Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

# Databases and web-based tools for studying structures of protein-nucleic acid complexes

Justas Dapkūnas and Česlovas Venclovas

Structural data on protein-DNA and protein-RNA interactions are indispensable in molecular biology research. In this article, we review available databases and other web-based resources devoted to 3D structures of protein-nucleic acid complexes. First, we describe the core databases that collect and disseminate experimental data. We then review derivative databases focused specifically on structural data on protein-nucleic acid interactions. Finally, we provide an overview of several useful web servers for structure prediction, analysis and comparison. Tools for investigating protein-nucleic acid complexes are relatively scarce. This is primarily because the methods that integrate structural information from both proteins and nucleic acids are in short supply. However, the emerging AI-driven techniques for structure prediction are expected to boost the development of such methods.

**Address**
Institute of Biotechnology, Life Sciences Center, Vilnius University, Saulėtekio 7, LT-10257 Vilnius, Lithuania

Corresponding author: Venclovas, Česlovas (ceslovas.venclovas@bti.vu.lt)

**Current Opinion in Structural Biology** 2025, **94**:103079

This review comes from a themed issue on **Protein Nucleic Acid Interactions (2025)**

Edited by **Shandar Ahmad** and **Elodie Laine**

For a complete overview see the Issue and the Editorial

Available online xxx

**https://doi.org/10.1016/j.sbi.2025.103079**



## Introduction

Protein-nucleic acid complexes play important roles in many biological processes, including DNA replication and repair, RNA processing, gene regulation, and protein synthesis. Three-dimensional (3D) structures of these complexes offer atomic-level insights into protein interactions with DNA and RNA, helping researchers decipher the underlying molecular mechanisms. Over the years, a variety of structures for protein-nucleic acid complexes have been determined. However, the number of experimentally determined protein-nucleic acid structures is still much smaller than that of protein–protein assemblies [1,2]. Consequently, fewer databases, web servers, and resources are dedicated to protein-nucleic acid structures compared to protein structures.

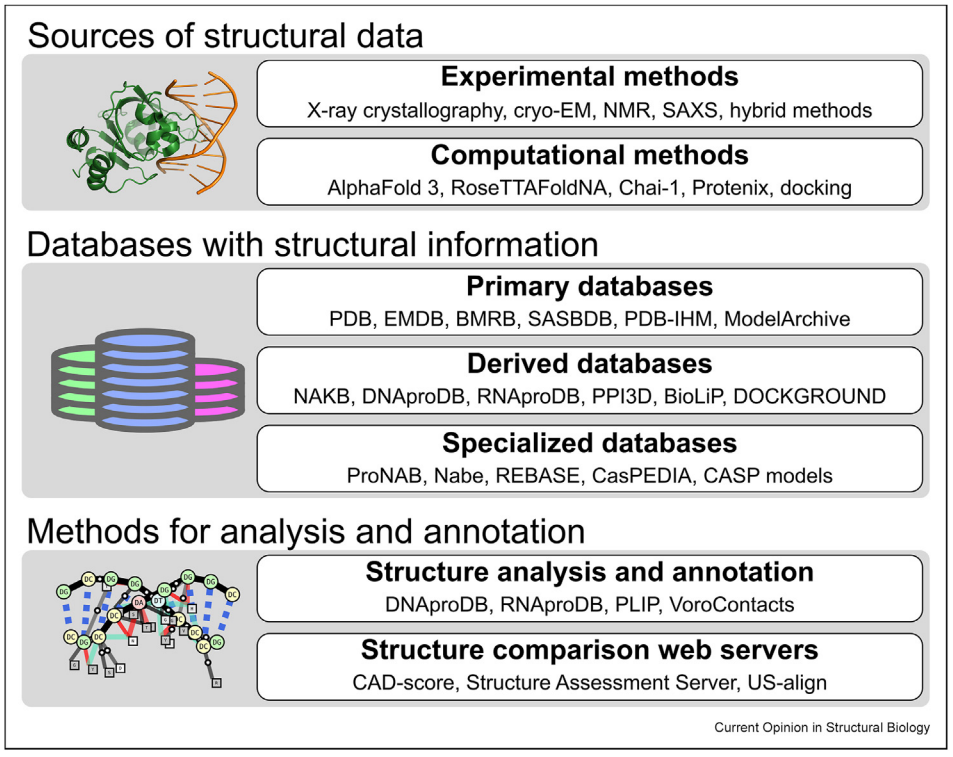
Recently, the structural coverage of the protein universe has greatly expanded due to AI-driven breakthroughs in structure prediction of monomeric [3] and multimeric [4] proteins. Millions of high-confidence structural models became available in addition to experimentally determined protein structures [5,6]. The achievements in protein structure prediction have also fueled similar research efforts for other types of biomolecules and their assemblies, including protein-nucleic acid complexes [7–9]. While structural models for protein-nucleic acid complexes remain less accurate than those for proteins, emerging modeling capabilities have already stimulated the development of specialized methods and databases (Figure 1).

Here, we review the available structural data resources for protein-nucleic acid complexes, focusing on recent developments and updates. We begin by describing the core databases that collect and disseminate experimental data, also briefly touching upon repositories of computational structure models (Table 1). Next, we provide an overview of several derivative databases and other useful data resources and finish with the web-based tools for prediction and analysis of structures for protein-nucleic acid complexes (Table 2).

## Primary databases of experimental and predicted structures

The key resource in structural biology is the Protein Data Bank (PDB) [1]. PDB collects and makes freely available structures and corresponding metadata for biological macromolecules and their complexes solved primarily by X-ray crystallography, cryo-EM and NMR. As the name implies, PDB was intended for proteins, but now it also hosts other structures, including protein-nucleic acid complexes. At the time of writing, PDB had close to 15,000 entries for protein-nucleic acid complexes, making them just over 6 % of the total PDB entries (>235,000) and much less abundant than protein–protein multimers (~200,000). The PDB data is collected and distributed by three regional data centers located in the United States (RCSB PDB) [10], Europe (PDBe) [11], and Japan (PDBj) [12].

**Figure 1**




Selected methods and databases for studying protein-nucleic acid complexes.

Importantly, these three sites serve the same data, and all three provide APIs to access the structural data programmatically [12–14]. The major differences between the three sites lie in the user interface, additional annotations and services provided on top of the original structural data. Therefore, depending on the user demands, one or the other site may be preferable.

While PDB is primarily valued for providing atomic-level experimental structures, there are complementary databases hosting raw experimental data across different structural biology techniques, including electron microscopy (EM), NMR, and small angle scattering (SAS).

Electron density maps are collected and distributed by the Electron Microscopy Data Bank (EMDB) [15]. In recent years, thanks to the 'resolution revolution' in the cryo-EM field [16], electron microscopy has become a highly important source of structural data. Currently, most structures of protein-nucleic acid complexes are solved by EM (>70 % in years 2023–2024), and they now comprise more than 20 % of total EM entries in the PDB. In EMDB, more than half of EM maps are associated with atomic structural models deposited in PDB, and this fraction is growing fast. Moreover, EM maps of lower resolution combined with high-confidence computational models may help reveal previously unattainable structural details. Even when an EM map is associated with the PDB structure, the analysis of raw EM data may help identify different conformational states or explore dynamic properties of a complex.

Raw NMR data, housed by the Biological Magnetic Resonance Data Bank (BMRB) [17], can provide important insights into structural dynamics and disorder. Similar to other core databases, BMRB is dominated by protein and peptide data. Protein-DNA and protein-RNA complexes represent a minor fraction, with just over 100 entries for each type of complexes. Most of these BMRB entries have corresponding atomic structures in PDB.

Small Angle Scattering Biological Data Bank (SASBDB) is a resource of low-resolution experimental data for biological structures, including protein-nucleic acid complexes [18]. Small angle scattering (SAS) of X-ray and neutrons provides structural information on biological macromolecules in solution at a resolution of 1–2 nm. Introduced in 2014, SASBDB has grown substantially in recent years. Currently it contains close to 5000 data sets, of which about 9 % represent 'hetero-complexes', mostly protein-DNA and protein-RNA complexes. In addition to raw data, SASBDB entries are in many cases supplemented by so-called 'ab initio'

**Table 1**

**Databases containing structural data on protein-nucleic acid interactions.**

| Database | Website address | Description | Number of protein-nucleic acid entries[a] |
|---|---|---|---|
| *Primary data resources:* | | | |
| **wwPDB [1]** | **https://www.wwpdb.org/** | **The main global archive of experimentally determined biomolecular structures** | **14,667 structures** |
| EMDB [15] | https://ebi.ac.uk/emdb | The archive of 3D electron microscopy maps of biological specimens | 5,446 EM maps |
| **BMRB [17]** | **https://bmrb.io/** | **Spectral and quantitative data derived from NMR investigations of biomolecules** | **276 entries** |
| SASBDB [18] | https://www.sasbdb.org/ | Curated repository for small angle scattering data and models | 451 entries[b] |
| **PDB-IHM [19]** | **https://pdb-ihm.org/** | **An archive of structures, determined using hybrid methods (experimental and computational)** | **N/A[c]** |
| ModelArchive [20] | https://modelarchive.org/ | Repository and archive for computational structure models that are not based on experimental data | N/A[c] |
| *Derived databases:* | | | |
| **NAKB [2]** | **https://nakb.org/** | **Subset of the PDB consisting of structures containing nucleic acids, enriched with nucleic acid-specific annotations** | **14,671 structures** |
| 3D-footprint [27] | https://3dfootprint.eead.csic.es/ | A database providing binding specificity data for protein-DNA complexes in the PDB | 10,803 complexes |
| **DNAproDB [26]** | **https://dnaprodb.usc.edu/** | **Advanced visualization of protein-DNA interactions in PDB and user-uploaded structures** | **7,229 structures** |
| RNAproDB [28] | https://rnaprodb.usc.edu/ | Advanced visualization of protein-nucleic acid interactions in PDB and user-uploaded structures | 16,230 structures[d] |
| **PPI3D [29]** | **https://bioinformatics.lt/ppi3d/** | **Clustered datasets of protein-nucleic acid interfaces with sequence search capability and detailed analysis of the interfaces** | **160,343 interfaces** |
| BioLiP2 [30] | https://zhanggroup.org/BioLiP/ | Databases of protein-ligand interactions that also include nucleic acids as a specific type of ligands, also having the possibility to search for structural data and analyze the identified interfaces | 206,486 DNA or RNA ligands |
| **Q-BioLiP [31]** | **https://yanglab.qd.sdu.edu.cn/Q-BioLiP/** | | 37,242 nucleic acid ligands |
| DOCKGROUND [32] | https://dockground.compbio.ku.edu/ | A data resource for development of methods for modeling protein interactions (including protein-RNA interactions) | 132,042 interactions |
| *Specialized databases:* | | | |
| **ProNAB [36]** | **https://web.iitm.ac.in/bioinfo2/pronab/** | **Manually curated database of binding affinities of protein-nucleic acid complexes** | **20,219 affinity entries, 1138 structures** |
| NABE [37] | http://nabe.denglab.org/ | A manually curated energetic database of amino acid mutations in protein-nucleic acid binding interfaces | 2,503 mutations in 473 structures |
| **REBASE [38]** | **https://rebase.neb.com/** | **A curated database on restriction-modification systems** | **695 structures for 177 enzymes[b]** |
| CASpedia [39] | http://caspedia.org/ | Encyclopedia of Class 2 CRISPR-Cas systems | 32 proteins, 21 structures |
| **CASP structure models [54]** | **https://predictioncenter.org/** | **Community-wide resource providing computationally derived models and their evaluation scores** | **2,690 models for 22 protein-nucleic acid targets[e]** |

[a] The numbers were collected from either the database web sites (accessed on 2025-05-07) or corresponding publications.

[b] Includes structures with and without nucleic acids bound.

[c] Not available; the number of protein-nucleic acid entries could not be estimated because of limited search and filtering options.

[d] Includes structures having only nucleic acids.

[e] Only the data for the most recent CASP16 experiment are provided.

**Table 2**

**Web servers for structure-based studies of protein-nucleic acid interactions.**

| Web server | Website address | Input | Output | Description |
|---|---|---|---|---|
| *Machine learning-based structure prediction:* | | | | |
| **AlphaFold Server (AlphaFold3) [9]** | **https://alphafoldserver.com/** | Sequences | Structure models | Web servers that can predict structures of biomolecular complexes, containing proteins, nucleic acids, ligands, ions and post-translational modifications |
| Chai-1 [42] | https://lab.chaidiscovery.com/ | | | |
| **Protenix [43]** | **https://protenix-server.com/** | | | |
| Boltz-1 [44] | https://huggingface.co/spaces/simonduerr/boltz-1 [a] | | | |
| *Protein-nucleic acid docking:* | | | | |
| **HADDOCK [45]** | **https://wenmr.science.uu.nl/haddock2.4** | Subunit structures and constraints (optional) | Structure models of complexes | Web servers for protein-nucleic acid docking |
| HDOCK [46] | http://hdock.phys.hust.edu.cn/ | | | |
| **LightDock [47]** | **https://lightdock.org/** | | | |
| NPDock [48] | https://genesilico.pl/NPDock/ | | | |
| **P3DOCK [49]** | **http://www.rnabinding.com/P3DOCK/P3DOCK.html** | | | |
| PyDockDNA [50] | https://model3dbio.csic.es/pydockdna | | | |
| *Analysis and annotation of protein-nucleic acid complexes:* | | | | |
| **PLIP [57]** | **https://plip-tool.biotec.tu-dresden.de/plip-web/plip/index** | Structures | Annotation of interactions | **Identification of non-covalent interactions between biological macromolecules and their ligands** |
| VoroContacts [58] | https://bioinformatics.lt/wtsam/vorocontacts | | | A web server for the Voronoi tessellation-based analysis of contacts in biomolecular structures |
| **DeepPBS [66]** | **https://deeppbs.usc.edu/** | | | **Deep learning-based prediction of protein-DNA binding specificity** |
| *Comparison of structures:* | | | | |
| **CAD-score [63]** | **https://bioinformatics.lt/cad-score/** | Structures | Comparison scores, visualization and associated data | **Superposition-free comparison of structures of proteins, nucleic acids and their complexes using interatomic contact areas** |
| Structure Assessment Server [64] | https://swissmodel.expasy.org/assess | | | Interactive evaluation and benchmarking of structural models of macromolecular complexes using an assortment of tools |
| **US-align [65]** | **https://zhanggroup.org/US-align/** | | | **Superposition-based comparison of biomolecular structures** |

[a] The Hugging Face space for Boltz-1 is an unofficial web app implementation of the original software (https://github.com/jwohlwend/boltz/).

bead models or even atomistic models. SAS data combined with AlphaFold or other structure prediction methods [7–9] may guide selection of an accurate model out of possible alternatives.

In general, structural characterization of many complex macromolecular assemblies is increasingly carried out using integrative or hybrid modeling, where the structure is obtained by combining complementary experimental and computational techniques. These types of models are collected and disseminated by PDB-IHM [19]. At least presently, protein-DNA/RNA complexes are not well-represented in PDB-IHM. However, rapidly increasing modeling capabilities make the hybrid

modeling a highly promising route towards resolving structures of protein-nucleic acid complexes.

Computational structure models not based on any experimental data are archived in ModelArchive [20]. It complements the PDB for experimental structures and PDB-IHM for integrative structures. ModelArchive accepts computational models in the standardized ModelCIF format [21] and collects model metadata, including details on modeled molecules, structure prediction methods and model accuracy estimations. Currently, ModelArchive includes a substantial number of models (over 60,000), but since it provides only browsing and basic search functionality, it is not clear how many of those correspond to the protein-nucleic acid complexes.

## PDB-derived databases/resources

While PDB serves as the primary repository, researchers are often interested in specific PDB subsets or enhanced annotation. Not surprisingly, there are multiple databases and resources derived from PDB, some directly related to the protein-nucleic acid assemblies. One such resource is the Nucleic Acid Knowledgebase (NAKB) [2]. NAKB contains the PDB structures corresponding to DNA, RNA, and protein-DNA/RNA complexes. The latter dominate NAKB as they comprise about three-quarters of all entries. In addition to the information from PDB, NAKB provides additional annotation using internal and external tools. For protein-nucleic acid complexes, NAKB provides various structural features of nucleic acids [22–25], protein-DNA interaction data [26], and estimated binding specificities [27].

Other resources are geared toward more detailed analysis of individual structures. DNAproDB [26] and RNAproDB [28] offer advanced visualization and interactive analysis of protein-DNA and protein-RNA interactions, respectively. Both databases contain regularly updated pre-processed PDB. Both sites offer rich and highly interactive visualization of nucleic acid base pairing and protein-nucleic acid interfaces. The newer RNAproDB provides somewhat richer data and more visualization features, including electrostatics of the molecular surfaces and interactive connections between different structure representations. Interestingly, RNAproDB also includes pre-calculated entries that contain DNA. In contrast, DNAproDB only provides data on protein-DNA interactions. In addition, both servers offer the possibility to analyze a user-provided structure.

A common recent trend is to include data on protein-nucleic acid structures as part of a more general set of protein complexes.

Thus, PPI3D, initially developed for the analysis of protein–protein interactions, now also includes protein-nucleic acid interactions [29]. Updated weekly, PPI3D uses PDB biological assemblies as the primary data source for all binary interactions that involve a protein chain. If DNA/RNA chains form double-stranded regions, they are joined into a single entity to enable the analysis of interactions with double-stranded nucleic acids. The interaction interfaces and binding sites are clustered by both sequence and structure similarity, ensuring that alternative interfaces are not lost. Interaction data, including protein-nucleic acid interfaces in PPI3D, can be queried either by PDB ID or by using sequence search. The platform provides detailed insights into interaction interfaces, including structures, interface residues, and residue–residue contacts derived from Voronoi tessellation. Pre-processed and clustered PDB structural data can also be downloaded in bulk for offline use.

BioLiP2 [30] and Q-BioLiP [31] databases focus on general protein-ligand interactions. In these databases, DNA and RNA represent just two of many types of protein ligands. Both databases collect structural data from PDB, including regular automatic updates. In general, BioLiP2 and Q-BioLiP share the common goal of enhancing the understanding of protein-ligand interactions, but they differ in data analysis, user interface and search options. One of the major differences is that BioLiP2 focuses on the interactions between individual protein and nucleic acids chains, whereas Q-BioLiP considers complete protein-nucleic acid complexes, thus providing a more holistic view. On the other hand, BioLiP2 has more advanced search functionality: the database can be queried not only by various IDs, but can also be searched using protein, DNA or RNA sequences or structures. Q-BioLiP does not have a sequence search option, but like BioLiP2 provides a possibility to query the database using structure. The datasets of both databases can be downloaded, making these resources useful for other applications.

Among emerging structural resources for protein-nucleic acid complexes is DOCKGROUND [32], initially established to foster development of protein–protein docking methods. In the latest update, DOCKGROUND provides a possibility to analyze experimental structures of protein-RNA complexes. Analysis can be done for either individual PDB entries or user-defined subsets. In addition, both full set and two clustered subsets of protein-RNA complexes can be downloaded.

## Specialized databases

In addition to general-purpose databases, there are databases developed for specific properties of complexes, protein families, or functions. For example, several databases contain manually curated data on thermodynamics of protein-nucleic acid interactions (reviewed in Refs. [33–35]). Among them, ProNAB [36] and Nabe [37] are two recently updated databases worth noting. ProNAB contains over 20,000 entries

with thermodynamic data on protein-nucleic acid binding, collected from old, no longer maintained, databases and scientific literature. About 1000 of these entries have corresponding PDB structures [36]. Nabe focuses on binding affinity changes upon mutation in protein-nucleic acid interfaces and contains data for about 2500 mutations in 473 protein complexes [37]. Both ProNAB and NABE web sites have data navigation options, basic structure visualization and a possibility to upload new experimental data for the database curators. REBASE [38] and CasPEDIA [39] are examples of databases devoted to specific protein families involved in nucleic acid metabolism. REBASE focuses on the components of restriction-modification systems [38]. It offers a curated dataset of restriction endonucleases and methyltransferases with their specificity information and also contains a summary of available experimental structures of these enzymes. CasPEDIA includes information on Class 2 CRISPR-associated enzymes [39]. This database offers expert-curated data on Cas9, Cas12 and Cas13 families, as well as evolutionarily related IscB and TnpB enzymes, including phylogeny and available structural data for these proteins.

## Structure prediction methods and resources for their development

In addition to experimental structures, protein-DNA/RNA complexes are increasingly often predicted computationally. Previously, structures of protein-DNA/RNA complexes were typically predicted using either homology modeling or docking [40,41]. The AlphaFold2-based breakthrough in modeling proteins and protein complexes [3,4] suggested that similar deep learning techniques may also be applied for the structural prediction of protein complexes with other biomolecules. Initial attempts showed that predicting structures for protein-DNA and protein-RNA complexes is indeed feasible [7,8], but the strongest impact was made by the release of AlphaFold3 [9]. Soon, AlphaFold3 was followed by several analogs, including Chai-1 [42], Protenix [43], and Boltz-1 [44]. AlphaFold3, Chai-1, Protenix are available as web servers for registered users, and Boltz-1 is available in the Hugging Face hub (Table 2). Additionally, the web-based interface for Chai-1, Boltz-1 and Protenix is available via bioinformatics platforms Neurosnap (https://neurosnap.ai/) and Tamarind Bio (https://www.tamarind.bio/).

Nonetheless, the accuracy of 3D structures for RNA and protein-DNA/RNA complexes predicted by AlphaFold3 and its analogs tends to be significantly lower than their structural models of proteins/protein complexes [9]. Therefore, classical homology modeling and docking remain still relevant. For example, partial template-based models generated using the PPI3D web server [29] might be useful for initial analysis. Web servers for protein-nucleic acid docking [45–50] might also be useful, but they usually work best if the conformations of both protein and nucleic acid components do not change significantly upon binding [40,51]. In general, the experience, gained from modeling protein–protein complexes, suggests that docking should be used with caution, preferably only as a complementary tool to other structure prediction methods [52,53].

The abilities to predict biomolecular structures are evaluated every two years by community-wide CASP experiments [54]. For a long time CASP focused only on proteins, but during the CASP15 experiment in 2022, prediction of structure for RNA and protein-RNA complexes was introduced for the first time [55]. In CASP15 there were only 12 RNA prediction targets and only two of them were protein-RNA complexes. While there were reasonable predictions of RNA-only structures, all the participating groups failed on protein-RNA complexes [56]. The number of nucleic acid-containing targets increased significantly during the most recent CASP16 experiment in 2024 and is expected to grow in the future. Therefore, the CASP models of protein-nucleic acid complexes and their assessment scores represent an emerging dataset, which might be useful for developing and testing structure prediction methods.

## Methods for structural analysis and comparison

With the increasing number of experimentally and computationally derived structures of protein-nucleic acid complexes, there is a growing need for tools dedicated to comprehensive annotation, analysis and comparison of these structures. Some web-based databases such as DNAproDB [26] and RNAproDB [28] also allow users to upload their own structures and to run these structures through the same analysis pipeline as for the database entries. Other web-based tools are developed specifically for the interactive analysis of user-supplied structures. For example, the analysis of interactions between proteins and other molecules, including DNA and RNA, can be performed using Protein–Ligand Interaction Profiler (PLIP) [57] and VoroContacts [58]. PLIP identifies interactions at the level of atoms by applying geometric criteria with knowledge-based thresholds. In contrast, VoroContacts identifies interactions based on contact surface areas derived using Voronoi tessellation. The VoroContacts server is highly configurable and can annotate contacts at the level of atoms, residues and subunits, compute solvent accessible areas and identify hydrogen and disulfide bonds.

Comparison of 3D structures can be done using superposition-free local scores such as CAD-score [59,60] or lDDT [61] and superposition-based global scores such as TM-score [62] or RMSD. CAD-score and lDDT are sequence-dependent scores, that is, they can compare

different instances of a structure (computational models or different conformations) against the reference. TM-score and RMSD can be used to compare structures with either the same or different sequences. As lDDT and TM-score are distance-based scores, initially developed for proteins, their adaptation for nucleic acids required significant re-calibration. On the other hand, CAD-score, which is based on contact areas, uses the same universal framework for all structure comparisons.

The CAD-score web server allows comparison of structures for proteins, nucleic acids and diverse protein complexes, including those with RNA or DNA [63]. Recently published 'structure assessment' web server [64] integrates several scores, including lDDT, TM-score and RMSD, for comparison of models with the reference structure. In addition, the server performs various structure quality checks, but these are primarily directed at protein structures. The Universal Structure Alignment (US-align) [65] represents a successful example of superposition-based global structure comparison platform capable of handling different types of macromolecules, including protein-nucleic acid complexes. US-align is built upon the TM-score metric and is available as both standalone software and a web server.

In addition to structure analysis and comparison, there are tools aiming to enrich the structural information with predictive annotations. One such example is the estimation of protein-DNA binding specificity based on the structure of corresponding complex. This type of predictive annotation is performed by 3D-footprint, organized as a pre-processed database [27], also accessible from NAKB. Another recently published annotation tool in this category is Deep Predictor of Binding Specificity (DeepPBS) [66], a geometric deep-learning model designed to predict binding specificity from protein−DNA structure. Unlike 3D-footprint, which only provides pre-processed data, DeepPBS accepts a user-provided complex structure for annotation.

## Concluding remarks

There is a clear tendency that protein-nucleic acid complexes are significantly less abundant than protein−protein complexes in PDB and all the other primary databases housing experimental or computational structural data. This might explain why databases focusing on protein-DNA/RNA structures are relatively scarce. The same trend applies for tools for structure prediction, analysis, annotation and comparison of protein-nucleic acid complexes. Furthermore, many tools and resources have been developed separately for proteins and nucleic acids (mostly RNA), reflecting the lack of interaction between research communities focusing primarily on different types of biological polymers. As we enter the post-AlphaFold era, this situation is bound to change. Even though currently the accuracy of predicted structures of protein-nucleic acid complexes lags behind protein-only structures, they are already making a strong impact on research in structural biology. In the nearest future the accuracy of computational structural models is expected to improve significantly due to steadily increasing amounts of experimental data and improvements of deep learning techniques. Consequently, the structural data on protein-nucleic acid complexes, primarily as computational models, are expected to become abundant. Therefore, the need to analyze, compare and annotate these structures will certainly drive the development of new databases and tools, especially those capable of integrating structural information on proteins, RNA and DNA into a common framework.

## 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.

## Acknowledgments

This work was in part supported by Lietuvos Mokslo Taryba [Grant No. S-IMPRESSU-24-5].

## Data availability

No data was used for the research described in the article.

## References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest

1. wwPDB consortium: **Protein Data Bank: the single global archive for 3D macromolecular structure data**. *Nucleic Acids Res* 2019, **47**:D520−D528.

2. * Lawson CL, Berman HM, Chen L, Vallat B, Zirbel CL: **The Nucleic Acid Knowledgebase: a new portal for 3D structural information about nucleic acids**. *Nucleic Acids Res* 2024, **52**: D245−D254.
The article describes the Nucleic Acid Knowledge Base (NAKB), a resource for analysis, annotation and dissemination of PDB entries containing nucleic acids. NAKB provides annotation focused on DNA and RNA components of the PDB structures.

3. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, *et al.*: **Highly accurate protein structure prediction with AlphaFold**. *Nature* 2021, **596**:583−589.

4. Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, Žídek A, Bates R, Blackwell S, Yim J, *et al.*: **Protein complex prediction with AlphaFold-Multimer**. *bioRxiv* 2022, https://doi.org/10.1101/2021.10.04.463034.

5. Varadi M, Bertoni D, Magana P, Paramval U, Pidruchna I, Radhakrishnan M, Tsenkov M, Nair S, Mirdita M, Yeo J, *et al.*: **AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences**. *Nucleic Acids Res* 2024, **52**:D368−D375.

6. Lin Z, Akin H, Rao R, Hie B, Zhu Z, Lu W, Smetanin N, Verkuil R, Kabeli O, Shmueli Y, *et al.*: **Evolutionary-scale prediction of atomic-level protein structure with a language model**. *Science* 2023, **379**:1123−1130.

7. Baek M, McHugh R, Anishchenko I, Jiang H, Baker D, DiMaio F: **Accurate prediction of protein-nucleic acid complexes using RoseTTAFoldNA**. *Nat Methods* 2024, **21**:117–121.

8. Krishna R, Wang J, Ahern W, Sturmfels P, Venkatesh P, Kalvet I, Lee GR, Morey-Burrows FS, Anishchenko I, Humphreys IR, *et al.*: **Generalized biomolecular modeling and design with RoseTTAFold All-Atom**. *Science* 2024, **384**, eadl2528.

9. ** Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L, Ballard AJ, Bambrick J, *et al.*: **Accurate structure prediction of biomolecular interactions with AlphaFold 3**. *Nature* 2024, **630**:493–500.

The article describes state-of-the-art structure prediction method AlphaFold 3, which is capable of modeling biomolecular complexes, containing proteins, nucleic acids, ions, organic ligands and post-translational modifications, demonstrating the possibility to handle diverse interactions within a single framework.

10. Burley SK, Bhatt R, Bhikadiya C, Bi C, Biester A, Biswas P, Bittrich S, Blaumann S, Brown R, Chao H, *et al.*: **Updated resources for exploring experimentally-determined PDB structures and computed structure models at the RCSB protein Data Bank**. *Nucleic Acids Res* 2025, **53**:D564–D574.

11. Armstrong DR, Berrisford JM, Conroy MJ, Gutmanas A, Anyango S, Choudhary P, Clark AR, Dana JM, Deshpande M, Dunlop R, *et al.*: **PDBe: improved findability of macromolecular structure data in the PDB**. *Nucleic Acids Res* 2020, **48**: D335–D343.

12. Kinjo AR, Bekker G-J, Wako H, Endo S, Tsuchiya Y, Sato H, Nishi H, Kinoshita K, Suzuki H, Kawabata T, *et al.*: **New tools and functions in data-out activities at Protein Data Bank Japan (PDBj)**. *Protein Sci* 2018, **27**:95–102.

13. Velankar S, van Ginkel G, Alhroub Y, Battle GM, Berrisford JM, Conroy MJ, Dana JM, Gore SP, Gutmanas A, Haslam P, *et al.*: **PDBe: improved accessibility of macromolecular structure data from PDB and EMDB**. *Nucleic Acids Res* 2016, **44**: D385–D395.

14. Bittrich S, Bhikadiya C, Bi C, Chao H, Duarte JM, Dutta S, Fayazi M, Henry J, Khokhriakov I, Lowe R, *et al.*: **RCSB protein Data Bank: efficient searching and simultaneous access to one million computed structure models alongside the PDB structures enabled by architectural advances**. *J Mol Biol* 2023, **435**, 167994.

15. wwPDB Consortium: **EMDB-The electron microscopy Data Bank**. *Nucleic Acids Res* 2024, **52**:D456–D465.

16. Kühlbrandt W: **Biochemistry. The resolution revolution**. *Science* 2014, **343**:1443–1444.

17. Hoch JC, Baskaran K, Burr H, Chin J, Eghbalnia HR, Fujiwara T, Gryk MR, Iwata T, Kojima C, Kurisu G, *et al.*: **Biological magnetic resonance Data Bank**. *Nucleic Acids Res* 2023, **51**: D368–D376.

18. Kikhney AG, Borges CR, Molodenskiy DS, Jeffries CM, Svergun DI: **SASBDB: towards an automatically curated and validated repository for biological scattering data**. *Protein Sci* 2020, **29**:66–75.

19. Vallat B, Webb BM, Zalevsky A, Tangmunarunkit H, Sekharan MR, Voinea S, Shafaeibejestan A, Sagendorf J, Hoch JC, Kurisu G, *et al.*: **PDB-IHM: a system for deposition, curation, validation, and dissemination of integrative structures**. *J Mol Biol* 2025, https://doi.org/10.1016/j.jmb.2025.168963.

20. Tauriello G, Waterhouse AM, Haas J, Behringer D, Bienert S, Garello T, Schwede T: **ModelArchive: a deposition database for computational macromolecular structural models**. *J Mol Biol* 2025, https://doi.org/10.1016/j.jmb.2025.168996.

21. Vallat B, Tauriello G, Bienert S, Haas J, Webb BM, Žídek A, Zheng W, Peisach E, Piehl DW, Anischanka I, *et al.*: **ModelCIF: an extension of PDBx/mmCIF data representation for computed structure models**. *J Mol Biol* 2023, **435**, 168021.

22. Petrov AI, Zirbel CL, Leontis NB: **Automated classification of RNA 3D motifs and the RNA 3D motif atlas**. *RNA* 2013, **19**: 1327–1340.

23. Lu X-J, Bussemaker HJ, Olson WK: **DSSR: an integrated software tool for dissecting the spatial structure of RNA**. *Nucleic Acids Res* 2015, **43**, e142.

24. Černý J, Božíková P, Malý M, Tykač M, Biedermannová L, Schneider B: **Structural alphabets for conformational analysis of nucleic acids available at dnatco.datmos.org**. *Acta Crystallogr D Struct Biol* 2020, **76**:805–813.

25. Mitra R, Cohen AS, Rohs R: **RNAscape: geometric mapping and customizable visualization of RNA structure**. *Nucleic Acids Res* 2024, **52**:W354–W361.

26. * Mitra R, Cohen AS, Sagendorf JM, Berman HM, Rohs R: **DNAproDB: an updated database for the automated and interactive analysis of protein-DNA complexes**. *Nucleic Acids Res* 2025, **53**:D396–D402.

The article presents DNAproDB – a database and a web server for structural analysis of protein-DNA complexes, providing information on DNA, protein and protein-DNA interactions. The server supports uploading of protein-DNA complex for detailed analysis, annotation and visualization.

27. Contreras-Moreira B: **3D-footprint: a database for the structural analysis of protein-DNA complexes**. *Nucleic Acids Res* 2010, **38**:D91–D97.

28. ** Mitra R, Cohen AS, Tang WY, Hosseini H, Hong Y, Berman HM, Rohs R: **RNAproDB: a webserver and interactive database for analyzing protein-RNA interactions**. *J Mol Biol* 2025, https://doi.org/10.1016/j.jmb.2025.169012.

Description of RNAproDB, a newly developed database and a web server, providing powerful analysis and rich visualization, thus enabling annotation of structures of protein-nucleic acid complexes. Database contains pre-processed PDB data, while the server enables the analysis of uploaded structures. Although the main focus of RNAproDB is protein-RNA structures, it also handles protein complexes with DNA and DNA/RNA hybrids.

29. * Dapkūnas J, Timinskas A, Olechnovič K, Tomkuvienė M, Venclovas Č: **PPI3D: a web server for searching, analyzing and modeling protein-protein, protein-peptide and protein-nucleic acid interactions**. *Nucleic Acids Res* 2024, **52**: W264–W271.

Description of an updated version of the PPI3D web server, which now also includes protein-nucleic acid interactions. PPI3D allows querying clustered structural data on interaction interfaces and binding sites using PDB IDs and sequence search. PPI3D provides tools for detailed analysis of the interactions and offers a possibility to download clustered datasets of protein-nucleic acid binding sites for offline analysis.

30. * Zhang C, Zhang X, Freddolino L, Zhang Y: **BioLiP2: an updated structure database for biologically relevant ligand-protein interactions**. *Nucleic Acids Res* 2024, **52**:D404–D412.

BioLiP2 is a semi-manually curated database on protein-ligand interactions, including nucleic acids as ligands, providing the ability to search and analyze these interactions as single chains.

31. * Wei H, Wang W, Peng Z, Yang J: **Q-BioLiP: a comprehensive resource for quaternary structure-based protein-ligand interactions**. *Genom Proteom Bioinform* 2024, **22**, qzae001.

The database Q-BioLiP provides the data on protein-ligand interactions, including interactions with nucleic acids, in the context of the quaternary structures of the resulting complexes, aiming to offer the set of biologically relevant interactions.

32. Collins KW, Copeland MM, Kundrotas PJ, Vakser IA: **Dockground: the resource expands to protein-RNA interactome**. *J Mol Biol* 2025, https://doi.org/10.1016/j.jmb.2025.169014.

33. Zhang X, Mei L-C, Gao Y-Y, Hao G-F, Song B-A: **Web tools support predicting protein-nucleic acid complexes stability with affinity changes**. *Wiley Interdiscip Rev RNA* 2023, **14**, e1781.

34. Mei L-C, Hao G-F, Yang G-F: **Protein-nucleic acid thermodynamic databases for specific uses**. *Trends Biotechnol* 2023, **41**:990–991.

35. Gromiha MM, Harini K: **Protein-nucleic acid complexes: docking and binding affinity**. *Curr Opin Struct Biol* 2025, **90**, 102955.

36. Harini K, Srivastava A, Kulandaisamy A, Gromiha MM: **ProNAB: database for binding affinities of protein-nucleic acid**

**complexes and their mutants**. *Nucleic Acids Res* 2022, **50**: D1528–D1534.

37. Liu J, Liu S, Liu C, Zhang Y, Pan Y, Wang Z, Wang J, Wen T, Deng L: **Nabe: an energetic database of amino acid mutations in protein-nucleic acid binding interfaces**. *Database* 2021, **2021**, baab050.

38. Roberts RJ, Vincze T, Posfai J, Macelis D: **REBASE: a database for DNA restriction and modification: enzymes, genes and genomes**. *Nucleic Acids Res* 2023, **51**:D629–D630.

39. Adler BA, Trinidad MI, Bellieny-Rabelo D, Zhang E, Karp HM, Skopintsev P, Thornton BW, Weissman RF, Yoon PH, Chen L, *et al.*: **CasPEDIA Database: a functional classification system for class 2 CRISPR-Cas enzymes**. *Nucleic Acids Res* 2024, **52**: D590–D596.

40. Nithin C, Ghosh P, Bujnicki JM: **Bioinformatics tools and benchmarks for computational docking and 3D structure prediction of RNA-protein complexes**. *Genes* 2018, **9**, E432.

41. Liu X, Duan Y, Hong X, Xie J, Liu S: **Challenges in structural modeling of RNA-protein interactions**. *Curr Opin Struct Biol* 2023, **81**, 102623.

42. Discovery C, Boitreaud J, Dent J, McPartlon M, Meier J, Reis V, Rogozhnikov A, Wu K: **Chai-1: decoding the molecular interactions of life**. *bioRxiv* 2024, https://doi.org/10.1101/2024.10.10.615955.

43. Team BAA, Chen X, Zhang Y, Lu C, Ma W, Guan J, Gong C, Yang J, Zhang H, Zhang K, *et al.*: **Protenix - advancing structure prediction through a comprehensive AlphaFold3 reproduction**. *bioRxiv* 2025, https://doi.org/10.1101/2025.01.08.631967.

44. Wohlwend J, Corso G, Passaro S, Reveiz M, Leidal K, Swiderski W, Portnoi T, Chinn I, Silterra J, Jaakkola T, *et al.*: **Boltz-1: democratizing biomolecular interaction modeling**. *bioRxiv* 2024, https://doi.org/10.1101/2024.11.19.624167.

45. Honorato RV, Trellet ME, Jiménez-García B, Schaarschmidt JJ, Giulini M, Reys V, Koukos PI, Rodrigues JPGLM, Karaca E, van Zundert GCP, *et al.*: **The HADDOCK2.4 web server for integrative modeling of biomolecular complexes**. *Nat Protoc* 2024, **19**:3219–3241.

46. Li H, Huang E, Zhang Y, Huang S-Y, Xiao Y: **HDOCK update for modeling protein-RNA/DNA complex structures**. *Protein Sci* 2022, **31**, e4441.

47. Jiménez-García B, Roel-Touris J, Barradas-Bautista D: **The LightDock server: artificial intelligence-powered modeling of macromolecular interactions**. *Nucleic Acids Res* 2023, **51**: W298–W304.

48. Tuszynska I, Magnus M, Jonak K, Dawson W, Bujnicki JM: **NPDock: a web server for protein-nucleic acid docking**. *Nucleic Acids Res* 2015, **43**:W425–W430.

49. Zheng J, Hong X, Xie J, Tong X, Liu S: **P3DOCK: a protein-RNA docking webserver based on template-based and template-free docking**. *Bioinformatics* 2020, **36**:96–103.

50. Rodríguez-Lumbreras LA, Jiménez-García B, Giménez-Santamarina S, Fernández-Recio J: **pyDockDNA: a new web server for energy-based protein-DNA docking and scoring**. *Front Mol Biosci* 2022, **9**, 988996.

51. Olechnovič K, Banciul R, Dapkūnas J, Venclovas Č: **FTDMP: a framework for protein-protein, protein-DNA, and protein-RNA docking and scoring**. *Proteins* 2025, https://doi.org/10.1002/prot.26792.

52. Ghani U, Desta I, Jindal A, Khan O, Jones G, Hashemi N, Kotelnikov S, Padhorny D, Vajda S, Kozakov D: **Improved docking of protein models by a combination of Alphafold2 and ClusPro**. *bioRxiv* 2022, https://doi.org/10.1101/2021.09.07.459290.

53. Olechnovič K, Valančauskas L, Dapkūnas J, Venclovas Č: **Prediction of protein assemblies by structure sampling followed by interface-focused scoring**. *Proteins* 2023, **91**:1724–1733.

54. Kryshtafovych A, Schwede T, Topf M, Fidelis K, Moult J: **Critical assessment of methods of protein structure prediction (CASP)-Round XV**. *Proteins* 2023, **91**:1539–1549.

55. Kryshtafovych A, Antczak M, Szachniuk M, Zok T, Kretsch RC, Rangan R, Pham P, Das R, Robin X, Studer G, *et al.*: **New prediction categories in CASP15**. *Proteins* 2023, **91**:1550–1557.

56. Das R, Kretsch RC, Simpkin AJ, Mulvaney T, Pham P, Rangan R, Bu F, Keegan RM, Topf M, Rigden DJ, *et al.*: **Assessment of three-dimensional RNA structure prediction in CASP15**. *Proteins* 2023, **91**:1747–1770.

57. Adasme MF, Linnemann KL, Bolz SN, Kaiser F, Salentin S, Haupt VJ, Schroeder M: **Plip 2021: expanding the scope of the protein-ligand interaction profiler to DNA and RNA**. *Nucleic Acids Res* 2021, **49**:W530–W534.

58. Olechnovič K, Venclovas Č: **VoroContacts: a tool for the analysis of interatomic contacts in macromolecular structures**. *Bioinformatics* 2021, **37**:4873–4875.

59. Olechnovič K, Kulberkytė E, Venclovas Č: **CAD-score: a new contact area difference-based function for evaluation of protein structural models**. *Proteins* 2013, **81**:149–162.

60. Olechnovič K, Venclovas Č: **The use of interatomic contact areas to quantify discrepancies between RNA 3D models and reference structures**. *Nucleic Acids Res* 2014, **42**:5407–5415.

61. Mariani V, Biasini M, Barbato A, Schwede T: **lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests**. *Bioinformatics* 2013, **29**:2722–2728.

62. Zhang Y, Skolnick J: **Scoring function for automated assessment of protein structure template quality**. *Proteins* 2004, **57**: 702–710.

63. Olechnovič K, Venclovas Č: **The CAD-score web server: contact area-based comparison of structures and interfaces of proteins, nucleic acids and their complexes**. *Nucleic Acids Res* 2014, **42**:W259–W263.

64. Waterhouse AM, Studer G, Robin X, Bienert S, Tauriello G, Schwede T: **The structure assessment web server: for proteins, complexes and more**. *Nucleic Acids Res* 2024, **52**: W318–W323.

65. Zhang C, Shine M, Pyle AM, Zhang Y: **US-align: universal structure alignments of proteins, nucleic acids, and macromolecular complexes**. *Nat Methods* 2022, **19**:1109–1115.

66. Mitra R, Li J, Sagendorf JM, Jiang Y, Cohen AS, Chiu T-P, Glasscock CJ, Rohs R: **Geometric deep learning of protein-DNA binding specificity**. *Nat Methods* 2024, **21**:1674–1683.