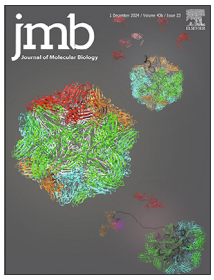

MERIT: Accurate Prediction of Multi Ligand-binding Residues with Hybrid Deep Transformer Network, Evolutionary Couplings and Transfer Learning

Jian Zhang [1,2*], Sushmita Basu [3], Fuhao Zhang [4], andLukasz Kurgan [3*]

*1 - **School of Computer and Information Technology,** Xinyang Normal University, Xinyang 464000, China*
*2 - **Yangtze Delta Region Institute (Quzhou),** University of Electronic Science and Technology of China, Quzhou 324003, China*
*3 - **Department of Computer Science,** Virginia Commonwealth University, Richmond, VA 23284, USA*
*4 - **College of Information Engineering,** Northwest A&F University, Yangling, Shaanxi 712100, China*

Correspondence to Jian ZhangLukasz Kurgan: School of Computer and Information Technology, Xinyang Normal University, Xinyang, 464000 China (J. Zhang); Department of Computer Science, Virginia Commonwealth University, Richmond, VA, 23284 USA (L. Kurgan). *jianzhang@xynu.edu.cn (J. Zhang), lkurgan@vcu.edu (L. Kurgan)*
https://doi.org/10.1016/j.jmb.2024.168872
Editor: Rita Casadio

Abstract

Multi-ligand binding residues (MLBRs) are amino acids in protein sequences that interact with multiple different ligands that include proteins, peptides, nucleic acids, and a variety of small molecules. MLBRs are implicated in a number of cellular functions and targeted in a context of multiple human diseases. There are many sequence-based predictors of residues that interact with specific ligand types and they can be collectively used to identify MLBRs. However, there are no methods that directly predict MLBRs. To this end, we conceptualize, design, evaluate and release MERIT (Multi-binding rEsidues pRedIcTor). This tool relies on a custom-crafted deep neural network that implements a number of innovative features, such as a multi-layered/step architecture with transformer modules that we train using a custom-designed loss function, computation of evolutionary couplings, and application of transfer learning. These innovations boost predictive performance, which we demonstrate using an ablation analysis. In particular, they reduce the number of cross-predictions, defined as residues that interact with a single ligand type that are incorrectly predicted as MLBRs. We compare MERIT against a representative selection of current and popular ligand-specific predictors, meta-predictors that combine their results to identify MLBRs, and a baseline regression-based predictor. These tests reveal that MERIT provides accurate predictions and statistically outperforms these alternatives. Moreover, using two test datasets, one with MLBRs and another with only the single ligand binding residues, we show that MERIT consistently produces relatively low false positive rates, including low rates of cross-predictions. The web server and datasets from this study are freely available at http://biomine.cs.vcu.edu/servers/MERIT/.

Introduction

Multi-ligand binding residues (MLBRs) are defined as amino acids in protein sequences that interact with multiple different types of ligands, i.e., different small molecules, peptides, proteins, DNA, RNA and/or lipids. The ability to bind multiple ligands in the same site can be explained by the existence of populations of different protein conformers in solution.[1] In the presence of a specific ligand, one of these conformations becomes energetically more favorable and the resulting

Please cite tts article as: J. Zhang, S. Basu, F. Zhang, et al., MERIT: Accurate Prediction of Multi Ligand-binding Residues with Hybrid Deep Transformer Network, Evolutionary Couplings and Transfer LearningJournal of Molecular Biology, https://doi.org/10.1016/j.jmb.2024.168872

protein–ligand complex is formed.[2] An "extreme" case of the underlying structural plasticity are the intrinsically disordered regions[3,4] that are capable of one-to-many binding, i.e., one disordered region binds multiple different partner molecules typically by folding into different structures.[5] Well-known proteins that include the one-to-many binding disordered regions are p53 and 14-3-3.[6,7] Other illustrative examples of MLBRs include G protein–coupled receptors and HIV-1 reverse transcriptase that have sites that interact with a variety of small ligands.[8,9] MLBRs also bind large biomolecules, such as nucleic acids, proteins and peptides. For instance, sequence regions that interact with both DNA and RNA.[10] Studies suggest that the multi-ligand binding and the associated conformational flexibility are involved in the allosteric regulation and facilitate evolution of new binding functions.[11,12] Moreover, some drugs that combat human diseases, which include cancers, mental disorders, cardiovascular diseases and HIV, rely on the binding sites composed of MLBRs.[13,14,8,9] These examples suggest that identification of MLBRs has biological and practical value.

Experimental methods, such as X-ray crystallography, nuclear magnetic resonance, and cryo-electron microscopy are used to identify binding residues in proteins.[15] However, they do not scale to match the exponential growth of the protein sequence space, which currently has about 300 million unique chains.[16,17] One viable option to reduce this annotation gap is to develop and use accurate computational predictors of binding residues. Dozens of these predictors were developed to date and they can be divided into methods that use protein structure vs. protein sequence as the input.[18–24] Some of the recently released structure-based predictors include GraphBind,[25] DeepPocket,[26] GraphSite,[27] LigBind,[28] GeoBind,[29] and DeepProSite.[30] We focus on methods that predict binding residues from the sequence since they are typically faster and require arguably easier to acquire sequences to make the predictions. A few popular and/or recently released methods that predict protein-binding residues include SCRIBER,[31] HybridPBRpred,[32] PROBselect,[33] DELPHI,[34] PITHIA,[35] DeepPROBind,[36] HN-PPISP,[37] and ISPRED-SEQ.[38] Representative sequence-based predictors of DNA-binding residues include TargetDNA,[39] DNAPred,[40] and HybridDBRpred.[41] Example predictors of the RNA-binding residues are FastRNABindR,[42] PredRBR,[43] and HybridRNAbind.[44] Popular methods that predict residues that bind small ligands include TargetS,[45] TargetVita,[46] HEMEsPred,[47] DeepATPseq,[48] SCAMPER,[49] LMetalSite,[50] and M-Ionic.[51] While many tools are limited to a specific ligand type, some predict binding for multiple types of ligands. Prediction of the DNA binding and the RNA binding residues is covered by BindN+,[52] DRNApred,[53] NucBind,[54] TSNAPred,[55] iNucRes-ASSH,[56] and MucLiPred.[57] The DNA, RNA and protein binding residues are predicted by DisoRDPbind,[58] HybridNAP,[24] ProNA2020,[59] and DeepDISOBind.[60] MTDsite covers DNA, RNA, carbohydrate, and peptide binding.[61]

However, to the best of our knowledge, none of the current predictors directly identifies MLBRs. While one could use multiple methods in tandem to predict MLBRs, this is rather tedious and difficult to accomplish. It requires identifying suitable predictors, running them using their web servers or implementations (if they are operational), collecting and standardizing their predictions, which could be in different ranges and formats, to properly combine them. To this end, we introduce MERIT (Multi-ligand binding rEsidues pRedIcTor), first-of-its-kind deep network model that accurately predicts MLBRs. We consider MLBRs as residues that interact with multiple biologically-relevant ligands, as defined in the popular BioLiP resource.[62,63] We developed and applied new training, validation and test datasets with annotations of MLBRs and residues that bind one ligand type. We utilize the latter annotations to consider and minimize cross-predictions, defined as the residues that bind one ligand type that are incorrectly predicted as MLBRs. Low cross-prediction values mean that the corresponding prediction properly differentiates single ligand binding residues vs. MLBRs. This is inspired by recent ligand-specific predictors of binding residues that similarly quantify and aim to combat cross-predictions, defined as the residues that bind other ligand types that are predicted as binding to the target ligand.[31,32,53,54,64–66] MERIT maximizes quality of the MLBR predictions by crafting a specialized deep network-based predictor that: (1) uses evolutionary couplings to model the fact that multiple residues bind the same ligand; (2) hybridizes transformer modules and fully connected feed-forward modules to adequately process different types of sequence-derived inputs; and (3) applies transfer learning and an advanced loss function to minimize the cross-predictions. We also empirically demonstrate that it generates predictions relatively quickly.

Materials and Methods

Selection of methods and construction of meta-predictors for comparative analysis

We select and use a collection of nine representative sequence-based predictors of binding residues to indirectly predict MLBRs and provide a baseline that should be substantially improved by MERIT. These tools are highly cited and/or recently released, relatively fast, available as server or code, and cover a broad spectrum of ligands. We provide details in the Supplement. The nine selected tools are BindN+,[52] TargetS,[45]

Table 1 Predictive performance of MERIT, the current predictors of binding residues, meta-predictors of MLBRs, and a baseline logistic regression model on the test dataset. We report averages ± the corresponding standard deviations based on the 100 tests; see the last paragraph in in the "Assessment of predictive performance" sub-section. We calibrate sensitivity so that the corresponding predictions maintain the same FPR = 5% or produce the number of putative MLBRs that is equal to the number of native MLBRs. We give the *p*-values when comparing against the MERIT model inside the round brackets. The best results are shown in bold font.

Type	Prediction target	Method name	Sensitivity		F1max	AUROC
			Number of putative MLBRs equals number of native MLBRs	FPR = 5%		
Current tools	DNA	BindN+	0.085 ± 0.011 (*p* = 2.0E−147)	0.126 ± 0.014 (*p* = 3.7E−156)	0.103 ± 0.009 (*p* = 8.8E−148)	0.608 ± 0.014 (*p* = 3.5E−162)
		DisoRDPbind	0.046 ± 0.007 (*p* = 8.8E−176)	0.065 ± 0.007 (*p* = 2.0E−191)	0.064 ± 0.005 (*p* = 1.3E−177)	0.502 ± 0.011 (*p* = 4.9E−216)
		MTDsite	0.064 ± 0.009 (*p* = 8.7E−163)	0.096 ± 0.012 (*p* = 5.7E−172)	0.096 ± 0.009 (*p* = 7.6E−153)	0.614 ± 0.012 (*p* = 5.9E−166)
		MucLiPred	0.115 ± 0.012 (*p* = 2.0E−123)	0.173 ± 0.013 (*p* = 2.4E−134)	0.136 ± 0.010 (*p* = 1.3E−121)	0.676 ± 0.012 (*p* = 1.3E−124)
	RNA	BindN+	0.104 ± 0.012 (*p* = 2.1E−130)	0.156 ± 0.014 (*p* = 5.0E−142)	0.123 ± 0.010 (*p* = 1.4E−131)	0.633 ± 0.013 (*p* = 8.0E−154)
		DisoRDPbind	0.028 ± 0.009 (*p* = 7.3E−180)	0.049 ± 0.011 (*p* = 1.4E−189)	0.065 ± 0.005 (*p* = 9.3E−177)	0.515 ± 0.014 (*p* = 2.5E−201)
		MTDsite	0.065 ± 0.007 (*p* = 2.6E−166)	0.100 ± 0.013 (*p* = 3.0E−168)	0.101 ± 0.008 (*p* = 5.6E−150)	0.620 ± 0.012 (*p* = 3.6E−166)
		MucLiPred	0.112 ± 0.015 (*p* = 5.7E−119)	0.180 ± 0.015 (*p* = 7.0E−126)	0.142 ± 0.009 (*p* = 1.6E−117)	0.687 ± 0.013 (*p* = 6.3E−112)
	Protein	DisoRDPbind	0.026 ± 0.006 (*p* = 3.9E−188)	0.043 ± 0.008 (*p* = 1.7E−197)	0.062 ± 0.005 (*p* = 2.0E−179)	0.477 ± 0.014 (*p* = 6.1E−212)
		SCRIBER	0.060 ± 0.011 (*p* = 1.5E−158)	0.090 ± 0.014 (*p* = 3.2E−169)	0.080 ± 0.009 (*p* = 5.1E−162)	0.539 ± 0.020 (*p* = 1.2E−171)
		ISPRED-SEQ	0.034 ± 0.008 (*p* = 5.7E−180)	0.058 ± 0.009 (*p* = 2.2E−191)	0.072 ± 0.006 (*p* = 4.5E−172)	0.554 ± 0.013 (*p* = 8.5E−193)
	Peptide	MTDsite	0.065 ± 0.007 (*p* = 3.3E−167)	0.096 ± 0.009 (*p* = 2.2E−177)	0.090 ± 0.007 (*p* = 9.8E−161)	0.591 ± 0.012 (*p* = 6.3E−178)
	ADP	TargetS	0.158 ± 0.018 (*p* = 1.2E−69)	0.230 ± 0.021 (*p* = 1.5E−80)	0.180 ± 0.020 (*p* = 4.2E−53)	0.656 ± 0.014 (*p* = 7.6E−133)
	AMP	TargetS	0.153 ± 0.015 (*p* = 6.6E−83)	0.212 ± 0.018 (*p* = 1.6E−98)	0.169 ± 0.013 (*p* = 2.4E−81)	0.683 ± 0.012 (*p* = 1.1E−118)
	ATP	TargetS	0.186 ± 0.017 (*p* = 4.9E−39)	0.248 ± 0.020 (*p* = 5.5E−66)	0.200 ± 0.019 (*p* = 1.9E−30)	0.677 ± 0.014 (*p* = 1.7E−118)
	GDP	TargetS	0.095 ± 0.012 (*p* = 2.4E−137)	0.161 ± 0.015 (*p* = 6.5E−138)	0.135 ± 0.011 (*p* = 5.4E−120)	0.690 ± 0.013 (*p* = 3.7E−109)
	GTP	TargetS	0.113 ± 0.013 (*p* = 3.6E−123)	0.187 ± 0.015 (*p* = 7.2E−122)	0.146 ± 0.010 (*p* = 1.2E−111)	0.657 ± 0.013 (*p* = 5.6E−139)
	Ca^{2+}	TargetS	0.089 ± 0.010 (*p* = 2.1E−146)	0.142 ± 0.011 (*p* = 1.8E−155)	0.112 ± 0.008 (*p* = 2.5E−143)	0.619 ± 0.010 (*p* = 3.2E−171)
		MIonic	0.111 ± 0.010 (*p* = 1.1E−130)	0.159 ± 0.013 (*p* = 3.3E−143)	0.122 ± 0.008 (*p* = 5.9E−136)	0.623 ± 0.011 (*p* = 1.1E−164)
	Fe^{2+}	MIonic	0.146 ± 0.011 (*p* = 1.1E−97)	0.197 ± 0.018 (*p* = 2.6E−110)	0.167 ± 0.009 (*p* = 3.2E−92)	0.537 ± 0.005 (*p* = 2.5E−225)
	Fe^{3+}	TargetS	0.110 ± 0.009 (*p* = 3.3E−135)	0.181 ± 0.012 (*p* = 8.2E−132)	0.143 ± 0.008 (*p* = 1.7E−119)	0.674 ± 0.011 (*p* = 2.6E−133)
		MIonic	0.142 ± 0.010 (*p* = 8.4E−105)	0.192 ± 0.015 (*p* = 1.9E−120)	0.151 ± 0.010 (*p* = 3.3E−108)	0.555 ± 0.006 (*p* = 1.7E−215)
	Mg^{2+}	TargetS	0.119 ± 0.012 (*p* = 1.2E−121)	0.207 ± 0.016 (*p* = 1.1E−109)	0.156 ± 0.011 (*p* = 1.8E−101)	0.664 ± 0.010 (*p* = 9.1E−144)
		MIonic	0.184 ± 0.013 (*p* = 1.4E−49)	0.247 ± 0.014 (*p* = 1.7E−77)	0.189 ± 0.012 (*p* = 4.1E−58)	0.690 ± 0.013 (*p* = 3.9E−109)
	Mn^{2+}	TargetS	0.124 ± 0.009 (*p* = 1.5E−124)	0.194 ± 0.014 (*p* = 8.3E−121)	0.154 ± 0.009 (*p* = 2.9E−106)	0.675 ± 0.011 (*p* = 3.3E−132)
		MIonic	0.183 ± 0.009 (*p* = 2.1E−57)	0.242 ± 0.013 (*p* = 1.1E−85)	0.187 ± 0.010 (*p* = 1.2E−66)	0.620 ± 0.008 (*p* = 1.7E−178)

(*continued on next page*)

Table 1 (*continued*)

Type	Prediction target	Method name	Sensitivity		F1max	AUROC
			Number of putative MLBRs equals number of native MLBRs	FPR = 5%		
	Zn^{2+}	TargetS	0.090 ± 0.008 (p = 7.4E−150)	0.138 ± 0.011 (p = 1.5E−155)	0.118 ± 0.007 (p = 1.9E−142)	0.627 ± 0.010 (p = 2.7E−167)
		MIonic	0.128 ± 0.008 (p = 1.5E−122)	0.177 ± 0.013 (p = 6.0E−133)	0.135 ± 0.007 (p = 3.5E−129)	0.596 ± 0.007 (p = 1.6E−194)
	Co^{2+}	MIonic	0.147 ± 0.009 (p = 5.8E−104)	0.194 ± 0.013 (p = 2.7E−121)	0.150 ± 0.008 (p = 4.6E−112)	0.586 ± 0.008 (p = 2.9E−197)
	Cu^{2+}	MIonic	0.123 ± 0.009 (p = 4.4E−124)	0.162 ± 0.014 (p = 1.6E−138)	0.137 ± 0.009 (p = 1.8E−123)	0.524 ± 0.004 (p = 3.8E−232)
	Po_4^{3-}	MIonic	0.215 ± 0.014 (p = 7.2E−05)	0.295 ± 0.017 (p = 2.2E−18)	0.220 ± 0.013 (p = 9.3E−11)	0.657 ± 0.010 (p = 3.9E−149)
	So_4^{2-}	MIonic	0.164 ± 0.015 (p = 1.2E−69)	0.241 ± 0.016 (p = 1.4E−80)	0.180 ± 0.013 (p = 3.1E−68)	0.653 ± 0.009 (p = 8.5E−153)
	Carbohydrate	MTDsite	0.061 ± 0.007 (p = 3.5E−169)	0.090 ± 0.010 (p = 1.8E−178)	0.085 ± 0.007 (p = 2.4E−163)	0.576 ± 0.012 (p = 4.2E−186)
	Heme	TargetS	0.063 ± 0.011 (p = 6.5E−158)	0.084 ± 0.013 (p = 6.5E−175)	0.093 ± 0.016 (p = 1.1E−156)	0.644 ± 0.009 (p = 3.6E−162)
	Vitamin	TargetVita	0.094 ± 0.010 (p = 4.7E−143)	0.130 ± 0.010 (p = 2.7E−161)	0.102 ± 0.008 (p = 3.2E−149)	0.585 ± 0.011 (p = 2.1E−183)
Meta predictor	MLBRs	$MinMax_{average}$	0.160 ± 0.010 (p = 3.1E−86)	0.222 ± 0.016 (p = 2.3E−95)	0.167 ± 0.010 (p = 3.2E−90)	0.720 ± 0.013 (p = 1.9E−74)
		$Map_{average}$	0.150 ± 0.014 (p = 1.4E−87)	0.239 ± 0.017 (p = 3.5E−79)	0.179 ± 0.011 (p = 7.9E−74)	0.730 ± 0.013 (p = 3.1E−61)
		$Binary_{average}$	0.137 ± 0.014 (p = 8.0E−99)	0.227 ± 0.018 (p = 5.2E−87)	0.171 ± 0.012 (p = 3.4E−82)	0.723 ± 0.014 (p = 3.5E−69)
Machine learning	MLBRs	Baseline regression	0.164 ± 0.015 (p = 2.3E−69)	0.241 ± 0.021 (p = 6.0E−70)	0.179 ± 0.015 (p = 1.5E−65)	0.702 ± 0.015 (p = 5.0E−92)
		MERIT	**0.223 ± 0.015**	**0.319 ± 0.019**	**0.233 ± 0.015**	**0.773 ± 0.011**

TargetVita,[46] DisoRDPbind,[58] SCRIBER,[31] MTDsite,[67] ISPRED-SEQ,[38] MucLiPred,[57] and MIonic[51] that collectively predict interactions with 22 types of ligands (proteins, peptides, DNA, RNA, ATP, ADP, AMP, GDP, GTP, Ca^{2+}, Mg^{2+}, Mn^{2+}, Fe^{2+}, Fe^{3+}, Zn^{2+}, Co^{2+}, Cu^{2+}, Po_4^{3-}, So_4^{2-}, carbohydrate, heme, and vitamins). We apply them in two alternative ways: by using their outputs directly to predict MLBRs, and by developing three meta-predictors that aim to identify residues predicted to bind multiple different ligands. The generation of the meta-predictions consists of two steps: normalize the putative propensities generated by the nine tools and calculate the meta-predictions from these normalized values. We formulate the three meta-predictors by considering three different normalizations, which we define in detail in the Supplement. They include a simple min–max normalization ($MinMax_{average}$ meta-predictor), distribution mapping-based normalization ($Map_{average}$ meta-predictor), and a normalization that uses the binary state predictions ($Binary_{average}$ meta-predictor). In the second step, we use the best predictor for a given ligand type which we select based on the highest AUROC on the test dataset for that ligand (Table 1). We average the two highest normalized propensities across the predictions for the 22 ligand types. This is motivated by the fact that MLBRs must interact with at least two different ligand types.

Datasets

We develop training, validation and test datasets with annotations of MLBRs and residues that interact with a single ligand type using BioLiP2. This resource provides access to annotations of residues that interact with biologically-relevant ligands (including peptides, proteins, nucleotides, nucleic acids and relevant small molecules) that are extracted from high-resolution structures (below 3 Å) of protein–ligand complexes that we collected from the Protein Data Bank.[68] We follow procedures in related works to construct these datasets.[24,31,69] We use the training dataset to compute the predictor, validation dataset to parametrize the trained model and ensure that it does not overfit the training dataset (i.e., we select parameters that maximize performance on the validation dataset), and test dataset to compare the trained model with alternative solutions (i.e., we exclude the test set

during the model design process). We also appropriately ensure that proteins in these datasets share low, below 25% sequence similarity to be consistent with the published studies. We further detail the dataset collection in the Supplement. Supplementary Table S1 summarizes datasets, which are available at http://biomine.cs.vcu.edu/servers/MERIT/. Supplementary Table S2 provides the breakdown of the number of binding residues and proteins for the ligands that interact with at least 100 amino acids across the three datasets. Moreover, Supplementary Figure S1A shows the breakdown of the binding residues according to the number of ligand types they interact with. About 60% of binding residues interact with one ligand type, 25% with two ligand types, <1% with over 5 ligands, and the maximal number of interacting ligand types is 9.

Assessment of predictive performance

The current ligand binding predictors, meta-predictors that we formulate, and MERIT generate two types of outputs: real-valued propensity for multi-ligand binding and a binary state (MLBR vs. nonMLBR). The binary states are typically generated from the propensities using a threshold, where residues with propensities $\geq$threshold are predicted as MLBRs, and otherwise as non-MLBRs. We assess the predicted propensities with the popular area under the receiver operating characteristic curve (AUROC) metric and we use sensitivity and F1max to assess the binary predictions. Moreover, motivated by related studies,[31,32,44,64–66] we evaluate cross-predictions (single ligand binding residues predicted as MLBRs) and over-predictions (non-binding residues predicted as MLBRs) using several metrics, CPRratio and OPRratio for the binary state predictions and the area under the cross-prediction curve (AUCPC) and the area under the over-prediction curve (AUOPC) for the putative propensities. We define these metrics and how they are computed in the Supplement.

As part of comparative analysis, we perform statistical significance tests to assess whether differences between the proposed method and other predictors are robust across a collection of diverse test sets. To do that, we compare 100 results collected for randomly picked subsets of 50% of the test proteins. We use the Anderson-Darling test at 0.05 significance to check whether the corresponding measurements are normal. For normal data, we use the student *t*-test, and otherwise we apply the Wilcoxon rank-sum test. We assume that differences are statistically significant if p-value < 0.01.

MERIT model

MERIT aims to accurately predict MLBRs (high AUROC) and minimize the cross-predictions (low

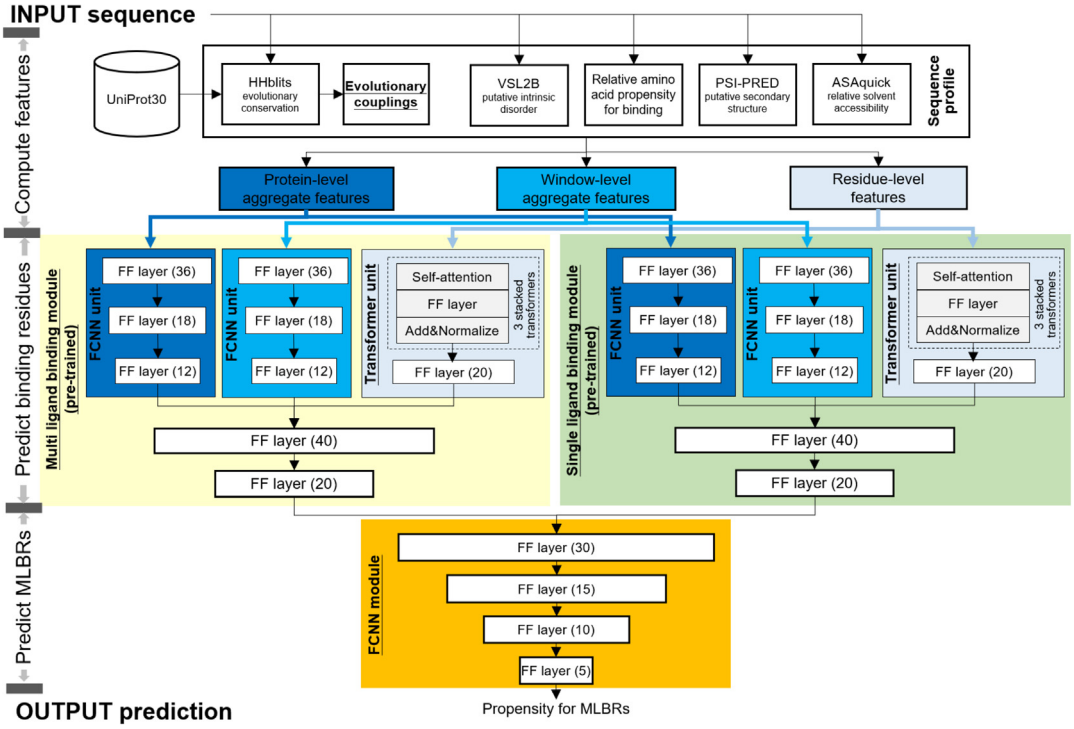

Figure 1. The architecture of MERIT. FCNN (fully connected feed-forward neural network); FF (feed-forward); numbers in the round brackets define the number of neurons used for a given network layer. Underlined text identifies major innovations.

AUCPC). It makes predictions in three main steps (Figure 1): (1) compute features; (2) predict binding residues using features; and (3) predict MLBRs.

In the first step, we use the input sequence to derive a sequence matrix which we utilize to compute a custom-designed feature set that is suitable for processing by the deep neural network model. The sequence matrix is composed of columns that correspond to the amino acids in the sequence and multiple rows that include the sequence itself; multiple sequence alignment generated by HHblits[70] using the UniProt30 dataset; relative amino acids propensity for binding with ligands that we compute using the training dataset with the Composition Profiler program,[71] putative intrinsic disorder produced by VSL2B,[72] which was recently shown to provide accurate results in the context of identifying disordered binding residues[73]; secondary structure predicted by the popular PSI-PRED[74]; and relative solvent accessibility predicted using fast and accurate ASAquick.[75] We use this sequence matrix to compute three distinct feature subsets (shown in Figure 1 by backgrounds in three shades of blue): protein-level, window-level, and the typically used residue-level features. The protein-level features aggregate the data in the matrix over the entire sequence to quantify an overall bias of a given protein to include MLBRs. The window-level features summarize information from the matrix in a sliding sequence window, which is motivated by the fact that binding residues cluster together in the sequence, i.e., some sequence segments include high density of binding residues vs. other that are devoid of binding residues. Finally, the residue-level features quantify inputs for individual amino acids including the predicted residue and its neighboring residues in the sequence. Development of these three distinct feature sets is motivated by their use in the flDPnn method, which produced accurate prediction of intrinsic disorder and disordered binding regions[76] in the recently completed CAID1[77] and CAID2[78] experiments. We enumerate and define individual features in the Supplement.

In step two, we input these three feature sets into a custom-designed deep neural network model that is composed of two modules that together predict ligand binding residues: the multi-ligand binding module (yellow background in Figure 1) and the single-ligand binding module (green background in Figure 1). Each module consists of three units that process the corresponding three feature sets. We input the protein-level and the window-level features into two fully connected feedforward neural networks (FCNNs) since these features are not ordered. These features are calculated by averaging the information from individual rows in the sequence matrix for a given sequence window and over the entire sequence. In other words, use of other, more complex network types is not warranted since these two feature sets quantify different types of input characteristics that do not follow a spatial or sequence arrangements. The two FCNNs include five layers where the last two layers are shared and the layers are gradually reduced in size to compact the resulting latent feature spaces. We process the residue-level features with a unit composed of three stacked transformers since these features follow the sequence order (i.e., they represent a short sequence window). The transformers are relatively fast to train and capable of exploiting the sequence order. Each transformer consists of a self-attention unit, a feedforward layer, and a normalization layer. The feature space generated by the transformer unit is merged with the spaces generated for the other two feature sets using the last two feedforward layers.

In the third step, we refine predictions of MLBRs by using a 4-layer FCNN that merges latent feature spaces generated in the second step by the single-ligand and the multi-ligand binding modules. The objective is to improve accuracy of the MLBR predictions by reducing the cross-predictions with the help of the single-ligand binding predictions. As in step two, the subsequent layers are gradually smaller to reduce the resulting latent feature space to a single output neuron that generates the propensity for multi-ligand binding. We use the ReLU activation function for all nodes in the feedforward layers, except for the output node where we apply the sigmoid function to properly scale the output propensities. We train this architecture using Pytorch with the popular Adam optimizer on the training dataset. We set the learning rate and the batch size to 0.001 and 256, respectively.

The MERIT's model includes five key innovations (Figure 1 identifies them using underline): (1) calculation of the evolutionary couplings in the sequence matrix; (2) use of transformer modules to process the residue-level features; (3) application of the transfer learning to develop the two modules in the second step; (4) design of a multi-step architecture where the third step is used to refine and improve prediction of the MLBRs from the multi ligand binding module in step two; and (5) development and use of an advanced loss function to train the network. The first two innovations aim to maximize overall performance of the MLBR predictions while the motivation for the latter three innovations is to minimize the cross-predictions.

The evolutionary couplings are evolutionarily conserved pairwise amino acid associations that are typically calculated from multiple sequence alignments. They link residues that coevolved together and which correspondingly may share similar purpose. One of these purposes could be

ligand binding,[79] in which case all residues that interact with the same ligand should be evolutionarily coupled. Thus, couplings should help with finding MLBRs that on average should have more couplings since they are included in multiple sets of binding ("coupled") residues that interact with different ligands when compared with the single-ligand binding residues. We compute the evolutionary coupling sores from the multiple sequence alignments generated with HHblits[70] against the UniClust30 database. We detail their calculation in the Supplement.

The transformer units are particularly suitable to model latent feature spaces for inputs that are sequence-ordered.[44,80,81] We posit that they can be effectively combined with the FCNN to improve predictive performance.

We use transfer learning to train the two modules, the multi-ligand binding and the single ligand binding, in the step two of our architecture (Figure 1). We pre-train the single ligand binding module using the training set annotated with the single ligand binding residues (i.e., we set MLBRs as negatives). We similarly pre-train the multi ligand binding module using the training set annotated with the MLBRs (i.e., we set the single ligand binding residues as negatives). We pre-trained these two modules separately using the cross-entropy loss function with the objective to maximize the AUROC value on the validation dataset. We also ensure that the difference in AUROC between the training and validation datasets is small (<0.03) to prevent potential overfitting into the training dataset. These modules specifically target predictions of the two distinct types of binding residues and use them to refine prediction of MLBRs by reducing the cross-predictions (prediction of MLBRs for the putative single ligand binding residues). We do that by freezing the two pre-trained networks when we subsequently train MERIT (i.e., we "transfer" the pre-trained networks into the MERIT model).

Loss function

The default loss function for training the network is the cross-entropy. Its main drawback is the inability to effectively differentiate between different types of errors, i.e., mispredicting single ligand binding for multi ligand binding, non-binding for multi ligand binding, and multi ligand binding for non-binding. We substitute the cross-entropy with the focal loss function[82,83] for training the MERIT model with the pre-trained single ligand binding and multi ligand binding modules. We reformulate the focal loss function to suit our prediction by introducing α and β coefficients that allow us to balance cross-predictions vs. over-predictions, with the objective to minimize cross-predictions while maintaining an overall high predictive performance:

$$\begin{aligned} Focal\ \ Loss = &-(1-pred_{multi})^{r} label_{multi} log\left(_{predmulti}\right) \\ &-\alpha \cdot pred_{multi}{}^{r} label_{non} log(1-pred_{multi}) \\ &-\beta \cdot pred_{multi}{}^{r} label_{single} log(1-pred_{multi}) \end{aligned}$$

where $pred_{multi}$ is the predicted MLBR propensity; $label_{multi}$, $label_{non}$, and $label_{single}$ stand for the multi ligand binding, non-binding and single ligand binding annotations, respectively; r is set to 2 to reduce impact of the errors for well-predicted MLBRs, i.e., difference between the label 1 that denotes MLBR vs. the predicted propensity is small; and α and β are parameters that quantify contribution of the over-prediction and cross-prediction errors, respectively. We fine-tune α and β using the training and validation datasets. We consider α = {0.5, 1, 2, 4} and β = {1, 5, 8, 10, 12, 15, 20}, where values of β are larger since we aim to reduce the cross-predictions. We run a grid search over these parameters and select α = 2 and β = 10 that provides high value of AUROC and low value of AUCPC on the validation dataset for the model trained on the training dataset using MLBRs as labels. Moreover, like for the pre-training, we ensure that the difference in AUROC between training and validation datasets is below 0.03 to avoid overfitting.

Results

Ablation analysis

We empirically test contributions of the five innovations to the predictive performance of MERIT. We perform an ablation analysis where we remove these innovations, one at the time, which leads to the following five configurations: (1) exclude evolutionary couplings; we re-train the model but when the evolutionary couplings-based inputs are removed; (2) replace transformer unit; we re-train the model where the transformer unit is replaced by a FCNN unit; (3) remove transfer learning; we re-train the model without pre-training the modules in step 2; in other words, we train the entire network at once; (4) remove step 3; we reduce this network to the multi-ligand binding module from step 2; and (5) use default loss function; we re-train using the default cross-entropy loss function.

Figure 2 compares performance of the MERIT model (dark blue bars) with the five ablation setups. We find that each of the five innovations provides statistically significant improvements to MERIT (p-values < 0.01) when considering the overall performance (AUROC in Fig 2A and F1max in Fig 2B), cross-predictions (AUCPC in Figure 2C) and over-predictions (AUOPC in Figure 2D). The exclusion of the step 3 of the model (yellow bars in Figure 2) leads to the largest drop in AUROC (from 0.773 to 0.736) and the biggest increase in the cross-predictions (from 0.303 to 0.383 in AUCPC) and in the over-predictions (from 0.223 to 0.257 in AUOPC). This

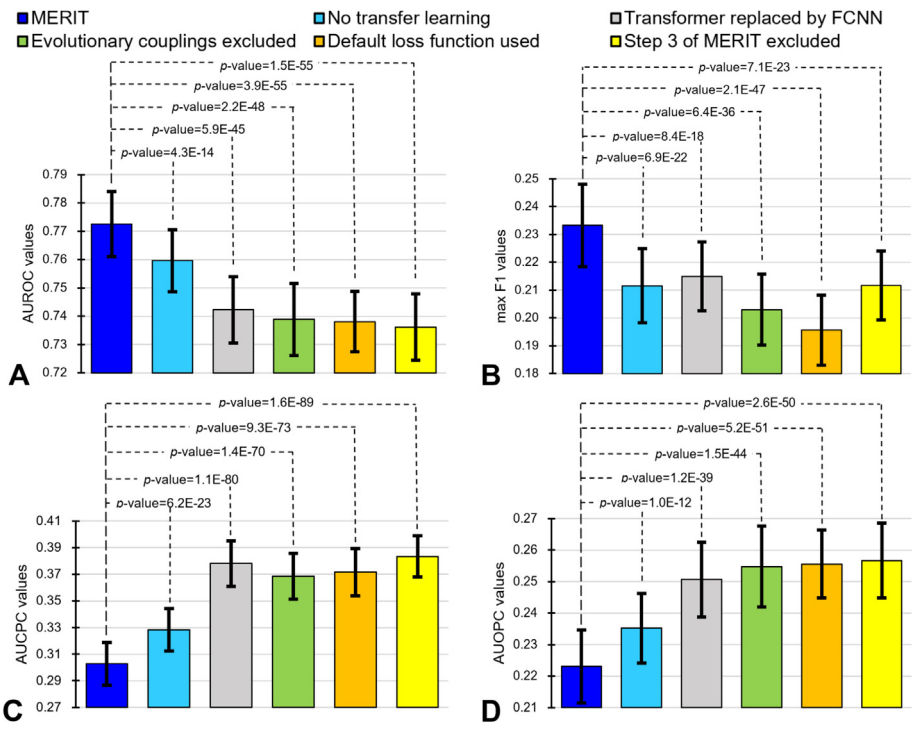

Figure 2. Ablation analysis on the test dataset that compares MERIT with its five variants where one of the five innovations is removed. We quantify the predictive performance with AUROC (panel A; higher values are better), F1max values (panel B; higher values are better), AUCPC values (panel C; lower values are better), and AUOPC values (panel D; lower values are better). We show averages (bars) with the corresponding standard deviations (error bars) based on the 100 tests; see the last paragraph in the "Assessment of predictive performance" sub-section. We report the corresponding *p*-values at the top of the bars by comparing against the complete MERIT model.

demonstrates the value of using the single ligand binding module in step 2 to limit the number of false positives (incorrectly predicted MLBRs), particularly considering the big drop in the cross-predictions. We also find that use of the optimized focal loss function (orange bars in Figure 2), evolutionary couplings (green bars in Figure 2), and transformer units (gray bars in Figure 2) also provide large reductions in cross-predictions and overpredictions, leading to the substantial increases in AUROC. The smallest magnitude of the change is associated with the use of the transfer learning (light blue bars in Figure 2); however, this innovation still provides statistically significant improvements over each of the four metrics.

Comparative assessment

Table 1 compares MERIT with the nine selected popular and/or recent predictors of DNA, RNA, protein, and small ligand binding residues, the three meta-predictors use their results to predict MLBRs, and a baseline predictor that applies the same inputs as MERIT and a simple logistic regression on the test dataset. The results produced by the current tools that predict interactions with specific ligand types provide modest levels of performance when applied to find MLBRs (AUROC $\leq$ 0.69). The meta- predictors provide more accurate results, with the best $Map_{average}$ model that secures AUROC of 0.730, F1max of 0.18, and sensitivity at FPR = 5% of 0.24. This is because they predict MLBRs by combining results across multiple ligands. This best meta-predictor outperforms the baseline regression model that obtains AUROC of 0.702, F1max of 0.18 and sensitivity at FPR = 5% of 0.24, which in turn improves over the predictions of the individual ligand types. MERIT generates the most accurate results, with AUROC of 0.77, F1max of 0.23 and sensitivity at FPR = 5% of 0.32, significantly outperforming all other methods (*p*-value < 0.01). The sensitivity reveals that MERIT secures 0.32/0.05 = 6.4 times higher true positive rate compared to its false positive rate. This and the AUROC > 0.75 suggest that MERIT generates relatively accurate predictions of MLBRs. The corresponding ROC curves (Supplementary Figure S2A) shows a wide margin between the MERIT's curve and the curves of the other methods. Comparison of MERIT with the baseline regression quantifies the overall contribution of using the multi-step deep neural network model. The differences have large magnitude (AUROC of 0.77 vs. 0.70; F1max of

0.23 vs. 0.18; sensitivity at FPR = 5% of 0.32 vs. 0.24) and they are statistically significant (p-value < 0.01).

Supplementary Figure S1B analyzes differences in the predictive performance, quantified with sensitivity, across MLBRs that interact with different number of distinct ligands. We compare MERIT, the most accurate $Map_{average}$ meta-predictor, and the baseline regressor. We consider MLBRs that bind 2, 3, 4, 5 and >5 ligands; we combine MLBRs that bind 6 and more distinct ligands due to relatively small sample size (i.e., <50 MLBRs for each number of ligands). Supplementary Figure S1B shows that MERIT's performance increases as the number of ligands grows, from 0.21 for MLBRs that interact with two ligand types to 0.29 for >5 ligands; the overall MERIT's sensitivity is 0.22 (Table 1). To compare, the performance of the best meta-predictor and the baseline regressor does not change in the function of the number of interacting ligands (Supplementary Figure S1B), staying around their overall sensitivity of 0.15 and 0.16 (Table 1), respectively. However, we note that these results should be considered accurate given that the overall fraction of MLBRs in the test dataset is 0.032.

We also performed this analysis for ligands that we grouped into five broad classes: (1) nucleic acids; (2) proteins and peptides; (3) metal ions; (4) nucleotides; and (5) other ligands. Supplementary Figure S1C compares predictive performance of MERIT against the best meta-predictor and baseline regressor for MLBRs that interact with ligands that belong to one ligand class, two classes and three ligand classes. There are only a few MLBRs that interact with four ligand classes and none that interact across the five classes, which is why exclude these cases from our analysis. Similar to the other result, MERIT's performance improves for MLBRs that cover more ligand classes. The two other methods follow similar trend but their predictions are consistently less accurate than MERIT's predictions. The consistent increase in the MERIT's performance as the number of interacting ligands and ligand classes grows can be explained by the use of the evolutionary couplings, which are not dependent on similarity between ligands that underlies the ligand classes, and transformers that can adequately take advantage of these inputs.

Altogether, we find that MERIT offers accurate predictions of MLBRs that outperform the current and alternative solutions, and makes modestly more accurate predictions of MLBRs that bind larger number of ligand types and classes.

Evaluation of the cross- and over-predictions

MLBRs and single ligand binding residues share some characteristics (e.g., they should be evolutionarily conserved and most of them should be localized on protein surface) when contrasted with the non-binding residues. Thus, MLBRs should be harder to differentiate from the other binding residues (cross-predictions) compared to the non-binding residues (over-predictions). This is why we expect that in relative terms the cross-predictions are likely to dominate false positives when compared to the over-predictions. Supplementary Table S3 compares the cross-prediction and over-prediction errors. The corresponding cross-prediction curves and over-prediction curves are in Supplementary Figures S2B and S2C, respectively.

We find that the cross-predictions for all methods, except MERIT, are at the near-random levels, with AUCPCs > 0.4. As expected, they are higher than the over-predictions, where the meta-predictors and the baseline regression secure AUOPCs of around 0.3. Similar observations are true when using the CPRratio and OPRratio values. CPRratios are lower than OPRratios and CPRratios for many of the methods are at near random levels, i.e., values are close to 1. OPRratios for the meta-predictors and regression are in the 5.2–6.3 range, suggesting that these solutions are 5–6 times better than random in the context of the over-predictions. The poor cross-prediction performance can be explained by the fact that the meta-predictors and the logistic regression baseline were not optimized to exclude single ligand binding residues among their predictions of MLBRs.

Importantly, MERIT secures relatively low rates of cross-predictions and over-predictions, which are significantly better than the results of all other methods (p-value < 0.01; Supplementary Table S3). It secures AUCPC of 0.303, which is reasonably good, and a low AUOPC of 0.223. The cross-prediction and over-prediction curves of MERIT are better/lower than the curves of the other methods by a wide margin (Supplementary Figures S2B and S2C). The MERIT's CPRratio and OPRratio measured when predicting the correct number of MLBRs (i.e., numbers of putative and native MLBRs are equal) reveal that our tool is 3.8 and 9.6 times better than a random predictor when considering the cross-predictions and over-predictions, respectively. These results suggest that MERIT performs very well in the context of the over- and cross-predictions, which can be attributed to the several innovations which specifically aim to reduce these errors.

Assessment on the single ligand binding proteins

Using the clustered sequences from the fourth step of the dataset selection process described in the Supplement, we randomly select 200 proteins that have single ligand binding residues and no MLBRs. This dataset shares low, below 25% similarity with the training and validation datasets,

and is available on the MERIT page at http://biomine.cs.vcu.edu/servers/MERIT/. We use it to assess whether MERIT improves over the other methods for these challenging "negative" proteins. Supplementary Figure S3 compares performance of MERIT, the three meta-predictors of MLBRs and the baseline regression model on this dataset. Supplementary Figure S3A, which quantifies the false positive rates (fraction of residues incorrectly predicted as MLBRs) on this dataset shows that MERIT generates substantially better results than the meta-predictors and the baseline. These improvements are consistent for the cross-predictions (Supplementary Figure S3B) and over-predictions (Supplementary Figure S3C). Using the threshold for the generation of the binary states that is calibrated to produce the correct number of MLBRs on the test dataset (i.e., the numbers of the predicted and the native MLBRs are equal), we find that MERIT generates 2.8% of residues as MLBRs on this dataset with the single ligand binding residues. To compare, the baseline regression and the best meta-predictor predict 4.7% and 4.9% residues as MLBRs, respectively. As expected, the cross-prediction rate is higher than the over-prediction rate; however, MERIT predicts 11% of the single ligand binding residues as MLBRs (cross-predictions), which corresponds to on average two cross-predicted residues per protein, compared to 14% for the baseline and 17% for the best meta-predictor.

A modest amount of cross-predictions could be explained because annotations of binding residues are potentially incomplete, i.e., interactions with some ligands might be missing in the source databases, which is true across all datasets used to develop predictors of the ligand binding residues. Thus, some of the single ligand binding residues could in fact bind multiple ligand types, suggesting that some cross predictions could actually be correct predictions. Correspondingly, MERIT shows higher rate of MLBR predictions for the single binding residues (Supplementary Figure S3B) than for the non-binding residues (Supplementary Figure S3C). We posit that its cross-predictions could be investigated to potentially uncover these missing/not-yet-annotated interactions. However, this analysis is beyond the scope of this method/web server article. Altogether, we argue that MERIT's predictions generalize well for the single ligand binding proteins, consistently improving over the alternatives.

Runtime analysis

We use a random selection of 100 proteins from the test dataset to investigate a relation between the MERIT runtime and sequence length. We sort the sequences in the ascending order by their length into four equally sized subsets. Supplementary Figure S4 plots the median per-protein runtimes measured in minutes against the median sequence length for the four protein sets. The overall median per protein runtime for MERIT is 8.9 min, with the first and third quartiles of 3.7 and 16.5 min, respectively. The runtime grows with the sequence length for shorter proteins and then the growth saturates for proteins with 500 or more residues. The main factor that drives runtime is the calculation of evolutionary couplings, which is done for conserved residues. The leveling of the runtime for the longer proteins is likely due to a drop in the fraction of conserved residues, relative to the sequence length, i.e., a larger fraction of residues is conserved for shorter chains compared to the fraction for longer proteins.

Web server

MERIT is available as a free web server at http://biomine.cs.vcu.edu/servers/MERIT/. We provide further details in the Supplement. The server page also provides access to the training, validation and test datasets, which can be used to facilitate future research in the area of analysis and prediction of MLBRs.

CRediT authorship contribution statement

Jian Zhang: Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation. **Sushmita Basu:** Visualization, Validation, Software, Resources. **Fuhao Zhang:** Validation, Resources, Methodology, Formal analysis. **Lukasz Kurgan:** Writing – original draft, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Funding

This work was supported in part by the Natural Science Foundation of Henan [242300421410] and the Nanhu Scholars Program for Young Scholars of the Xinyang Normal University to J.Z; the Northwest A&F University [Z1090124074] to F. Z; and the National Science Foundation [2146027, 2125218] and Robert J. Mattauch Endowed Chair funds to L.K.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary material

Supplementary material to this article can be found online at https://doi.org/10.1016/j.jmb.2024.168872.

Received 13 September 2024;
Accepted 15 November 2024;
Available online xxxx

Keywords:
protein-ligand interactions;
protein function;
prediction;
transformer;
web server

References

1. Ma, B., Shatsky, M., Wolfson, H.J., Nussinov, R., (2002). Multiple diverse ligands binding at a single protein site: a matter of pre-existing populations. *Protein Sci.* **11**, 184–197.
2. Nobeli, I., Favia, A.D., Thornton, J.M., (2009). Protein promiscuity and its implications for biotechnology. *Nature Biotechnol.* **27**, 157–167.
3. Oldfield, C.J., Uversky, V.N., Dunker, A.K., Kurgan, L., (2019). Introduction to intrinsically disordered proteins and regions. In: Salvi, N. (Ed.), *Intrinsically Disordered Proteins*. Academic Press, pp. 1–34.
4. Lieutaud, P., Ferron, F., Uversky, A.V., Kurgan, L., Uversky, V.N., Longhi, S., (2016). How disordered is my protein and what is its disorder for? A guide through the "dark side" of the protein universe. *Intrinsically Disord. Proteins* **4**, e1259708
5. Uversky, V.N., (2013). Intrinsic disorder-based protein interactions and their modulators. *Curr. Pharm Design.* **19**, 4191–4213.
6. Uversky, V.N., (2016). p53 proteoforms and intrinsic disorder: an illustration of the protein structure-function continuum concept. *Int. J. Mol. Sci.* **17**
7. Oldfield, C.J., Meng, J., Yang, J.Y., Yang, M.Q., Uversky, V.N., Dunker, A.K., (2008). Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. *BMC Genomics* **9** (Suppl 1), S1.
8. Rutigliano, G., Braunig, J., Del Grande, C., Carnicelli, V., Masci, I., Merlino, S., et al., (2019). Non-functional trace amine-associated receptor 1 variants in patients with mental disorders. *Front. Pharmacol.* **10**, 1027.
9. Ivetac, A., McCammon, J.A., (2011). Molecular recognition in the case of flexible targets. *Curr. Pharm. Des.* **17**, 1663–1671.
10. Hudson, W.H., Ortlund, E.A., (2014). The structure, function and evolution of proteins that bind DNA and RNA. *Nature Rev. Mol. Cell Biol.* **15**, 749–760.
11. James, L.C., Tawfik, D.S., (2003). Conformational diversity and protein evolution–a 60-year-old hypothesis revisited. *Trends Biochem. Sci* **28**, 361–368.
12. Wang, C., Karpowich, N., Hunt, J.F., Rance, M., Palmer, A. G., (2004). Dynamics of ATP-binding cassette contribute to allosteric control, nucleotide binding and energy transduction in ABC transporters. *J. Mol. Biol.* **342**, 525–537.
13. Choudhary, S., Lopus, M., Hosur, R.V., (2022). Targeting disorders in unstructured and structured proteins in various diseases. *Biophys. Chem.* **281**
14. Biesaga, M., Frigole-Vivas, M., Salvatella, X., (2021). Intrinsically disordered proteins and biomolecular condensates as drug targets. *Curr. Opin. Chem. Biol.* **62**, 90–100.
15. Du, X., Li, Y., Xia, Y.L., Ai, S.M., Liang, J., Sang, P., et al., (2016). Insights into Protein-Ligand Interactions: Mechanisms, Models, and Methods. *Int. J. Mol. Sci.* **17**
16. UniProt, C., (2023). UniProt: the universal protein knowledgebase in 2023. *Nucleic Acids Res.* **51**, D523–D531.
17. Haft, D.H., Badretdin, A., Coulouris, G., DiCuccio, M., Durkin, A.S., Jovenitti, E., et al., (2023). RefSeq and the prokaryotic genome annotation pipeline in the age of metagenomes. *Nucleic Acids Res.*.
18. Zhang, Y., Bao, W., Cao, Y., Cong, H., Chen, B., Chen, Y., (2022). A survey on protein-DNA-binding sites in computational biology. *Brief. Funct. Genomics* **21**, 357–375.
19. Zhang, J., Kurgan, L., (2018). Review and comparative assessment of sequence-based predictors of protein-binding residues. *Brief. Bioinform.* **19**, 821–837.
20. Wang, K., Hu, G., Wu, Z., Su, H., Yang, J., Kurgan, L., (2020). Comprehensive survey and comparative assessment of RNA-binding residue predictions with analysis by RNA type. *Int. J. Mol. Sci.* **21**
21. Macari, G., Toti, D., Polticelli, F., (2019). Computational methods and tools for binding site recognition between proteins and small molecules: from classical geometrical approaches to modern machine learning strategies. *J. Comput. Aided Mol. Des.* **33**, 887–903.
22. Yan, J., Friedrich, S., Kurgan, L., (2016). A comprehensive comparative review of sequence-based predictors of DNA- and RNA-binding residues. *Brief. Bioinform.* **17**, 88–105.
23. Dhakal, A., McKay, C., Tanner, J.J., Cheng, J., (2022). Artificial intelligence in the prediction of protein–ligand interactions: recent advances and future directions. *Brief. Bioinform.* **23**, bbab476
24. Zhang, J., Ma, Z., Kurgan, L., (2019). Comprehensive review and empirical analysis of hallmarks of DNA-, RNA- and protein-binding residues in protein chains. *Brief. Bioinform.* **20**, 1250–1268.
25. Xia, Y., Xia, C.Q., Pan, X., Shen, H.B., (2021). GraphBind: protein structural context embedded rules learned by hierarchical graph neural networks for recognizing nucleic-acid-binding residues. *Nucleic Acids Res.* **49**, e51.
26. Aggarwal, R., Gupta, A., Chelur, V., Jawahar, C.V., Priyakumar, U.D., (2022). DeepPocket: ligand binding site detection and segmentation using 3D convolutional neural networks. *J. Chem. Inf. Model.* **62**, 5069–5079.
27. Yuan, Q., Chen, S., Rao, J., Zheng, S., Zhao, H., Yang, Y., (2022). AlphaFold2-aware protein-DNA binding site prediction using graph transformer. *Brief. Bioinform.* **23**
28. Xia, Y., Pan, X., Shen, H.B., (2023). LigBind: identifying binding residues for over 1000 ligands with relation-aware graph neural networks. *J. Mol. Biol.* **435**, 168091
29. Li, P., Liu, Z.P., (2023). GeoBind: segmentation of nucleic acid binding interface on protein surface with geometric deep learning. *Nucleic Acids Res.* **51**, e60.

30. Fang, Y.T., Jiang, Y., Wei, L.Y., Ma, Q., Ren, Z.X., Yuan, Q.M., et al., (2023). DeepProSite: structure-aware protein binding site prediction using ESMFold and pretrained language model. *Bioinformatics* **39**
31. Zhang, J., Kurgan, L., (2019). SCRIBER: accurate and partner type-specific prediction of protein-binding residues from proteins sequences. *Bioinformatics* **35**, i343–i353.
32. Zhang, J., Ghadermarzi, S., Kurgan, L., (2020). Prediction of protein-binding residues: dichotomy of sequence-based methods developed using structured complexes versus disordered proteins. *Bioinformatics* **36**, 4729–4738.
33. Zhang, F., Shi, W., Zhang, J., Zeng, M., Li, M., Kurgan, L., (2020). PROBselect: accurate prediction of protein-binding residues from proteins sequences via dynamic predictor selection. *Bioinformatics* **36**, i735–i744.
34. Li, Y., Golding, G.B., Ilie, L., (2021). DELPHI: accurate deep ensemble model for protein interaction sites prediction. *Bioinformatics* **37**, 896–904.
35. Hosseini, S., Ilie, L., (2022). PITHIA: protein interaction site prediction using multiple sequence alignments and attention. *Int. J. Mol. Sci.* **23**, 12814.
36. Zhang, F., Li, M., Zhang, J., Shi, W., DeepPRObind, K.L., (2023). Modular deep learner that accurately predicts structure and disorder-annotated protein binding residues. *J. Mol. Biol.* 167945.
37. Kang, Y., Xu, Y., Wang, X., Pu, B., Yang, X., Rao, Y., et al., (2023). HN-PPISP: a hybrid network based on MLP-Mixer for protein–protein interaction site prediction. *Brief. Bioinform.* **24**, bbac480
38. Manfredi, M., Savojardo, C., Martelli, P.L., Casadio, R., (2023). ISPRED-SEQ: deep neural networks and embeddings for predicting interaction sites in protein sequences. *J. Mol. Biol.* **435**, 167963
39. Hu, J., Li, Y., Zhang, M., Yang, X., Shen, H.-B., Yu, D.-J., (2016). Predicting protein-DNA binding residues by weightedly combining sequence-based features and boosting multiple SVMs. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **14**, 1389–1398.
40. Zhu, Y.-H., Hu, J., Song, X.-N., Yu, D.-J., (2019). DNAPred: accurate identification of DNA-binding sites from protein sequence by ensembled hyperplane-distance-based support vector machines. *J. Chem. Inf. Model.* **59**, 3057–3071.
41. Zhang, J., Basu, S., Kurgan, L., (2024). HybridDBRpred: improved sequence-based prediction of DNA-binding amino acids using annotations from structured complexes and disordered proteins. *Nucleic Acids Res.* **52**, e10.
42. El-Manzalawy, Y., Abbas, M., Malluhi, Q., Honavar, V., (2016). FastRNABindR: fast and accurate prediction of protein-RNA interface residues. *PLoS One* **11**, e0158445
43. Tang, Y., Liu, D., Wang, Z., Wen, T., Deng, L., (2017). A boosting approach for prediction of protein-RNA binding residues. *BMC Bioinf.* **18**, 47–58.
44. Zhang, F., Li, M., Zhang, J., Kurgan, L., (2023). HybridRNAbind: prediction of RNA interacting residues across structure-annotated and disorder-annotated proteins. *Nucleic Acids Res.* **51**, e25.
45. Yu, D.-J., Hu, J., Yang, J., Shen, H.-B., Tang, J., Yang, J.-Y., (2013). Designing template-free predictor for targeting protein-ligand binding sites with classifier ensemble and spatial clustering. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **10**, 994–1008.
46. Yu, D.-J., Hu, J., Yan, H., Yang, X.-B., Yang, J.-Y., Shen, H.-B., (2014). Enhancing protein-vitamin binding residues prediction by multiple heterogeneous subspace SVMs ensemble. *BMC Bioinf.* **15**, 1–14.
47. Zhang, J., Chai, H., Gao, B., Yang, G., Ma, Z., (2016). HEMEsPred: structure-based ligand-specific heme binding residues prediction by using fast-adaptive ensemble learning scheme. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **15**, 147–156.
48. Hu, J., Zheng, L.-L., Bai, Y.-S., Zhang, K.-W., Yu, D.-J., Zhang, G.-J., (2021). Accurate prediction of protein-ATP binding residues using position-specific frequency matrix. *Anal. Biochem.* **626**, 114241
49. Zhang, J., Zhou, F., Liang, X., Yang, G., (2022). SCAMPER: accurate type-specific prediction of calcium-binding residues using sequence-derived features. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **20**, 1406–1416.
50. Yuan, Q., Chen, S., Wang, Y., Zhao, H., Yang, Y., (2022). Alignment-free metal ion-binding site prediction from protein sequence through pretrained language model and multi-task learning. *Brief. Bioinform.* **23**, bbac444
51. Shenoy, A., Kalakoti, Y., Sundar, D., Elofsson, A., (2024). M-Ionic: prediction of metal-ion-binding sites from sequence using residue embeddings. *Bioinformatics* **40**, btad782
52. Wang, L., Huang, C., Yang, M.Q., Yang, J.Y., (2010). BindN+ for accurate prediction of DNA and RNA-binding residues from protein sequence features. *BMC Syst. Biol.* **4**, 1–9.
53. Yan, J., Kurgan, L., (2017). DRNApred, fast sequence-based method that accurately predicts and discriminates DNA- and RNA-binding residues. *Nucleic Acids Res.* **45**, e84.
54. Su, H., Liu, M., Sun, S., Peng, Z., Yang, J., (2019). Improving the prediction of protein-nucleic acids binding residues via multiple sequence profiles and the consensus of complementary methods. *Bioinformatics* **35**, 930–936.
55. Nie, W., Deng, L., (2022). TSNAPred: predicting type-specific nucleic acid binding residues via an ensemble approach. *Brief. Bioinform.* **23**, bbac244
56. Zhang, J., Chen, Q., Liu, B., (2023). iNucRes-ASSH: Identifying nucleic acid-binding residues in proteins by using self-attention-based structure-sequence hybrid neural network. *Proteins*.
57. Zhang, J., Wang, R., Wei, L., (2024). MucLiPred: multi-level contrastive learning for predicting nucleic acid binding residues of proteins. *J. Chem. Inf. Model.*.
58. Peng, Z., Kurgan, L., (2015). High-throughput prediction of RNA, DNA and protein binding regions mediated by intrinsic disorder. *Nucleic Acids Res.* **43**, e121-e
59. Qiu, J., Bernhofer, M., Heinzinger, M., Kemper, S., Norambuena, T., Melo, F., et al., (2020). ProNA2020 predicts protein–DNA, protein–RNA, and protein–protein binding proteins and residues from sequence. *J. Mol. Biol.* **432**, 2428–2443.
60. Zhang, F., Zhao, B., Shi, W., Li, M., Kurgan, L., (2022). DeepDISOBind: accurate prediction of RNA-, DNA- and protein-binding intrinsically disordered residues with deep multi-task learning. *Brief. Bioinform.* **23**
61. Sun, Z., Zheng, S., Zhao, H., Niu, Z., Lu, Y., Pan, Y., et al., (2021). To improve prediction of binding residues with DNA, RNA, carbohydrate, and peptide via multi-task deep

neural networks. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **19**, 3735–3743.
62. Yang, J., Roy, A., Zhang, Y., (2013). BioLiP: a semi-manually curated database for biologically relevant ligand-protein interactions. *Nucleic Acids Res.* **41**, D1096–D1103.
63. Zhang, C., Zhang, X., Freddolino Peter, L., Zhang, Y., (2023). BioLiP2: an updated structure database for biologically relevant ligand–protein interactions. *Nucleic Acids Res.*.
64. Zhang, J., Ghadermarzi, S., Katuwawala, A., Kurgan, L., (2021). DNAgenie: accurate prediction of DNA-type-specific binding residues in protein sequences. *Brief. Bioinform.* **22**
65. Zhang, J., Chen, Q., Liu, B., (2021). NCBRPred: predicting nucleic acid binding residues in proteins based on multilabel learning. *Brief. Bioinform.* **22**
66. Du, X., Hu, J., (2022). Deep multi-label joint learning for RNA and DNA-binding proteins prediction. *IEEE/ACM Trans. Comput. Biol. Bioinf.*, PP.
67. Sun, Z., Zheng, S., Zhao, H., Niu, Z., Lu, Y., Pan, Y., et al., (2022). To improve prediction of binding residues with DNA, RNA, carbohydrate, and peptide via multi-task deep neural networks. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **19**, 3735–3743.
68. wwPDB consortium, (2019). Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic Acids Res.* **47**, D520–D528.
69. Zhang, F., Li, M., Zhang, J., Kurgan, L., (2023). HybridRNAbind: prediction of RNA interacting residues across structure-annotated and disorder-annotated proteins. *Nucleic Acids Res.* **51**, e25
70. Remmert, M., Biegert, A., Hauser, A., Söding, J., (2011). HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nature Methods* **9**, 173–175.
71. Vacic, V., Uversky, V.N., Dunker, A.K., Lonardi, S., (2007). Composition Profiler: a tool for discovery and visualization of amino acid composition differences. *BMC Bioinf.* **8**, 211.
72. Peng, K., Radivojac, P., Vucetic, S., Dunker, A.K., Obradovic, Z., (2006). Length-dependent prediction of protein intrinsic disorder. *BMC Bioinf.* **7**, 208.
73. Katuwawala, A., Kurgan, L., (2020). Comparative assessment of intrinsic disorder predictions with a focus on protein and nucleic acid-binding proteins. *Biomolecules* **10**
74. Buchan, D.W., Jones, D.T., (2019). The PSIPRED protein analysis workbench: 20 years on. *Nucleic Acids Res.* **47**, W402–W407.
75. Faraggi, E., Zhou, Y.Q., Kloczkowski, A., (2014). Accurate single-sequence prediction of solvent accessible surface area using local and global features. *Proteins* **82**, 3170–3176.
76. Hu, G., Katuwawala, A., Wang, K., Wu, Z., Ghadermarzi, S., Gao, J., et al., (2021). flDPnn: accurate intrinsic disorder prediction with putative propensities of disorder functions. *Nature Commun.* **12**, 4438.
77. Necci, M., Piovesan, D., Predictors, C., DisProt, C., Tosatto, S.C.E., (2021). Critical assessment of protein intrinsic disorder prediction. *Nature Methods* **18**, 472–481.
78. Conte, A.D., Mehdiabadi, M., Bouhraoua, A., Miguel Monzon, A., Tosatto, S.C.E., Piovesan, D., (2023). Critical assessment of protein intrinsic disorder prediction (CAID) – results of round 2. *Proteins*.
79. Schelling, M., Hopf, T.A., Rost, B., (2018). Evolutionary couplings and sequence variation effect predict protein binding sites. *Proteins* **86**, 1064–1074.
80. Hong, Y., Song, J., Ko, J., Lee, J., Shin, W.H., (2022). S-Pred: protein structural property prediction using MSA transformer. *Sci. Rep.* **12**, 13891.
81. Huang, B., Fan, T., Wang, K., Zhang, H., Yu, C., Nie, S., et al., (2023). Accurate and efficient protein sequence design through learning concise local environment of residues. *Bioinformatics* **39**
82. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollar, P., (2020). Focal loss for dense object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **42**, 318–327.
83. Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P., (2017). Focal loss for dense object detection. *In: Proceedings of the IEEE International Conference on Computer Vision*, pp. 2980–2988.