# Machine learning methods for predicting protein-nucleic acids interactions

Min Li[1]*, Fuhao Zhang[1], Lukasz Kurgan[2]*

[1]Hunan Provincial Key Lab on Bioinformatics, School of Computer Science and Engineering, Central South University, Changsha, China, 410083
[2]Department of Computer Science, Virginia Commonwealth University, Richmond, Virginia, United States
* Corresponding authors: Min Li at limin@mail.csu.edu.cn ; Lukasz Kurgan at lkurgan@vcu.edu

## Abstract

Protein-nucleic acids drive many key cellular functions, such as regulation of gene expression, transcription and translation. Experimental characterization of the molecular-level details of these interactions is relatively expensive and time consuming since it requires application of complex methods, such as X-ray crystallography and/or NMR. Given the relatively low coverage of the experimental molecular-level data on the protein-nucleic acids interactions, many computational methods that predict these interactions from the readily available protein sequences were developed. We introduce and describe a comprehensive collection of 51 methods that predict nucleic acid interacting amino acids in protein sequences. These methods include 20 DNA-binding predictors, 20 RNA-binding predictors and 11 methods that predict both DNA- and RNA-binding residues. We briefly summarize their inputs, predictive architectures, outputs and availability. We find that most of these methods were trained using protein-nucleic acids structures, compared to a more limited number of methods that predict these interactions in the intrinsically disordered regions. We observe that these methods rely almost exclusively on classical/shallow machine learning and deep learning algorithms. Finally, we endorse five recent, readily available and arguably more useful predictors.

## 1 Introduction

Proteins carry out their cellular functions by interacting with nucleic acids [1-12], proteins [13-17] and a variety of other ligands including small molecules and lipids [18-20]. The protein-nucleic acid interactions are instrumental for a wide spectrum of cellular functions, such as regulation of gene expression, transcription, and translation, to name but a few. Molecular-level knowledge of these interactions is largely derived from structural studies of protein-nucleic acid complexes, which are often sourced from the Protein Data Bank database [21]. The structural data are used to categorize the protein-nucleic acids interactions, characterize the underlying physics, and decipher patterns that define molecular recognition and specificity of interactions [22-25]. Moreover, recent studies reveal that the protein-DNA and protein-RNA interactions are also a common function of intrinsically disordered

regions (IDRs) [11, 26-30], which are defined as sequence segments that lack a stable equilibrium structure under physiological conditions [31-33].

The experimental methods to study these interactions are relatively time-consuming and labor-intensive, and consequently they cannot keep up with the rapid accumulation of protein sequences. One solution is to use the available experimental data on the protein-nucleic acids interactions to devise computational models that accurately predict these interactions from protein sequences [6, 34-42]. These computational methods can be classified into two categories: protein-level vs. residue-level. The protein-level methods identify whether a given protein sequence binds DNA and/or RNA, while residue-level methods predict whether and which residues in a given protein sequence interact with DNA and/or RNA. We focus on the residue-level methods that provide a higher level of details. The sequence-based predictors of nucleic-acid binding residues require only a protein sequence as the input and thus are able to provide predictions for the over 200 million of currently available protein sequences [43].

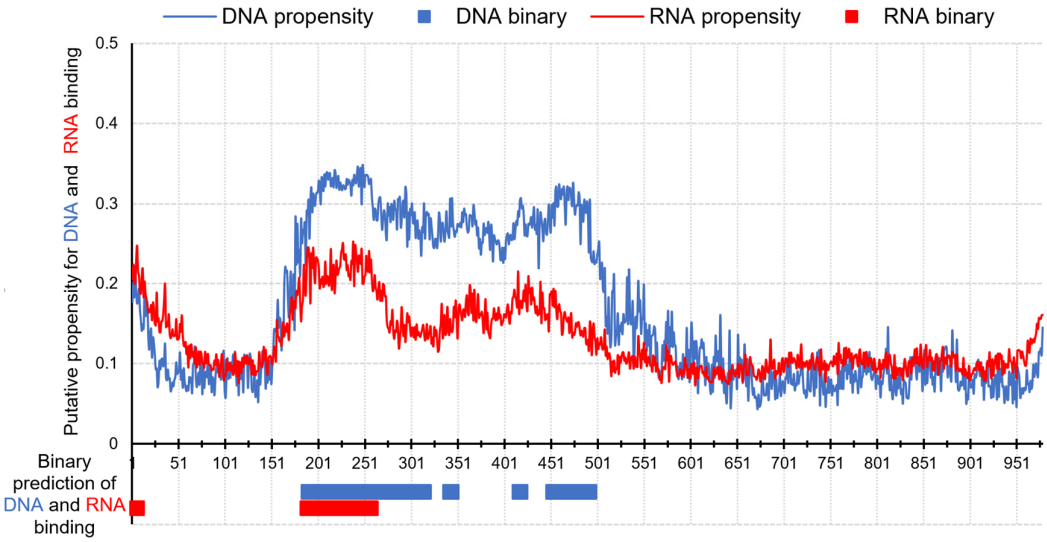


**Figure 1**. Prediction of the DNA binding and RNA-binding residues generated by DeepDISObind for the silent information regulator Sir3p from yeast (DisProt: DP00533; UniProt: P06701). The prediction was generated using the DeepDISObind's webserver located at https://www.csuligroup.com/DeepDISOBind/.

We survey and describe over 50 sequence-based predictors of the DNA and/or RNA binding residues. We identify these methods by scanning past surveys [6, 35, 37-40, 42, 44, 45] supplemented with a manual literature search. These methods output binary predictions and numeric propensities for each residue in an input protein sequence. The binary predictions denote whether a given residue interacts with DNA and/or RNA (0 for non-binding residue vs. 1 means for a DNA and/or RNA binding residue) while the propensities express likelihood of these interactions. Figure 1 shows an example result generated by DeepDISOBind, one of the most recent methods that predict DNA and RNA interactions in IDRs [46], for the silent information regulator Sir3p from budding yeast (DisProt: DP00533; UniProt: P06701). This protein is instrumental for modulating chromatin [47, 48]. It includes

disordered DNA-binding region (positions 216 to 549) that is flanked by structured segments that extend to both sequence termini, as shown based on the experimental annotations available in the DisProt database [49]. The blue and red plots in Figure 1 represent the predicted propensities for interactions with DNA and RNA partners, respectively. The binary predictions are shown underneath using horizontal color-coded bars. This example illustrates the format and value of the sequence-based predictions. In this particular case, DeepDISOBind identifies DNA-binding residues in the segments between positions 190 and 500, which is in good agreement with the location of the experimentally identified DNA-binding IDR (positions 216 to 549). At the same time, the RNA-binding prediction generated by DeepDISOBind suggests a much smaller likelihood of interactions with RNA for this protein, i.e., the propensities shown in red are relatively low.

## 2 Prediction of the protein-nucleic acid binding residues from sequence

Majority of the sequence-based methods for the predictions of protein-nucleic binding residues rely on predictive models that are generated from training data using classical/shallow machine learning algorithms and deep learning algorithms. The training process employs the experimentally annotated data, which is typically collected from publicly available databases, such as PDB[21], BioLip [50] and DisProt [49, 51], to optimize the architectures and parameters of the machine learning-generated models. This is done by minimizing differences between the outputs of these models and the native annotations. After the training process is completed, the models can be used to predict the DNA-binding and/or RNA-binding residues for proteins sequences outside of the training set.

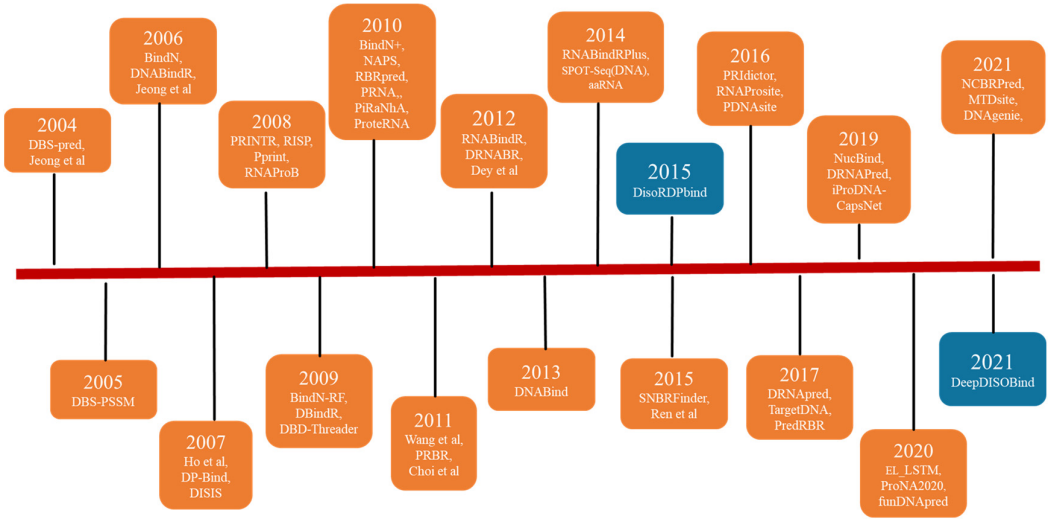


**Figure 2**. Timeline of the sequence-based predictors of the protein-nucleic acid binding residues. The structure-trained predictors are in orange while the disorder-trained predictors are in blue.

Figure 2 shows a timeline of the sequence-based predictors, which are categorized into methods that rely on training data collected from structured protein-RNA and

protein-DNA complexes, typically collected from PDB [21, 52] or BioLip [50] (in orange) vs. those that use training data concerning interactions in IDRs, which are usually obtained from DisProt [49, 51]. The first predictors were developed around 2004 [53, 54] and over 50 were published since. This results in the average rate of 3 new methods per year. Figure 2 reveals that these development efforts are relatively even across the years. However, we note that only two predictors focus on the interactions in the disordered regions, DisoRDPbind that was released in 2015 [55-57] and DeepDISObind that was published in 2021 [46]. Interestingly, recent research that investigates structure-trained vs. disorder-trained methods for the prediction of the protein-binding residues shows that these two classes of methods produce complementary results [58]. Similar observations are expected in the case of the protein-nucleic acids interactions, highlighting the importance of developing both classes of predictors.

## 2.1 Overview of the sequence-based predictors

Table 1 summarizes key aspects of the 51 sequence-based predictors of the DNA and/or RNA binding residues including their predictive target (DNA binding, RNA binding, and DNA and RNA binding) and availability. This comprehensive table reveals several interesting insights. First, significant majority of these computational methods target prediction of either DNA-binding or RNA-binding residues. To be more exact, we find 20 methods that predict the DNA-binding residues and another 20 predictors of the RNA-binding residues. However, we also identify 11 methods that concomitantly predict RNA-binding and DNA-binding residues, with the first one being BindN that was published in 2006 [59]. These methods are arguably more convenient to use since they provide both predictions, typically with a single execution, and since their use does not require sometimes painful conversions between different formats of input and outputs that the use of different individual methods often requires. Interestingly, 5 of these 11 predictors were released in the last three years including NucBind[60], ProNA2020 [61], NCBRPred [62], MTDsite [63] and DeepDISOBind [46]. Moreover, four methods simultaneously predict protein/peptide-binding, DNA-binding and RNA-binding residues: DisoRDPbind [55-57], ProNA2020 [61], MTDsite [63], and DeepDISOBind [46], providing further advantages when compared to the other currently available options.

Table 1 also lists the URLs where the 51 methods are available, as described in the original articles. Most of these predictors are provided as web servers or source code. Only 11 predictors did not provide web servers nor source code at the time of the publication. The web servers are arguably more convenient and primarily target users who are not programmers or computer experts. The predictions are performed on the web server side, which means that users do not need to install software or run the predictions on their local hardware. The users simply utilize a web browser to arrive at a given URL, input their proteins sequence(s), typically in the FASTA format, click start, and collect the resulting predictions after the web server completes the work. The results are returned via the web page and/or are sent back by email. However, one of the common limitations of web servers is the number of input protein sequences,

which is often limited to one or a few for a single request. On the other hand, the source code is better suited for programmers or bioinformaticians. It has to be downloaded, installed and run on user's hardware. This option is particularly attractive if the predictions must be done on a larger scale (large protein families or genomes) and when embedding these predictors into larger bioinformatics platforms. Table 1 reveals that source code is available for only five methods and three of them, iProDNA-CapsNet [64], ProNA2020 [61] and NCBRPred [62], are also available as web servers. These three tools were released in the three years. We find that while 38 of the 51 predictors provided web servers when they were originally published, only half of the URLs of these web servers were working as December 2021 when we tested them. We note that in some cases these methods could be moved to another URL, while we rely on the addresses provided in the original article.

**Table 1**. Summary of the 51 sequence-based predictors of the protein-nucleic acid binding residue. The table considers the prediction target DNA binding, RNA binding, and DNA and RNA binding) and availability. The availability indicates whether the published predictors are provided as web servers, source code or in both modes. The URL provides the location of the implementation/web server that was published in the original article. The "accessible" column indicates whether the URL was available as 15 December 2021. N/A means that a given predictor did not provide web server and source codes information when it was published.

| Method | Ref | Predictive Target | Availability | | |
|---|---|---|---|---|---|
| | | | Type | URL | Accessible |
| DBS-pred | [54] | DNA | web server | http://www.abren.net/dbs-pred/ | NO |
| Jeong *et al*. 2004 | [53] | RNA | N/A | N/A | N/A |
| DBS-PSSM | [65] | DNA | web server | http://dbspssm.netasa.org | NO |
| BindN | [59] | DNA and RNA | web server | http://bioinfo.ggc.org/bindn/ | NO |
| DNABindR | [66] | DNA | web server | http://turing.cs.iastate.edu/PredDNA/index.html | NO |
| Jeong *et al*. 2006 | [67] | RNA | N/A | N/A | N/A |
| Ho *et al*. | [68] | DNA | N/A | N/A | N/A |
| DP-Bind | [69] | DNA | web server | http://lcg.rit.albany.edu/dp-bind | YES |
| DISIS | [70] | DNA | web server | http://cubic.bioc.columbia.edu/services/disis | NO |
| PRINTR | [71] | RNA | web server | http://210.42.106.80/printr/ | NO |
| RISP | [72] | RNA | web server | http://grc.seu.edu.cn/RISP | NO |
| Pprint | [73] | RNA | web server | http://www.imtech.res.in/raghava/pprint/ | YES |
| RNAProB | [74] | RNA | N/A | N/A | N/A |
| BindN-RF | [75] | DNA | web server | http://bioinfo.ggc.org/bindn-rf/ | NO |
| DBindR | [76] | DNA | web server | http://www.cbi.seu.edu.cn/DBindR/DBindR.htm | NO |
| DBD-Threader | [77] | DNA | web server | http://cssb.biology.gatech.edu/skolnick/webservice/DBD-Threader/index.html | NO |
| BindN+ | [78] | DNA and RNA | web server | http://bioinfo.ggc.org/bindn+/ | NO |
| NAPS | [79] | DNA and RNA | web server | http://proteomics.bioengr.uic.edu/NAPS/ | NO |
| PiRaNhA | [80] | RNA | web server | http://www.bioinformatics.sussex.ac.uk/PIRANHA | NO |
| ProteRNA | [81] | RNA | N/A | N/A | N/A |
| RBRpred | [10] | RNA | N/A | N/A | N/A |
| PRNA | [82] | RNA | source code | http://www.aporc.org/doc/wiki/PRNA<br>http://www.sysbio.ac.cn/datatools.asp | NO |
| Wang *et al*. | [83] | RNA | N/A | N/A | N/A |
| PRBR | [84] | RNA | web server | http://www.cbi.seu.edu.cn/PRBR/ | NO |
| Choi *et al*. | [85] | RNA | N/A | N/A | N/A |

| | | | | | |
|---|---|---|---|---|---|
| RNABindR | [86] | RNA | web server | http://einstein.cs.iastate.edu/RNABindR/ | NO |
| DNABR | [87] | DNA | web server | http://www.cbi.seu.edu.cn/DNABR/ | NO |
| Dey *et al*. | [88] | DNA | N/A | N/A | N/A |
| DNABind | [89] | DNA | web server | http://mleg.cse.sc.edu/DNABind/ | YES |
| RNABindRPlus | [90] | RNA | web server | http://einstein.cs.iastate.edu/RNABindRPlus/ | NO |
| SPOT-Seq (DNA) | [91] | DNA | web server | http://sparks-lab.org | YES |
| aaRNA | [92] | RNA | web server | http://sysimm.ifrec.osaka-u.ac.jp/aarna/ | YES |
| SNBRFinder | [93] | DNA and RNA | web server | http://ibi.hzau.edu.cn/SNBRFinder | NO |
| DisoRDPbind | [57] | DNA and RNA | web server | http://biomine.ece.ualberta.ca/DisoRDPbind/ | YES |
| Ren *et al*. | [94] | RNA | N/A | N/A | N/A |
| PRIdictor | [95] | RNA | web server | http://bclab.inha.ac.kr/pridictor | YES |
| RNAProSite | [96] | RNA | web server | http://lilab.ecust.edu.cn/NABind | YES |
| PDNAsite | [97] | DNA | web server | http://hlt.hitsz.edu.cn:8080/PDNAsite/ | NO |
| DRNApred | [98] | DNA and RNA | web server | http://biomine.cs.vcu.edu/servers/DRNApred/ | YES |
| TargetDNA | [99] | DNA | web server | http://csbio.njust.edu.cn/bioinf/TargetDNA/ | YES |
| PredRBR | [100] | RNA | N/A | N/A | N/A |
| NucBind | [60] | DNA and RNA | web server | http://yanglab.nankai.edu.cn/NucBind | YES |
| DNAPred | [101] | DNA | web server | http://csbio.njust.edu.cn/bioinf/dnapred/ | YES |
| iProDNA-CapsNet | [64] | DNA | web server and source code | http://45.117.83.253/problem-iProDNA-CapsNet<br>https://github.com/ngphubinh/iProDNA-CapsNet | YES |
| EL_LSTM | [102] | DNA | source code | http://hlt.hitsz.edu.cn/EL_LSTM/ | NO |
| ProNA2020 | [61] | DNA and RNA | web server and source code | https://github.com/Rostlab/ProNA2020.git<br>http://www.predictprotein.org | YES |
| funDNApred | [103] | DNA | web server | http://biomine.cs.vcu.edu/servers/funDNApred/ | YES |
| NCBRPred | [62] | DNA and RNA | web server and source code | http://bliulab.net/NCBRPred | YES |
| MTDsite | [63] | DNA and RNA | web server | http://biomed.nscc-gz.cn/server/MTDsite/ | YES |
| DNAgenie | [104] | DNA | web server | http://biomine.cs.vcu.edu/servers/DNAgenie/ | YES |
| DeepDISOBind | [46] | DNA and RNA | web server | https://csuligroup.com/DeepDISOBind/ | YES |

## 2.2 Architectures of the sequence-based predictors

Table 2 provides further insights concerning the 51 sequence-based predictors of the DNA and/or RNA binding residues by summarizing and comparing their predictive architectures. This includes details about the predictive inputs that are produced from the protein sequence, predictive models, and the outputs that they produce.

Based on a recent study [6], we identify several common types of inputs that these predictors use including the amino acid sequence (AAS) itself; the evolutionary information (EVO) that is derived from the sequence using multiple sequence alignment algorithms [105], such as PSI-BLAST [106] or HHblits [107]; and the secondary structure (SS) and relative solvent accessibility (RSA) that are predicted from the input sequence using third-party predictors, such as PSIPRED [108], ASAquick [109], SPINE X[110], SPOT-1D [111] and SPIDER [112]. Recent surveys of the sequence-based SS and RSA predictors provide a broader overview of these methods [113-120]. The AAS input is typically expressed using 1-hot encoding, where each amino acid is represented by 20-dimensional binary vector. We find that each of the 51 predictors utilizes at least one of these four input types, and virtually all of them use the AAS input. Furthermore, Table 2 reveals only 1 out of the 16 methods that were published until 2009 uses all of the four inputs [53, 54, 59, 65, 67-77]. In contrast, 10 out of the 13 methods that were published in the last 5 years (since 2017) utilize the four inputs [46, 60-64, 98-104]. This transition clearly demonstrates that these four inputs are very useful for the prediction of the protein-nucleic acid interactions.

Interestingly, we note that nearly all of the 51 methods rely on machine learning algorithms to produce their predictive models. The only two exceptions are DBD-Threader [77] and SPOT-Seq [91] that predict protein-DNA binding residues based on template-based/homology prediction. The most commonly used classical/shallow machine learning methods are support vector machines (24 predictors), neural networks (11 predictors), random forest (7 predictors) and logistic regression (3 predictors). We also find that 10 methods utilize more than one machine learning algorithm. Moreover, several methods, including DNABind [89], SNBRFinder [93] and NucBind [60] combine a machine learning-generated model with the template-based approach.

Three recent methods utilize deep learning [46, 62, 63]. The deep learning algorithms produce neural networks with multiple/many hidden layers which often rely on sophisticated network topologies that include recurrent, convolutional and transformer modules. Deep learners are nowadays used to develop predictors of numerous aspects of protein structure and function. Examples include the state-of-the-art protein structure predictor, AlphaFold [121, 122], methods that predict residue-residue contacts [123], secondary structure [111, 124-126] protein function [127-129], protein-drug interactions [130, 131], and functional sites [58, 132]. The introduction of the deep learning into the prediction of the nucleic-acid binding residues stems from the recent popularity of these models, as shown above, but also from the fact that recent works find them to produce more accurate models when compared to the shallow/classical machine learning algorithms [46, 62].

**Table 2**. Summary of predictive models, inputs and outputs of the 51 sequence-based predictors of the protein-nucleic acid binding residues. The inputs include amino acid sequence (AAS), secondary structure (SS) predicted from sequence, solvent accessibility (RSA) predicted from sequence and evolutionary information (EVO) computed from multiple-sequence alignment. The predictive models are divided into two categories: classical/shallow machine learning (ML) and deep learning (DL). Specific ML algorithms include logistic regression (LR), neural network (NN), support vector machine (SVM), and random forest (RF). N/A in the outputs means that the information could not be checked since the implementation is not available.

| Method | Ref | Inputs | | | | Predictive model | Outputs | |
|---|---|---|---|---|---|---|---|---|
| | | AAS | EVO | SS | RSA | | Binary prediction | Predictive propensity |
| DBS-pred | [54] | √ | × | × | × | ML(NN) | N/A | N/A |
| Jeong *et al*. 2004 | [53] | √ | × | √ | × | ML(NN) | N/A | N/A |
| DBS-PSSM | [65] | √ | √ | × | × | ML(NN) | √ | √ |
| BindN | [59] | √ | × | × | × | ML(SVM) | √ | √ |
| DNABindR | [66] | √ | × | × | × | ML(Naïve Bayes) | √ | × |
| Jeong *et al*. 2006 | [67] | √ | × | × | × | ML(NN) | N/A | N/A |
| Ho *et al*. | [68] | √ | √ | × | × | ML(SVM) | N/A | N/A |
| DP-Bind | [69] | √ | √ | × | × | ML(ensemble learning) | √ | √ |
| DISIS | [70] | √ | √ | √ | √ | ML(SVM, NN) | N/A | N/A |
| PRINTR | [71] | √ | × | √ | × | ML(SVM) | N/A | N/A |
| RISP | [72] | √ | √ | × | × | ML(NN) | N/A | N/A |
| Pprint | [73] | √ | √ | × | × | ML(SVM) | √ | √ |
| RNAProB | [74] | √ | √ | × | × | ML(SVM) | N/A | N/A |
| BindN-RF | [75] | √ | √ | × | × | ML(RF) | N/A | N/A |
| DBindR | [76] | √ | √ | √ | × | ML(RF) | N/A | N/A |
| DBD-Threader | [77] | √ | × | × | × | Template-based | √ | × |
| BindN+ | [78] | √ | √ | × | × | ML(SVM) | √ | √ |
| NAPS | [79] | √ | √ | × | × | ML(C4.5 ) | N/A | N/A |
| PiRaNhA | [80] | √ | √ | × | √ | ML(SVM) | N/A | N/A |
| ProteRNA | [81] | √ | √ | √ | × | ML(SVM) | N/A | N/A |
| RBRpred | [10] | √ | √ | √ | √ | ML(SVM) | N/A | N/A |
| PRNA | [82] | √ | × | √ | × | ML(RF) | √ | √ |
| Wang *et al.* | [83] | √ | √ | × | √ | ML(SVM) | N/A | N/A |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| PRBR | [84] | √ | √ | √ | × | ML(RF) | √ | √ |
| Choi *et al*. | [85] | √ | √ | × | × | ML(SVM), Homology-Based | √ | √ |
| RNABindR | [86] | √ | × | √ | × | ML(SVM) | √ | × |
| DNABR | [87] | √ | √ | × | × | ML(SVM) | √ | √ |
| Dey *et al*. | [88] | √ | √ | × | × | ML(RF) | N/A | N/A |
| DNABind | [89] | √ | √ | √ | × | ML(SVM) | N/A | N/A |
| RNABindRPlus | [90] | √ | √ | √ | √ | ML, Template-based | √ | √ |
| SPOT-Seq (DNA) | [91] | √ | × | × | × | Template-based | √ | √ |
| aaRNA | [92] | √ | √ | √ | √ | ML(NN) | √ | √ |
| SNBRFinder | [93] | √ | √ | √ | √ | ML(HMM,SVM), Template-based | N/A | N/A |
| DisoRDPbind | [57] | √ | × | √ | √ | ML(LR) | √ | √ |
| Ren *et al*. | [94] | × | √ | × | × | ML(ensemble learning) | N/A | N/A |
| PRIdictor | [95] | √ | × | × | × | ML(SVM) | √ | √ |
| RNAProSite | [96] | √ | √ | √ | √ | ML(RF) | √ | √ |
| PDNAsite | [97] | √ | √ | √ | √ | ML(ensemble learning) | N/A | N/A |
| DRNApred | [98] | √ | √ | √ | √ | ML(ensemble learning) | √ | √ |
| TargetDNA | [99] | √ | √ | √ | √ | ML(SVM) | √ | √ |
| PredRBR | [100] | √ | √ | √ | √ | ML(Gradient Boosting) | √ | √ |
| NucBind | [60] | √ | √ | √ | × | ML(SVM), homologous templates | √ | √ |
| DNAPred | [101] | √ | √ | √ | √ | ML(SVM) | √ | √ |
| iProDNA-CapsNet | [64] | √ | √ | × | × | ML(NN) | √ | √ |
| EL_LSTM | [102] | √ | √ | √ | √ | ML(NN, Bagging) | √ | √ |
| ProNA2020 | [61] | √ | √ | √ | √ | ML(SVM, NN) | √ | √ |
| funDNApred | [103] | √ | √ | √ | √ | ML(FCM) | × | √ |
| NCBRPred | [62] | √ | √ | √ | √ | DL | √ | √ |
| MTDsite | [63] | √ | √ | × | × | DL | √ | √ |
| DNAgenie | [104] | √ | √ | √ | √ | ML(SVM) | √ | √ |
| DeepDISOBind | [46] | √ | √ | √ | √ | DL | √ | √ |

Table 2 also summarizes the outputs of the 51 predictors, which may include the binary score and/or numeric propensity, as we explain in the introduction. We cannot identify the outputs for 20 methods since we have no access to their implementations or web servers. Among the remaining predictors, over 80% (27 out of 31) produce both types of outputs. DNABindR [66], DBD-Threader [77] and the method by Choi *et al*. [85] output just the binary predictions. While funDNApred [103] provides only the propensities, user can utilize these scores to produce binary predictions, i.e., residues with propensity > a given threshold can be predicted in binary as binding. Altogether, we suggest that predictors should generate both types of outputs since the predictive propensities provide a useful context for the arguably easier to comprehend binary predictions.

## 3 Summary

This chapter summarizes a comprehensive collection of 51 sequence-based predictors of protein-nucleic acid binding residues. We find that while the early methods would typically target prediction of the DNA-binding or RNA-binding residues, five methods that were published in the last three years simultaneously predict DNA and RNA binding residues [46, 60-63]. Moreover, three of these methods, including ProNA2020, MTDsite [63] and DeepDISOBind [46], extend this scope by predicting protein/peptide-binding residues. This observation suggests a recent trend to develop tools that offer a wider scope of predictions.

We observe that many methods do not have source code or web servers. Some are unavailable because the authors did not maintain the originally published URLs. This problem is especially acute for the methods published before 2016. To better serve the community, we encourage the authors to keep the web servers running and to provide source code for the end users.

We also identify a shortage of methods that predict protein-nucleic acids interactions in intrinsically disordered regions, with only two currently available methods: DisoRDPbind [57] and DeepDISObind [46]. This is an important issue since recent work demonstrates that structure-trained and disorder-trained predictors of the protein-binding residues produce complementary results [58]. While a comparable study for the prediction of the nucleic acid binding residues is missing, we believe that similar conclusions would be drawn.

Given the above observations, we endorse a few predictors, focusing on the methods that cover multiple types of ligands (DNA, RNA and proteins/peptides), are currently available, and which consider both structure- and disorder-annotated interactions. Our recommendations include NCBRPred [62], DisoRDPbind [55-57], MTDsite [63], ProNA2020 [61] and DeepDISOBind [46].

We conclude this chapter with a brief discussion of future directions for this active research areas. We find that a few recent methods apply deep learners [46, 62, 63]. Given the recent empirical results which demonstrate that deep learners provide more accurate predictions compared to the shallow machine learning algorithms [46, 62], we anticipate that future methods will continue to utilize deep neural networks. We note that three of our recommended predictors, namely NCBRPred [62], MTDsite [63] and DeepDISOBind [46], use deep networks. The use of more sophisticated deep learners could help to combat the cross-prediction problem [60, 62, 98], which means

that residues that are predicted to bind RNA in fact often interact with DNA, and vice versa. In other words, current methods relatively often mis-predict the type of the interacting ligand. The cross-predictions also happens between protein-binding and nucleic acids-binding residues [58, 133-135]. Lastly, the current predictors are agnostic to the DNA and RNA types, with one exception, DNAgenie [104]. DNAgenie is capable to accurately identify the type of the interacting DNA, covering A-DNA, B-DNA and single-stranded DNA. This means DNAgenie predicts which residues bind A-DNA, B-DNA vs. single stranded DNA, providing more details when compared to the other current tools. Tools that would provide RNA-type specific predictions would be a welcome addition. This is motivated by a recent study that finds that current RNA type-agnostic methods are deficient for the predictions of some RNA types, such as tRNA [42].

## References


1. Siggers, T. and R. Gordan, *Protein-DNA binding: complexities and multi-protein codes.* Nucleic Acids Res, 2014. **42**(4): p. 2099-111.
2. Cook, K.B., T.R. Hughes, and Q.D. Morris, *High-throughput characterization of protein-RNA interactions.* Brief Funct Genomics, 2015. **14**(1): p. 74-89.
3. Sathyapriya, R., M.S. Vijayabaskar, and S. Vishveshwara, *Insights into Protein-DNA Interactions through Structure Network Analysis.* Plos Comp Biology, 2008. **4**(9).
4. Steffen, N.R., et al., *DNA sequence and structure: direct and indirect recognition in protein-DNA binding.* Bioinformatics, 2002. **18 Suppl 1**: p. S22-30.
5. Pugh, B.F. and D.S. Gilmour, *Genome-wide analysis of protein-DNA interactions in living cells.* Genome Biol, 2001. **2**(4): p. REVIEWS1013.
6. Zhang, J., Z. Ma, and L. Kurgan, *Comprehensive review and empirical analysis of hallmarks of DNA-, RNA- and protein-binding residues in protein chains.* Brief Bioinform, 2019. **20**(4): p. 1250-1268.
7. Marchese, D., et al., *Advances in the characterization of RNA-binding proteins.* Wiley Interdiscip Rev RNA, 2016. **7**(6): p. 793-810.
8. Nahalka, J., *Protein-RNA recognition: cracking the code.* J Theor Biol, 2014. **343**: p. 9-15.
9. Konig, J., et al., *Protein-RNA interactions: new genomic technologies and perspectives.* Nat Rev Genet, 2012. **13**(2): p. 77-83.
10. Zhang, T., et al., *Analysis and prediction of RNA-binding residues using sequence, evolutionary conservation, and predicted secondary structure and solvent accessibility.* Curr Protein Pept Sci, 2010. **11**(7): p. 609-28.
11. Chowdhury, S., J. Zhang, and L. Kurgan, *In Silico Prediction and Validation of Novel RNA Binding Proteins and Residues in the Human Proteome.* Proteomics, 2018: p. e1800064.
12. Cozzolino, F., et al., *Protein-DNA/RNA Interactions: An Overview of Investigation Methods in the -Omics Era.* J Proteome Res, 2021. **20**(6): p. 3018-3030.
13. Sudha, G., R. Nussinov, and N. Srinivasan, *An overview of recent advances in structural bioinformatics of protein-protein interactions and a guide to their principles.* Prog Biophys Mol Biol, 2014. **116**(2-3): p. 141-50.
14. Vakser, I.A., *Protein-protein docking: from interaction to interactome.* Biophys J, 2014.

**107**(8): p. 1785-1793.
15. Hsu, W.L., et al., *Intrinsic protein disorder and protein-protein interactions.* Pac Symp Biocomput, 2012: p. 116-27.
16. De Las Rivas, J. and C. Fontanillo, *Protein-protein interaction networks: unraveling the wiring of molecular machines within the cell.* Brief Funct Genomics, 2012. **11**(6): p. 489-96.
17. Meng, F., et al., *Compartmentalization and Functionality of Nuclear Disorder: Intrinsic Disorder and Protein-Protein Interactions in Intra-Nuclear Compartments.* Int J Mol Sci, 2015. **17**(1).
18. Chen, K. and L. Kurgan, *Investigation of atomic level patterns in protein--small ligand interactions.* PLoS One, 2009. **4**(2): p. e4473.
19. Dudev, T. and C. Lim, *Competition among Metal Ions for Protein Binding Sites: Determinants of Metal Ion Selectivity in Proteins.* Chemical Reviews, 2014. **114**(1): p. 538-556.
20. Peng, T., X. Yuan, and H.C. Hang, *Turning the spotlight on protein-lipid interactions in cells.* Curr Opin Chem Biol, 2014. **21**: p. 144-53.
21. wwPDB consortium, *Protein Data Bank: the single global archive for 3D macromolecular structure data.* Nucleic Acids Res, 2019. **47**(D1): p. D520-D528.
22. Nagarajan, R., et al., *Structure based approach for understanding organism specific recognition of protein-RNA complexes.* Biol Direct, 2015. **10**: p. 8.
23. Ellis, J.J., M. Broom, and S. Jones, *Protein-RNA interactions: structural analysis and functional classes.* Proteins, 2007. **66**(4): p. 903-11.
24. Prabakaran, P., et al., *Classification of protein-DNA complexes based on structural descriptors.* Structure, 2006. **14**(9): p. 1355-67.
25. Lejeune, D., et al., *Protein-nucleic acid recognition: statistical analysis of atomic interactions and influence of DNA structure.* Proteins, 2005. **61**(2): p. 258-71.
26. Wang, C., V.N. Uversky, and L. Kurgan, *Disordered nucleiome: Abundance of intrinsic disorder in the DNA- and RNA-binding proteins in 1121 species from Eukaryota, Bacteria and Archaea.* Proteomics, 2016. **16**(10): p. 1486-98.
27. Wu, Z., et al., *In various protein complexes, disordered protomers have large per-residue surface areas and area of protein-, DNA- and RNA-binding interfaces.* FEBS Lett, 2015. **589**(19 Pt A): p. 2561-9.
28. Varadi, M., et al., *Functional Advantages of Conserved Intrinsic Disorder in RNA-Binding Proteins.* PLoS One, 2015. **10**(10): p. e0139731.
29. Dyson, H.J., *Roles of intrinsic disorder in protein-nucleic acid interactions.* Mol Biosyst, 2012. **8**(1): p. 97-104.
30. Peng, Z., et al., *A creature with a hundred waggly tails: intrinsically disordered proteins in the ribosome.* Cell Mol Life Sci, 2014. **71**(8): p. 1477-504.
31. Lieutaud, P., 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): p. e1259708.
32. Oldfield, C.J., et al., *Introduction to intrinsically disordered proteins and regions*, in *Intrinsically Disordered Proteins*, N. Salvi, Editor. 2019, Academic Press. p. 1-34.
33. Habchi, J., et al., *Introducing protein intrinsic disorder.* Chem Rev, 2014. **114**(13): p.

6561-88.
34. Si, J., R. Zhao, and R. Wu, *An overview of the prediction of protein DNA-binding sites.* Int J Mol Sci, 2015. **16**(3): p. 5194-215.
35. Yan, J., S. Friedrich, and L. Kurgan, *A comprehensive comparative review of sequence-based predictors of DNA- and RNA-binding residues.* Brief Bioinform, 2016. **17**(1): p. 88-105.
36. Katuwawala, A. and L. Kurgan, *Comparative Assessment of Intrinsic Disorder Predictions with a Focus on Protein and Nucleic Acid-Binding Proteins.* Biomolecules, 2020. **10**(12).
37. Meng, F., V.N. Uversky, and L. Kurgan, *Comprehensive review of methods for prediction of intrinsic disorder and its molecular functions.* Cell Mol Life Sci, 2017. **74**(17): p. 3069-3090.
38. Si, J., et al., *Computational Prediction of RNA-Binding Proteins and Binding Sites.* Int J Mol Sci, 2015. **16**(11): p. 26303-17.
39. Zhao, H., Y. Yang, and Y. Zhou, *Prediction of RNA binding proteins comes of age from low resolution to high resolution.* Mol Biosyst, 2013. **9**(10): p. 2417-25.
40. Walia, R.R., et al., *Protein-RNA interface residue prediction using machine learning: an assessment of the state of the art.* BMC Bioinformatics, 2012. **13**: p. 89.
41. Gromiha, M.M. and R. Nagarajan, *Computational approaches for predicting the binding sites and understanding the recognition mechanism of protein-DNA complexes.* Adv Protein Chem Struct Biol, 2013. **91**: p. 65-99.
42. Wang, K., 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): p. 6879.
43. Li, W., et al., *RefSeq: expanding the Prokaryotic Genome Annotation Pipeline reach with protein family model curation.* Nucleic Acids Res, 2021. **49**(D1): p. D1020-D1028.
44. Katuwawala, A., S. Ghadermarzi, and L. Kurgan, *Computational prediction of functions of intrinsically disordered regions.* Prog Mol Biol Transl Sci, 2019. **166**: p. 341-369.
45. Kurgan, L., M. Li, and Y. Li, *The Methods and Tools for Intrinsic Disorder Prediction and their Application to Systems Medicine*, in *Systems Medicine*, O. Wolkenhauer, Editor. 2021, Academic Press: Oxford. p. 159-169.
46. Zhang, F., et al., *DeepDISOBind: accurate prediction of RNA-, DNA- and protein-binding intrinsically disordered residues with deep multi-task learning.* Brief Bioinform, 2022. **23**(1).
47. Georgel, P.T., et al., *Sir3-dependent assembly of supramolecular chromatin structures in vitro.* Proc Natl Acad Sci U S A, 2001. **98**(15): p. 8584-9.
48. McBryant, S.J., V.H. Adams, and J.C. Hansen, *Chromatin architectural proteins.* Chromosome Res, 2006. **14**(1): p. 39-51.
49. Quaglia, F., et al., *DisProt in 2022: improved quality and accessibility of protein intrinsic disorder annotation.* Nucleic Acids Res, 2022. **50**(D1): p. D480-D487.
50. Yang, J., A. Roy, and Y. Zhang, *BioLiP: a semi-manually curated database for biologically relevant ligand-protein interactions.* Nucleic Acids Res, 2013. **41**(Database issue): p. D1096-103.
51. Sickmeier, M., et al., *DisProt: the Database of Disordered Proteins.* Nucleic Acids Res,

2007. **35**(Database issue): p. D786-93.
52. Berman, H.M., et al., *The Protein Data Bank.* Nucleic Acids Res, 2000. **28**(1): p. 235-42.
53. Jeong, E., I.F. Chung, and S. Miyano, *A neural network method for identification of RNA-interacting residues in protein.* Genome Inform, 2004. **15**(1): p. 105-16.
54. Ahmad, S., M.M. Gromiha, and A. Sarai, *Analysis and prediction of DNA-binding proteins and their binding residues based on composition, sequence and structural information.* Bioinformatics, 2004. **20**(4): p. 477-486.
55. Oldfield, C.J., Z. Peng, and L. Kurgan, *Disordered RNA-Binding Region Prediction with DisoRDPbind.* Methods Mol Biol, 2020. **2106**: p. 225-239.
56. Peng, Z., et al., *Prediction of Disordered RNA, DNA, and Protein Binding Regions Using DisoRDPbind.* Methods Mol Biol, 2017. **1484**: p. 187-203.
57. Peng, Z. and L. Kurgan, *High-throughput prediction of RNA, DNA and protein binding regions mediated by intrinsic disorder.* Nucleic Acids Res, 2015. **43**(18): p. e121.
58. Zhang, J., S. Ghadermarzi, and L. Kurgan, *Prediction of protein-binding residues: dichotomy of sequence-based methods developed using structured complexes versus disordered proteins.* Bioinformatics, 2020. **36**(18): p. 4729-4738.
59. Wang, L. and S.J. Brown, *BindN: a web-based tool for efficient prediction of DNA and RNA binding sites in amino acid sequences.* Nucleic Acids Res, 2006. **34**(Web Server issue): p. W243-8.
60. Su, H., et al., *Improving the prediction of protein–nucleic acids binding residues via multiple sequence profiles and the consensus of complementary methods.* Bioinformatics, 2019. **35**(6): p. 930-936.
61. Qiu, J., et al., *ProNA2020 predicts protein–DNA, protein–RNA, and protein–protein binding proteins and residues from sequence.* Journal of molecular biology, 2020. **432**(7): p. 2428-2443.
62. Zhang, J., Q. Chen, and B. Liu, *NCBRPred: predicting nucleic acid binding residues in proteins based on multilabel learning.* Brief Bioinform, 2021. **22**(5).
63. Sun, Z., et al., *To improve the predictions of binding residues with DNA, RNA, carbohydrate, and peptide via multi-task deep neural networks.* IEEE/ACM transactions on computational biology and bioinformatics, 2021.
64. Nguyen, B.P., et al., *iProDNA-CapsNet: identifying protein-DNA binding residues using capsule neural networks.* Bmc Bioinformatics, 2019. **20**(23): p. 1-12.
65. Ahmad, S. and A. Sarai, *PSSM-based prediction of DNA binding sites in proteins.* BMC bioinformatics, 2005. **6**(1): p. 1-6.
66. Yan, C., et al., *Predicting DNA-binding sites of proteins from amino acid sequence.* BMC bioinformatics, 2006. **7**(1): p. 1-10.
67. Jeong, E. and S. Miyano, *A weighted profile based method for protein-RNA interacting residue prediction*, in *Transactions on Computational Systems Biology IV*. 2006, Springer. p. 123-139.
68. Ho, S.-Y., et al., *Design of accurate predictors for DNA-binding sites in proteins using hybrid SVM–PSSM method.* Biosystems, 2007. **90**(1): p. 234-241.
69. Hwang, S., Z. Gou, and I.B. Kuznetsov, *DP-Bind: a web server for sequence-based prediction of DNA-binding residues in DNA-binding proteins.* Bioinformatics, 2007. **23**(5): p. 634-636.

70. Ofran, Y., V. Mysore, and B. Rost, *Prediction of DNA-binding residues from sequence.* Bioinformatics, 2007. **23**(13): p. i347-i353.
71. Wang, Y., et al., *PRINTR: prediction of RNA binding sites in proteins using SVM and profiles.* Amino acids, 2008. **35**(2): p. 295-302.
72. Tong, J., P. Jiang, and Z.-h. Lu, *RISP: a web-based server for prediction of RNA-binding sites in proteins.* Computer methods and programs in biomedicine, 2008. **90**(2): p. 148-153.
73. Kumar, M., M.M. Gromiha, and G.P.S. Raghava, *Prediction of RNA binding sites in a protein using SVM and PSSM profile.* Proteins: Structure, Function, and Bioinformatics, 2008. **71**(1): p. 189-194.
74. Cheng, C.-W., et al., *Predicting RNA-binding sites of proteins using support vector machines and evolutionary information.* BMC bioinformatics, 2008. **9**(12): p. 1-19.
75. Wang, L., M.Q. Yang, and J.Y. Yang, *Prediction of DNA-binding residues from protein sequence information using random forests.* Bmc Genomics, 2009. **10**(1): p. 1-9.
76. Wu, J., et al., *Prediction of DNA-binding residues in proteins from amino acid sequences using a random forest model with a hybrid feature.* Bioinformatics, 2009. **25**(1): p. 30-35.
77. Gao, M. and J. Skolnick, *A threading-based method for the prediction of DNA-binding proteins with application to the human genome.* PLoS computational biology, 2009. **5**(11): p. e1000567.
78. Wang, L., et al., *BindN+ for accurate prediction of DNA and RNA-binding residues from protein sequence features.* BMC Systems Biology, 2010. **4**(1): p. 1-9.
79. Carson, M.B., R. Langlois, and H. Lu, *NAPS: a residue-level nucleic acid-binding prediction server.* Nucleic Acids Res, 2010. **38**(Web Server issue): p. W431-5.
80. Murakami, Y., et al., *PiRaNhA: a server for the computational prediction of RNA-binding residues in protein sequences.* Nucleic acids research, 2010. **38**(suppl_2): p. W412-W416.
81. Huang, Y.-F., et al. *Predicting RNA-binding residues from evolutionary information and sequence conservation*. in *BMC genomics*. 2010. Springer.
82. Liu, Z.-P., et al., *Prediction of protein–RNA binding sites by a random forest method with combined features.* Bioinformatics, 2010. **26**(13): p. 1616-1622.
83. Wang, C.-c., et al., *Identification of RNA-binding sites in proteins by integrating various sequence information.* Amino acids, 2011. **40**(1): p. 239-248.
84. Ma, X., et al., *Prediction of RNA-binding residues in proteins from primary sequence using an enriched random forest model with a novel hybrid feature.* Proteins: Structure, Function, and Bioinformatics, 2011. **79**(4): p. 1230-1239.
85. Choi, S. and K. Han. *Prediction of RNA-binding amino acids from protein and RNA sequences*. in *Bmc Bioinformatics*. 2011. BioMed Central.
86. Terribilini, M., et al., *RNABindR: a server for analyzing and predicting RNA-binding sites in proteins.* Nucleic acids research, 2007. **35**(suppl_2): p. W578-W584.
87. Ma, X., et al., *Sequence-based prediction of DNA-binding residues in proteins with conservation and correlation information.* IEEE/ACM transactions on computational biology and bioinformatics, 2012. **9**(6): p. 1766-1775.
88. Dey, S., et al., *Characterization and prediction of the binding site in DNA-binding proteins: improvement of accuracy by combining residue composition, evolutionary conservation and structural parameters.* Nucleic acids research, 2012. **40**(15): p. 7150-

7161.
89. Liu, R. and J. Hu, *DNABind: A hybrid algorithm for structure-based prediction of DNA-binding residues by combining machine learning-and template-based approaches.* Proteins: Structure, Function, and Bioinformatics, 2013. **81**(11): p. 1885-1899.
90. Walia, R.R., et al., *RNABindRPlus: a predictor that combines machine learning and sequence homology-based methods to improve the reliability of predicted RNA-binding residues in proteins.* PloS one, 2014. **9**(5): p. e97725.
91. Zhao, H., et al., *Predicting DNA-binding proteins and binding residues by complex structure prediction and application to human proteome.* PloS one, 2014. **9**(5): p. e96694.
92. Li, S., et al., *Quantifying sequence and structural features of protein–RNA interactions.* Nucleic acids research, 2014. **42**(15): p. 10086-10098.
93. Yang, X., et al., *SNBRFinder: a sequence-based hybrid algorithm for enhanced prediction of nucleic acid-binding residues.* PloS one, 2015. **10**(7): p. e0133260.
94. Ren, H. and Y. Shen, *RNA-binding residues prediction using structural features.* BMC bioinformatics, 2015. **16**(1): p. 1-10.
95. Tuvshinjargal, N., et al., *PRIdictor: protein–RNA interaction predictor.* Biosystems, 2016. **139**: p. 17-22.
96. Sun, M., et al., *Accurate prediction of RNA-binding protein residues with two discriminative structural descriptors.* BMC bioinformatics, 2016. **17**(1): p. 1-14.
97. Zhou, J., et al., *PDNAsite: identification of DNA-binding site from protein sequence by incorporating spatial and sequence context.* Scientific reports, 2016. **6**(1): p. 1-15.
98. Yan, J. and L. Kurgan, *DRNApred, fast sequence-based method that accurately predicts and discriminates DNA- and RNA-binding residues.* Nucleic Acids Res, 2017. **45**(10): p. e84.
99. Hu, J., et al., *Predicting Protein-DNA Binding Residues by Weightedly Combining Sequence-Based Features and Boosting Multiple SVMs.* IEEE/ACM Trans Comput Biol Bioinform, 2017. **14**(6): p. 1389-1398.
100. Tang, Y., et al., *A boosting approach for prediction of protein-RNA binding residues.* BMC Bioinformatics, 2017. **18**(Suppl 13): p. 465.
101. Zhu, Y.-H., et al., *DNAPred: accurate identification of DNA-binding sites from protein sequence by ensembled hyperplane-distance-based support vector machines.* Journal of chemical information and modeling, 2019. **59**(6): p. 3057-3071.
102. Zhou, J., et al., *EL_LSTM: prediction of DNA-binding residue from protein sequence by combining long short-term memory and ensemble learning.* IEEE/ACM transactions on computational biology and bioinformatics, 2018. **17**(1): p. 124-135.
103. Amirkhani, A., et al., *Prediction of DNA-binding residues in local segments of protein sequences with Fuzzy Cognitive Maps.* IEEE/ACM transactions on computational biology and bioinformatics, 2018. **17**(4): p. 1372-1382.
104. Zhang, J., et al., *DNAgenie: accurate prediction of DNA-type-specific binding residues in protein sequences.* Briefings in Bioinformatics, 2021. **22**(6): p. bbab336.
105. Hu, G. and L. Kurgan, *Sequence Similarity Searching.* Curr Protoc Protein Sci, 2019. **95**(1): p. e71.
106. Altschul, S.F., et al., *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.* Nucleic acids research, 1997. **25**(17): p. 3389-3402.

107. Remmert, M., et al., *HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment.* Nature methods, 2012. **9**(2): p. 173-175.
108. Buchan, D.W.A. and D.T. Jones, *The PSIPRED Protein Analysis Workbench: 20 years on.* Nucleic Acids Research, 2019. **47**(W1): p. W402-W407.
109. Faraggi, E., Y.Q. Zhou, and A. Kloczkowski, *Accurate single-sequence prediction of solvent accessible surface area using local and global features.* Proteins, 2014. **82**(11): p. 3170-3176.
110. Faraggi, E., et al., *SPINE X: improving protein secondary structure prediction by multistep learning coupled with prediction of solvent accessible surface area and backbone torsion angles.* J Comput Chem, 2012. **33**(3): p. 259-67.
111. Singh, J., et al., *SPOT-1D-Single: Improving the Single-Sequence-Based Prediction of Protein Secondary Structure, Backbone Angles, Solvent Accessibility and Half-Sphere Exposures using a Large Training Set and Ensembled Deep Learning.* Bioinformatics, 2021.
112. Yang, Y., et al., *SPIDER2: A Package to Predict Secondary Structure, Accessible Surface Area, and Main-Chain Torsional Angles by Deep Neural Networks.* Methods Mol Biol, 2017. **1484**: p. 55-63.
113. Zhang, H., et al., *Critical assessment of high-throughput standalone methods for secondary structure prediction.* Brief Bioinform, 2011. **12**(6): p. 672-88.
114. Ho, H.K., et al., *A survey of machine learning methods for secondary and supersecondary protein structure prediction.* Methods Mol Biol, 2013. **932**: p. 87-106.
115. Oldfield, C.J., K. Chen, and L. Kurgan, *Computational Prediction of Secondary and Supersecondary Structures from Protein Sequences.* Methods Mol Biol, 2019. **1958**: p. 73-100.
116. Meng, F. and L. Kurgan, *Computational Prediction of Protein Secondary Structure from Sequence.* Curr Protoc Protein Sci, 2016. **86**: p. 2 3 1-2 3 10.
117. Chen, K. and L. Kurgan, *Computational prediction of secondary and supersecondary structures.* Methods Mol Biol, 2013. **932**: p. 63-86.
118. Kurgan, L. and F.M. Disfani, *Structural protein descriptors in 1-dimension and their sequence-based predictions.* Curr Protein Pept Sci, 2011. **12**(6): p. 470-89.
119. Jiang, Q., et al., *Protein secondary structure prediction: A survey of the state of the art.* Journal of Molecular Graphics & Modelling, 2017. **76**: p. 379-402.
120. Pirovano, W. and J. Heringa, *Protein secondary structure prediction.* Methods Mol Biol, 2010. **609**: p. 327-48.
121. AlQuraishi, M., *AlphaFold at CASP13.* Bioinformatics, 2019. **35**(22): p. 4862-4865.
122. Jumper, J., et al., *Highly accurate protein structure prediction with AlphaFold.* Nature, 2021. **596**(7873): p. 583-589.
123. Schaarschmidt, J., et al., *Assessment of contact predictions in CASP12: Co-evolution and deep learning coming of age.* Proteins, 2018. **86 Suppl 1**: p. 51-66.
124. Guo, Z.Y., J. Hou, and J.L. Cheng, *DNSS2: Improved ab initio protein secondary structure prediction using advanced deep learning architectures.* Proteins-Structure Function and Bioinformatics, 2021. **89**(2): p. 207-217.
125. Lyu, Z., et al., *Protein Secondary Structure Prediction With a Reductive Deep Learning Method.* Frontiers in Bioengineering and Biotechnology, 2021. **9**(404).

126. Hanson, J., et al., *Improving prediction of protein secondary structure, backbone angles, solvent accessibility and contact numbers by using predicted contact maps and an ensemble of recurrent and residual convolutional neural networks.* Bioinformatics, 2019. **35**(14): p. 2403-2410.
127. Zhang, F., et al., *DeepFunc: A Deep Learning Framework for Accurate Prediction of Protein Functions from Protein Sequences and Interactions.* Proteomics, 2019. **19**(12): p. e1900019.
128. Kulmanov, M., et al., *DeepGO: predicting protein functions from sequence and interactions using a deep ontology-aware classifier.* Bioinformatics, 2018. **34**(4): p. 660-668.
129. Littmann, M., et al., *Embeddings from deep learning transfer GO annotations beyond homology.* Sci Rep, 2021. **11**(1): p. 1160.
130. Muller, C., O. Rabal, and C. Diaz Gonzalez, *Artificial Intelligence, Machine Learning, and Deep Learning in Real-Life Drug Design Cases.* Methods Mol Biol, 2022. **2390**: p. 383-407.
131. Kim, J., et al., *Comprehensive Survey of Recent Drug Discovery Using Deep Learning.* Int J Mol Sci, 2021. **22**(18).
132. Li, F., et al., *DeepCleave: a deep learning predictor for caspase and matrix metalloprotease substrates and cleavage sites.* Bioinformatics, 2020. **36**(4): p. 1057-1065.
133. Zhang, F., et al., *PROBselect: accurate prediction of protein-binding residues from proteins sequences via dynamic predictor selection.* Bioinformatics, 2020. **36**(Supplement_2): p. i735-i744.
134. Zhang, J. and L. Kurgan, *SCRIBER: accurate and partner type-specific prediction of protein-binding residues from proteins sequences.* Bioinformatics, 2019. **35**(14): p. i343-i353.
135. Zhang, J. and L. Kurgan, *Review and comparative assessment of sequence-based predictors of protein-binding residues.* Brief Bioinform, 2018. **19**(5): p. 821-837.