Two decades of advances in sequence-based prediction of MoRFs, disorder-to-order transitioning binding regions

Jiangning Song[1,2], and Lukasz Kurgan[3*]

[1]Biomedicine Discovery Institute and Department of Biochemistry and Molecular Biology, Monash University, Melbourne, VIC 3800, Australia

[2]Monash Data Futures Institute, Monash University, VIC 3800, Australia

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

*Corresponding author; Address: Department of Computer Science, Virginia Commonwealth University, 401 West Main Street, Room E4225, Richmond, Virginia 23284, USA; Email: lkurgan@vcu.edu; Phone: (804) 827-3986

Abstract

Introduction: Molecular recognition features (MoRFs) are regions in protein sequences that undergo induced folding upon binding partner molecules. MoRFs are common in nature and can be predicted from sequences based on their distinctive sequence signatures.

Areas covered: We overview twenty years of progress in the sequence-based prediction of MoRFs which resulted in the development of 25 predictors of MoRFs that interact with proteins, peptides and lipids. These methods range from simple discriminant analysis to sophisticated deep transformer networks that use protein language models. They generate relatively accurate predictions as evidenced by the results of a recently published community-driven assessment.

Expert opinion: MoRFs prediction is a mature field of research that is poised to continue at a steady pace in the foreseeable future. We anticipate further expansion of the scope of MoRF predictions to additional partner molecules, such as nucleic acids, and continued use of recent machine learning advances. Other future efforts should concentrate on improving availability of MoRF predictions by releasing, maintaining and popularizing web servers and by depositing MoRF predictions to large databases of protein structure and function predictions. Furthermore, accurate MoRF predictions should be coupled with the equally accurate prediction and modelling of the resulting structures of complexes.

Keywords

Intrinsic disorder; MoRF; Molecular recognition feature; prediction; deep learning; protein-protein interactions; protein-lipid interactions; protein-peptide interactions

1 Introduction

Intrinsically disordered regions (IDRs) are regions in protein sequences that lack stable structure under physiological conditions [1-4]. Several bioinformatics studies demonstrate that proteins with IDRs are present across the entire taxonomy, with much higher rates of occurrence in the eukaryotic organisms [5-9]. Many IDRs interact with ligands that include a broad spectrum of biomolecules, such as proteins, peptides, DNA, RNA, lipids, and a variety of small molecules that include drugs [10-18]. The conformational flexibility of IDRs offers certain advantages when compared to ordered (structured) regions, including the one-to-many binding where one IDR interacts with multiple different molecules [19-23]. One of the most common types of interacting IDRs is the molecular recognition feature (MoRF) [11,24]. MoRFs are relatively short sequence segments that are embedded in longer IDRs and that typically undergo disorder-to-order transitions when interacting with ligands [11,24,25]. Some MORF

regions can remain partly or even fully disordered in the bound state [26,27]. The limits on the MoRF region lengths differs across studies, with some defining them as in the 10 to 70 consecutive amino acids range [24,25] while other works considering shorter segments that span between 5 and 25 residues in length [11,24,28]. Moreover, MoRFs were classified into several types that are defined based on the primary type of the secondary structure that they fold into upon binding, i.e., α-MoRFs (composed mostly of α-helices), β-MoRFs (mostly β-sheets), γ-MoRFs (irregular structures), and complex-MoRFs (mixed secondary structures) [25]. A recent bioinformatics study that investigated nearly 900 species suggested that 20 to 30% of IDRs, depending on the taxonomic assignment, include MoRF regions [11]. Importantly, sequences of the MoRF segments have unique signatures that differ from other types of disordered and ordered regions [11]. These differences motivated the development of computational sequence-based predictors of MoRFs [29,30]. Scientists use these predictive tools in a wide range of investigations. For instance, one of the most popular MoRF predictors, MoRFpred [31], has been utilized recently to investigate cell signaling pathways [32], 20S proteasome substrates [33], a variety of viral proteomes including SARS-CoV-2 [34], rotavirus [35], and hepatitis E [36], and interactomes of YY1 [37], SNED1 [38], and G0S2 [39] proteins. Moreover, MoRF predictions have clinical relevance, as the dysfunction of proteins with binding IDRs was found to be associated with a number of human diseases [40-42]. One of the results of this disfunction is misfolding that may induce a range of conformational illnesses including the prion, Alzheimer's, polyQ and Parkinson's diseases and the Down's syndrome [41,43]. Furthermore, MoRF-containing proteins have regulatory and signaling functions that fundamentally rely on the protein-ligand interactions, and their dysfunction was linked to cardiovascular diseases, cancers, and viral pathogenesis [34,35,40,44,45]. A few specific examples include the tau and Aβ proteins that are associated with conformational diseases, BRCA-1, p53 and AFP proteins that are involved in cancers, and viral capsid proteins that were shown to be implicated in the pathogenesis of viral infections [34-36,40,45,46]. We note that these findings relied on bioinformatics analyses that took advantage of high-quality predictions of binding IDRs [34-36,40,45,46].

Dozens of MoRF predictors have been developed to date, prompting the need to survey them. The last time the MoRF predictors were comprehensively surveyed was in 2019 [30]. That review provided a brief historical overview, covered 13 predictors (shown in bold font in Table 1), and discussed their predictive performance by relying on results collected from several articles that introduce individual predictors [30]. Several more recent surveys that focused on a broader collection of methods that predict binding IDRs also listed and briefly summarized the MoRF predictors [47-49]. These broader predictors target binding regions that are not limited in length and are not necessarily embedded in longer IDRs, and which interact with specific ligand types. Some popular examples include ANCHOR [50] and ANCHOR2 [51] that target protein and peptide binding IDRs; DisoRDPbind [52], DeepDISObind [53], and DisoFLAG [54] that predict DNA and RNA binding IDRs; and DisoLipPred [55], MemDis [56], and DisoFLAG that focus on the lipid binding IDRs. Importantly, the recent surveys discussed MoRF predictors in passing and lacked coverage of the newest tools, beyond 2020 [47] and 2021 [48,49]. Motivated by the promiscuity and functional importance of MoRFs in nature, substantial amount of recent efforts towards the development of MoRF predictors, and a number of modern machine learning advances that were utilized in these efforts, herein we provide an updated, comprehensive and practical overview of the MoRF prediction area. In particular, we cover 25 methods, provide an insightful historical overview that spans the 20 years of these development efforts, highlight recent advances that include use of deep learning algorithms and protein language models, and summarize evaluation of representative methods based on arguably more objective results from a large community-organized assessment (compared to the past survey). In addition, as developers of these tools and authors of some of the past surveys, we also offer our opinion on the current issues and future progress in this active area of research.

Table 1. Detailed summary of MoRF predictors. Methods are sorted chronologically; bold font denotes methods covered in the 2019 survey [30]. 'Predictive model' column includes Feed Forward Network (FNN), Convolutional Network (CN), Bidirectional Long Short-Term Memory (BLSTM) network, and Support Vector Machine (SVM). 'Availability' column includes web server (WS), source code (SC), both (WS+SC), never available (NA; original article does not provide information on availability), and no longer available (NLA; original article provides links to WS and/or SC but these links no longer work). 'URL' gives pages where a given method was available as of September 2024. 'Citations' column includes total citations with annual citations inside brackets; these data were collected from Google Scholar in September of 2024. For methods published in multiple articles, we use the reference with the highest citation count to avoid duplicate counting.

Method name (year published)	Ref.	Uses machine learning (ML)	Predictive model	Uses deep learning (protein language model)	Availability as of Sept 2024	URL	Citations total (per year)
α-MoRFpred (2005)	[28]	Yes	Discriminant analysis	No	NA	NA	703 (35.2)
α-MoRFpred II (2007)	[57]	Yes	FFN	No	NA	NA	355 (19.7)
retro-MoRFs (2010)	[58]	No	Sequence alignment	No	NA	NA	53 (3.5)
MoRFpred (2012)	[31,59]	Yes	SVM	No	WS	http://biomine.cs.vcu.edu/servers/MoRFpred/	367 (28.2)
MFSPSSMpred (2013)	[60]	Yes	SVM	No	NLA		68 (5.7)
MoRF$_{CHiBi}$ (2015)	[61]	Yes	SVM	No	WS+SC	https://morf.msl.ubc.ca/index.xhtml (WS) https://gsponerlab.msl.ubc.ca/software/morf_chibi/ (SC)	83 (8.3)
DISOPRED3 (2015)	[62]	Yes	SVM	No	WS+SC	http://bioinf.cs.ucl.ac.uk/psipred/ (WS) http://bioinfadmin.cs.ucl.ac.uk/downloads/DISOPRED/ (SC)	886 (88.6)
MoRF$_{CHiBi}$ SYSTEM (2015)	[63,64]	Yes	Meta predictor that combines MoRF$_{CHiBi}$ [61] and ESpritz [65]	No	WS+SC	https://morf.msl.ubc.ca/index.xhtml (WS) https://gsponerlab.msl.ubc.ca/software/morf_chibi/ (SC)	146 (14.6)
fMoRFpred (2016)	[66]	Yes	SVM	No	WS	http://biomine.cs.vcu.edu/servers/fMoRFpred/	156 (17.3)
Predict-MoRFs (2016)	[67]	Yes	SVM	No	SC	https://github.com/roneshsharma/Predict-MoRFs (SC)	34 (3.8)
Fang et al. (2018)	[68]	Yes	SVM	No	NA	NA	8 (1.1)
MoRFPred-plus (2018)	[69]	Yes	SVM	No	SC	https://github.com/roneshsharma/MoRFpred-plus/wiki/MoRFpred-plus (SC)	52 (7.4)
OPAL (2018)	[70]	Yes	SVM	No	WS+SC	http://www.alok-ai-lab.com/tools/opal/ (WS) https://github.com/roneshsharma/OPAL/wiki/OPAL-download (SC)	69 (9.9)
OPAL+ (2019)	[71]	Yes	SVM	No	WS+SC	http://www.alok-ai-lab.com/tools/opal_plus/ (WS) https://github.com/roneshsharma/OPAL-plus/wiki/OPAL-plus-Download (SC)	41 (6.8)
en_DCNNMoRF (2019)	[72]	Yes	CN	Yes	NLA		17 (2.8)
MoRF$_{MLP}$ (2019)	[73]	Yes	Hybrid of FFN and Naïve Bayes	Yes	NA		10 (1.7)
MoRF$_{MPM}$ (2019)	[74]	Yes	Minimax probability machine	No	SC	https://github.com/HHJHgithub/MoRFs_MPM	8 (1.3)
MoRFPred_en (2019)	[75]	Yes	Hybrid of CNs and SVM	Yes	NLA		9 (1.5)
SPOT-MoRF (2020)	[76]	Yes	Hybrid of Inception-Residual-Squeeze and Excitation network and BLSTM network	Yes	WS+SC	https://sparks-lab.org/server/spot-morf/ (WS) http://zhouyq-lab.szbl.ac.cn/download/ (SC)	51 (10.2)
MoRF$_{CNN}$ (2021)	[77]	Yes	CN	Yes	NA		6 (1.5)
Res-BiLstm (2021)	[78]	Yes	BLSTM network	Yes	SC	https://github.com/Yanzziang/Transition_Disorder_Prediction (SC)	0 (0)
MoRF-FUNCpred (2022)	[79]	Yes	Ensemble of SVM, Logistic Regression, Decision Tree and Random Forest	No	SC	https://github.com/LiangYu-Xidian/MoRF-FUNCpred (SC)	5 (1.7)
CoMemMoRFPred (2023)	[80]	Yes	Meta predictor that combines flDPnn [81,82], DisoLipPred [55] and MoRF$_{CHiBi}$ [61]	No	WS	http://biomine.cs.vcu.edu/servers/CoMemMoRFPred/ (WS)	2 (1)
MoRF_ESM (2024)	[83]	Yes	Transformer network	Yes (ESM-2)	NA		0 (0)
IDBindT5 (2024)	[84]	Yes	FFN	Yes (ProtT5)	SC	https://github.com/jahnl/binding_in_disorder (SC)	2 (2)

2 Historical overview

Table 1 summarizes key characteristics for the 25 MoRF predictors that include 7 methods that were released since 2020 and 12 methods that were not covered in the last survey [30]. This comprehensive list of methods was established by analyzing past surveys [30,47-49], manually scanning citations to the articles that introduce the listed predictors, and performing manual analysis of relevant PubMed searches. We focus our discussion of this active field of research on three important and complementary aspects. First, we provide a chronological historical overview that highlights major milestones. Second, we discuss availability of these 25 predictors and analyze relation of this aspect with their impact measured using citations. Third, we discuss recent community-driven efforts in measuring predictive performance and runtime and highlight the corresponding results for the MoRF predictors.

Figure 1 presents a chronological record of the 20-years long development efforts and includes annotations of the five major milestones. The first milestone in 2005 marks publication of the first α-MoRFpred method [28]. This method is limited to the prediction of the α-MoRFs and it was designed using a small dataset of 12 proteins with 14 α-MoRF regions. This design was improved two years later with the publication of α-MoRFpred-II by the same research group headed by Prof. Dunker [57]. MoRFpred-II used a larger training dataset with 99 proteins and 102 α-MoRFs and applied machine learning algorithm to produce the predictive model in a form of a shallow feed-forward neural network [57]. The second milestone (Figure 1) is the release of MoRFpred [31,59], the first tool that addresses prediction of generic MoRFs that are not limited to a particular MoRF type (as compared to the α-MoRFs). This method was trained on a relatively large dataset with over 400 proteins and features a more advanced design that includes several sequence-derived inputs, such as an evolutionary profile and prediction of intrinsic disorder and solvent accessibility, which are input to a support vector machine model. MoRFpred was released as a free webserver that is available and operational to this date.

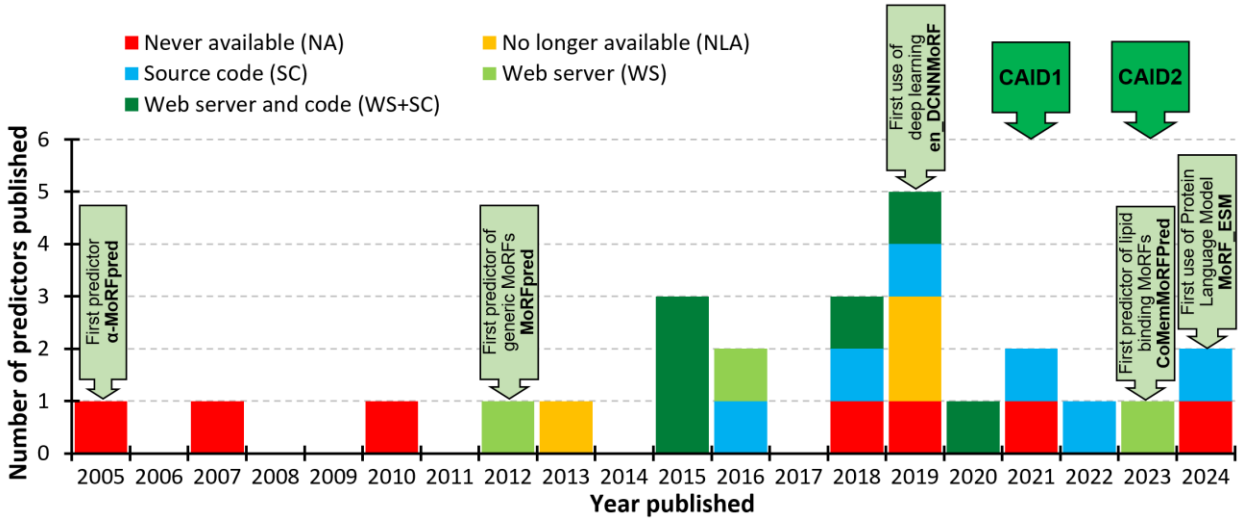

Figure 1. Timeline of the development of MoRF predictors. Color-coded bars denote modes of availability, which include Never Available (NA), No Longer Available (NLA), Source Code (SC), Web Server (WS), and Web Server and Code (WS+SC). Dark green callouts show major community assessment events. Light green callouts identify major milestones.

The third milestone is defined by the first use of a deep learning-based model in the en_DCNNMoRF predictor that was published in early 2019 [72]. This marks a major shift in the design of the MoRF predictors since the substantial majority of the subsequently developed methods also rely on the deep neural network models, i.e., 8 out of 11 released since 2019 (Table 1). The en_DCNNMoRF's model includes two deep convolutional neural networks which results are averaged to produce the final MoRF predictions [72]. The other deep learning-based MoRF predictors utilize a wide range of network topologies including feed forward networks [73,84], convolutional networks [75,77], recurrent networks

[76,78], and transformers [83]. The fourth milestone marks the recent expansion of the scope of the MoRF predictors. Until 2023 these methods targeted prediction of MoRFs that interact with proteins and peptides. This can be explained by the fact that the ground truth annotations of these protein and peptide binding MoRFs, which were used for training and assessment of these methods, were relatively easy to collect from existing databases, such as Protein Data Bank [85] and MobiDB [86]. The first method that considers other types of partner molecules is CoMemMoRFPred, which predicts lipid binding MoRFs [80]. Development of this method was possible because of the preceding release of the MemMoRF database in 2021 [87], which was used to source the corresponding ground truth annotations. The most recent milestone is associated with the first use of the protein language models (PLMs), which occurred in 2024 [83]. PLMs are used to generate inputs into the predictive models and they are typically applied in conjunction with deep neural networks, which is the case for both MoRF predictors that applied PLMs [83,84]. More specifically, MoRF_ESM uses the ESM-2 PLM [88] and a deep transformer network [83], and IDBindT5 uses the ProtT5 PLM [89] and a deep feed forward network [84]. We believe that this field of research has reached a mature stage, as evidenced by the steady rate of the development efforts over the last five year, after a spike between 2015 and 2019 (Figure 1). The new methods will continue to be released at a steady pace that will be fueled by the last three milestones, in particular the development of new deep network architectures and new PLMs, and expansion of the scope.

3 Availability and impact

We summarize the availability of 25 MoRF predictors in Table 1. We enumerated five scenarios: available as a web server (WS; 3 predictors); available as downloadable source code (SC; 6 predictors); available as both server and code (WS+SC; 6 predictors); never available (NA; 7 predictors) when the corresponding article that introduced a given method did not provide information on availability; and no longer available (NLA; 3 predictors) when the links to the code or server that were provided in the original article no longer work. The WS option is arguably convenient since users can easily access servers using a web browser and the entire prediction process is typically done in the server side without installing software on the user's side. However, servers typically limit individual prediction requests to one protein or a small batch of proteins (for load balancing between users) and the runtime of a given prediction is affected by the current server load. The SC option is less convenient since the code has to be downloaded and installed by users and the computations have to be done on the user's hardware. Some of these installations can be challenging since they rely on multiple third-party applications and may require specific hardware and/or software infrastructure. On the other hand, the SC option facilitates generation of predictions at a large scale and embedding of the corresponding predictor into other bioinformatics pipelines. Altogether, 15 of the 25 methods are available to the end users (60% availability rate), with 6 of them available as both WS and SC. This is similar to the recently estimated 65% availability rate for the predictors of the intrinsic disorder [90] and a bit higher than the below 50% availability for predictors of protein and nucleic acid binding residues [91,92].

We investigated whether the mode (lack) of availability is associated with impact of MoRF predictors, which we approximate based on their citations in Google Scholar as of September 2024 (Table 1). We quantified the total number of citations and the annual number of citations (total divided by the number of years since publication), and we used the latter to compare impact across methods. We excluded predictors from 2024 since their citation data is not reliable. The 25 MoRF predictors were cited altogether about 3100 times. More importantly, we found that predictors that offer WS were cited at a much higher rate, i.e., median annual citations of 17.3 for the methods available as only WS and 10.1 for the tools available as code and web server, when compared with the other three options, i.e., median annual citations of 2.6, 2.8 and 1.7 for the predictors that were never available, no longer available, and available as only SC, respectively. Our observation that availability of the WS option substantially boosts citations agrees with a recently released broader analysis of the availability and impact of sequence based predictors of protein structure and function [93]. We hypothesize that tools available as WSs are more popular because users may need their predictions in an *ad hoc* manner that would not justify the

installation effort and/or may not have the computational resources and experience needed to install and run the predictors locally.

4 Predictive performance

Assessments of the predictions of ligand binding IDRs were included in the two recently completed community-organized Critical Assessment of Intrinsic disorder (CAID) events: CAID1 in 2021 [94] and CAID2 in 2023 [95] (Figure 1). This inclusion demonstrates the importance and relevance of MoRF predictors. These evaluations were performed by independent assessors who evaluated predictors that were provided by their authors before the event started. A large number of predictors was tested on blind test datasets (authors of predictors did not have access to the test proteins) using community-accepted metrics that quantify predictive quality. The CAID evaluations are arguably more objective when compared to the smaller-scale tests that are performed when individual predictive tools are published. Moreover, the fact that the participating predictors are run by the same assessors on the same hardware platform facilitates reliable and consistent comparison of runtime.

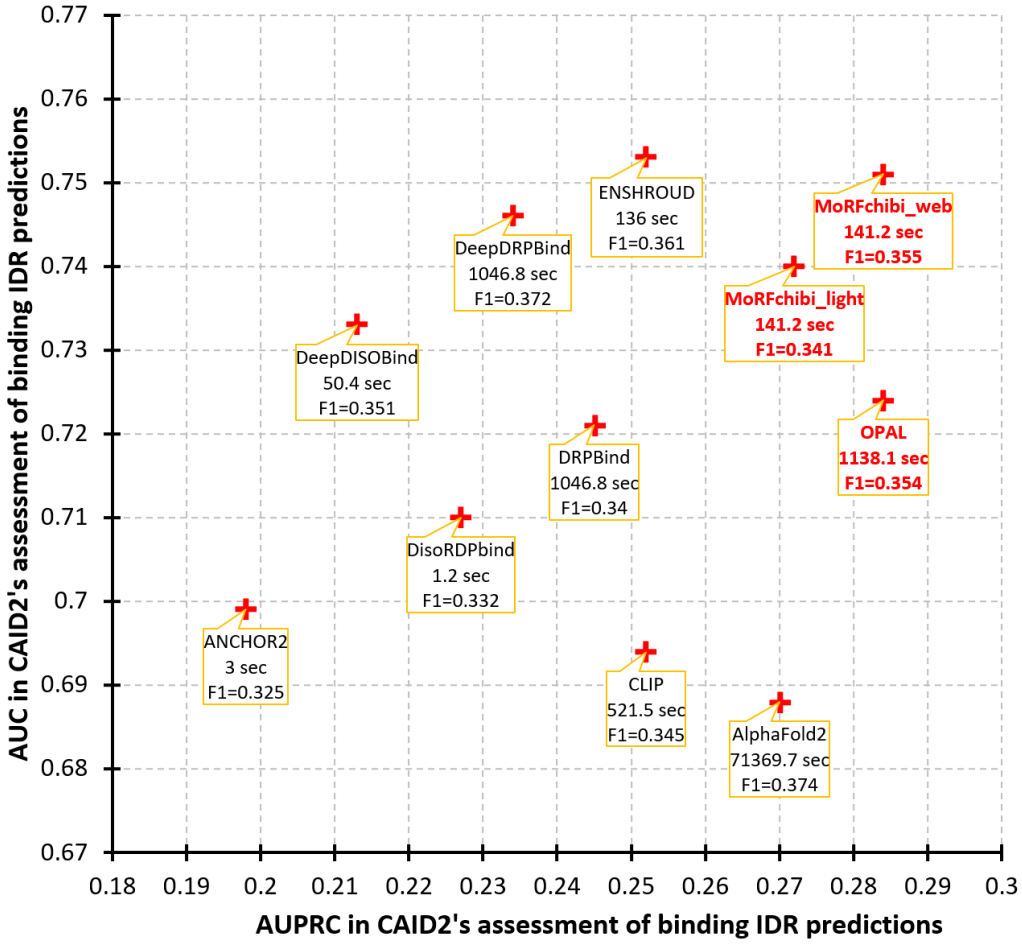

Figure 2. Comparison of predictive performance for the prediction of binding IDRs in the CAID2 experiment [95]. The performance was measured using AUC (*y*-axis), AUPRC (*x*-axis), F1 (callouts) and runtime measured per 1000 amino acids long protein (callouts). MoRF predictors are highlighted by bold font in red color in the callouts.

CAID2 evaluated 32 predictors of binding IDRs that included 4 MoRF predictors: DISOPRED3 [62], MoRFchibi_light and MoRFchibi_web that are part of the MoRFchibi SYSTEM [63,64], and OPAL [70]. Figure 2 summarizes these results by comparing the top 10 predictors of binding IDRs that were ranked based on two popular metrics: Area Under the ROC Curve (AUC; *y*-axis in Figure 2) and Area Under the

Prediction-Recall Curve (AUPRC, *x*-axis in Figure 2). Following CAID2, we also include the F1 metric that quantifies performance based on the highest point on the precision–recall curve, i.e., maximal F1 values that can be obtained by a given predictor [95] (callouts in Figure 2). We observed that 3 of the 4 MoRF predictors were ranked among the top 10 predictors of binding IDRs in CAID2 (Figure 2). These three methods secured the highest AUPRC values and relatively high AUC values, which placed them in the best top-right quadrant in Figure 2. Moreover, their F1 scores were 0.36 for MoRFchibi_web, 0.35 for OPAL, and 0.34 for MoRFchibi_light. MoRFchibi_web was arguably the best predictor when considering both the predictive performance and runtime. It generated predictions in about 2.5 minutes per protein, secured the highest AUPRC of 0.284, the second highest AUC of 0.751, and MCC of 0.36, behind only the ENSHROUD method that obtained nearly identical AUC of 0.753 and MCC of 0.36 but much lower AUPRC of 0.252. Altogether, these results demonstrate that current MoRF predictors offer competitive levels of predictive performance.

5 Expert opinion

Computational prediction of MoRFs in protein sequences is a mature field of research with deep historical roots that stretch over 20 years. We show that the current tools are relatively accurate and that recently developed methods already took advantage of recent machine learning advances including the use of sophisticated deep neural networks (e.g., transformers) and protein language models (e.g., EMS-2 and ProtT5). We believe that these efforts will continue at a steady pace in the foreseeable future as new deep network architectures and PLMs will be developed and released. In particular, we observe a recent trend in the development of PLMs that began to target specific classes/families of proteins, with examples of ProGen that focuses on certain families of lysozymes [96] and IgLM on antibodies [97]. Similar efforts towards developing PLMs that target proteins with MoRFs should drive further improvements in accuracy for the MoRF predictors. We also foresee further expansion of the scope of the MoRF predictions to additional types of partner molecules, such as DNAs and RNAs.

Given that this field has reached the mature stage, we believe that efforts should be shifted to improving the availability of the MoRF predictions to the end users. This could be done in three complementary ways. **First**, the authors of the new predictors should be required to offer and maintain a web server for their tools. This should substantially increase impact, as we demonstrated empirically for the already published predictors, and we argue that the corresponding cost is relatively low. We believe that the requirement to support web servers for an extended period of time should be enforced at the point of their publication. Several venues stipulate these requirements including the *Bioinformatics* journal (application notes articles; minimum of two years of support), *Journal of Molecular Biology* ("Computation Resources for Molecular Biology" issue; three years of support), and *Nucleic Acids Research* journal (web server issue; five years of support). These requirements should be unified and potentially extended to over five years, which in our view would benefit both the developers (boosted impact) and users (improved access). **Second**, the web servers of the leading MoRF predictors should be popularized via inclusion into centralized predictive resources, which provide easy access to multiple predictors that cover a broad spectrum of structural and functional aspects of proteins. Several such resources are available including (alphabetically) Brewery [98,99], CAID prediction portal [100], DEPICTER [101,102], MULTICOM [103,104], PredictProtein [105,106], RIDAO [107] and PSIPRED workbench [108,109]. As of October 2024, the CAID portal includes three MoRF predictors (DISOPRED3, MoRFchibi SYSTEM, and OPAL) [100], DEPICTER covers the MoRFchibi SYSTEM [102], and PSIPRED workbench includes DISOPRED3 [109]. These efforts should be strengthened by expanding into other resources. **Third**, pre-computed results generated by MoRF predictors should be made available via the existing databases of the intrinsic disorder predictions, which include D^2P^2 [110], MobiDB [86,111] and DescribePROT [112,113]. These resources offer access to large collections of pre-computed predictions that span hundreds and even thousands of organisms, and which can be conveniently searched and obtained nearly instantly via a web interface. These databases address several issues related to the direct use of predictors which could be difficult (i.e., finding server or code could be challenging and making

predictions could be time-consuming) and wasteful (different users make the same predictions when studying the same proteins). However, predictors still have to be used when attempting to obtain results for proteins that are not included in these databases. We note that as of October 2024 DescribePROT includes prediction of the MoRFchibi SYSTEM [102] for 2.3 million proteins from 273 organisms while the other two databases do not cover MoRF predictions. Adding MoRF predictors to the other resources, particularly MobiDB that covers 245 million proteins, would substantially improve the availability of the MoRF predictions.

Prediction of the MoRFs in protein sequences should be subsequently followed by modelling structures of the resulting protein-protein, protein-peptide, protein-lipid complexes (i.e., MoRFs typically fold upon binding). Modelling these interactions for IDRs, including MoRF regions, is rather challenging and relatively few suitable tools are currently available. One of the first methods that can handle docking for intrinsically disordered regions is IDP-LZerD [114,115]. Importance of docking for modelling these interactions can be supported with numerous examples, such as the work on the intrinsically disordered NUPR1 protein [116-118]. A relatively recent investigation of methods for docking with IDRs reveals that three tools produce relatively good results [119]: IDP-LZerD [114,115], CABS-Dock [120] and AlphaFold-Multimer [121]. However, the atomic-level details of the structures that they produce require further improvements [119]. Coupling accurate sequence-based MoRF predictions with an equally accurate subsequent predictions of the complex structure would provide powerful means to enable a more comprehensive understanding of the protein-ligand interactions. These investigations, particularly when performed jointly between these two research communities, deserve more attention.

Article highlights

- 25 sequence-based MoRF predictors were developed over the last two decades
- Current MoRF predictors address interactions with proteins, peptides and lipids
- Recently developed MoRF predictors utilize sophisticated deep neural networks and protein language models
- Predictors available as web servers are much more impactful than those that are available as source code or that are not available
- Accessibility of MoRF predictions should be improved by releasing, maintaining and popularizing web servers and including pre-computed MoRF predictions in databases of disorder predictions
- Accurate MoRF predictions should be coupled with equally accurate predictions of the resulting structures of protein-ligand complexes

Declaration of interest

The authors declare no conflicts of interest.

Funding

JS was supported by the National Health and Medical Research Council of Australia (APP1127948, APP1144652, APP2036864). LK was funded in part by the National Science Foundation (DBI2146027 and IIS2125218) and the Robert J. Mattauch Endowment funds.

References

1. Dunker AK, Lawson JD, Brown CJ, et al. Intrinsically disordered protein. Journal of Molecular Graphics and Modelling. 2001 2001/02/01/;19(1):26-59.
2. Oldfield CJ, Uversky VN, Dunker AK, et al. Introduction to intrinsically disordered proteins and regions. Intrinsically Disordered Proteins: Dynamics, Binding, and Function. 2019 2019.

3. Dunker AK, Babu MM, Barbar E, et al. What's in a name? Why these proteins are intrinsically disordered: Why these proteins are intrinsically disordered. Intrinsically Disord Proteins. 2013 Jan-Dec;1(1):e24157.
4. Lieutaud P, Ferron F, Uversky AV, et al. How disordered is my protein and what is its disorder for? A guide through the "dark side" of the protein universe. Intrinsically Disord Proteins. 2016;4(1):e1259708.
5. Peng ZL, Yan J, Fan X, et al. Exceptionally abundant exceptions: comprehensive characterization of intrinsic disorder in all domains of life. Cell Mol Life Sci. 2015 Jan;72(1)(1):137-151.
6. Ward JJ, Sodhi JS, McGuffin LJ, et al. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. Journal of Molecular Biology. 2004 Mar 26

Mar;337(3)(3):635-645.
7. 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(2):137-49.
8. Basile W, Salvatore M, Bassot C, et al. Why do eukaryotic proteins contain more intrinsically disordered regions? PLoS Comput Biol. 2019 Jul;15(7):e1007186.
9. Necci M, Piovesan D, Tosatto SC. Large-scale analysis of intrinsic disorder flavors and associated functions in the protein sequence universe. Protein Sci. 2016 Dec;25(12):2164-2174.
10. Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J Mol Biol. 1999 Oct 22;293(2):321-331.
11. Yan J, Dunker AK, Uversky VN, et al. Molecular recognition features (MoRFs) in three domains of life. Mol Biosyst. 2016 Mar;12(3):697-710.
12. Kjaergaard M, Kragelund BB. Functions of intrinsic disorder in transmembrane proteins. Cellular and Molecular Life Sciences. 2017 Sep;74(17):3205-3224.
13. Wang C, Uversky VN, Kurgan L. Disordered nucleiome: Abundance of intrinsic disorder in the DNA- and RNA-binding proteins in 1121 species from Eukaryota, Bacteria and Archaea. Proteomics. 2016 May;16(10):1486-98.
14. Zhao B, Katuwawala A, Oldfield CJ, et al. Intrinsic Disorder in Human RNA-Binding Proteins. J Mol Biol. 2021 Oct 15;433(21):167229.
15. Dyson HJ. Roles of intrinsic disorder in protein-nucleic acid interactions. Mol Biosyst. 2012 Jan;8(1):97-104.
16. Luo S, Wohl S, Zheng W, et al. Biophysical and Integrative Characterization of Protein Intrinsic Disorder as a Prime Target for Drug Discovery. Biomolecules. 2023 Mar 14;13(3).
17. Hosoya Y, Ohkanda J. Intrinsically Disordered Proteins as Regulators of Transient Biological Processes and as Untapped Drug Targets. Molecules. 2021 Apr 7;26(8).
18. Blundell TL, Gupta MN, Hasnain SE. Intrinsic disorder in proteins: Relevance to protein assemblies, drug design and host-pathogen interactions. Prog Biophys Mol Biol. 2020 Oct;156:34-42.
19. Uversky VN. Intrinsic Disorder-based Protein Interactions and their Modulators. Curr Pharm Design. 2013 Jul;19(23):4191-4213.
20. Hu G, Wu ZH, Uversky VN, et al. Functional Analysis of Human Hub Proteins and Their Interactors Involved in the Intrinsic Disorder-Enriched Interactions. Int J Mol Sci. 2017 Dec;18(12).
21. Hsu WL, Oldfield CJ, Xue B, et al. Exploring the binding diversity of intrinsically disordered proteins involved in one-to-many binding. Protein Sci. 2013 Mar;22(3):258-73.
22. Dyson HJ, Wright PE. Coupling of folding and binding for unstructured proteins. Curr Opin Struct Biol. 2002 Feb;12(1):54-60.

23. Oldfield CJ, Meng J, Yang JY, et al. Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. BMC Genomics. 2008;9 Suppl 1:S1.
24. Vacic V, Oldfield CJ, Mohan A, et al. Characterization of molecular recognition features, MoRFs, and their binding partners. J Proteome Res. 2007 Jun;6(6):2351-66.
25. Mohan A, Oldfield CJ, Radivojac P, et al. Analysis of molecular recognition features (MoRFs). J Mol Biol. 2006 Oct 06;362(5):1043-59.
26. Wang D, Wu SW, Wang DD, et al. The importance of the compact disordered state in the fuzzy interactions between intrinsically disordered proteins. Chemical Science. 2022 Feb 23;13(8):2363-2377.
27. Roterman I, Stapor K, Konieczny L. Engagement of intrinsic disordered proteins in protein-protein interaction. Front Mol Biosci. 2023;10:1230922.
28. Oldfield CJ, Cheng Y, Cortese MS, et al. Coupled Folding and Binding with α-Helix-Forming Molecular Recognition Elements. Biochemistry. 2005;44(37):12454-12470.
29. Meng F, Uversky VN, Kurgan L. Comprehensive review of methods for prediction of intrinsic disorder and its molecular functions. Cell Mol Life Sci. 2017 Sep;74(17):3069-3090.
30. Katuwawala A, Peng Z, Yang J, et al. Computational Prediction of MoRFs, Short Disorder-to-order Transitioning Protein Binding Regions. Comput Struct Biotechnol J. 2019;17:454-462.
31. Disfani FM, Hsu WL, Mizianty MJ, et al. MoRFpred, a computational tool for sequence-based prediction and characterization of short disorder-to-order transitioning binding regions in proteins. Bioinformatics. 2012 Jun 15;28(12):i75-83.
32. Bondos SE, Dunker AK, Uversky VN. Intrinsically disordered proteins play diverse roles in cell signaling. Cell Commun Signal. 2022 Feb 17;20(1):20.
33. Pepelnjak M, Rogawski R, Arkind G, et al. Systematic identification of 20S proteasome substrates. Mol Syst Biol. 2024 Apr;20(4):403-427.
34. Giri R, Bhardwaj T, Shegane M, et al. Understanding COVID-19 via comparative analysis of dark proteomes of SARS-CoV-2, human SARS and bat SARS-like coronaviruses. Cell Mol Life Sci. 2020 Jul 25.
35. Kumar D, Singh A, Kumar P, et al. Understanding the penetrance of intrinsic protein disorder in rotavirus proteome. Int J Biol Macromol. 2020 Feb 1;144:892-908.
36. Shafat Z, Ahmed A, Parvez MK, et al. Role of ORF4 in Hepatitis E virus regulation: analysis of intrinsically disordered regions. Journal of Proteins and Proteomics. 2021 2021/12/01;12(4):289-306.
37. Donald H, Blane A, Buthelezi S, et al. Assessing the dynamics and macromolecular interactions of the intrinsically disordered protein YY1. Biosci Rep. 2023 Oct 31;43(10).
38. Vallet SD, Davis MN, Barque A, et al. Computational and experimental characterization of the novel ECM glycoprotein SNED1 and prediction of its interactome. Biochem J. 2021 Apr 16;478(7):1413-1434.
39. Paez-Perez ED, Llamas-Garcia ML, Benitez-Cardoza CG, et al. Bioinformatic Analysis and Biophysical Characterization Reveal Structural Disorder in G0S2 Protein. ACS Omega. 2020 Oct 13;5(40):25841-25847.
40. Uversky VN, Oldfield CJ, Dunker AK. Intrinsically disordered proteins in human diseases: introducing the D2 concept. Annu Rev Biophys. 2008;37:215-46.
41. Uversky VN. The triple power of D(3): protein intrinsic disorder in degenerative diseases. Front Biosci (Landmark Ed). 2014;19:181-258.
42. Uversky VN. Intrinsic Disorder, Protein-Protein Interactions, and Disease. Adv Protein Chem Struct Biol. 2018;110:85-121.

43. Gadhave K, Gehi BR, Kumar P, et al. The dark side of Alzheimer's disease: unstructured biology of proteins from the amyloid cascade signaling pathway. Cell Mol Life Sci. 2020 Oct;77(20):4163-4208.
44. Cheng Y, LeGall T, Oldfield CJ, et al. Abundance of intrinsic disorder in protein associated with cardiovascular disease. Biochemistry. 2006 Sep 5;45(35):10448-60.
45. Sundar A, Umashankar P, Sankar P, et al. Intrinsic disorder in flaviviral capsid proteins and its role in pathogenesis. J Biosci. 2024;49.
46. Uversky VN, Oldfield CJ, Midic U, et al. Unfoldomics of human diseases: linking protein intrinsic disorder with diseases. BMC Genomics. 2009;10 Suppl 1:S7.
47. Tamburrini KC, Pesce G, Nilsson J, et al. Predicting Protein Conformational Disorder and Disordered Binding Sites. Methods Mol Biol. 2022;2449:95-147.
48. Han B, Ren C, Wang W, et al. Computational Prediction of Protein Intrinsically Disordered Region Related Interactions and Functions. Genes (Basel). 2023 Feb 8;14(2).
49. Basu S, Kihara D, Kurgan L. Computational prediction of disordered binding regions. Comput Struct Biotechnol J. 2023;21:1487-1497.
50. Dosztanyi Z, Meszaros B, Simon I. ANCHOR: web server for predicting protein binding regions in disordered proteins. Bioinformatics. 2009 Oct 15;25(20):2745-6.
51. Meszaros B, Erdos G, Dosztanyi Z. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. Nucleic Acids Res. 2018 Jul 2;46(W1):W329-W337.
52. Peng Z, Wang C, Uversky VN, et al. Prediction of Disordered RNA, DNA, and Protein Binding Regions Using DisoRDPbind. Methods Mol Biol. 2017;1484:187-203.
53. Zhang F, Zhao B, Shi W, et al. DeepDISOBind: accurate prediction of RNA-, DNA- and protein-binding intrinsically disordered residues with deep multi-task learning. Brief Bioinform. 2022 Jan 17;23(1).
54. Pang Y, Liu B. DisoFLAG: accurate prediction of protein intrinsic disorder and its functions using graph-based interaction protein language model. BMC Biol. 2024 Jan 2;22(1):3.
55. Katuwawala A, Zhao B, Kurgan L. DisoLipPred: accurate prediction of disordered lipid-binding residues in protein sequences with deep recurrent networks and transfer learning. Bioinformatics. 2021 Dec 22;38(1):115-124.
56. Dobson L, Tusnady GE. MemDis: Predicting Disordered Regions in Transmembrane Proteins. Int J Mol Sci. 2021 Nov 12;22(22).
57. Cheng YG, Oldfield CJ, Meng JW, et al. Mining alpha-helix-forming molecular recognition features with cross species sequence alignments. Biochemistry. 2007 Nov;46(47):13468-13477.
58. Xue B, Dunker AK, Uversky VN. Retro-MoRFs: Identifying Protein Binding Sites by Normal and Reverse Alignment and Intrinsic Disorder Prediction. International Journal of Molecular Sciences. 2010 Oct;11(10):3725-3747.
59. Oldfield CJ, Uversky VN, Kurgan L. Predicting Functions of Disordered Proteins with MoRFpred. Methods Mol Biol. 2019;1851:337-352.
60. Fang C, Noguchi T, Tominaga D, et al. MFSPSSMpred: identifying short disorder-to-order binding regions in disordered proteins based on contextual local evolutionary conservation. Bmc Bioinformatics. 2013 Oct 4;14.
61. Malhis N, Gsponer J. Computational identification of MoRFs in protein sequences. Bioinformatics. 2015 Jun 1;31(11):1738-44.
62. Jones DT, Cozzetto D. DISOPRED3: precise disordered region predictions with annotated protein-binding activity. Bioinformatics. 2015 Mar 15;31(6):857-63.
63. Malhis N, Jacobson M, Gsponer J. MoRFchibi SYSTEM: software tools for the identification of MoRFs in protein sequences. Nucleic Acids Research. 2016 Jul 8;44(W1):W488-W493.

64. Malhis N, Wong ETC, Nassar R, et al. Computational Identification of MoRFs in Protein Sequences Using Hierarchical Application of Bayes Rule. Plos One. 2015 Oct 30;10(10).
65. Walsh I, Martin AJ, Di Domenico T, et al. ESpritz: accurate and fast prediction of protein disorder. Bioinformatics. 2012 Feb 15;28(4):503-9.
66. Yan J, Dunker AK, Uversky VN, et al. Molecular recognition features (MoRFs) in three domains of life. Mol Biosyst. 2016;12(3):697-710.
67. Sharma R, Kumar S, Tsunoda T, et al. Predicting MoRFs in protein sequences using HMM profiles. Bmc Bioinformatics. 2016 Dec 22;17.
68. Fang C, Moriwaki Y, Zhu DM, et al. Identifying MoRFs in Disordered Proteins Using Enlarged Conserved Features. Proceedings of 2018 6th International Conference on Bioinformatics and Computational Biology (Icbcb 2018). 2018:50-54.
69. Sharma R, Bayarjargal M, Tsunoda T, et al. MoRFPred-plus: Computational Identification of MoRFs in Protein Sequences using Physicochemical Properties and HMM profiles. J Theor Biol. 2018 Jan 21;437:9-16.
70. Sharma R, Raicar G, Tsunoda T, et al. OPAL: prediction of MoRF regions in intrinsically disordered protein sequences. Bioinformatics. 2018 Jun 1;34(11):1850-1858.
71. Sharma R, Sharma A, Raicar G, et al. OPAL plus : Length-Specific MoRF Prediction in Intrinsically Disordered Protein Sequences. Proteomics. 2019 Mar;19(6).
72. Fang C, Moriwaki Y, Tian AK, et al. Identifying short disorder-to-order binding regions in disordered proteins with a deep convolutional neural network method. J Bioinf Comput Biol. 2019 Feb;17(1).
73. He H, Zhao JX, Sun GL. Prediction of MoRFs in Protein Sequences with MLPs Based on Sequence Properties and Evolution Information. Entropy-Switz. 2019 Jul;21(7).
74. He H, Zhao JX, Sun GL. Computational prediction of MoRFs based on protein sequences and minimax probability machine. Bmc Bioinformatics. 2019 Oct 28;20(1).
75. Fang C, Moriwaki Y, Li CH, et al. MoRFPred_en: Sequence-based prediction of MoRFs using an ensemble learning strategy. J Bioinf Comput Biol. 2019 Dec;17(6).
76. Hanson J, Litfin T, Paliwal K, et al. Identifying molecular recognition features in intrinsically disordered regions of proteins by transfer learning. Bioinformatics. 2020 Feb 15;36(4):1107-1113.
77. He H, Zhou YT, Chi Y, et al. Prediction of MoRFs based on sequence properties and convolutional neural networks. Biodata Min. 2021 Aug 14;14(1).
78. Yan Z, Omori S, Yamada KD, et al. Prediction and Characterization of Disorder-Order Transition Regions in Proteins by Deep Learning. bioRxiv. 2021:2021.06.11.448022.
79. Li H, Pang Y, Liu B, et al. MoRF-FUNCpred: Molecular Recognition Feature Function Prediction Based on Multi-Label Learning and Ensemble Learning. Front Pharmacol. 2022;13:856417.
80. Basu S, Hegedus T, Kurgan L. CoMemMoRFPred: Sequence-based Prediction of MemMoRFs by Combining Predictors of Intrinsic Disorder, MoRFs and Disordered Lipid-binding Regions. J Mol Biol. 2023 Nov 1;435(21):168272.
81. Hu G, Katuwawala A, Wang K, et al. flDPnn: Accurate intrinsic disorder prediction with putative propensities of disorder functions. Nature Communications. 2021 Jul 21;12(1):4438.
82. Wang K, Hu G, Basu S, et al. flDPnn2: Accurate and fast predictor of intrinsic disorder in proteins. Journal of Molecular Biology. 2024 2024/05/08/:168605.
83. Fang C, He J, Yamana H. MoRF_ESM: Prediction of MoRFs in disordered proteins based on a deep transformer protein language model. J Bioinform Comput Biol. 2024 Apr;22(2):2450006.
84. Jahn LR, Marquet C, Heinzinger M, et al. Protein embeddings predict binding residues in disordered regions. Sci Rep-Uk. 2024 Jun 12;14(1).

85. Burley SK, Bhikadiya C, Bi C, et al. RCSB Protein Data Bank (RCSB.org): delivery of experimentally-determined PDB structures alongside one million computed structure models of proteins from artificial intelligence/machine learning. Nucleic Acids Res. 2023 Jan 6;51(D1):D488-D508.
86. Piovesan D, Del Conte A, Clementel D, et al. MobiDB: 10 years of intrinsically disordered proteins. Nucleic Acids Res. 2023 Jan 6;51(D1):D438-D444.
87. Csizmadia G, Erdos G, Tordai H, et al. The MemMoRF database for recognizing disordered protein regions interacting with cellular membranes. Nucleic Acids Res. 2021 Jan 8;49(D1):D355-D360.
88. Lin Z, Akin H, Rao R, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. Science. 2023 Mar 17;379(6637):1123-1130.
89. Elnaggar A, Heinzinger M, Dallago C, et al. ProtTrans: Toward Understanding the Language of Life Through Self-Supervised Learning. Ieee T Pattern Anal. 2022 Oct 1;44(10):7112-7127.
90. Zhao B, Kurgan L. Surveying over 100 predictors of intrinsic disorder in proteins. Expert Review of Proteomics. 2021 Dec;18(12):1019-1029.
91. Zhang J, Kurgan L. Review and comparative assessment of sequence-based predictors of protein-binding residues. Brief Bioinform. 2018 Sep 28;19(5):821-837.
92. Wang K, Hu G, Wu Z, et al. Comprehensive Survey and Comparative Assessment of RNA-Binding Residue Predictions with Analysis by RNA Type. International Journal of Molecular Sciences. 2020;21(18):6879.
93. Song J, Kurgan L. Availability of web servers significantly boosts citations rates of bioinformatics methods for protein function and disorder prediction. Bioinform Adv. 2023 Jan 5;3(1):vbad184.
94. Necci M, Piovesan D, Predictors C, et al. Critical assessment of protein intrinsic disorder prediction. Nat Methods. 2021 May;18(5):472-481.
95. Conte AD, Mehdiabadi M, Bouhraoua A, et al. Critical assessment of protein intrinsic disorder prediction (CAID) - Results of round 2. Proteins. 2023 Aug 25.
96. Madani A, Krause B, Greene ER, et al. Large language models generate functional protein sequences across diverse families. Nat Biotechnol. 2023 Aug;41(8):1099-+.
97. Shuai RW, Ruffolo JA, Gray JJ. IgLM: Infilling language modeling for antibody sequence design. Cell Syst. 2023 Nov 15;14(11):979-+.
98. Torrisi M, Pollastri G. Brewery: deep learning and deeper profiles for the prediction of 1D protein structure annotations. Bioinformatics. 2020 Jun 1;36(12):3897-3898.
99. Bau D, Martin AJ, Mooney C, et al. Distill: a suite of web servers for the prediction of one-, two- and three-dimensional structural features of proteins. BMC Bioinformatics. 2006 Sep 5;7:402.
100. Del Conte A, Bouhraoua A, Mehdiabadi M, et al. CAID prediction portal: a comprehensive service for predicting intrinsic disorder and binding regions in proteins. Nucleic Acids Res. 2023 Jul 5;51(W1):W62-W69.
101. Barik A, Katuwawala A, Hanson J, et al. DEPICTER: Intrinsic Disorder and Disorder Function Prediction Server. J Mol Biol. 2020 May 15;432(11):3379-3387.
102. Basu S, Gsponer J, Kurgan L. DEPICTER2: a comprehensive webserver for intrinsic disorder and disorder function prediction. Nucleic Acids Research. 2023.
103. Hou J, Wu T, Guo Z, et al. The MULTICOM Protein Structure Prediction Server Empowered by Deep Learning and Contact Distance Prediction. Methods Mol Biol. 2020;2165:13-26.
104. Cheng J, Li J, Wang Z, et al. The MULTICOM toolbox for protein structure prediction. BMC Bioinformatics. 2012;13:65.
105. Bernhofer M, Dallago C, Karl T, et al. PredictProtein - Predicting Protein Structure and Function for 29 Years. Nucleic Acids Res. 2021 Jul 2;49(W1):W535-W540.
106. Rost B, Yachdav G, Liu JF. The PredictProtein server. Nucleic Acids Research. 2004 Jul 1;32:W321-W326.

107. Dayhoff GW, 2nd, Uversky VN. Rapid prediction and analysis of protein intrinsic disorder. Protein Sci. 2022 Dec;31(12):e4496.
108. Buchan DWA, Jones DT. The PSIPRED Protein Analysis Workbench: 20 years on. Nucleic Acids Res. 2019 Jul 2;47(W1):W402-W407.
109. Buchan DWA, Moffat L, Lau A, et al. Deep learning for the PSIPRED Protein Analysis Workbench. Nucleic Acids Res. 2024 Jul 5;52(W1):W287-W293.
110. Oates ME, Romero P, Ishida T, et al. (DP2)-P-2: database of disordered protein predictions. Nucleic Acids Research. 2013 Jan;41(D1):D508-D516.
111. Piovesan D, Necci M, Escobedo N, et al. MobiDB: intrinsically disordered proteins in 2021. Nucleic Acids Research. 2021 Jan 8;49(D1):D361-D367.
112. Zhao B, Katuwawala A, Oldfield CJ, et al. DescribePROT: database of amino acid-level protein structure and function predictions. Nucleic Acids Research. 2021 Jan 8;49(D1):D298-D308.
113. Basu S, Zhao B, Biro B, et al. DescribePROT in 2023: more, higher-quality and experimental annotations and improved data download options. Nucleic Acids Res. 2024 Jan 5;52(D1):D426-D433.
114. Peterson LX, Roy A, Christoffer C, et al. Modeling disordered protein interactions from biophysical principles. PLoS Comput Biol. 2017 Apr;13(4):e1005485.
115. Christoffer C, Kihara D. IDP-LZerD: Software for Modeling Disordered Protein Interactions. Methods Mol Biol. 2020;2165:231-244.
116. Neira JL, Rizzuti B, Jimenez-Alesanco A, et al. The Paralogue of the Intrinsically Disordered Nuclear Protein 1 Has a Nuclear Localization Sequence that Binds to Human Importin alpha3. Int J Mol Sci. 2020 Oct 8;21(19).
117. Neira JL, Rizzuti B, Jimenez-Alesanco A, et al. A Phosphorylation-Induced Switch in the Nuclear Localization Sequence of the Intrinsically Disordered NUPR1 Hampers Binding to Importin. Biomolecules. 2020 Sep 11;10(9).
118. Santofimia-Castano P, Rizzuti B, Pey AL, et al. Intrinsically disordered protein NUPR1 binds to the armadillo-repeat domain of Plakophilin 1. Int J Biol Macromol. 2021 Feb 15;170:549-560.
119. Verburgt J, Zhang Z, Kihara D. Multi-level analysis of intrinsically disordered protein docking methods. Methods. 2022 Aug;204:55-63.
120. Kurcinski M, Pawel Ciemny M, Oleniecki T, et al. CABS-dock standalone: a toolbox for flexible protein-peptide docking. Bioinformatics. 2019 Oct 15;35(20):4170-4172.
121. Evans R, O'Neill M, Pritzel A, et al. Protein complex prediction with AlphaFold-Multimer. bioRxiv. 2022:2021.10.04.463034.