G, and B states).

We compared the modified SPOT-Disorder2 to other methods for 31 large proteins (consisting of >700 AA residues) that

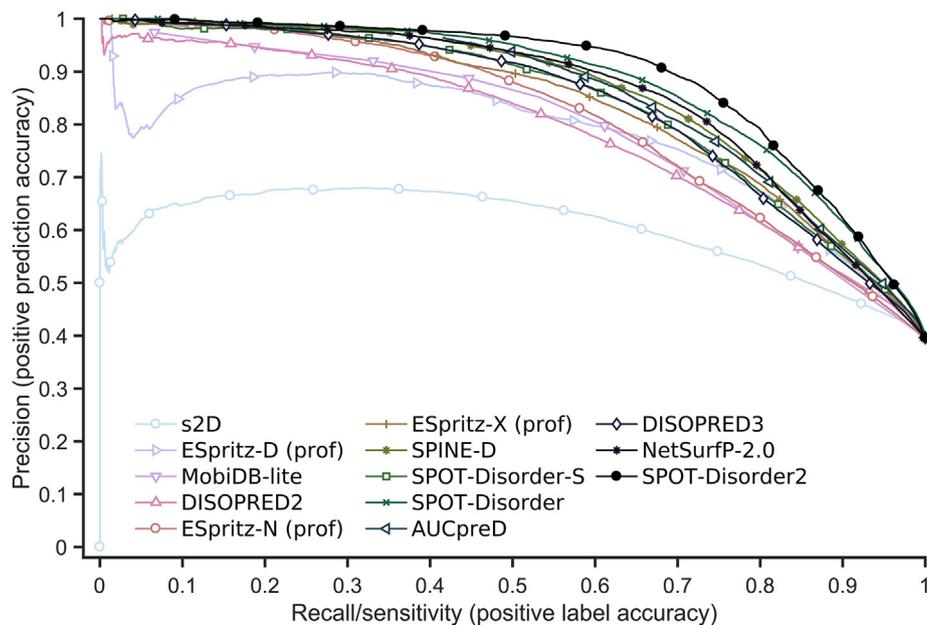


Figure 4 Precision–recall curves of 13 predictors for the SL250 dataset.

were initially omitted from the DisProt complement set from Necci et al. [44] in DisProt228. Table S4 shows that s2D is the top predictor for long proteins although it was the worst predictor for Mobi4730 and SL250, indicating that the disordered residues in the large-protein dataset tend to be in a coil state. However, the MCC of s2D is poor. SPOT-Disorder2 drops in the rankings, as is expected due to the learned distribution of the secondary structure inputs changing from SPOT-1D to SPIDER3, as well as losing the information from the 8-state secondary structure. The higher performance of SPOT-Disorder-S (highest MCC of 0.457) for this set of 31 proteins with > 700 AA residues might be explained by the fact that profile-based models are not well-trained for large proteins consisting of > 700 AA residues. A single-sequence-based method, on the other hand, is less dependent on sequence length. This is also echoed in the performance of single sequence-based ESpritz-D (seq) against the sequence profile-based ESpritz-D (prof) method (MCC of 0.382 vs. 0.228, respectively). Nevertheless, SPOT-Disorder2 is still one of the higher-ranking predictors, indicating that it is useful for long protein chains as well.

SPOT-Disorder2 is less accurate for the proteins with few sequence homologies

Robust performance of SPOT-Disorder2 across different datasets can be attributed to the evolutionary information derived from multiple sequence alignments in PSI-BLAST and HHblits. To examine the contribution of evolutionary information, we evaluated the performance of disorder prediction according to AUC_{PR} as a function of Neff. The larger Mobi9414 set is used, so that we have sufficient statistics for different values of Neff. As **Figure 5** shows, SPOT-Disorder2 performs more accurately for proteins with $Neff > 5$, below which there is a sharp decline in performance. However, there is a drop for proteins with $Neff > 6$. Significant homology between sequences seems to introduce noise into our predic-

tion of disordered regions, indicating that these regions might not be conserved like structured regions. Another possible cause is the sensitivity of disorder prediction to false positives in the homolog search. More studies are needed to isolate the cause of this pattern. Nevertheless, SPOT-Disorder2 makes significant improvement over SPOT-Disorder at all Neff values even for sequences with little evolution information ($Neff \approx 1$). This suggests that improvement is possible even at the single sequence level when several advanced machine learning techniques are integrated for consensus prediction.

Application of SPOT-Disorder2 to prediction of binding regions in disordered regions

Some intrinsically disordered regions can fold when interacting with other molecules including proteins, while others are structureless under any circumstances. Separating these foldable and non-foldable disordered regions is important for identification of functional regions, or MoRFs. Previously, we have proposed that foldable disordered regions are in a semi-disordered state with predicted disordered probabilities ranging from fully disordered [$p(D) = 1$] to fully structured [$p(D) = 0$] [56]. We tested this hypothesis using the output predictions from SPOT-Disorder2.

We have downloaded the Test and Test2012 datasets from the MoRFpred server [57] (<http://biomine.cs.vcu.edu/servers/MoRFpred/>) for validation and independent testing, respectively. We removed redundant sequences between the Test2012 and Test datasets at 25% sequence similarity using BLASTClust and the proteins with > 700 AA residues. As a result, 220 and 22 chains from Test2012 and Test datasets were retained for further analysis, respectively. The smoothing window size, along with the upper and lower thresholds, are optimized on the Test dataset. For comparison with other models, besides the web servers of MoRFpred, fMoRFpred, and DisoRDPbind [58,59] (<http://biomine.cs.vcu.edu/#webservers>), we also used ANCHOR2 [60] (<https://iupred2a.elte.hu/>),

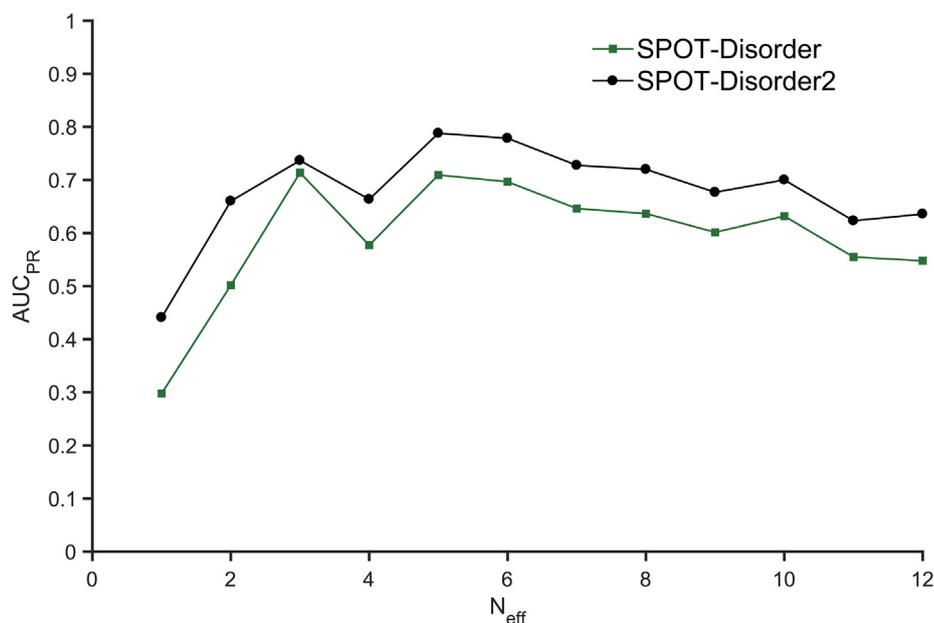


Figure 5 AUC_{PR} for proteins with different Neff values generated from HHblits
AUC_{PR}, area under precision-recall curve; Neff, number of effective homologous sequences of a given protein.

MoRFchibi [61] (https://gsponerlab.msl.ubc.ca/software/morf_chibi/), and the local version of MoRFPred-plus [62] (<https://github.com/roneshsharma/MoRFPred-plus>). In addition, we compared our results to the prediction by DISOPRED3 [49].

The performance of all predictors on the subset (220 chains) of the Test2012 dataset is shown in Table S5. With only three parameters trained for the Test dataset, SPOT-Disorder2 outperforms the second best MoRFPred-plus for MoRF prediction of the Test2012 dataset in terms of MCC (0.155 by SPOT-Disorder2 compared to 0.143 by MoRFPred-plus). Unlike SPOT-Disorder2, all other methods were specifically trained for MoRF regions. However, the performance of all methods is low, with MCC < 0.2. More data might be needed to further improve these methods for predicting binding residues in disordered regions.

Conclusion

In this paper, we have introduced a new method for predicting protein intrinsic disorder by taking advantage of recent progress in image recognition. With regard to the neural network architecture, we implemented two recent developments for an extension on residual convolutional neural networks, *i.e.*, multiple inception-style pathways [30] and signal Squeeze-and-Excitation [29]. We have also updated our feature set from our previous work [19] to include the latest state-of-the-art predictions for protein secondary structure from SPOT-1D [27]. Finally, the use of an ensemble of these methods has been again demonstrated effective in increasing accuracy through the removal of spurious false predictions. These enhancements over our previous and other disorder predictors enables SPOT-Disorder2 to achieve more robust and higher performance across different datasets with varied disorder to order ratios.

Consequently, SPOT-Disorder2 achieves the best performance over all metrics analyzed among the predictors tested.

MMSeq2 [63] is considered in this study due to its speedup in generating profiles over PSI-BLAST. However, MMSeq2 produces a less accurate prediction if its profiles are directly used to replace the profiles from PSI-BLAST, partially because SPOT-Disorder2 is trained on PSI-BLAST profiles. We hope to train a model for disorder prediction based on MMSeq2 profiles in a future work.

Availability

SPOT-Disorder2 is available at <https://sparks-lab.org/server/spot-disorder2/> as a server and downloadable package for local implementation.

Authors' contributions

JH, KKP, and YZ conceived and designed the experiments. TL prepared the data. JH conducted the experiments. JH, TL, and YZ analyzed the data and wrote the paper. All authors read, revised, and approved the final manuscript.

Competing interests

The authors declare no conflicts of interest.

Acknowledgments

This work was supported by Australian Research Council (Grant No. DP180102060) to YZ and KP, and in part by

the National Health and Medical Research Council (Grant No. 1121629) of Australia to YZ. We also gratefully acknowledge the use of the High Performance Computing Cluster ‘Gowonda’ to complete this study, and the aid of the research cloud resources provided by the Queensland Cyber Infrastructure Foundation (QCIF), Australia. The Titan V used for this research was donated by the NVIDIA Corporation.

Supplementary material

Supplementary material to this article can be found online at <https://doi.org/10.1016/j.gpb.2019.01.004>.

References

- [1] Uversky VN, Oldfield CJ, Dunker AK. Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J Mol Recognit* 2005;18:343–84.
- [2] Wright PE, Dyson HJ. Intrinsically unstructured proteins: reassessing the protein structure-function paradigm. *J Mol Biol* 1999;293:321–31.
- [3] Uversky VN. p53 proteoforms and intrinsic disorder: an illustration of the protein structure-function continuum concept. *Int J Mol Sci* 2016;17:1874.
- [4] Uversky VN. Functions of short lifetime biological structures at large: the case of intrinsically disordered proteins. *Brief Funct Genomics* 2018, Ely023.
- [5] Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* 2005;6:197–208.
- [6] Uversky VN, Oldfield CJ, Dunker AK. Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu Rev Biophys* 2008;37:215–46.
- [7] Hu G, Wang K, Song J, Uversky VN, Kurgan L. Taxonomic landscape of the dark proteomes: whole-proteome scale interplay between structural darkness, intrinsic disorder, and crystallization propensity. *Proteomics* 2018;18:1800243.
- [8] Peng Z, Yan J, Fan X, Mizianty MJ, Xue B, Wang K, et al. Exceptionally abundant exceptions: comprehensive characterization of intrinsic disorder in all domains of life. *Cell Mol Life Sci* 2015;72:137–51.
- [9] Xue B, Dunker AK, Uversky VN. Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. *J Biomol Struct Dyn* 2012;30:137–49.
- [10] Yu JF, Cao Z, Yang Y, Wang CL, Su ZD, Zhao YW, et al. Natural protein sequences are more intrinsically disordered than random sequences. *Cell Mol Life Sci* 2016;73:2949–57.
- [11] Uversky VN. Intrinsic disorder here, there, and everywhere, and nowhere to escape from it. *Cell Mol Life Sci* 2017;74:3065–7.
- [12] Receveur-Bréchet V, Bourhis JM, Uversky VN, Canard B, Longhi S. Assessing protein disorder and induced folding. *Proteins* 2006;62:24–45.
- [13] Konrat R. NMR contributions to structural dynamics studies of intrinsically disordered proteins. *J Magn Reson* 2014;241:74–85.
- [14] Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK. Sequence complexity of disordered protein. *Proteins* 2001;42:38–48.
- [15] Dosztányi Z, Csizmek V, Tompa P, Simon I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* 2005;21:3433–4.
- [16] Zhang T, Faraggi E, Xue B, Dunker AK, Uversky VN, Zhou Y. SPINE-D: accurate prediction of short and long disordered regions by a single neural-network based method. *J Biomol Struct Dyn* 2012;29:799–813.
- [17] Walsh I, Martin AJ, Di Domenico T, Tosatto SC. ESpritz: accurate and fast prediction of protein disorder. *Bioinformatics* 2012;28:503–9.
- [18] Wang S, Ma J, Xu J. AUCpreD: proteome-level protein disorder prediction by AUC-maximized deep convolutional neural fields. *Bioinformatics* 2016;32:i672–9.
- [19] Hanson J, Yang Y, Paliwal K, Zhou Y. Improving protein disorder prediction by deep bidirectional long short-term memory recurrent neural networks. *Bioinformatics* 2017;33:685–94.
- [20] Klausen MS, Jespersen MC, Nielsen H, Jensen KK, Jurtz VI, Soenderby CK, et al. NetSurfP- 2.0: improved prediction of protein structural features by integrated deep learning. *Proteins* 2019;87:520–7.
- [21] Liu Y, Wang X, Liu B. A comprehensive review and comparison of existing computational methods for intrinsically disordered protein and region prediction. *Brief Bioinform* 2019;18:330–46.
- [22] Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput* 1997;9:1735–80.
- [23] Schuster M, Paliwal KK. Bidirectional recurrent neural networks. *IEEE Trans Sign Proc* 1997;45:2673–81.
- [24] He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. In: European conference on computer vision. Berlin: Springer Publishing Company; 2016. p. 630–45.
- [25] Hanson J, Paliwal K, Litfin T, Yang Y, Zhou Y. Accurate prediction of protein contact maps by coupling residual two-dimensional bidirectional long short-term memory with convolutional neural networks. *Bioinformatics* 2018;34:4039–45.
- [26] Singh J, Hanson J, Heffernan R, Paliwal K, Yang Y, Zhou Y. Detecting proline and non-proline cis isomers in protein structures from sequences using deep residual ensemble learning. *J Chem Info Model* 2018;58:2033–42.
- [27] Hanson J, Paliwal K, Litfin T, Yang Y, Zhou Y. Improving prediction of protein secondary structure, backbone angles, solvent accessibility, and contact numbers by using predicted contact maps and an ensemble of recurrent and residual convolutional neural networks. *Bioinformatics* 2019;35:2403–10.
- [28] Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, inception-resnet and the impact of residual connections on learning. *Proc 31st AAAI Conf Artif Intell* 2017;4278–84.
- [29] Hu J, Shen L, Sun G. Squeeze-and-Excitation networks. *Proc IEEE Conf Comput Vision Pattern Recognit* 2018;7132–41.
- [30] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. *ICML’15 Proc 32nd Inter Conf Mach Learn* 2015;448–56.
- [31] Clevert DA, Unterthiner T, Hochreiter S. Fast and accurate deep network learning by exponential linear units (elus). *Inter Conf Learn Representations*, arXiv 2016;151107289.2015.
- [32] Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res* 2014;15:1929–58.
- [33] Dahl GE, Sainath TN, Hinton GE. Improving deep neural networks for LVCSR using rectified linear units and dropout. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing 2013;8609–13.
- [34] Hansen LK, Salamon P. Neural network ensembles. *IEEE Trans Pattern Anal Mach Intel* 1990;12:993–1001.
- [35] Kingma DP, Ba J. Adam: a method for stochastic optimization. *CoRR arXiv* 2014;abs/1412.6980.
- [36] Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. TensorFlow: large-scale machine learning on heterogeneous distributed systems. *CoRR arXiv* 2016;abs/1603.04467.
- [37] Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389–402.
- [38] Rimmert M, Biegert A, Hauser A, Söding J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat Methods* 2012;9:173–5.

- [39] Mirdita M, von den Driesch L, Galiez C, Martin MJ, Söding J, Steinegger M. Uniclust databases of clustered and deeply annotated protein sequences and alignments. *Nucleic Acids Res* 2016;45:D170–6.
- [40] Heffernan R, Paliwal K, Lyons J, Dehzangi A, Sharma A, Wang J, et al. Improving prediction of secondary structure, local backbone angles, and solvent accessible surface area of proteins by iterative deep learning. *Sci Rep* 2015;5:11476.
- [41] Heffernan R, Dehzangi A, Lyons J, Paliwal K, Sharma A, Wang J, et al. Highly accurate sequence-based prediction of half-sphere exposures of amino acid residues in proteins. *Bioinformatics* 2016;32:843–9.
- [42] Vucetic S, Obradovic Z, Vacic V, Radivojac P, Peng K, Iakoucheva LM, et al. DisProt: a database of protein disorder. *Bioinformatics* 2005;21:137–40.
- [43] Potenza E, Di Domenico T, Walsh I, Tosatto SC. MobiDB 2.0: an improved database of intrinsically disordered and mobile proteins. *Nucleic Acids Res* 2015;43:D315–20.
- [44] Necci M, Piovesan D, Dosztányi Z, Tompa P, Tosatto SC. A comprehensive assessment of long intrinsic protein disorder from the DisProt database. *Bioinformatics* 2018;34:445–52.
- [45] Piovesan D, Tabaro F, Mičetić I, Necci M, Quaglia F, Oldfield CJ, et al. DisProt 7.0: a major update of the database of disordered proteins. *Nucleic Acids Res* 2016;45:D219–27.
- [46] Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143:29–36.
- [47] Davis J, Goadrich M. The relationship between precision-recall and ROC curves. *ICML'06 Proc 23rd Inter Conf Mach Learn* 2006;233–40.
- [48] Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol* 2004;337:635–45.
- [49] Jones DT, Cozzetto D. DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics* 2015;31:857–63.
- [50] Necci M, Piovesan D, Dosztányi Z, Tosatto SCE. MobiDB-lite: fast and highly specific consensus prediction of intrinsic disorder in proteins. *Bioinformatics* 2017;33:1402–4.
- [51] Sormanni P, Camilloni C, Fariselli P, Vendruscolo M. The s2D method: simultaneous sequence-based prediction of the statistical populations of ordered and disordered regions in proteins. *J Mol Biol* 2015;427:982–96.
- [52] Hanson J, Paliwal K, Zhou Y. Accurate single-sequence prediction of protein intrinsic disorder by an ensemble of deep recurrent and convolutional architectures. *J Chem Info Model* 2018;58:2369–76.
- [53] Gerhold CB, Winkler DD, Lakomek K, Seifert FU, Fenn S, Kessler B, et al. Structure of actin-related protein 8 and its contribution to nucleosome binding. *Nucleic Acids Res* 2012;40:11036–46.
- [54] Bateman A, Martin MJ, Orchard S, Magrane M, Alpi E, Bely B, et al. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* 2018;47:D506–15.
- [55] Heffernan R, Yang Y, Paliwal K, Zhou Y. Capturing non-local interactions by long short term memory bidirectional recurrent neural networks for improving prediction of protein secondary structure. *Bioinformatics* 2017;33:2842–9.
- [56] Zhang T, Faraggi E, Li Z, Zhou Y. Intrinsically semi-disordered state and its role in induced folding and protein aggregation. *Cell Biochem Biophys* 2013;67:1193–205.
- [57] Disfani FM, Hsu WL, Mizianty MJ, Oldfield CJ, Xue B, Dunker AK, et al. MoRFPred, a computational tool for sequence based prediction and characterization of short disorder-to-order transitioning binding regions in proteins. *Bioinformatics* 2012;28:i75–83.
- [58] Yan J, Dunker AK, Uversky VN, Kurgan L. Molecular recognition features (MoRFs) in three domains of life. *Mol BioSyst* 2016;12:697–710.
- [59] Peng Z, Kurgan L. High-throughput prediction of RNA, DNA and protein binding regions mediated by intrinsic disorder. *Nucleic Acids Res* 2015;43:e121.
- [60] Mészáros B, Erdős G, Dosztányi Z. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res* 2018;46:329–37.
- [61] Malhis N, Jacobson M, Gsponer J. MoRFChibi SYSTEM: software tools for the identification of MoRFs in protein sequences. *Nucleic Acids Res* 2016;44:W488–93.
- [62] Sharma R, Bayarjargal M, Tsunoda T, Patil A, Sharma A. MoRFPred-plus: computational identification of MoRFs in protein sequences using physicochemical properties and HMM profiles. *J Theor Biol* 2018;437:9–16.
- [63] Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol* 2017;35:1026.
- [64] Linding R, Russell RB, Neduva V, Gibson TJ. GlobPlot: exploring protein sequences for globularity and disorder. *Nucleic Acids Res* 2003;31:3701–8.
- [65] Kozłowski LP, Bujnicki JM. MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins. *BMC Bioinf* 2012;13:111.
- [66] Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ, Russell RB. Protein disorder prediction: implications for structural proteomics. *Structure* 2003;11:1453–9.
- [67] Yang ZR, Thomson R, McNeil P, Esnouf RM. RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics* 2005;21:3369–76.
- [68] Mizianty MJ, Peng Z, Kurgan L. MFDp2: accurate predictor of disorder in proteins by fusion of disorder probabilities, content and profiles. *Intrinsically Disord Proteins* 2013;1:e24428.
- [69] Mizianty MJ, Stach W, Chen K, Kedarisetti KD, Disfani FM, Kurgan L. Improved sequence-based prediction of disordered regions with multilayer fusion of multiple information sources. *Bioinformatics* 2010;26:i489–96.
- [70] Peng K, Radivojac P, Vucetic S, Dunker AK, Obradovic Z. Length-dependent prediction of protein intrinsic disorder. *BMC Bioinf* 2006;7:208.