# BindUP-Alpha: A Webserver for Predicting DNA-and RNA-binding Proteins based on Experimental and Computational Structural Models☆

**Dina Alexandrovich [1], Shani Kagan [2], and Yael Mandel-Gutfreund [1,2,*]**

*1 - **Technion-Israel Institute of Technology,** Faculty of Biology, Emerson Building, Haifa, Israel*
*2 - **Technion- Israel Institute of Technology,** Faculty of Computer Science, Taub Building, Haifa, Israel*

***Correspondence to Yael Mandel-Gutfreund:*** Technion-Israel Institute of Technology, Faculty of Biology, Emerson Building, Haifa, Israel. *yaelmg@technion.ac.il (Y. Mandel-Gutfreund)*
https://doi.org/10.1016/j.jmb.2025.169240
***Edited by David Mathews***

## Abstract

Structural data provides important information on the proteins' function. Recent development of advanced machine learning and artificial intelligence tools, such as AlphaFold, have led to an explosion of predicted protein structures. However, many of the computed protein models contain unstructured and disordered regions, posing challenges in protein function characterization. Here we present BindUP-Alpha, an upgraded webserver for predicting nucleic acid binding proteins. Our structure-based algorithm utilizes the electrostatic features of the protein surface and other physiochemical and structural properties extracted from the protein sequence. Using a Support Vector Machine (SVM) learning approach, BindUP-Alpha successfully predicts DNA- and RNA-binding proteins from both experimentally solved structures and predicted models. In addition, BindUP-Alpha identifies electrostatic patches on the protein's surface that represent potential nucleic-acid binding interfaces. BindUP-Alpha is freely accessible at https://bindup.technion.ac.il, providing interactive three-dimensional visualizations and downloadable text-based results.


## Introduction

Nucleic Acid Binding Proteins (NABPs), encompassing both DNA and RNA binding proteins, are essential for regulating gene expression and influence every stage of the process. DNA-binding proteins (DBPs) serve as transcriptional regulators, playing crucial roles in epigenetic regulation and genome organization [1,2], while RNA-binding proteins (RBPs) are primarily involved in post-transcriptional regulatory mechanisms such as splicing, RNA stability, polyadenylation, and translation [3].

High-resolution structures of DBPs, RBPs, and their complexes with nucleic acids, determined through experimental methods (i.e. X-ray crystallography, NMR and cryo-electron microscopy), have provided crucial information on the unique properties of NABPs and their diverse binding modes [4]. The advent of Artificial Intelligence-based structure prediction tools, such as AlphaFold [5–7] and RoseTTAFold [8] has further expanded our knowledge, offering unprecedented insights into how these proteins recognize and interact with their molecular partners. Despite the significant advancements of computational approaches in predicting large biomolecular structures, predicting nucleic acid-binding proteins and their binding surfaces remains a challenge. Nevertheless, the availability of a large and diverse set of high-resolution predicted protein structures enhances our ability to infer their functions and interactions within cellular networks. Over time, computational methods utilizing machine learning

and more recently, deep learning approaches such as Graph Convolutional Networks (GNNs), have been developed to predict DNA- and RNA-binding sites on proteins and NA-binding functions (reviewed in [9,10]). Most of these approaches leverage features derived from protein sequences, including motifs, physicochemical properties, and evolutionary conservation [11,12]. Some models incorporate three-dimensional (3D) structural information, broadening their predictive power [13,14].

Molecular recognition is fundamentally driven by the electrostatic properties of the protein surface [15]. These properties have been utilized for predicting nucleic acid binding interfaces [16,17]. Previously, we introduced BindUP [18] a tool designed to extract the largest electrostatic (positive and negative) patches on the protein surfaces, utilizing the PatchFinder algorithm [19,20]. Furthermore, BindUP facilitates the prediction of nucleic acid-binding function relying on the electrostatic and physiochemical properties associated with the assigned positive and negative patches on the protein surface. Until recently, this prediction approach was constrained to a limited set of proteins experimentally resolved or inferred through homology, either available from the Protein Data Bank (PDB) [21] or provided by the user. With the expansion of the PDB, which now contains approximately 200,000 structures, and the release of more than 200 million predicted protein structures available through the AlphaFold Protein Structure Database (AlphaFold DB) [22], we have significantly enhanced the scope of BindUP. The upgraded version, now called BindUP-Alpha, leverages these extensive resources to offer improved predictions and broader applicability to any protein of interest. In addition to the predictive value, the unique utility of BindUP-Alpha to extract electrostatic patches from high resolution structures as well as from structural models provides the user additional insight on the protein's NA-binding function and mode of binding. BindUP-Alpha is freely accessible through the website https://bindup.technion.ac.il.

## BindUP-Alpha Methodology

The BindUP-Alpha website provides information on the electrostatic patches on protein surfaces and predicts NA-binding propensity given a 3D structure or a structural model of a protein. It is based on the BindUP algorithm, which was previously developed for extracting surface electrostatic patches and predicting NA-binding function from high resolution protein or domain structures [18]. The new BindUP-Alpha algorithm was developed for extracting patches and predicting NABPs from a structural model of the protein, such as those from AlphaFold DB [22] and is applicable for any protein in the protein database. The new algorithm incorporates two independent structure-based machine learning methods.

Both methods calculate the structural and electrostatic features of the protein, based on its 3D representation. To calculate the surface features, we first extract the positive and negative surface patches using the PatchFinder algorithm [20] originally developed for annotating DNA- [23] and RNA-binding [19] proteins. For experimentally solved protein structures, physiochemical features and surface patches are calculated per individual protein chain, as defined in the PDB. A detailed description of the patch and structure-based features can be found in Shazman et al., [20]. For predicted protein structures, when only a protein sequence and a structural model are available, the physiochemical features and the surface patches are calculated from the largest well-defined domain, obtained from the 3D structural model of the protein. Such domains are defined as a continuous sequence of at least 50 amino acids, with a Predicted Local Distance Difference Test (pLDDT) score greater than 0.65. In addition to the structure-based features, we calculate diverse properties derived from the protein sequence. For normalization purposes we split each protein to 10 equal segments, as well as the C' and N' terminals (100 amino acids each). For each segment, we calculate the intrinsic disorder propensities (IDR) using AIUpred [24], the percentage of positive, negative, and aromatic amino acids, as well as other sequence and biochemical features.

The calculated features are combined into a feature vector, which is then fed into a Support Vector Machine (SVM) to predict whether the protein is an NABP. For experimental models, we employ the SVM algorithm as applied in BindUP [18]. For predicted models, we use the SVM classifier as implemented in the Scikit-learn Python module https://scikit-learn.org/stable/. The classifier was trained and tested on a highly curated dataset, as described in the results section below. An overview of the BindUP-Alpha pipeline is depicted in Figure 1.

## Web Interface

BindUP-Alpha has two main modes – an "Experimental Model" mode and "Predicted Model" mode. Each mode can be used either for a single protein or in batch processing. For experimental structures, users can provide either a PDB ID or a coordinate file in PDB/mmCIF format (https://www.rcsb.org/). In case the input is provided as a PDB ID, BindUP-Alpha retrieves the results from a precalculated database. Users can choose to process all protein chains in the structure (each chain is processed separately) or to select a specific chain identifier. For predicted structures, users can provide any protein from the Universe Protein knowledgebase (UNIPROT) [26] either as a UniProt ID (e.g. Q9UBC3) or an AlphaFold model ID, named also Computed Structure

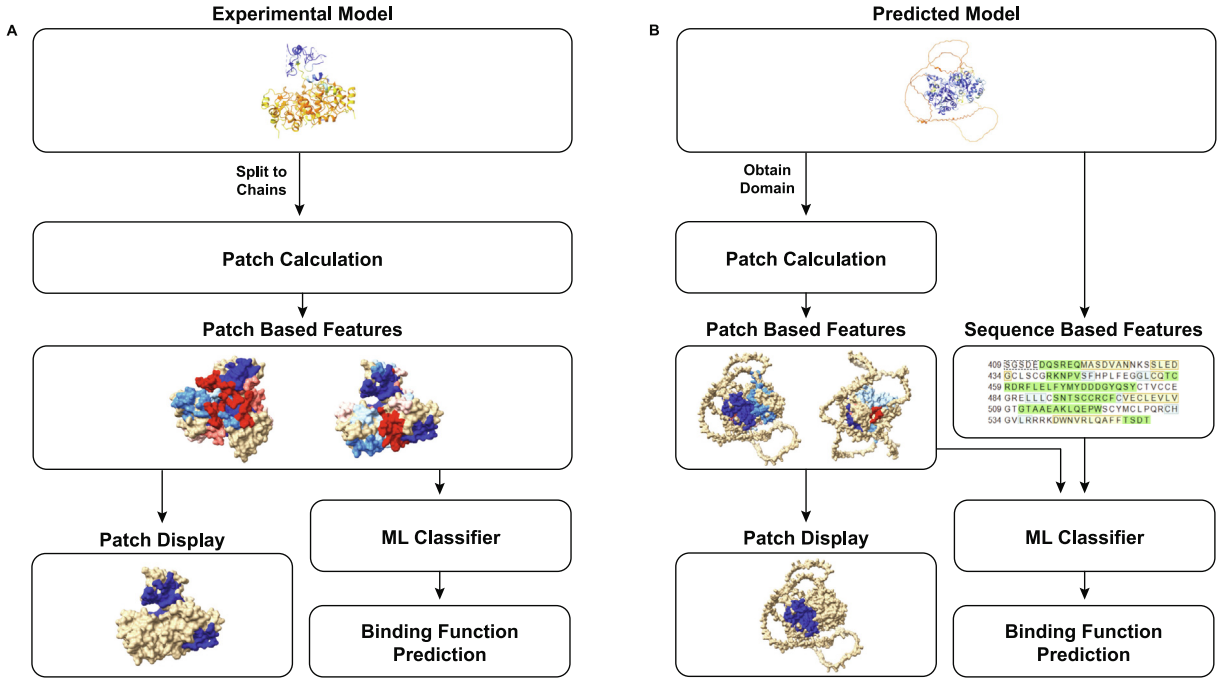


**Figure 1.** An overview of the BindUP-Alpha algorithm. (A) A presentation of the "Experimental Model" mode. PDB files are first split into separate chains. For each chain, we calculate structure-based features to determine the electrostatic patches on the protein's surface. The patches are displayed on the 3D representation of the protein. The features are fed to a Linear SVM, which predicts the chain's binding function. (B) A presentation of the "Predicted Model" mode. First, we obtain a well-defined domain and calculate structure-based features to determine the electrostatic patches on the protein's surface. The patches are displayed on the 3D representation of the protein. Then, we calculate sequence-based features. Both structure-based and sequence-based features are fed to an RBF SVM, which predicts the protein's binding function.

Model (CSM) (e.g AF-Q9UBC3-F1-v4). By default, BindUP-Alpha outputs the largest positive electrostatic patch and the NA-binding prediction. Additional processing options allow users to customize the type and number of electrostatic patches displayed in the results. In batch processing, it is possible to input a list of PDB IDs, UniProt IDs or CSM IDs, pasted into the browser or uploaded as a text file. A PDB ID can be followed by a chain identifier to specify a specific chain on which the calculation is required.

For both modes, BindUP-Alpha outputs the electrostatic patches and the NA-binding prediction in both graphic and text representations. The 3D graphics are generated using ChimeraX [27]. In the "Experimental Model" mode, BindUP-Alpha provides results for the entire protein and for each protein chain separately, while in the "Predicted Model" mode it provides a single result based on calculations performed on the largest well-defined domain. In addition to the graphical output, BindUP-Alpha also provides the patch results as downloadable files. Three files are provided for the entire protein: one showing the NA-binding prediction, its confidence score, and the residues composing the requested electrostatic patches; the second is a coordinate mmCIF file with the patch annotation in the B-factor column (the color-coding is described in the manual section of the website); and the third file includes the ChimeraX script used to create the graphic visualization. Additionally, in the "Experimental Model" mode, these three files are provided for each chain. In batch processing, results for each protein structure are available only as downloadable files.

## Results

As aforementioned, BindUP-Alpha predicts NABPs using a machine learning model trained and tested on two independent, highly curated datasets. For experimental models, where features are calculated per individual chain, we implement the GIST SVM classifier [25] that was trained with a linear kernel on a manually curated non-redundant set (<25 % sequence identity) of 450 protein chains, acquired from the PDB database. This set, generated originally in [18], includes 90 DNA-binding chains and 60 RNA-binding chains extracted from protein-NA complexes and 300 non-NA-binding chains (see Supplementary Table 1). For predicted models, where features are calculated based on the largest well-defined domain, we implement an SVM with an RBF kernel that

was trained on a non-redundant set (<35 % sequence identity) of 9301 proteins from [14], including 3365 DNA-binding sequences, 2617 RNA-binding sequences and 3319 non-NA-binding sequences (see Supplementary Table 1). The coordinate files used in the latter set were acquired from the AlphaFold DB (https://alphafold.ebi.ac.uk/). Cross-validation was applied to determine the parameters for this classifier and grid search to find the best combination of those parameters. The linear SVM was tested on five independent datasets, achieving Area Under ROC Curve (AUC) values ranging from 0.83 to 0.96 (for details see [18]). The RBF SVM was tested using 10-fold cross-validation, achieving an Area Under ROC Curve (AUC) value of 0.87 (Figure 2). Analysis of feature importance revealed that molecular weight, patch surface accessibility, and the ratio of large hydrophobic amino acids were among the features that contributed most significantly to the classification decision. A full list of feature importance scores is provided in Supplementary Table 2.

An example of an RBP predicted by BindUP-Alpha is RBM25. RBM25 is a highly conserved splicing factor [28] that is highly expressed in embryonic stem cells and cancer cells [29]. RBM25 contains a PWI domain, an RNA/DNA-binding domain, which was solved experimentally by X-ray crystallography (PDB ID: 3v53) [30] and is successfully predicted in BindUP-Alpha as NA-binding using the "Experimental Model" mode (Figure 3A). Another example of an experimentally solved protein structure predicted by BindUP-Alpha correctly as an NABP is the DNA methyltransferase 3B (DNMT3B). DNMT3B is a DBP required for genome-wide de-novo methylation and is essential for the establishment of DNA methylation patterns during early development [31]. Figure 3B demonstrates the results of BindUP-Alpha for the crystal structure of the human DNMT3B/DNMT3L complex with DNA (PDB ID: 6kda) [32]. As shown, the structures assigned as chains A and D, representing DNMT3B, are predicted as NA-binding, while chains B and C, which represent DNMT3L that does not bind directly to the DNA, are predicted as non-NA-binding. Further, when applying BindUP-Alpha on the AlphaFold structural model of the entire DNMT3B protein (CSM ID: AF-Q9UBC3-F1-v4), using the "Predicted Model" mode, we successfully predict it as a NA-binding protein (Figure 3C). In contrast, a protein for which only a sequence and a predicted model are available, such as the E3 ubiquitin-protein ligase TM129 (UniProt ID: A0AVI4) [33], is successfully predicted as non-NA-binding, as shown in Figure 3D. Moreover, consistent with the prediction of TM129 as a non-NABP, BindUP-Alpha obtained two large negative patches on the surface of the predicted protein structure (Figure 3D).

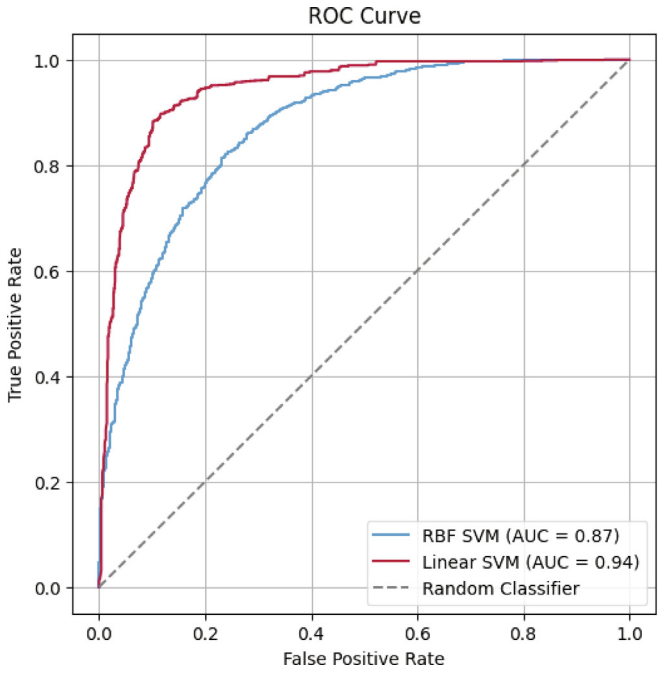


**Figure 2.** ROC curves representing the performances of two SVM models for predicting nucleic-acid binding proteins. The blue line represents the results of the linear SVM model trained on a manually curated dataset of 450 experimentally solved protein chains and tested on an independent dataset, with an AUC value of 0.94. The red line represents the results of the predicted model trained and tested on a dataset of 9436 predicted protein structures using an RBF SVM with 10-fold cross-validation, achieving an AUC value of 0.87 (for detail see Supplementary Table 1).

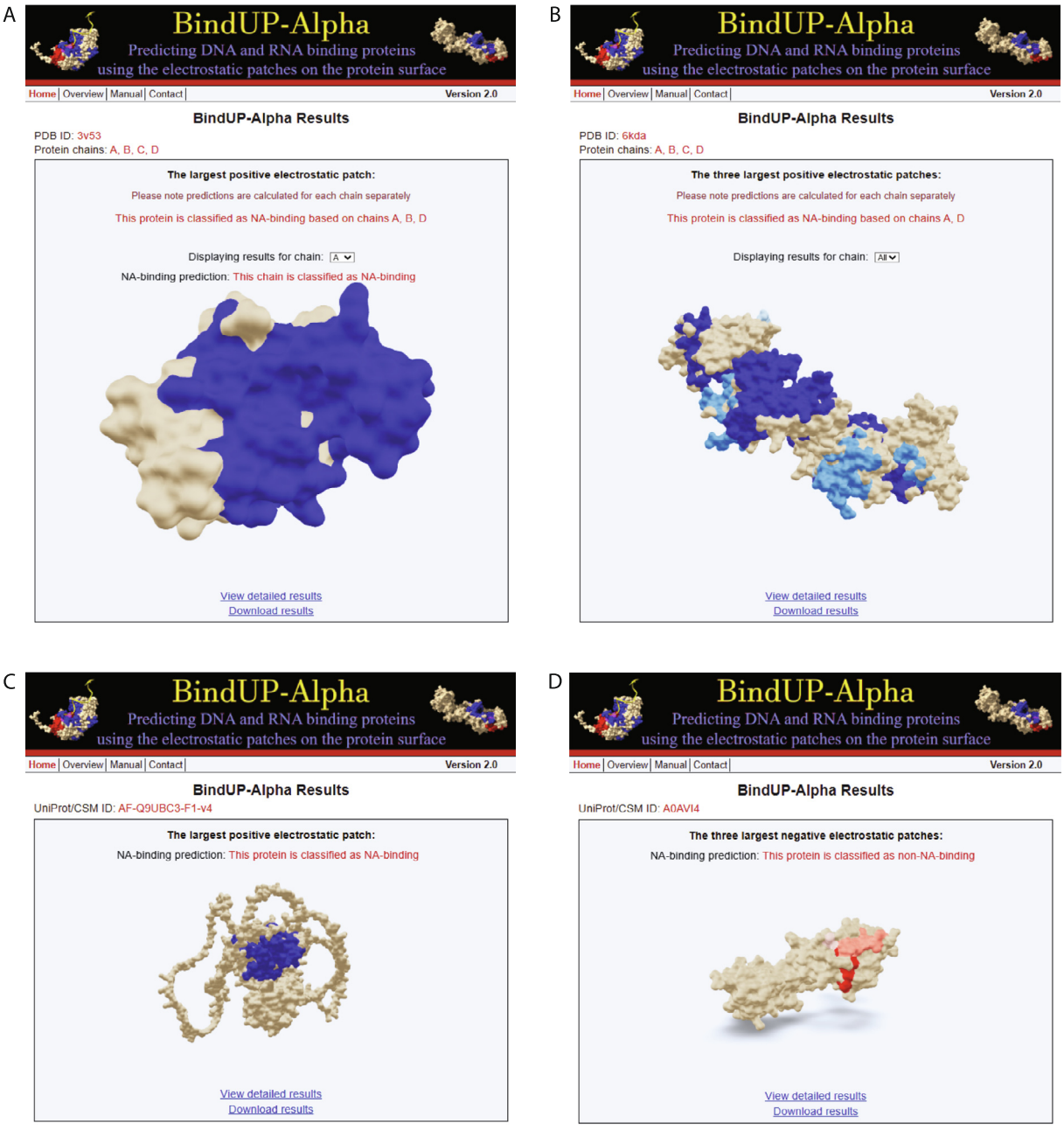


**Figure 3.** Selected examples of BindUP-Alpha webserver results. (A) A presentation of the largest positive patch, calculated on an experimental model of RBM25 using PDB ID 3v53A. The chain is predicted to be NA-binding. (B) A presentation of the 3 largest positive patches, calculated on an experimental model of DNMT3B (PDB ID 6kda). (C) A presentation of the largest positive patch, calculated on the AlphaFold predicted model of DNMT3B (CSM ID AF-Q9UBC3-F1-v4). The protein is predicted to be NA-binding. (D) A presentation of the three largest negative patches, calculated on predicted model of TM129, using UniProt ID A0AVI4. The protein is predicted to be non-NA-binding.

Overall, BindUP-Alpha is a highly valuable web server offering a user-friendly way to predict DNA- and RNA-binding proteins and their putative binding interfaces using experimental and computational structural models. We anticipate that the inclusion of more model types and prediction methods will enhance the prediction ability, resulting in more accurate DNA- and RNA-binding annotations, further solidifying BindUP-Alpha's role as an essential resource in the study of NA-protein interactions.

## Funding Sources

This research was supported by the ISF grant number 1556/22.

## CRediT authorship contribution statement

**Dina Alexandrovich:** Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Shani Kagan:** Validation, Software, Methodology, Formal analysis, Data curation. **Yael Mandel-Gutfreund:** Writing – review & editing, Supervision, Methodology, Investigation, Funding acquisition, Conceptualization.

### DECLARATION OF COMPETING INTEREST

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

## Acknowledgements

This work was supported by the Israel Science Foundation (ISF), grant number 1556/22 granted to YMG. We would like to thank Mais Khoury and Fabian Glaser for their contribution to data curation. We would also like to thank Ohad Aharon for his invaluable assistance.

## Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmb.2025.169240.

*Received 13 February 2025;*
*Accepted 26 May 2025;*
*Available online 29 May 2025*

***Keywords***:
DNA-binding proteins;
RNA-binding proteins;
function prediction;
machine learning

## References

[1]. Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., Weirauch, M.T., (2018). The human transcription factors. *Cell* **172**, 650–665. https://doi.org/10.1016/j.cell.2018.01.029.

[2]. Sivakumar, A., de las Heras, J.I., Schirmer, E.C., (2019). Spatial genome organization: from development to disease. *Front. Cell Dev. Biol.* **7** https://doi.org/10.3389/fcell.2019.00018.

[3]. Corley, M., Burns, M.C., Yeo, G.W., (2020). How RNA-binding proteins interact with RNA: molecules and mechanisms. *Mol. Cell* **78**, 9–29. https://doi.org/10.1016/j.molcel.2020.03.011.

[4]. Wolberger, C., (2021). How structural biology transformed studies of transcription regulation. *J. Biol. Chem.* **296**, 100741. https://doi.org/10.1016/j.jbc.2021.100741.

[5]. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., Bodenstein, S.W., Evans, D.A., Hung, C.-C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., Beattie, C., Bertolli, O., Bridgland, A., Cherepanov, A., Congreve, M., Cowen-Rivers, A.I., Cowie, A., Figurnov, M., Fuchs, F.B., Gladman, H., Jain, R., Khan, Y.A., Low, C.M.R., Perlin, K., Potapenko, A., Savy, P., Singh, S., Stecula, A., Thillaisundaram, A., Tong, C., Yakneen, S., Zhong, E. D., Zielinski, M., Žídek, A., Bapst, V., Kohli, P., Jaderberg, M., Hassabis, D., Jumper, J.M., (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500. https://doi.org/10.1038/s41586-024-07487-w.

[6]. Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G.R., Wang, J., Cong, Q., Kinch, L. N., Schaeffer, R.D., Millán, C., Park, H., Adams, C., Glassman, C.R., DeGiovanni, A., Pereira, J.H., Rodrigues, A.V., van Dijk, A.A., Ebrecht, A.C., Opperman, D.J., Sagmeister, T., Buhlheller, C., Pavkov-Keller, T., Rathinaswamy, M.K., Dalwadi, U., Yip, C.K., Burke, J.E., Garcia, K.C., Grishin, N.V., Adams, P.D., Read, R.J., Baker, D., (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876. https://doi.org/10.1126/science.abj8754.

[7]. Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, Y., Dos Santos Costa, A., Fazel-Zarandi, M., Sercu, T., Candido, S., Rives, A., (2023). Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **379**, 1123–1130. https://doi.org/10.1126/science.ade2574.

[8]. Lisanza, S.L., Gershon, J.M., Tipps, S.W.K., Sims, J.N., Arnoldt, L., Hendel, S.J., Simma, M.K., Liu, G., Yase, M., Wu, H., Tharp, C.D., Li, X., Kang, A., Brackenbrough, E., Bera, A.K., Gerben, S., Wittmann, B.J., McShan, A.C., Baker, D., (2024). Multistate and functional protein design using RoseTTAFold sequence space diffusion. *Nat. Biotechnol.*. https://doi.org/10.1038/s41587-024-02395-w.

[9]. Zhang, Y., Bao, W., Cao, Y., Cong, H., Chen, B., Chen, Y., (2022). A survey on protein-DNA-binding sites in computational biology. *Brief Funct. Genomics* **21**, 357–375. https://doi.org/10.1093/bfgp/elac009.

[10]. Dvir, S., Argoetti, A., Mandel-Gutfreund, Y., (2018). Ribonucleoprotein particles: advances and challenges in computational methods. *Curr. Opin. Struct. Biol.* **53** https://doi.org/10.1016/j.sbi.2018.08.002.

[11]. Hu, S., Ma, R., Wang, H., (2019). An improved deep learning method for predicting DNA-binding proteins based on contextual features in amino acid sequences. *PLoS One* **14**, e0225317. https://doi.org/10.1371/journal.pone.0225317.

[12]. Wu, S., Guo, J.-T., (2024). Improved prediction of DNA and RNA binding proteins with deep learning models. *Brief. Bioinform.* **25** https://doi.org/10.1093/bib/bbae285.

[13]. Gligorijević, V., Renfrew, P.D., Kosciolek, T., Leman, J.K., Berenberg, D., Vatanen, T., Chandler, C., Taylor, B.C.,

Fisk, Vlamakis, H., Xavier, R.J., Knight, R., Cho, K., Bonneau, R., (2021). Structure-based protein function prediction using graph convolutional networks. *Nat. Commun.* **12**, 3168. https://doi.org/10.1038/s41467-021-23303-9.

[14]. Sagendorf, J.M., Mitra, R., Huang, J., Chen, X.S., Rohs, R., (2024). Structure-based prediction of protein-nucleic acid binding using graph neural networks. *Biophys. Rev.* **16**, 297–314. https://doi.org/10.1007/s12551-024-01201-w.

[15]. Honig, B., Nicholls, A., (1995). Classical electrostatics in biology and chemistry. *Science* **268**, 1144–1149. https://doi.org/10.1126/science.7761829.

[16]. Jones, S., Shanahan, H.P., Berman, H.M., Thornton, J. M., (2003). Using electrostatic potentials to predict DNA-binding sites on DNA-binding proteins. *Nucleic Acids Res.* **31**, 7189–7198. https://doi.org/10.1093/nar/gkg922.

[17]. Pérez-Cano, L., Fernández-Recio, J., (2010). Optimal protein-RNA area, OPRA: a propensity-based method to identify RNA-binding sites on proteins. *Proteins* **78**, 25–35. https://doi.org/10.1002/prot.22527.

[18]. Paz, I., Kligun, E., Bengad, B., Mandel-Gutfreund, Y., (2016). BindUP: a web server for non-homology-based prediction of DNA and RNA binding proteins. *Nucleic Acids Res.* **44** https://doi.org/10.1093/NAR/GKW454.

[19]. Shazman, S., Mandel-Gutfreund, Y., (2008). Classifying RNA-binding proteins based on electrostatic properties. *PLoS Comput. Biol.* **4** https://doi.org/10.1371/journal.pcbi.1000146.

[20]. Shazman, S., Celniker, G., Haber, O., Glaser, F., Mandel-Gutfreund, Y., (2007). Patch finder plus (PFplus): a web server for extracting and displaying positive electrostatic patches on protein surfaces. *Nucleic Acids Res.* **35**, W526–W530.

[21]. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., (2000). The protein data bank. *Nucleic Acids Res.* **28**, 235–242 http://www.ncbi.nlm.nih.gov/pmc/articles/PMC102472/pdf/gkd090.pdf.

[22]. Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., Žídek, A., Green, T., Tunyasuvunakool, K., Petersen, S., Jumper, J., Clancy, E., Green, R., Vora, A., Lutfi, M., Figurnov, M., Cowie, A., Hobbs, N., Kohli, P., Kleywegt, G., Birney, E., Hassabis, D., Velankar, S., (2022). AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444. https://doi.org/10.1093/nar/gkab1061.

[23]. Stawiski, E.W., Gregoret, L.M., Mandel-Gutfreund, Y., (2003). Annotating nucleic acid-binding function based on protein structure. *J. Mol. Biol.* **326** https://doi.org/10.1016/S0022-2836(03)00031-7.

[24]. Erdős, G., Dosztányi, Z., (2024). AIUPred: combining energy estimation with deep learning for the enhanced prediction of protein disorder. *Nucleic Acids Res.* **52**, W176–W181. https://doi.org/10.1093/nar/gkae385.

[25]. Pavlidis, P., Wapinski, I., Noble, W.S., (2004). Support vector machine classification on the web. *Bioinformatics* **20**, 586–587. https://doi.org/10.1093/bioinformatics/btg461.

[26]. UniProt Consortium, (2025). UniProt: the universal protein knowledgebase in 2025. *Nucleic Acids Res.* **53**, D609–D617. https://doi.org/10.1093/nar/gkae1010.

[27]. Meng, E.C., Goddard, T.D., Pettersen, E.F., Couch, G.S., Pearson, Z.J., Morris, J.H., Ferrin, T.E., ChimeraX, U.C. S.F., (2023). Tools for structure building and analysis. *Protein Sci.* **32**, e4792. https://doi.org/10.1002/pro.4792.

[28]. Carlson, S.M., Soulette, C.M., Yang, Z., Elias, J.E., Brooks, A.N., Gozani, O., (2017). RBM25 is a global splicing factor promoting inclusion of alternatively spliced exons and is itself regulated by lysine mono-methylation. *J. Biol. Chem.* **292**, 13381–13390. https://doi.org/10.1074/jbc.M117.784371.

[29]. Dvir, S., Argoetti, A., Lesnik, C., Roytblat, M., Shriki, K., Amit, M., Hashimshony, T., Mandel-Gutfreund, Y., (2021). Uncovering the RNA-binding protein landscape in the pluripotency network of human embryonic stem cells. *Cell Rep.* **35** https://doi.org/10.1016/j.celrep.2021.109198.

[30]. Gong, D., Yang, F., Li, F., Qian, D., Wu, M., Shao, Z., Wu, M., Wu, J., Shi, Y., (2013). Crystal structure and functional characterization of the human RBM25 PWI domain and its flanking basic region. *Biochem. J.* **450**, 85–94. https://doi.org/10.1042/BJ20121382.

[31]. Tóth, D.M., Szeri, F., Ashaber, M., Muazu, M., Székvölgyi, L., Arányi, T., (2025). Tissue-specific roles of de novo DNA methyltransferases. *Epigenet. Chromatin* **18**, 5. https://doi.org/10.1186/s13072-024-00566-2.

[32]. Lin, C.-C., Chen, Y.-P., Yang, W.-Z., Shen, J.C.K., Yuan, H.S., (2020). Structural insights into CpG-specific DNA methylation by human DNA methyltransferase 3B. *Nucleic Acids Res.* **48**, 3949–3961. https://doi.org/10.1093/nar/gkaa111.

[33]. van de Weijer, M.L., Bassik, M.C., Luteijn, R.D., Voorburg, C.M., Lohuis, M.A.M., Kremmer, E., Hoeben, R.C., LeProust, E.M., Chen, S., Hoelen, H., Ressing, M.E., Patena, W., Weissman, J.S., McManus, M.T., Wiertz, E.J.H.J., Lebbink, R.J., (2014). A high-coverage shRNA screen identifies TMEM129 as an E3 ligase involved in ER-associated protein degradation. *Nat. Commun.* **5**, 3832. https://doi.org/10.1038/ncomms4832.