



RNAproDB: A Webserver and Interactive Database for Analyzing Protein–RNA Interactions [☆]

Raktim Mitra^{1,†,‡}, Ari S. Cohen^{1,†}, Wei Yu Tang^{1,†}, Hiran Hosseini¹,
Yongchan Hong¹, Helen M. Berman^{1,2}, and Remo Rohs^{1,3,4,5,6,*}

1 - Department of Quantitative and Computational Biology, University of Southern California, Los Angeles, CA 90089, USA

2 - Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, 174 Frelinghuysen Road, Piscataway, NJ 08854, USA

3 - Department of Chemistry, University of Southern California, Los Angeles, CA 90089, USA

4 - Department of Physics & Astronomy, University of Southern California, Los Angeles, CA 90089, USA

5 - Thomas Lord Department of Computer Science, University of Southern California, Los Angeles, CA 90089, USA

6 - Department of Medicine, Division of Medical Oncology, University of Southern California, Los Angeles, CA 90033, USA

Correspondence to Remo Rohs: *Department of Quantitative and Computational Biology, University of Southern California, 1050 Childs Way, Los Angeles, CA 90089, USA. rohs@usc.edu (R. Rohs)

<https://doi.org/10.1016/j.jmb.2025.169012>

Editor: David Mathews

Abstract

We present RNAproDB (<https://rnaprodb.usc.edu/>), a new webserver, analysis pipeline, database, and highly interactive visualization tool, designed for protein–RNA complexes, and applicable to all forms of nucleic acid containing structures. RNAproDB computes several mapping schemes to place nucleic acid components and present protein–RNA interactions appropriately. Various structural annotations are computed including non-canonical base-pairing geometries, hydrogen bonds, and protein–RNA and RNA–RNA water-mediated interactions. This information is presented through integrated visualization and data tools. Subgraph selection facilitates studying smaller components of the interface. Molecular surface electrostatic potential can be visualized. RNAproDB enables analyzing and exploring experimentally determined, predicted, and designed protein–nucleic acid complexes. We present a quantitative analysis of pre-analyzed protein–RNA structures in RNAproDB revealing statistical patterns of molecular binding and recognition.

© 2025 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

The structural complexity of RNA molecules is vast as are their modes of interacting with proteins¹. Although there is an ever-expanding repertoire of structural data in the Protein Data Bank (PDB)^{2,3}, there is also a lack of data resources which systematically analyze these structures. In

addition, recent advances in artificial intelligence have made high-throughput prediction of protein–RNA complex structures possible^{4,5}. RNAproDB is a webserver and database where a user can upload a protein–RNA complex and explore various results and analyses (e.g., direct and water-mediated hydrogen bonds, base-pairing annotations, nucleotide modifications, secondary structure features, and surface electrostatics; [Figure 1](#), [Figure S1](#)).

Compared to existing resources, which often focus on the RNA structure alone and provide a single tool for visualization^{6,7}, or use a more coarse-grained

[☆] This article is part of a special issue entitled: ‘Computation Resources (2025)’ published in *Journal of Molecular Biology*.

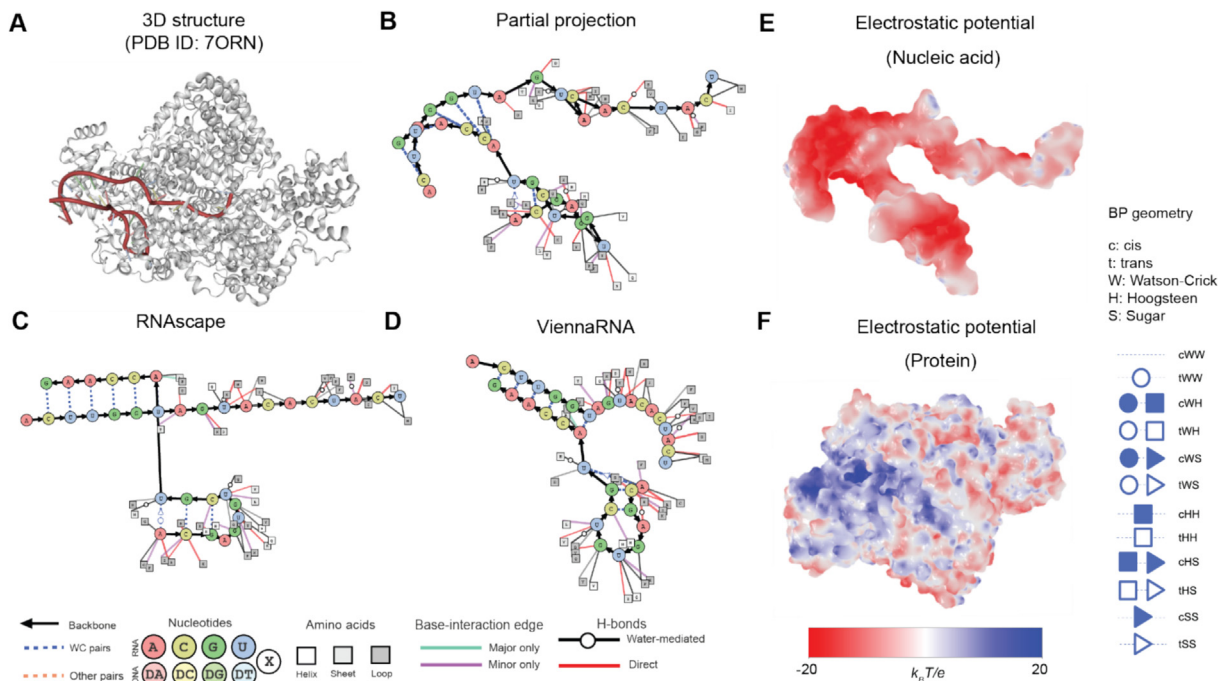


Figure 1. RNAProDB interactive mapping and surface electrostatics visualization of a protein–RNA complex. **A.** Three-dimensional (3D) structure view including solvent (PDB ID 7ORN). **B.** Mapping produced based on partial projection. **C.** Mapping produced based on RNAscape algorithm. **D.** Mapping produced after applying ViennaRNA secondary structure layout algorithm. All mappings are interactive and modifiable by the user. **E.** Electrostatics 3D surface visualization of nucleic acid components. **F.** Electrostatics 3D surface visualization of protein components.

representation⁸, RNAProDB provides three different algorithms for visualizing the nucleic acid layout and its interacting protein residues. The algorithms include a new layout, based on a partial projection of the structure that maximizes the spatial variance of nucleotides, a tertiary structure-aware mapping⁹, and a secondary structure-based mapping¹⁰. This information is presented via a highly interactive ‘Interface explorer’ integrated with a ‘Sequence viewer’, a ‘3D viewer’, and a ‘Secondary structure selector’. Tabular data corresponding to the ‘Interface explorer’ and a list of potential steric clashes (Figure S2), is also available. Another novel functionality, subgraph exploration, allows a user to explore parts of the structure (e.g., a particular junction region) through the ‘Interface explorer’ (Figure 2). Subgraph selections can be manually specified or visually selected from the ‘Secondary structure selector’. To the best of our knowledge, no existing tool offers such a capability.

In addition to the webserver, we provide analysis data for pre-analyzed structures containing RNA from the PDB in the form of a searchable collection. With a cutoff of 10,000 residues (nucleotides or amino acids) per biological assembly and a molecular weight cutoff of 800 kDa on the asymmetric unit, the initial RNAProDB collection contains more than 3,500

biological assemblies containing RNA molecules. We apply these cutoffs to exclude very large structures, allowing the web interface to perform efficiently and ensuring reasonable processing times for each uploaded structure within our server resources. Unlike DNAProDB¹¹, which focuses on the double-helical properties of DNA as its underlying framework for the analysis of protein–DNA recognition, RNAProDB is designed to be more broadly applicable to the much more diverse conformations of RNA compared to a DNA double-helix. As such, the definitions of helical parameters are not standardized (although some tools¹² attempt to compute analogous features) for RNA molecules¹³ (Figure S3C).

The broadly applicable design allowed us to also process pre-analyzed nucleic acid containing structures of other types (e.g., DNA, DNA/RNA hybrids (Figure S4)). RNAProDB annotates water-mediated RNA–RNA and protein–RNA interactions. This allows annotating base-pairings that occur purely through water-mediated interactions, which may not be detected as a pair by existing tools¹² (Figure S5). The RNAProDB collection is automatically updated weekly with new PDB entries. We believe that this tool will be a valuable resource for the nucleic acid community interested in protein–nucleic acid interaction and

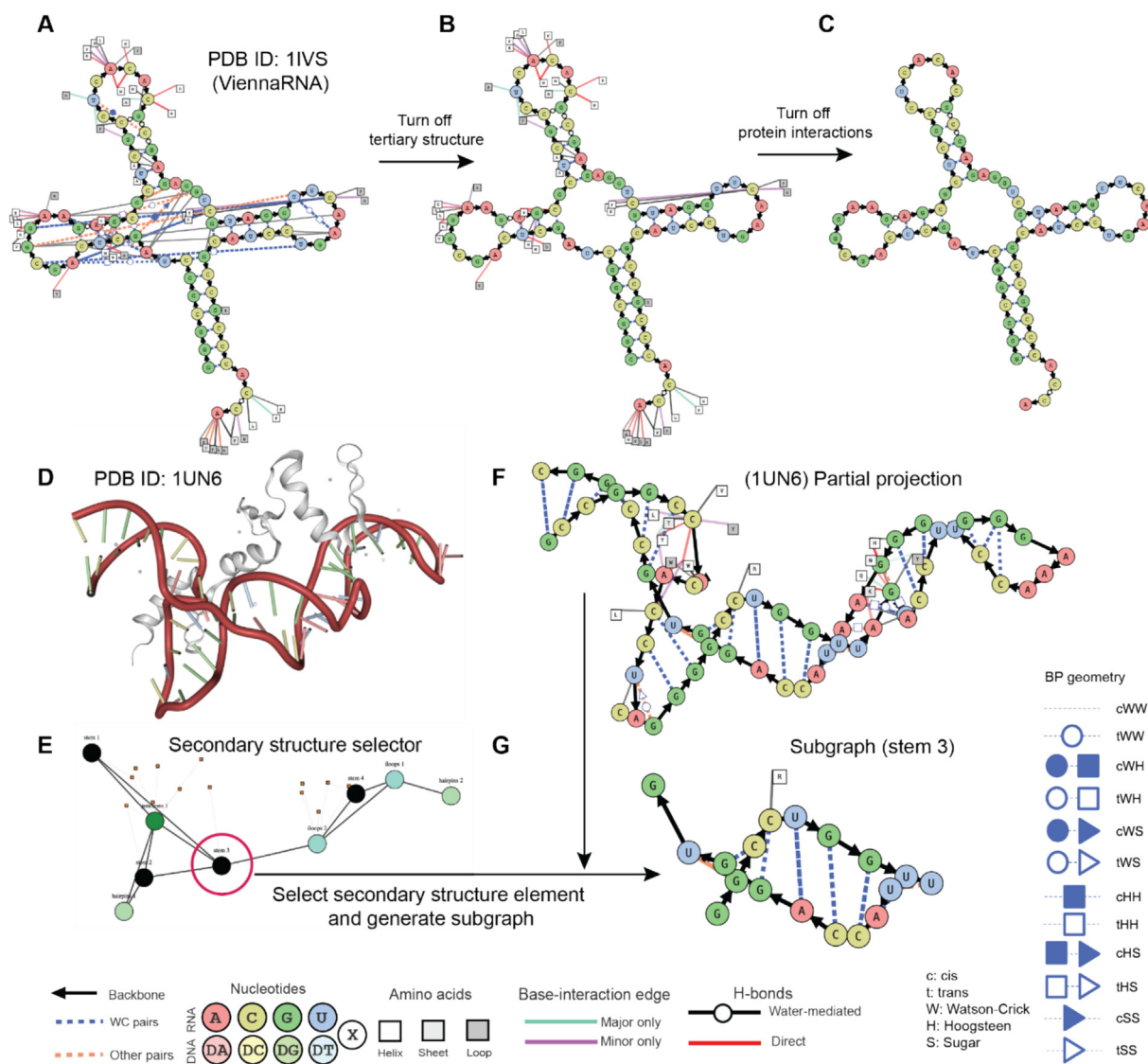


Figure 2. Selected projection and subgraph exploration features available in RNAproDB. **A.** Secondary structure layout obtained via the ViennaRNA¹⁰ tool for PDB ID 1IVS, as displayed in RNAproDB with added tertiary and protein interactions. **B.** Layout obtained after removing tertiary interactions. **C.** RNA-only secondary structure layout obtained after removing both tertiary and protein interactions. **D.** 3D structure view of PDB ID 1UN6 in RNAproDB. **E.** Selecting a specific secondary structure element (stem 3) in the secondary structure selector adds its nodes and their first-order neighbors to the subgraph dialogue box. **F.** RNAproDB partial projection mapping of PDB ID 1UN6. **G.** Subgraph generated from the secondary structure selector in E, corresponding to stem 3.

function, structure prediction⁴, analysis¹⁴ and design^{15,16}. RNAproDB is freely available for all users at <https://rnaprodb.usc.edu/>.

Processing pipeline

The RNAproDB processing pipeline processes a structure file and computes multiple visualizations and interaction features. As part of the processing pipeline, multiple software is run including DSSR¹² (base-pairing geometries, protein–RNA hydrogen bonds, and RNA secondary structure), HBPLUS¹⁷

(hydrogen bonds involving water molecules), RNAscape⁹ (tertiary structure-aware nucleotide mapping), ViennaRNA¹⁰ (secondary structure-based nucleotide mapping), DSSP^{18,19} (protein secondary structure analysis), and TABI-PB²⁰ (surface electrostatics). In addition, a new ‘Partial projection’ based mapping algorithm was designed to position nucleotides. The processing pipeline combines mapping algorithms and processed data to create highly interactive and integrated representations of a structure, namely, an ‘Interface explorer’, ‘3D viewer’, ‘Secondary structure selector’,

'Sequence viewer', 'Electrostatics viewer', and access to 'Tabular data'. For the pre-analyzed collection, the selection of biological assembly 1 for each PDB entry was used. In the following sections, we expand on the functionalities offered by these various components.

Interface explorer

The RNAproDB 'Interface explorer' presents an interactive graph of an RNA structure and contacting protein residues. The explorer offers three different layout algorithm options selectable by the user. The first (default) option, partial projection, is a new mapping scheme produced by using Principal Component Analysis (PCA)²¹ to project residues onto the 2D plane that maximizes the spatial variance of the nucleotides. We call this method 'Partial projection' because we exclude the protein residues from calculating the direction of projection as they tend to skew the visualization, making it less useful and cluttered (Figure S6). In our observation, this scheme is the most visually intuitive for exploring the corresponding 3D structure. Two other options include a secondary structure-based mapping (computed using ViennaRNA¹⁰) and a tertiary structure-aware mapping (computed using RNAscape⁹). Nucleic acid residues are placed using these algorithms, while protein residues are placed using a force-directed layout scheme. The force-directed layout is implemented using D3.js library²².

We demonstrate the three layout schemes for the La Crosse virus polymerase at replication initiation stage²³ (PDB ID 7ORN) (Figure 1A). The partial projection-based layout is shown in Figure 1B. This layout reflects the helical turns, corresponding with the 3D structure. The RNAscape⁹ layout Figure 1C is more intuitive and suitable for users who prefer a ladder-like helical representation. The secondary structure-based layout, shown in Figure 1D, although familiar, is usually obscured by tertiary interactions crisscrossing the plot. We present this information juxtaposed with the 3D structure (along with computed surface electrostatics) for a few more PDB entries: the structure of SARS-CoV-2 nsp10-nsp14 (WT)-RNA²⁴ (PDB ID 7N0B) is presented in Figure S3A, which is important for understanding mismatch recognition in A-form double-helical RNA structure by a SARS-CoV-2 proofreading enzyme²⁴ (a case study is available on the RNAproDB documentation page). Telomerase reverse transcriptase from the fungus *Candida tropicalis* demonstrates a more complicated topology of the RNA structure²⁵ (PDB ID 6ZDP), and its corresponding layout schemes are presented by RNAproDB (Figure S3B). Finally, we show the layouts for a more complicated structural variation demonstrated through the type III-E CRISPR-Cas7-11 effector complex¹³ (PDB ID 7WAH, Figure S3C).

DNA/RNA-hybrid structures are also supported, and example layouts for a Cas9-DNA/RNA complex are presented in Figure S4. Modified nucleotides are shown either as an 'x' or a lower-case letter version of their parent standard nucleotide, if available. The RNAproDB interface explorer and provided layouts are more detailed and flexible compared to existing options which attempt to show protein interactions, e.g., TBI-Forna⁶ (see Supplementary information). We also demonstrate the protein-RNA representation of PDB ID 7ORN computed by Forna in Figure S7 (analogous to RNAproDB representations in Figure 1).

Options to turn off tertiary interaction edges and protein interactions were implemented to provide a more intuitive view of the secondary structure. This is demonstrated in (Figure 2A–C) for the secondary structure layout (Figure 2A) of a tRNA molecule⁵⁵ (PDB ID 11VS). The user can turn off the tertiary interactions within the nucleic acids to produce a representation as shown in Figure 2B. The protein interactions can be turned off to produce a nucleic acid-only layout (Figure 2C).

The protein-RNA interactions in the 'Interface explorer' are determined using an atom-atom cutoff distance of 6 Å. This distance is a conservative choice for identifying molecular interactions. Typical hydrogen bonds range from 2.7 to 3.3 Å²⁶. Beyond hydrogen bonds, van der Waals (vdW) and electrostatic interactions contribute to binding. In addition, in the 'Interface explorer', users can adjust the distance threshold of visible protein-RNA interactions, which is based on the centroid-centroid distance (0–15 Å) between a protein residue and a nucleotide. This value can be changed using a slider, and upon modification, edge distances that fall beyond the threshold are automatically hidden.

Hydrogen bonds, including water-mediated hydrogen bonds, are detected in the 'Interface explorer' using HBPLUS¹⁷ with a 3 Å hydrogen-acceptor cutoff ('-h') and a 3.5 Å donor-acceptor cutoff ('-d'). We retain the default value for the bond angle parameters. These parameters reflect typical hydrogen bond distances²⁶ and maintain consistency with our other established tools¹¹. In addition, protein-RNA interaction edges have a distance-dependent opacity setting, making more distant interactions less prominent and reducing visual clutter. When a protein residue interacts only with the major groove edge or the minor groove edge (as in a Watson-Crick conformation) of a base pair, it is assigned a special color. Alternate conformations are displayed in gray.

At any point, the full graph can be turned into a force-directed layout via a toggle button ('Relax graph' option). This provides additional ability to move larger sections of a structure, with the tradeoff of lifting pre-computed layout constraints. Turning off the force-directed layout allows a user

to reposition specific nucleotides and residues by clicking and dragging. Other graph customization features include rotating and reflecting the graph across the *X* and *Y* axes and toggling the indication of hydrogen bonds.

Leontis–Westhof^{27,57} base pair annotations (as computed by DSSR¹²) and water-mediated hydrogen bonds are displayed via different shapes present on the graph edges. Hovering over any residue or edge will display additional relevant information for the user. At any point during exploration, a user can download the visualization as a static picture or scalable vector graphic format image along with the corresponding graphical data.

Sequence viewer and 3D viewer

Alongside the ‘Interface explorer’, we provide a ‘3D viewer’ and ‘Sequence viewer’ to facilitate structure exploration (Figure S8). The ‘Sequence viewer’ lists the chains within each structure. After selecting a chain, the corresponding sequence will be displayed to the user. Residue numbers are displayed above the sequence, and hovering over a specific residue will show its number. Clicking a specific residue/nucleotide in the ‘Sequence viewer’ highlights the corresponding residue/nucleotide in the ‘Interface explorer’, while simultaneously zooming and orienting the ‘3D viewer’ to focus on the specific residue/nucleotide (shown in a ball-and-stick view). Finally, if the subgraph selection dialogue box is open, it will populate with the ID of the selected residue/nucleotide. With these features, users more comfortable working with a sequence can quickly orient themselves and visualize relevant areas of a structure. Individual or multiple residues/nucleotides can also be selected from the ‘Interface explorer’ via a single click or shift-click, respectively. Upon doing so, corresponding residue or residues in the ‘3D viewer’ will simultaneously come into focus. Moreover, right-clicking on consecutive atoms in the ‘3D viewer’ allows a user to visualize distance measurements (Figure S5CD). Buttons to show/hide cartoon representation and solvent are also available. The 3D viewer is implemented using ngl.js²⁸, a web-based library for molecular graphics.

Secondary structure selector and subgraph exploration

A more coarse-grained diagram, ‘Secondary structure selector’, is computed in the RNAproDB processing pipeline. The structural elements (stems, loops, hairpins, junctions, etc.) are detected using DSSR¹² and mapped to the partial projection layout (via averaging corresponding residue coordinates). Thus, this diagram presents each secondary structure element as one node which can be interacted with by the user. An example⁵⁶

is shown for PDB ID 1UN6 (Figure 2D). The corresponding ‘Secondary structure selector’ is shown in Figure 2E. The partial projection-based layout for this structure is presented in Figure 2F.

Whenever a user clicks a particular node (e.g., ‘Stem 3’ in Figure 2E), the nucleic acid residue IDs that the node encompasses are used for subgraph generation. A user can then click the ‘Generate subgraph’ button to create and explore a subgraph (containing up to first-order neighbors) via the ‘Interface explorer’ (Figure 2G). Subgraphs can also be created from the ‘Sequence viewer’ and ‘Interface explorer’ panels, or they can be manually specified.

Analysis of various protein–RNA and RNA–RNA interactions

We analyzed protein–RNA entries in the RNAproDB collection as of October 15, 2024, to produce a quantitative summary of various interactions. The cumulative count of these entries, based on their release dates, is presented in Figure 3A, demonstrating a continuous exponential growth. We explored the relative abundance of different standard nucleotides (A, C, G, and U) in helical vs. non-helical regions (as computed by DSSR¹²). For each nucleotide, we determined whether it was part of a helix and plotted the frequency ratio of each nucleotide type within these regions, across all entries, calculated by:

$$\text{Relative Abundance } (b, R) = \frac{|b \text{ in region } R|}{|\text{nucleotides}|}$$

where $b \in \{A, C, G, U\}$ and $R \in \{\text{helical, non-helical}\}$.

Helical regions in general contain more nucleotides (Figure 3B). However, the proportions of different bases within helical regions differ from non-helical regions (which shows a lower proportion of G or C compared to A or U). Frequency of different amino acids involved in direct hydrogen bond interactions with RNA bases were also computed, as shown in Figure 3C. For each amino acid, the bar plot was proportionately divided by the corresponding interacting RNA bases. Differences in interaction patterns are visualized. While Arg and Lys drive most of these interactions, notably, some amino acids (Glu, Asp, etc.) are disproportionately favored to interact with non-helical regions.

Non-canonical base-pairing is very common in RNA structures and often influences the structural organization of a molecule and its interaction with other molecules²⁹. We quantified the propensity of base-pairing (as detected by DSSR¹²) between different RNA bases (Figure 3D). The ratio of each base-pair type was determined relative to the total number of pairings, with symmetrical pairings (e.g., C/G and G/C) considered equivalent. Specifically, the ratio is defined by

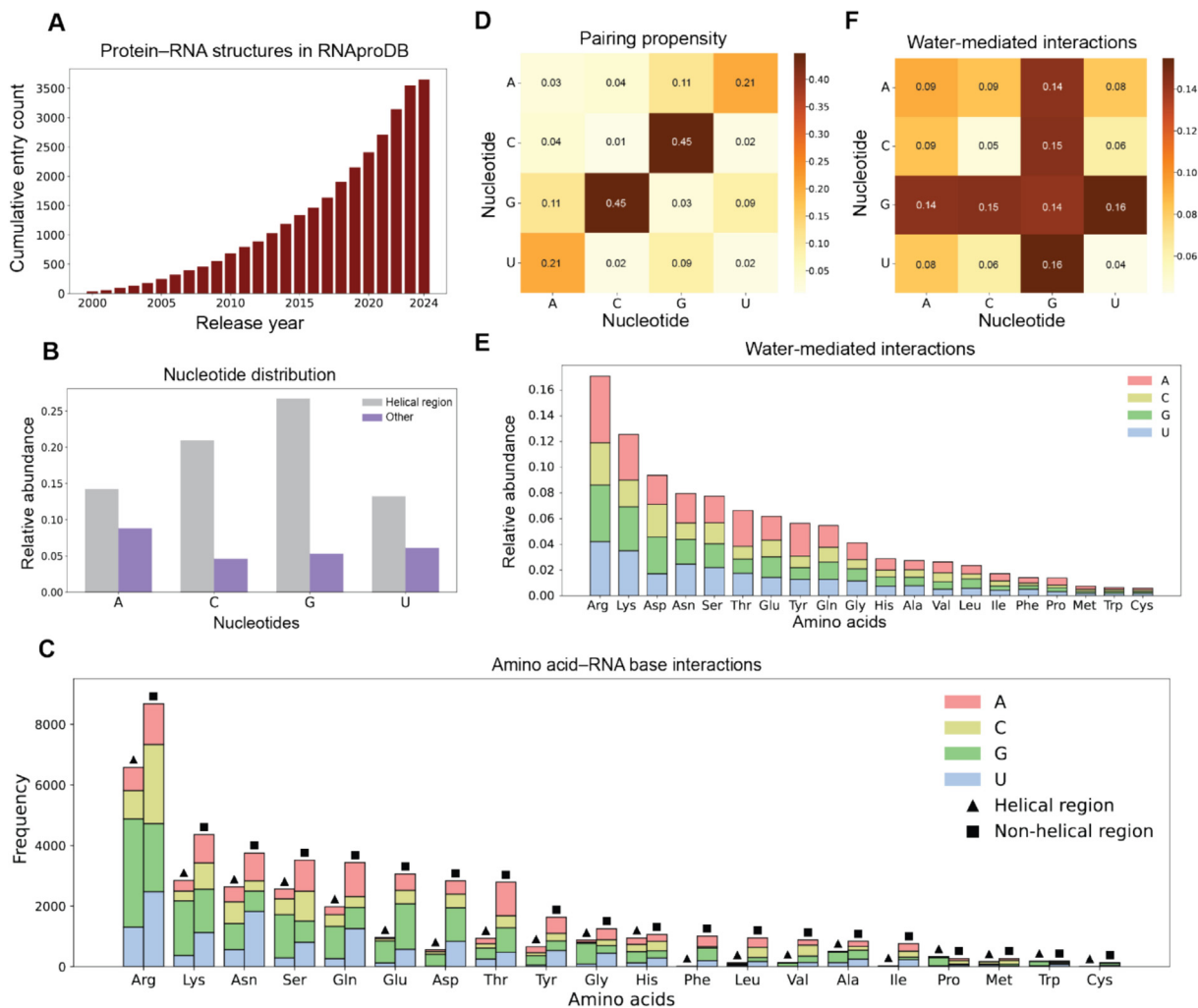


Figure 3. Statistical analysis of the protein–RNA interactions in RNAProDB collection. **A.** Cumulative count of protein–RNA entries within the RNAProDB collection per release year in the PDB. **B.** Relative abundance of nucleotides in helical and non-helical regions. **C.** Frequency of amino acid interactions (direct hydrogen bonds) with RNA bases, segregated by helical and non-helical regions (as detected by DSSR¹²). **D.** Propensity of base pairing between nucleotides. **E.** Relative abundance of water-mediated protein–RNA interactions involving each amino acid. Each bar is proportionally divided into abundances of corresponding nucleotide partners. **F.** Propensity of nucleotide–nucleotide water-mediated interactions.

$$\text{Heatmap Ratio } (x, y) = \frac{|pairing(x, y)| + (1 - \delta_{x,y})|pairing(y, x)|}{\sum_{x', y'} |pairing(x', y')|}$$

where $pairing(x, y)$ denotes the pairing between nucleotides of type x and type y and $\delta_{x,y} = 1$ if $x = y$ and 0 otherwise.

Water-mediated hydrogen bonds play a role in protein–nucleic acid recognition³⁰ (and sometimes form very crucial interactions^{31,32}). We compute a quantitative summary of these interactions as shown in Figure 3E. Their relative abundance was computed by

$$\text{Relative Abundance } (a) = \frac{|interactions \text{ involving amino acid } a|}{\sum_a |interactions \text{ involving amino acid } a|}$$

where a denotes one of the 20 standard amino acids. Only water-mediated interactions with RNA bases were considered (not with the RNA backbone). Polar, charged amino acids (Arg, Lys, and Asp) appear to contribute most in terms of water-mediated interactions.

Another important aspect to discuss is RNA–RNA water-mediated interactions.^{33,34} Various combinations of base pairings can lead to sub-optimal direct hydrogen bonds between these pairs or adjacent bases. Water molecules appear to compensate for such situations by making pairing possible through water-mediated hydrogen bonds. One such example is the CUG repeat structure from PDB ID 7Y2B³⁵ (Figure S5A). The U/U mismatches in this

structure are often unable to form direct hydrogen bonds (specifically, the central U/U mismatch forms no direct hydrogen bond). Therefore, DSSR¹² does not classify it as a base pair. However, two water molecules form water-mediated hydrogen bonds between the two U bases. This has a noticeable impact on the structure, as the U-U mismatch region has a distance across the base pair that is similar to the C/G paired regions (Figure S5D), while it would otherwise be shorter. RNAproDB computes RNA–RNA water-mediated hydrogen bonds and displays them in the ‘Interface explorer’ (Figure S5B). We demonstrate the molecular conformation of the central U/U mismatch (Figure S5C) and an off-center U/U mismatch (Figure S5D). We present a quantitative summary of such phenomena in Figure 3F, where we quantify the proportions of water-mediated interactions between different RNA nucleotides (in an analogous manner to Figure 3D).

We also explored the RNA family diversity of the pre-analyzed set of protein–RNA entries. For each protein–RNA entry, we extracted the RNA sequence and matched it with families in the Rfam (RNA families) database.³⁶ This matching process was conducted using the Rfam sequence search tool, where every query is compared with the Rfam library of RNA sequence families. Entries for which the query produced a result were compiled to compute the distribution of RNA families (depicted in Figure S9). Among top representatives, tRNA and rRNAs are prevalent. CRISPR-related, RNases, and microRNAs are also well represented.

Electrostatics

Surface electrostatics of proteins and nucleic acids play a critical role in binding.^{37,38} For example, electrostatic interactions are crucial for both guiding a protein toward RNA at long range (global landscape) and stabilizing the protein–RNA complex upon close contact (local landscape).³⁷ Nucleic acids, being negatively charged, can attract positively charged proteins from a comparably long range for molecular interaction. More intricately, global asymmetries and local differences in nucleic acid shape can affect electrostatic potential, leading to differences in protein selectivity.^{1,39–41}

To facilitate studying these phenomena, RNAproDB offers an ‘Electrostatics’ panel to view the electrostatic potential on the molecular surface of a nucleic acid, protein, and their full complex in three dimensions. Tree Accelerated Boundary Integral method (TABI-PB)²⁰ for solving the linear Poisson-Boltzmann equation is used to obtain this data (with no explicit solvent present, solvent dielectric constant set to 78, solute dielectric constant set to 8, temperature set to 310 K, see Data availability for further details), which computes the potential on a triangulated surface mesh of the structure⁴². This choice was made because of the efficiency of this method (as opposed to solving the non-linear

Poisson-Boltzmann equation with a 3D boundary box⁴³) and viability for on-demand use of the web-server. The electrostatics panel supports full rotation and zooming of a structure, allowing pockets and less accessible areas to be explored. Differences in electrostatic potential can be intuitively seen in red (negative potential) and blue (positive potential) color shading.

Electrostatic potential visualizations are pre-computed for structures in the RNAproDB collection. For user-uploaded structures, an option to compute the electrostatic potential is provided in the report page after initial processing is complete. We show here various examples of computed electrostatics data (Figure 1EF, Figure S3A–C), demonstrating a general trend of a largely positive binding site on the protein surface and strong negative electrostatic potential at nucleic acid backbones and minor grooves of double-helical regions. The colorized surface mesh file visualized in the ‘Electrostatics’ panel can be downloaded by the user for further downstream use. We note that the electrostatic potential calculation in RNAproDB provides a qualitative analysis of positively vs. negatively charged surface patches, whereas the calculation of electrostatic potential values requires solving the non-linear Poisson-Boltzmann equation due to the substantial negative charges of nucleic acids,^{39,44} which goes beyond a database approach.

Search functionalities and upload

The pre-analyzed RNAproDB collection contains more than 3,500 protein–RNA structures. In addition, DNA/RNA-hybrid structures, RNA structures without proteins, and DNA structures are also pre-processed and available to search. A specific PDB ID can be viewed using the quick search box available at the top-right of the website header. More complex searches can be performed using the ‘Search’ page. The ‘Search’ page supports various search queries including author names and keywords related to the structure of interest. Additional filters include molecule type (polymer entity type: RNA, DNA, or DNA/RNA hybrid), experimental modality, resolution range, publication year, number of nucleic acids polymers, number of protein polymers, molecular weight of the biological assembly, etc. The search results are presented in a compact ‘Table view’. Optionally, a ‘Card view’ showing large image previews of the structure is available. The resulting PDB IDs can then be copied to the clipboard, and the search results can be downloaded in either comma-separated value (CSV) or JavaScript object notation (JSON) formats.

Structure files containing a biological assembly in mmCIF format (up to 10 MB) can be uploaded to the RNAproDB webserver via the upload page. Upon upload, the server will process the structure and create a custom link displaying its analysis. We

expect that this feature will be valuable for researchers looking to analyze predicted complexes (e.g., from AlphaFold 3⁴) of proteins and nucleic acids. The ‘Interface explorer’ in an example output page for an uploaded AlphaFold 3 predicted structure (model 0) (based on sequences from PDB ID 1IVS) is shown in [Figure S8](#).

Discussion

Currently, a lack of a modern web-software exists enabling structural analysis and highly interactive visualizations of protein–RNA interactions. We developed RNAproDB to analyze these interactions while also supporting a broader class of nucleic acid containing structures, which include RNA, DNA, and RNA/DNA hybrid structures. RNAproDB detects water-mediated hydrogen bonds^{45,46} and enables the calculation and visualization of electrostatic potential.^{47,48} RNAproDB analysis required innovating ways of presenting data and interfaces such as the partial projection layout and subgraph exploration, paired with secondary structure selection. RNAproDB provides access to DNA/RNA-hybrid structures such as Cas9 bound to target DNA and guide RNA⁴⁹ and structures related to the newly developed bridge editing technique for genome editing,^{50,51} which previously lacked an interface for analysis and interactive visualization and falls beyond the scope of tools like DNAProDB¹⁰ or Forna⁶ (see [Supplementary Information](#)).

RNAproDB is also suitable for analyzing simulated⁵² or predicted complexes enabled by recent advances such as AlphaFold 3,⁴ RoseTTAfold-NA,⁵³ or RoseTTAfold-All-Atom⁵⁴ predictions, which can be uploaded to the web-server. RNAproDB also analyzes and visualizes surface electrostatics, enabling users to study this data in tandem with the 3D structure and the ‘Interface explorer’. RNAproDB also supports search capabilities to explore its diverse collection. It offers a ‘Sequence viewer’ and subgraph features that may help researchers more familiar with sequence-based representations to use RNAproDB. The visualization and exploration features are extensively integrated, and the multiple layout algorithms offered by the ‘Interface explorer’ cater to a variety of users and use cases. In addition to the list of capabilities described above, RNAproDB requires no programming experience by users. We hope that RNAproDB serves as a valuable web-server and database for the structural and cellular biology communities.

Funding sources

The authors acknowledge financial support by an Andrew J. Viterbi Fellowship in Computational

Biology and Bioinformatics (to R.M.) and the National Institutes of Health [R35GM130376 to R.R.].

CRedit authorship contribution statement

Raktim Mitra: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Formal analysis, Conceptualization. **Ari S. Cohen:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Formal analysis, Conceptualization. **Wei Yu Tang:** Writing – review & editing, Writing – original draft, Visualization, Software, Formal analysis, Conceptualization, Methodology. **Hirad Hosseini:** Visualization, Software. **Yongchan Hong:** Visualization, Software. **Helen M. Berman:** Writing – review & editing, Conceptualization. **Remo Rohs:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Conceptualization, Funding acquisition.

DATA AVAILABILITY

RNAproDB is freely available to users at <https://rnaprodb.usc.edu/>, and via GitHub at <https://github.com/timkartar/RNAproDB> and https://github.com/ariscohen/RNAproDB_frontend.

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

The authors thank current and former Rohs Lab members including Jared M. Sagendorf for valuable discussions and feedback. The authors acknowledge Luigi Manna for setup and maintenance of the RNAproDB webserver.

Appendix A. Supplementary material

Supplementary material to this article can be found online at <https://doi.org/10.1016/j.jmb.2025.169012>.

Received 23 November 2024;
Accepted 12 February 2025;
Available online 15 February 2025

Keywords:

RNA structure;
DNA structure;

protein-nucleic acid binding;
molecular recognition;
nucleic acids

† These authors should be regarded as Joint First Authors.

‡ Present address: Institute for Protein Design, University of Washington, Seattle, WA 98195, USA.

References

- Jones, S., (2001). Protein-RNA interactions: a structural analysis. *Nucleic Acids Res.* **29**, 943–954. <https://doi.org/10.1093/nar/29.4.943>.
- wwPDB consortium, (2019). Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic Acids Res.* **47**, D520–D528. <https://doi.org/10.1093/nar/gky949>.
- 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. <https://doi.org/10.1093/nar/28.1.235>.
- Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., et al., (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500. <https://doi.org/10.1038/s41586-024-07487-w>.
- Watson, J.L., Juergens, D., Bennett, N.R., Trippie, B.L., Yim, J., Eisenach, H.E., Ahern, W., Borst, A.J., Ragotte, R. J., Milles, L.F., et al., (2023). De novo design of protein structure and function with RFdiffusion. *Nature* **620**, 1089–1100. <https://doi.org/10.1038/s41586-023-06415-8>.
- Kerpedjiev, P., Hammer, S., Hofacker, I.L., (2015). Forna (force-directed RNA): Simple and effective online RNA secondary structure diagrams. *Bioinformatics* **31**, 3377–3379. <https://doi.org/10.1093/bioinformatics/btv372>.
- Yang, H., Jossinet, F., Leontis, N., Chen, L., Westbrook, J., Berman, H., Westhof, E., (2003). Tools for the automatic identification and classification of RNA base pairs. *Nucleic Acids Res.* **31**, 3450–3460. <https://doi.org/10.1093/nar/gkg529>.
- Chojnowski, G., Waleń, T., Bujnicki, J.M., (2014). RNA Bricks—a database of RNA 3D motifs and their interactions. *Nucleic Acids Res.* **42**, D123–D131. <https://doi.org/10.1093/nar/gkt1084>.
- Mitra, R., Cohen, A.S., Rohs, R., (2024). RNAscape: geometric mapping and customizable visualization of RNA structure. *Nucleic Acids Res.* **52**, W354–W361. <https://doi.org/10.1093/nar/gkae269>.
- Lorenz, R., Bernhart, S.H., Höner zu Siederdisen, C., Tafer, H., Flamm, C., Stadler, P.F., Hofacker, I.L., (2011). ViennaRNA Package 2.0. *Mol. Biol.* **6**, 26. <https://doi.org/10.1186/1748-7188-6-26>.
- Mitra, R., Cohen, A.S., Sagendorf, J.M., Berman, H.M., Rohs, R., (2025). DNAProDB: an updated database for the automated and interactive analysis of protein–DNA complexes. *Nucleic Acids Res.* **53**, D396–D402. <https://doi.org/10.1093/nar/gkae970>.
- Lu, X.-J., Bussemaker, H.J., Olson, W.K., (2015). DSSR: an integrated software tool for dissecting the spatial structure of RNA. *Nucleic Acids Res.* **43**, e142. <https://doi.org/10.1093/nar/gkv716>.
- Kato, K., Zhou, W., Okazaki, S., Isayama, Y., Nishizawa, T., Gootenberg, J.S., Abudayyeh, O.O., Nishimasu, H., (2022). Structure and engineering of the type III-E CRISPR-Cas7-11 effector complex. *Cell* **185**, 2324–2337. e16. <https://doi.org/10.1016/j.cell.2022.05.003>.
- Mitra, R., Li, J., Sagendorf, J.M., Jiang, Y., Cohen, A.S., Chiu, T.P., Glasscock, C.J., Rohs, R., (2024). Geometric deep learning of protein–DNA binding specificity. *Nature Methods* **21**, 1674–1683. <https://doi.org/10.1038/s41592-024-02372-w>.
- Mohr, S.E., Hu, Y., Ewen-Campen, B., Housden, B.E., Viswanatha, R., Perrimon, N., (2016). CRISPR guide RNA design for research applications. *FEBS J.* **283**, 3232–3238. <https://doi.org/10.1111/febs.13777>.
- Glasscock, C.J., Pecoraro, R., McHugh, R., Doyle, L.A., Chen, W., Boivin, O., Lonquist, B., Na, E., Politanska, Y., Haddox, H.K., et al., (2023). Computational design of sequence-specific DNA-binding proteins. *bioRxiv*. <https://doi.org/10.1101/2023.09.20.558720>.
- McDonald, I.K., Thornton, J.M., (1994). Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238**, 777–793. <https://doi.org/10.1006/jmbi.1994.1334>.
- Kabsch, W., Sander, C., (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637. <https://doi.org/10.1002/bip.360221211>.
- Joosten, R.P., te Beek, T.A.H., Krieger, E., Hekkelman, M. L., Hooft, R.W.W., Schneider, R., Sander, C., Vriend, G., (2011). A series of PDB related databases for everyday needs. *Nucleic Acids Res.* **39**, D411–D419. <https://doi.org/10.1093/nar/gkq1105>.
- Wilson, L., Geng, W., Krasny, R., (2022). TABI-PB 2.0: an improved version of the treecode-accelerated boundary integral Poisson–Boltzmann solver. *J. Phys. Chem. B* **126**, 7104–7113. <https://doi.org/10.1021/acs.jpcc.2c04604>.
- Abdi, H., Williams, L.J., (2010). Principal component analysis. *WIREs Comp. Stats.* **2** (4), 433–459. <https://doi.org/10.1002/wics.101>.
- Bostock, M., Ogievetsky, V., Heer, J., (2011). D³ data-driven documents. *IEEE Trans. Visualiz. Comp. Graph.* **17** (12), 2301–2309. <https://doi.org/10.1109/TVCG.2011.185>.
- Arragain, B., Durieux Trouilleton, Q., Baudin, F., Provaznik, J., Azevedo, N., Cusack, S., Schoehn, G., Malet, H., (2022). Structural snapshots of La Crosse virus polymerase reveal the mechanisms underlying Peribunyaviridae replication and transcription. *Nature Commun.* **13**, 902. <https://doi.org/10.1038/s41467-022-28428-z>.
- Liu, C., Shi, W., Becker, S.T., Schatz, D.G., Liu, B., Yang, Y., (2021). Structural basis of mismatch recognition by a SARS-CoV-2 proofreading enzyme. *Science* **373**, 1142–1146. <https://doi.org/10.1126/science.abi9310>.
- Zhai, L.-T., Rety, S., Chen, W.-F., Song, Z.-Y., Auguin, D., Sun, B., Dou, S.-X., Xi, X.-G., (2021). Crystal structures of N-terminally truncated telomerase reverse transcriptase from fungi. *Nucleic Acids Res.* **49**, 4768–4781. <https://doi.org/10.1093/nar/gkab261>.
- Stickle, D.F., Presta, L.G., Dill, K.A., Rose, G.D., (1992). Hydrogen bonding in globular proteins. *J. Mol. Biol.* **226** (4), 1143–1159. [https://doi.org/10.1016/0022-2836\(92\)91058-w](https://doi.org/10.1016/0022-2836(92)91058-w).
- Hermann, T., Westhof, E., (1999). Non-Watson-Crick base pairs in RNA-protein recognition. *Chem. Biol.* **6**, R335–R343. [https://doi.org/10.1016/s1074-5521\(00\)80003-4](https://doi.org/10.1016/s1074-5521(00)80003-4).
- Rose, A.S., Bradley, A.R., Valasatava, Y., Duarte, J.M., Prilič, A., Rose, P.W., (2016). Web-based molecular

- graphics for large complexes. In: *Proceedings of the 21st international conference on Web3D technology*, pp. 185–186. <https://doi.org/10.1145/2945292.2945324>.
29. Olson, W.K., Li, S., Kaukonen, T., Colasanti, A.V., Xin, Y., Lu, X.-J., (2019). Effects of noncanonical base pairing on RNA folding: structural context and spatial arrangements of G-A pairs. *Biochemistry* **58**, 2474–2487. <https://doi.org/10.1021/acs.biochem.9b00122>.
30. Barik, A., Bahadur, R.P., (2014). Hydration of protein–RNA recognition sites. *Nucleic Acids Res.* **42**, 10148–10160. <https://doi.org/10.1093/nar/gku679>.
31. Schirle, N.T., Sheu-Gruttadauria, J., Chandradoss, S.D., Joo, C., MacRae, I.J., (2015). Water-mediated recognition of t1-adenosine anchors Argonaute2 to microRNA targets. *Elife* **4** <https://doi.org/10.7554/eLife.07646>.
32. Otwinowski, Z., Schevitz, R.W., Zhang, R.-G., Lawson, C. L., Joachimiak, A., Marmorstein, R.Q., Luisi, B.F., Sigler, P. B., (1988). Crystal structure of trp repressor/operator complex at atomic resolution. *Nature* **335**, 321–329. <https://doi.org/10.1038/335321a0>.
33. Singh, O., Venugopal, P.P., Chakraborty, D., (2023). Effect of water models on the stability of RNA: role of counterions. *Chem. Phys. Impact* **7**, 100313. <https://doi.org/10.1016/j.chphi.2023.100313>.
34. Fingerhut, B.P., (2021). The mutual interactions of RNA, counterions and water – quantifying the electrostatics at the phosphate–water interface. *Chem. Commun.* **57**, 12880–12897. <https://doi.org/10.1039/D1CC005367A>.
35. Wang, S.-C., Chen, Y.-T., Satange, R., Chu, J.-W., Hou, M.-H., (2023). Structural basis for water modulating RNA duplex formation in the CUG repeats of myotonic dystrophy type 1. *J. Biol. Chem.* **299**, 104864. <https://doi.org/10.1016/j.jbc.2023.104864>.
36. Kalvari, I., Nawrocki, E.P., Ontiveros-Palacios, N., Argasinska, J., Lamkiewicz, K., Marz, M., Griffiths-Jones, S., Toffano-Nioche, C., Gautheret, D., Weinberg, Z., et al., (2021). Rfam 14: expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Res.* **49**, D192–D200. <https://doi.org/10.1093/nar/gkaa1047>.
37. Ghaemi, Z., Guzman, I., Gnutt, D., Luthey-Schulten, Z., Gruebele, M., (2017). Role of electrostatics in protein–RNA binding: the global vs the local energy landscape. *J. Phys. Chem. B* **121** (36), 8437–8446. <https://doi.org/10.1021/acs.jpcc.7b04318>.
38. Masoumzadeh, E., Grozdanov, P.N., Jetly, A., MacDonald, C.C., Latham, M.P., (2022). Electrostatic Interactions between CSTF2 and pre-mRNA drive cleavage and polyadenylation. *Biophys. J.* **121** (4), 607–619. <https://doi.org/10.1016/j.bpj.2022.01.005>.
39. Rohs, R., West, S.M., Sosinsky, A., Liu, P., Mann, R.S., Honig, B., (2009). The role of DNA shape in protein–DNA recognition. *Nature* **461**, 1248–1253. <https://doi.org/10.1038/nature08473>.
40. 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>.
41. Stefl, R., Oberstrass, F.C., Hood, J.L., Jourdan, M., Zimmermann, M., Skrisovska, L., Maris, C., Peng, L., Hofr, C., Emeson, R.B., et al., (2010). The solution structure of the ADAR2 dsRBM–RNA complex reveals a sequence-specific readout of the minor groove. *Cell* **143**, 225–237. <https://doi.org/10.1016/j.cell.2010.09.026>.
42. Decherchi, S., Rocchia, W., (2013). A general and robust ray-casting-based algorithm for triangulating surfaces at the nanoscale. *PLoS One* **8**, e59744. <https://doi.org/10.1371/journal.pone.0059744>.
43. Jurrus, E., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L.E., Brookes, D.H., Wilson, L., Chen, J., Liles, K., et al., (2018). Improvements to the APBS biomolecular solvation software suite. *Prot. Sci.* **27**, 112–128. <https://doi.org/10.1002/pro.3280>.
44. Honig, B., Nicholls, A., (1995). Classical electrostatics in biology and chemistry. *Science* **268**, 1144–1149. <https://doi.org/10.1126/science.7761829>.
45. Jayaram, B., Jain, T., (2004). The role of water in protein–DNA recognition. *Annu. Rev. Biophys. Biomol. Struct.* **33**, 343–361. <https://doi.org/10.1146/annurev.biophys.33.110502.140414>.
46. Kretsch, R.C., Li, S., Pintilie, G., Palo, M.Z., Case, D.A., Das, R., Zhang, K., Chiu, W., (2025). Complex water networks visualized through 2.2–2.3 angstrom cryogenic electron microscopy of RNA. *bioRxiv*. <https://doi.org/10.1101/2025.01.23.634578>.
47. Shazman, S., Mandel, Y.-G., (2008). Classifying RNA-binding proteins based on electrostatic properties. *PLoS Comput. Biol.* **4**, (8)e1000146 <https://doi.org/10.1371/journal.pcbi.1000146>.
48. Harris, R.C., Mackoy, T., Dantas Machado, A.C., Xu, D., Rohs, R., Fenley, M.O., (2012). Opposites attract: shape and electrostatic complementarity in protein–DNA complexes. In: Schlick, T. (Ed.), *Innovations in biomolecular modeling and simulations*. The Royal Society of Chemistry, pp. 53–80. <https://doi.org/10.1039/9781849735056-00053>.
49. Nishimasu, H., Ran, F.A., Hsu, P.D., Konermann, S., Shehata, S.I., Dohmae, N., Ishitani, R., Zhang, F., Nureki, O., (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935–949. <https://doi.org/10.1016/j.cell.2014.02.001>.
50. Hiraizumi, M., Perry, N.T., Durrant, M.G., Soma, T., Nagahata, N., Okazaki, S., Athukoralage, J.S., Isayama, Y., Pai, J.J., Pawluk, A., et al., (2024). Structural mechanism of bridge RNA-guided recombination. *Nature* **630**, 994–1002. <https://doi.org/10.1038/s41586-024-07570-2>.
51. Durrant, M.G., Perry, N.T., Pai, J.J., Jangid, A.R., Athukoralage, J.S., Hiraizumi, M., McSpedon, J.P., Pawluk, A., Nishimasu, H., Konermann, S., et al., (2024). Bridge RNAs direct programmable recombination of target and donor DNA. *Nature* **630**, 984–993. <https://doi.org/10.1038/s41586-024-07552-4>.
52. Jiang, Y., Chiu, T.P., Mitra, R., Rohs, R., (2023). Probing the role of the protonation state of a minor groove-linker histidine in Exd–Hox–DNA binding. *Biophys. J.* **123**, 248–259. <https://doi.org/10.1016/j.bpj.2023.12.013>.
53. Baek, M., McHugh, R., Anishchenko, I., Jiang, H., Baker, D., DiMaio, F., (2024). Accurate prediction of protein–nucleic acid complexes using RoseTTAFoldNA. *Nature Methods* **21**, 117–121. <https://doi.org/10.1038/s41592-023-02086-5>.

54. Krishna, R., Wang, J., Ahern, W., Sturmfels, P., Venkatesh, P., Kalvet, I., Lee, G.R., Morey-Burrows, F.S., Anishchenko, I., Humphreys, I.R., et al., (2024). Generalized biomolecular modeling and design with RoseTTAFold All-Atom. *Science* **384**, eadl2528. <https://doi.org/10.1126/science.adl2528>.
55. Fukai, S., Nureki, O., Sekine, S.I., Shimada, A., Vassylyev, D.G., Yokoyama, S., (2003). Mechanism of molecular interactions for tRNA^{Val} recognition by valyl-tRNA synthetase. *RNA* **9**, 100–111. <https://doi.org/10.1261/rna.2760703>.
56. Lu, D., Searles, A., Klug, A., (2003). Crystal structure of a zinc-finger–RNA complex reveals two modes of molecular recognition. *Nature* **426**, 96–100. <https://doi.org/10.1038/nature02088>.
57. Leontis, N.B., Westhof, E., (2001). Geometric nomenclature and classification of RNA base pairs. *RNA* **7**, 499–512. <https://doi.org/10.1017/s1355838201002515>.