Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Protein-nucleic acid complexes: Docking and binding affinity

M. Michael Gromiha and K. Harini

Protein-nucleic interactions play essential roles in several biological processes, such as gene regulation, replication, transcription, repair and packaging. The knowledge of three-dimensional structures of protein-nucleic acid complexes and their binding affinities helps to understand these functions. In this review, we focus on two major aspects namely, (i) deciphering the three-dimensional structures of protein-nucleic acid complexes and (ii) predicting their binding affinities. The first part is devoted to the state-of-the-art methods for predicting the native structures and their performances including recent CASP targets. The second part is focused on different aspects of investigating the binding affinity of protein-nucleic acid complexes: (i) databases for thermodynamic parameters to understand the binding affinity, (ii) important features determining protein-nucleic acid binding affinity, (iii) predicting the binding affinity of protein-nucleic acid complexes using sequence and structure-based parameters and (iv) change in binding affinity upon mutation. It includes the latest developments in protein-nucleic acid docking algorithms and binding affinity predictions along with a list of computational resources for understanding protein-DNA and protein-RNA interactions.

Addresses
Department of Biotechnology, Bhupat and Jyoti Mehta School of Biosciences, Indian Institute of Technology Madras, Chennai 600036, Tamil Nadu, India

Corresponding author: Gromiha, M. Michael (gromiha@iitm.ac.in)

Current Opinion in Structural Biology 2025, **90**:102955

This review comes from a themed issue on **Folding and Binding (2025)**

Edited by **Yana Bromberg** and **Ben Schuler**

For complete overview of the section, please refer the article collection - Folding and Binding (2025)

Available online 30 November 2024

https://doi.org/10.1016/j.sbi.2024.102955

0959-440X/© 2024 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

Keywords
Protein-DNA interactions, Protein-RNA interactions, Three-dimensional structures, Docking, Binding affinity, Mutation, Thermodynamics.

Introduction

Protein-nucleic complexes perform diverse functions in living organisms such as transcription, gene regulation, replication, repair, packaging, methylation, control and genome assembly [1]. These functions are well understood owing to the availability of experimentally known three-dimensional structures of protein-DNA and protein-RNA complexes as well as their binding affinities. The three-dimensional structures of protein-nucleic complexes are determined using X-ray crystallography, NMR spectroscopy and cryo-electron microscopy. On the other hand, binding affinities of these complexes are obtained from isothermal titration calorimetry (ITC), surface plasmon resonance (SPR) and fluorescence spectroscopy. These experimental data provide insights into the mechanism of binding in terms of amino acid residues/nucleotides at the interface, as well as the binding affinity of protein-nucleic acid complexes with dissociation constant and binding free energy [2].

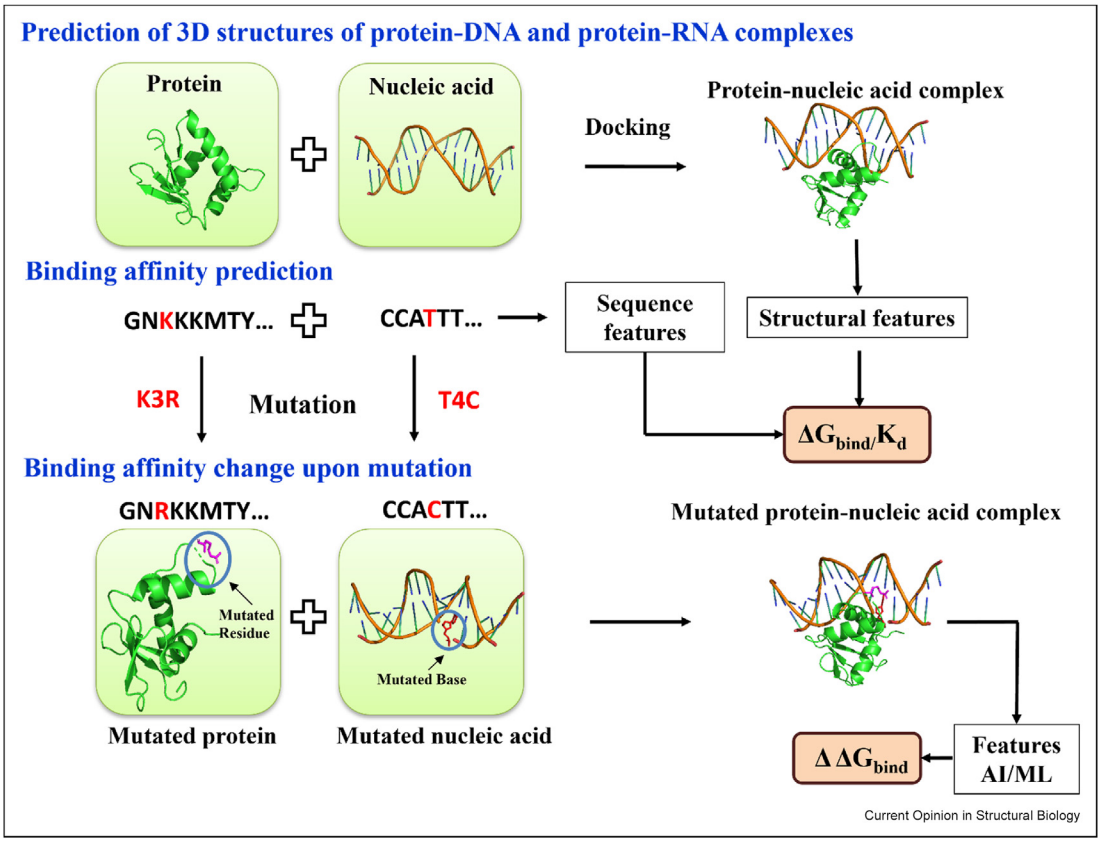
Computationally, several investigations have been carried out to explore the binding of protein-nucleic acid complexes: (i) identification of DNA/RNA binding proteins, (ii) prediction of binding sites from either known three-dimensional structure of an unbound protein or amino acid sequence, (iii) deciphering the three-dimensional structure of a protein-DNA or protein-RNA complex, (iv) binding affinity of protein-DNA and protein-RNA complexes using their structures, (v) binding affinity of a complex from the sequences of interacting protein and nucleic acid, and (vi) change in binding affinity upon mutation [3,4]. We have witnessed advancements in these areas on two convergent aspects, namely (i) analysis and prediction of amino acid residues at the binding interface and (ii) docking, and predicting the binding affinity of protein-DNA and protein-RNA complexes [5].

In this review, we focus on protein-nucleic acid docking and binding affinity. We survey the development of computational methods for predicting the three-dimensional structures of protein-DNA and protein-RNA complexes as well as binding affinities of protein-nucleic acid complexes and their mutants. The outline of the review is illustrated in Figure 1.

Protein-nucleic acid docking

Predicting the three-dimensional structure of a protein-DNA or a protein-RNA complex enhances our

Figure 1



Workflow for predicting the three-dimensional structures of protein-DNA and protein-RNA complexes and their binding affinities.

knowledge to understand the principles of biomolecular recognition at the atomic level, molecular functions and binding affinity, which has potential biomedical applications in computer-aided and structure-based drug design.

Protein-nucleic acid complexes are mainly generated using template-based docking. In addition, *ab initio*, artificial intelligence (AI), and machine learning (ML) are used to predict the three-dimensional structures of protein-nucleic acid complexes. Template-based modeling utilizes structures or homology models of protein and DNA/RNA, which are aligned on the structure of a protein-nucleic acid complex containing DNA/RNA and protein homologous to the target DNA/RNA and protein, respectively [6]. On the other hand, *ab initio* docking methods utilize atomic coordinates of a protein and nucleic acid to predict its complex three-dimensional structure. As it does not depend on available structural data for homologous complexes, it is useful for predicting more protein-DNA/RNA complexes, that lack homologous structures [7••]. AI or ML-based methods consider existing protein-nucleic acid complex structures in the Protein Data Bank (PDB) and utilize the information on frequency profiles, evolutionary coupling, conservation scores, and structural constraints, including covariation patterns from the sequence alignments, for prediction [8••].

Computational methods for protein-nucleic acid docking have been developed in two directions: (i) adapting/transferring the knowledge/protocol developed for protein−protein complexes and (ii) specifically designed for predicting the three-dimensional structures of protein-nucleic acid complexes [9−13]. It includes various steps such as sampling, scoring of poses, clustering and refinement. Most of the currently available approaches for predicting protein-nucleic acid complex structures are based on developing predictive models for the protein and nucleic acid (DNA/RNA) separately and constructing complexes using molecular docking.

Tuszynska et al. [9] developed the first web server specific for predicting protein-nucleic acid complex structures. It performs rigid body docking to generate protein-nucleic acid complex structures, and these structures are ranked using statistical potentials to select the final docked complex. Later, Rodríguez-Lumbreras et al. [10] proposed a method for obtaining

protein-DNA complex structures based on three steps: (i) generating rigid-body docking poses between protein and DNA using shape complementarity, electrostatics and biochemical information, (ii) scoring with energy-based parameters, which includes desolvation, electrostatics and van der Waals energy along with nucleotide atom-types and (iii) external restraints to amino acid-nucleotide distance to rescore docking models.

Li et al. [12] developed a method, HDOCK, for docking protein−RNA/DNA complexes, which include ssRNA, ssDNA, dsRNA, and dsDNA in addition to protein−protein complexes. It accepts both sequence and structure input and utilizes a hybrid protocol to support both template-based and template-free docking. Initially, HDOCK searches the PDB for homologous complexes of two molecules (E.g., protein and DNA), and if a reliable template is found, the structure will be constructed by homology modeling and used for docking. If no homologous complex is found, input structures will be directly used for docking.

Esmaeeli et al. [13] reported a computational approach, MELD-DNA, to predict protein-DNA complex structures, which combines molecular dynamics simulations with experimental information. It is shown to have several applications such as (i) sampling multiple binding modes; (ii) identifying the preferred binding mode from the ensembles and (iii) providing qualitative binding preferences between DNA sequences.

Recently, Baek et al. [7••] developed a method, RoseTTAFoldNA, which predicts three-dimensional structures with confidence scores for both protein-DNA and protein-RNA complexes. It utilizes a three-track architecture, which simultaneously refines three representations namely, sequence (1D), distance between residue-pairs (2D) and cartesian coordinates (3D). Further, RoseTTAFoldNA incorporated physical interactions using Lennard-Jones and hydrogen-bonding energies as input features for the final refinement and part of the loss function during fine-tuning. Abramson et al. [8••] reported the development of an AI-based method, AlphaFold 3, which includes predicting protein-nucleic acid and protein-ligand complex structures. The architecture has different blocks: (i) embedding, (ii) Pairformer and (iii) diffusion block. First, individual residues in the molecules are represented by their physicochemical characteristics, conservation, and structural information, encoded as a single representation. Further, it considers the pairwise representation encoding the relationship between residues. In addition, information about the evolutionary relationship with existing complexes is obtained from multiple sequence alignment and encoded into the pairwise representation. In the second stage, the single and pair representations are given as input to the "pairformer" to predict the atom coordinates of complex structures. Finally, these noised coordinates are refined in the diffusion block to obtain highly confident structures. They claimed that the prediction accuracy of protein-nucleic acid complexes is higher than nucleic acid specific predictors. However, this method has a few limitations, such as violation of chirality and producing misleading structural orders during hallucination in disordered regions.

Scoring functions for identifying the best poses obtained from conformational sampling

Several scoring functions have been reported to select the best native-like protein-nucleic acid complex structure from a pool of modelled structures obtained by docking. Generally, contributions from energy-based parameters (electrostatic, hydrogen bonds and van der Waals interactions, effects of solvation, contact potentials) and geometric or physical principles (predicted binding sites, shape or chemical complementarity) are used for developing scoring functions [14].

Li et al. [12] utilized a shape-based pairwise scoring function for scoring binding modes of protein-nucleic acid complexes obtained in sampling. Rodríguez-Lumbreras [10] ranked protein-DNA docking poses using a scoring function, which is composed of electrostatics, desolvation, and van der Waals energy. Further, interface hydrogen bond energy and strength are considered for ranking, besides the distance-based metrics for ranking the models of protein-DNA complexes [6].

Software for predicting the structures of protein-nucleic acid complexes

Several computational methods and algorithms have been reported in the literature for predicting the three-dimensional structures of protein-DNA and protein-RNA complexes from their sequences or structures. These methods are based on (i) replication of empirical evidence (template-based modeling, experimentally known biochemical or biophysical interactions, and identifying native-like interface), (ii) prediction of potentially novel interfaces (sampling binding sites, shape complementarity to maximize the binding interface, and global range molecular matching (GRAMM)), and (iii) favorable energetic contributions (minimizing Lennard-Jones potential, reducing the clashes, atomic desolvation energy, and statistical potentials). Table 1 provides a list of available web servers for predicting the three-dimensional structures of protein-DNA and protein-RNA complexes along with important features used to develop the method.

Critical Assessment of protein structure prediction (CASP)

CASP is a community-wide experiment in which participants are provided with the sequences of

Table 1

Webserver for predicting the three-dimensional structures of protein-nucleic acid complexes.

Tools	Description	Link
HDOCK	Protein−protein and protein-DNA/RNA docking; hybrid algorithm with template-based modeling and *ab initio* docking.	http://hdock.phys.hust.edu.cn/
GRAMM-X	Mapping energy landscape between molecules of sampled docked complexes.	https://gramm.compbio.ku.edu/gramm
FTDock	Shape complementarity between grids of molecules and Fourier transform-based docking algorithm.	http://www.sbg.bio.ic.ac.uk/docking/ftdock.html
PatchDock	Mapping the surface into patches using Connolly dot representation and scoring based on the complementarity and energy	https://bioinfo3d.cs.tau.ac.il/PatchDock/
HADDOCK	*ab-initio* flexible docking by encoding the interfaces identified as restraints	https://rascar.science.uu.nl/haddock2.4/
NPDock	Obtain a cluster of complexes using GRAMM docking, and rank the docked complexes based on statistical potentials	https://genesilico.pl/NPDock/
TFmodeller	Contact matrix, binding interface, and atomic coordinates	https://hub.docker.com/r/eeadcsiccompbio/tfmodeller
pyDockDNA	Energy-based protein-DNA docking and scoring of orientations generated using FTDOCK	https://model3dbio.csic.es/pydockdna
MELD-DNA	Protein-DNA docking with the knowledge from statistical mechanics of known structures along with interface dynamics.	github.com/maccallumlab/meld
RoseTTAFoldNA	*ab initio* method using a three-track architecture incorporating interaction and structural patterns in proteins and bases.	https://github.com/uw-ipd/RoseTTAFold2NA
AlphaFold3	AI method using diffusion-based architecture with single and pair representations of residues	https://alphafoldserver.com/

accessed on 4th July 2024.

biomolecules or their complexes and expected to submit predicted structures. The methods are evaluated by comparing the predicted structure with experimentally known three-dimensional structure. CASP mainly focused on proteins and ligands, and the previous CASP15 in 2022 introduced RNA structure predictions. The evaluation of RNA tertiary structure prediction methods showed that ALCHEMY_RNA2 ranked as the best method [15]. Recently, CASP16, in 2024, provided the sequences of protein and RNA/DNA for predicting protein-RNA/DNA complex structures. Notably, about 25 % of targets released in CASP16 contain nucleic acids or their complexes. We observed that only a few servers could reliably predict protein-nucleic acid complex structures, emphasizing the necessity for reliable and accurate structure prediction methods for protein-DNA/RNA complexes.

Different measures are used to compare the performance of predicted protein-nucleic acid complex structures such as Global Distance test_Total score (GDT_TS), RMSD of CA atoms and Interface LDDT, which measures the distance between the interface atoms across different chains in the complex. CAPRI (Critical Assessment of PRediction of Interactions) used (i) the fraction of the contacts in the native structure, which are correctly predicted in the docking model (F_{nat}), (ii) root mean square deviation of DNA/RNA atoms with respect to the optimally aligned protein structures (lRMSD) and (iii) RMSD between DNA/RNA and protein atoms at the interface (iRMSD) for evaluation. In addition, CASP efficiently considers individual domains, interfaces, interface contact score (ICS) and interface path score (IPS) for evaluating the performance of prediction methods for multimeric complexes at high precision. Recently, the number of conserved hydrogen bonds, average Local Distance Difference Test (lDDT) scores, and the combination of weighted hydrogen bond energy and distance-based metrics [6] are used to assess the performance.

Binding affinity of protein-nucleic acid complexes

Databases for binding affinity

Developing databases of experimentally determined thermodynamic data for protein-DNA and protein-RNA complexes is necessary for understanding the influence of sequence and structure-based parameters on their binding affinities and designing robust prediction algorithms. Hence, several databases have been developed to accumulate thermodynamic data (K_d and ΔG) for protein-nucleic acid complexes, along with sequence and

structure information of the protein and nucleic acid (DNA/RNA) in a complex, experimental methods and conditions, and literature. Further, few databases contain information on the change in binding affinity upon mutations (ΔΔG). Table 2 lists the available thermodynamic databases for protein-nucleic acid complexes. Harini et al. [16••] developed the largest database, ProNAB for thermodynamic data on protein-nucleic acid complexes based on experimentally determined binding affinity reported in the literature. It has more than 20,000 entries with options for searching, displaying, sorting, downloading, and uploading the data. ProNAB has several advantages: (i) contain the highest number of data for alanine and non-alanine mutations compared to other databases (ProNIT: 12174, dbAMEPNI: 578, PDBbind: 973, PNATDB: 12635), (ii) improved functionality (advanced search option to obtain data with various search conditions, including UniProt ID, mutation type, secondary structure, solvent accessibility, range of ΔG or ΔΔG values and PubMed ID), (iii) comprehensive information (links to other related sequence, structure and functional databases, and tools), (iv) feasibility of retrieving data for analysis (efficient search, display, sort, and download options), and (v) specific design with detailed tutorials, as well as data upload option [17]. Mei et al. [18] developed a database containing thermodynamic data and details of molecular interactions in protein-nucleic acid complexes.

Structure-based parameters for relating binding affinity

The binding affinities of protein-nucleic complexes are related to several structure-based parameters such as buried surface area, number of hydrogen bonds, electrostatic interactions, interatomic contacts, and interaction energy [19]. The conformational change of

Table 2

Online resources for binding free energy of protein-nucleic acid complexes.

Databases for binding affinity of protein-nucleic acid complexes and mutants		
Database	Remarks	Link
dbAMEPNI[a]	Alanine mutations	http://zhulab.ahu.edu.cn/dbAMEPNI.
PNATDB	Includes molecular interactions	http://chemyang.ccnu.edu.cn/ccb/database/PNAT/.
ProNAB	The largest available thermodynamic database for protein-nucleic acid complexes and mutants	https://web.iitm.ac.in/bioinfo2/pronab/.
PDBBind	Binding data for protein complexes with known structures	http://www.pdbbind-cn.org/
Web servers for predicting binding free energy of protein-nucleic acid complexes		
Method	Features	Link
PredPRBA	Interface hydrophobicity, hydration pattern, and change in the conformation due to binding	http://predprba.denglab.org/
PRA-Pred	Contact-based features, interaction energies, RNA base step parameters, and hydrogen bonding.	https://web.iitm.ac.in/bioinfo2/prapred/
DNAffinity	Physics-based machine learning using molecular dynamics simulation-based features	https://github.com/Jalbiti/DNAffinity
PNAB	Protein-sequence-based features, physicochemical properties, nucleic acid sequence features	http://pnab.denglab.org/
PreDBA	An ensemble model using sequence and structural features of the protein and DNA.	http://predba.denglab.org/
PDA-Pred	Interaction features, volume and surface area of the interface, DNA base step parameters, and atom contacts	https://web.iitm.ac.in/bioinfo2/pdapred/
emPDBA	Sequence, structure, and interface features of the complex and the individual partners	https://github.com/ChunhuaLiLab/emPDBA/
Web servers for predicting binding free energy change of protein-nucleic acid complexes upon mutation		
mCSM−NA	Graph-based signatures utilizing the encoded amino acid residue	https://biosig.lab.uq.edu.au/mcsm_na/prediction
PEMPNI	Energy-based and structural interface features, such as contacts and residue−nucleotide pairs	http://liulab.hzau.edu.cn/PEMPNI
PremPRI	Interface interactions and graph-based features	https://lilab.jysw.suda.edu.cn/research/PremPRI/
PremPDI	Molecular mechanics, statistical potentials and accessibility	https://lilab.jysw.suda.edu.cn/research/PremPDI/
SAMPDI-3D	Structural features and knowledge-based terms (protein and DNA).	http://compbio.clemson.edu/SAMPDI-3D/

[a] currently not available; accessed on 4th July 2024.

protein and/or nucleic acid (DNA/RNA) upon binding is also an important parameter to estimate the binding free energy [20]. Further, local structural motifs and secondary structures are reported to be important for DNA recognition, specifically in the major groove of the DNA [21]. Pant et al. [22] showed that the bicyclo-nucleotide modifications in DNA enhance the affinity of proteins-DNA binding.

On the other hand, intrinsic disorder regions are shown to be important in determining the affinity of DNA-binding proteins. The disordered regions assist in forming transient and dynamic interactions by supplementing the interactions in a folded globular protein with DNA, thereby improving the binding affinity of the complex [23]. In protein-RNA interactions, RNA binding proteins are often extended with intrinsically disordered regions, which mediate interactions with RNA and increase the affinity of binding [24].

Recently, Harini et al. [25•] revealed that protein-RNA binding affinity is influenced by the type of RNA and function of the protein. Further, several structure-based features including RNA base-step parameters, number of hydrogen bonds, number of atomic contacts in a complex, contact potentials and interaction energies between protein and RNA play a significant role in understanding and predicting the binding affinities of protein-RNA complexes. These features are important for understanding the binding affinity of protein-DNA complexes as well [26].

Prediction of binding free energy using structure-based parameters

Structure-based parameters and energy potentials (physicochemical features, RNA base step parameters, protein-RNA contacts, energy-based features, hydration, binding site volume, fraction of non-polar interface residues, and structural entropy) have been widely used to predict the binding affinities of protein-nucleic acid complexes. These parameters are utilized as input features for developing statistical methods and machine learning techniques to predict binding affinity. Further, molecular dynamics simulations are used to explore the conformational changes in binding affinity. Table 2 includes a list of methods for predicting the binding affinity of protein-DNA and protein-RNA complexes and a few recent methods are discussed below.

Barissi et al. [27] computed structural and mechanical properties of DNA, including base step parameters, electrostatic patterns, sequence composition, geometry, and flexibility using atomistic molecular dynamics simulations and developed a random forest method for predicting the DNA binding affinities of transcription factors. Harini et al. [26] considered a set of protein-DNA complexes and derived various structure-based features using the structure of DNA (base step parameters), protein (volume and surface area of binding site residues), and protein-DNA complex (interaction energy and contact potentials). These features are incorporated in multiple regression equations for predicting the binding affinities of protein-DNA complexes belonging to different structural and functional classes. Yang et al. [28•] proposed an ensemble model for predicting the binding affinities of protein-DNA complexes under four categories, based on DNA structure and percentage of binding site residues, using sequence, structure, and energy-based features.

Hong et al. [29] extracted structural features from protein-RNA complex structures and utilized them to train regression models for predicting the binding affinity. These features include hydrogen bond energy, solvation energy, salt bridges, noninteracting interface, and interface contacts. Recently, Harini et al. [25] implemented a structure-based regression model using RNA base step parameters, protein-RNA contacts, and energy-based features for predicting the binding affinities of protein-RNA complexes.

Sequence-based prediction of binding affinity

Apart from the structure-based methods, Pandey et al. [30•] proposed a deep learning-based model for predicting the binding affinity in terms of dissociation constants (K_d) in protein-nucleic acid complexes using a large dataset obtained from ProNAB database and sequence-based descriptors of proteins and nucleic acids.

Prediction of binding affinity change upon mutation

The mutation of amino acid residues in a protein-nucleic acid complex changes its structure, binding affinity, stability and function, and some of them lead to diseases [31]. The influence of mutations on functions of protein-DNA complexes is studied with their change in binding affinities [32]. Consequently, several methods have been proposed to understand the principles and predict the binding free energy change due to mutation using features obtained from sequence, structure, and interactions as well as combinations of them.

Pires et al. [33] developed a method based on the mutational cut-off scanning matrices (mCSM) and graph-based signatures for predicting the change in binding affinity upon mutation in protein-DNA and protein-RNA complexes. Zhang et al. [34,35] developed multiple linear regression-based models for predicting the effect of mutations in protein-DNA and protein-RNA complexes using sequence and structure-based features. Specifically, differences in number of hydrogen bonds, van der Waals interaction energies, and electrostatic interaction energies between the wild-type and mutant proteins, as well as amino acids at the interface, ASA features, hydrophobicity, and closeness of

the nodes in the residue interaction network are used for prediction.

Jiang et al. [36] developed a random forest regressor-based model using energy and non-energy-based interaction features to predict the change in binding affinity upon mutation in protein-nucleic acid complexes. It utilizes the information on interface and non-interface regions and the parameters including residue−residue pairs, residue−nucleotide pairs, nucleotide−nucleotide pairs, solvent accessibilities, hydrogen bonds, contacts, and evolutionary information for predicting the binding affinity.

Peng et al. [37] utilized Molecular Mechanics Poisson-Boltzmann Surface Area (MM/PBSA) approach for predicting the effect of mutations in protein-DNA binding affinity. Li et al. [38] developed a gradient-boosting decision tree algorithm to predict the effect of single-point protein using structural and interaction features such as secondary structure propensity, solvent accessibility, backbone torsion angles, inter-atomic contacts, hydrogen bonds, and stacking interactions. Pandey et al. [30] proposed a sequence-based method to predict the change in binding free energy upon mutation in protein-nucleic acid complexes.

Influence of nucleic acid mutation in protein-nucleic acid binding affinity

Most of the prediction methods are focused on predicting the effect of amino acid mutations on protein-nucleic acid affinity. Recently, Li et al. [38] developed a model for predicting the binding affinity change upon DNA mutation using a gradient-boosting decision tree algorithm and various features from the protein (secondary structure propensities, backbone torsions, accessibility), DNA (base-step and helical parameters) and complex (hydrogen bonding, stacking interaction, atom contacts).

Conclusions and future perspectives

Several docking algorithms are developed to predict the three-dimensional structure of protein-nucleic acid complexes, which are crucial for understanding the function and have potential applications in structure-based drug design. These methods are mainly based on homology modelling, *ab initio* docking, and AI based protocol, which have advantages and limitations. The advantages are generating high-quality structures if the sequence identity between the target and template is high for the former and obtaining accurate model structures without templates for the latter. The major limitations are (i) requirement of proper templates for homology-based docking, (ii) necessity of high computing power for *ab initio* based docking and (iii) less confidence in AI or ML based methods as well as non-availability of source codes [39]. Significant progress has been made to improve the quality of predictions using optimized parameters and a large volume of data. However, the performance is still limited, and improvements are necessary for accurately predicting protein-nucleic acid complex structures as evidenced from recent CASP experiments.

On the other hand, the thermodynamics of binding for protein-DNA and protein-RNA complexes has been studied with their three-dimensional structures to relate structure-based parameters with binding affinity. Although few methods are available in the literature for predicting the binding affinity of protein-DNA and protein-RNA complexes, it requires significant improvement in the performance uniformly for different types of complexes based on their structure and function. Further, it is also important to develop accurate methods for predicting the binding affinity from sequences of interacting proteins and nucleic acids in a complex. While few methods have been recently reported in literature for the prediction of binding affinity change upon mutation, the performance is not appealing, and significant efforts are necessary to achieve the task. In addition, most of the protein-nucleic acid complexes have multiple copies of proteins (i.e, dimer/trimer) binding to nucleic acids, which have to be treated appropriately. Further, most of the methods focus on single amino acid mutations, and considering the effect of mutation in nucleic acid is also necessary. Accurately predicting the binding affinity of protein-nucleic acid complexes upon mutation and disease-causing mutations at the interface as well as integrating binding affinity change with diseases will have potential applications in drug discovery.

Declaration of competing interest

We state that there is no conflict of interest.

Acknowledgement

MMG thanks Dr. Yana Bromberg for her invitation to contribute the review article. We thank Sowmya Krishnan for her valuable suggestions and Indian Institute of Technology Madras for infrastructure facilities. The work was partially supported by the Science and Engineering Research Board, Department of Science and Technology, Government of India to MMG (CRG/2020/000314).

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. Jeon P, Ham HJ, Park S, Lee JA: **Regulation of cellular ribonucleoprotein granules: from assembly to degradation via post-translational modification**. *Cells* 2022, **11**:2063.

2. Jarmoskaite I, AlSadhan I, Vaidyanathan PP, Herschlag D: **How to measure and evaluate binding affinities**. *Elife* 2020, **9**, e57264.

3. Harini K, Phogat A, Gromiha MM: **Bioinformatics approaches for understanding the consequences of mutations on the binding affinity of protein-DNA complexes**. In *Protein mutations: consequences on structure, function and diseases*. Edited by Gromiha MM, Singapore: World Scientific; 2024:123–150.

4. Harini K, Krishnan SR, Sekijima M, Gromiha MM: **Computational resources for understanding the effect of mutations in binding affinities of protein–RNA complexes**. In *Protein mutations: consequences on structure, function and diseases*. Edited by Gromiha MM, Singapore: World Scientific; 2024:151–170.

5. Gromiha MM: *Protein Interactions: computational methods, analysis and applications*. Singapore: World Scientific; 2020.

6. Malik FK, Guo JT: **ComparePD: improving protein-DNA complex model comparison with hydrogen bond energy-based metrics**. *Proteins* 2023, **91**:1077–1088.

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.

One of the recently developed methods that used a three-track neural network architecture to predict the structures of protein-nucleic acid complexes. It incorporated different patterns in protein and DNA sequences, interactions between (i) amino acids, (ii) bases as well as (iii) amino acids and bases, and structural features to predict the structure of a protein-DNA/RNA complex.

8. •• Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L, Ballard AJ, Bambrick J, Bodenstein SW, Evans DA, Hung CC, 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 AI, Cowie A, Figurnov M, Fuchs FB, Gladman H, Jain R, Khan YA, Low CMR, Perlin K, Potapenko A, Savy P, Singh S, Stecula A, Thillaisundaram A, Tong C, Yakneen S, Zhong ED, Zielinski M, Žídek A, Bapst V, Kohli P, Jaderberg M, Hassabis D, Jumper JM: **Accurate structure prediction of biomolecular interactions with AlphaFold 3**. *Nature* 2024, **630**:493–500.

The method used diffusion-based architecture to predict the three-dimensional structure of a biomolecular complex with high accuracy. The method uses single residue properties, interactions, and evolutionary information for predicting the structures of any biomolecular complex including proteins, nucleic acids and small molecules.

9. 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.

10. Rodriguez-Lumbreras LA, Jimenez-Garcia B, Gimenez-Santamarina S, Fernandez-Recio J: **pyDockDNA: a new web server for energy-based protein-DNA docking and scoring**. *Front Mol Biosci* 2022, **9**, 988996.

11. Oliveira R, Pinho E, Sousa AL, Dias Ó, Azevedo NF, Almeida C: **Modelling aptamers with nucleic acid mimics (NAM): from sequence to three-dimensional docking**. *PLoS One* 2022, **17**, e0264701.

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

13. Esmaeeli R, Bauzá A, Perez A: **Structural predictions of protein-DNA binding: meld-dna**. *Nucleic Acids Res* 2023, **51**:1625–1636.

14. Qiu L, Zou X: **Scoring functions for protein-RNA complex structure prediction: advances, applications, and future directions**. *Commun Inf Syst* 2020, **20**:1–22.

15. Chen K, Zhou Y, Wang S, Xiong P: **RNA tertiary structure modeling with BRiQ potential in CASP15**. *Proteins* 2023, **91**:1771–1778.

16. •• 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.

ProNAB is the currently available largest database, which accumulates the experimentally determined thermodynamic parameters of protein-nucleic complexes and their mutants. It is integrated with sequence and structure of the protein and DNA/RNA, experimental conditions and literature for understanding the recognition mechanism of protein-nucleic complexes.

17. Gromiha MM, Harini K: **Comment on 'Thermodynamic database supports deciphering protein-nucleic acid interactions'**. *Trends Biotechnol* 2023, **41**:988–989.

18. Mei LC, Hao GF, Yang GF: **Thermodynamic database supports deciphering protein-nucleic acid interactions**. *Trends Biotechnol* 2023, **41**:140–143.

19. Chen J, Sawyer N, Regan L: **Protein-protein interactions: general trends in the relationship between binding affinity and interfacial buried surface area**. *Protein Sci* 2013, **22**:510–515.

20. Michielssens S, de Groot BL, Grubmüller H: **Binding affinities controlled by shifting conformational equilibria: opportunities and limitations**. *Biophys J* 2015, **108**:2585–2590.

21. Rohs R, Jin X, West SM, Joshi R, Honig B, Mann RS: **Origins of specificity in protein-DNA recognition**. *Annu Rev Biochem* 2010, **79**:233–269.

22. Pant P, Pathak A, Jayaram B: **Bicyclo-DNA mimics with enhanced protein binding affinities: insights from molecular dynamics simulations**. *J Biomol Struct Dyn* 2023, **41**:4040–4047.

23. Vuzman D, Levy Y: **Intrinsically disordered regions as affinity tuners in protein-DNA interactions**. *Mol Biosyst* 2012, **8**:47–57.

24. Ottoz DSM, Berchowitz LE: **The role of disorder in RNA binding affinity and specificity**. *Open Biol* 2020, **10**, 200328.

25. • Harini K, Sekijima M, Gromiha MM: **PRA-Pred: structure-based prediction of protein-RNA binding affinity**. *Int J Biol Macromol* 2024, **259**, 129490.

PRA-Pred is a reliable method with high performance for predicting the protein-RNA binding affinity, which uses multiple linear regression technique trained on interaction energies, RNA base step parameters, and atom contacts in the interface of the complex.

26. Harini K, Kihara D, Gromiha MM: **PDA-Pred: predicting the binding affinity of protein-DNA complexes using machine learning techniques and structural features**. *Methods* 2023, **213**:10–17.

27. Barissi S, Sala A, Wieczór M, Battistini F, Orozco M: **DNAffinity: a machine-learning approach to predict DNA binding affinities of transcription factors**. *Nucleic Acids Res* 2022, **50**:9105–9114.

28. • Yang S, Gong W, Zhou T, Sun X, Chen L, Zhou W, Li C: **emPDBA: protein-DNA binding affinity prediction by combining features from binding partners and interface learned with ensemble regression model**. *Briefings Bioinf* 2023, **24**, bbad192.

emPDBA is a regression model that predicts the binding affinity of the complex using the sequence, structure, and interface features of the protein based on single/double stranded DNA and percentage of residues in the interface.

29. Hong X, Tong X, Xie J, Liu P, Liu X, Song Q, Liu S, Liu S: **An updated dataset and a structure-based prediction model for protein-RNA binding affinity**. *Proteins* 2023, **91**:1245–1253.

30. • Pandey U, Behara SM, Sharma S, Patil RS, Nambiar S, Koner D, Bhukya H: **DeePNAP: a deep learning method to predict protein-nucleic acid binding affinity from their sequences**. *J Chem Inf Model* 2024, **64**:1806–1815.

This is the first sequence-based method to predict the binding affinity of protein-nucleic acid complexes and the affinity changes upon mutation using deep neural network architecture trained on the largest dataset from the ProNAB database.

31. Taniguchi H, Fujimoto A, Kono H, Furuta M, Fujita M, Nakagawa H: **Loss-of-function mutations in Zn-finger DNA-binding domain of HNF4A cause aberrant transcriptional regulation in liver cancer**. *Oncotarget* 2018, **9**:26144–26156.

32. Nichols NM, Matthews KS: **Human p53 phosphorylation mimic, S392E, increases nonspecific DNA affinity and thermal stability**. *Biochemistry* 2002, **41**:170–178.

33. Pires DEV, Ascher DB: **mCSM-NA: predicting the effects of mutations on protein-nucleic acids interactions**. *Nucleic Acids Res* 2017, **45**:W241–W246.

34. Zhang N, Chen Y, Zhao F, Yang Q, Simonetti FL, Li M: **PremPDI estimates and interprets the effects of missense mutations on protein-DNA interactions**. *PLoS Comput Biol* 2018, **14**, e1006615.

35. Zhang N, Lu H, Chen Y, Zhu Z, Yang Q, Wang S, Li M: **PremPRI: predicting the effects of missense mutations on protein-RNA interactions**. *Int J Mol Sci* 2020, **21**:5560.

36. Jiang Y, Liu HF, Liu R: **Systematic comparison and prediction of the effects of missense mutations on protein-DNA and protein-RNA interactions**. *PLoS Comput Biol* 2021, **17**, e1008951.

37. Peng Y, Sun L, Jia Z, Li L, Alexov E: **Predicting protein-DNA binding free energy change upon missense mutations using modified MM/PBSA approach: SAMPDI webserver**. *Bioinformatics* 2018, **34**:779–786.

38. Li G, Panday SK, Peng Y, Alexov E: **SAMPDI-3D: predicting the effects of protein and DNA mutations on protein-DNA interactions**. *Bioinformatics* 2021, **37**:3760–3765.

39. Callaway E: **Who will make AlphaFold3 open source? Scientists race to crack AI model**. *Nature* 2024, **630**:14–15.