SUPPLEMENTARY MATERIAL

RNAproDB: a webserver and interactive database for analyzing protein–RNA interactions

Raktim Mitra[1,†,‡], Ari S. Cohen[1,†], Wei Yu Tang[1,†], Hirad Hosseini[1], Yongchan Hong[1], 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]Division of Medical Oncology, Department of Medicine, University of Southern California, Los Angeles, CA 90033, USA

*To whom correspondence should be addressed. Tel: +1 213 740 0552; Email: rohs@usc.edu

[†]These authors should be regarded as Joint First Authors.

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

Supplementary Information

Comparison with TBI-Forna and other visualization tools

TBI-Forna[1] is an RNA secondary structure visualization tool based on a force-directed graph layout. While Forna supports visualization of RNA structures from the PDB, it has several limitations that hinder structural analysis. RNAproDB addresses these limitations in the following ways:

- **3D Layout Options**: RNAproDB's 'Interface explorer' provides three different layout options (partial projection, RNAscape [2], and ViennaRNA[3]), allowing users to choose the representation that best reflects the structural details. In contrast, Forna's layout does not always reflect the overall shape of the structure (e.g., 7ORN, Fig. S5).

- **Protein Residues**: RNAproDB's 'Interface explorer' displays individual protein residues and the nucleotides it interacts with, whereas Forna represents all residues as a single node, potentially obscuring crucial protein–RNA interactions. RNAproDB also allows layout customization for protein residues, such as setting distance thresholds and applying force-directed layout.

- **Chemical Interactions**: RNAproDB provides detailed information and annotations for node interactions: base pairs, mismatches, water-mediated hydrogen bonds, and protein–RNA bonds. This detailed annotation provides a deeper understanding of interactions, which Forna lacks.

- **3D Structure Visualization**: RNAproDB provides a 3D view of RNA and protein structures, displaying the geometry of nucleotides, protein residues, and water molecules. In addition, the '3D viewer' interacts with other tools like the 'Interface explorer' and 'Sequence viewer'.

- **Additional Tools**: RNAproDB offers many other tools such as electrostatics 3D surface visualization, 'Secondary structure selector', and tabular data to help with analyzing structure.

ChimeraX[4] and PyMOL[5] are widely used visualization tools for analyzing molecular structures; each uses their own proprietary engineering tools to achieve relevant functionalities. We do not view RNAproDB as a competitor to these tools; rather, it is complementary to these tools and can help further analyze structural data. Nevertheless, RNAproDB offers several advantages:

- **Unique Features**: RNAproDB provides 'Interface explorer' and 'Secondary structure selector' that integrate directly with the '3D viewer'. These capabilities are not found in other tools.

- **Accessibility**: Because RNAproDB is entirely browser-based, it requires no setup or installation, making it more accessible than the other tools.

Steric clash detection

Steric clashes occur when two non-bonded atoms are positioned unrealistically close to each other. To identify potential steric clashes, we follow the protocols established by tools such as ChimeraX[4] and MolProbity[6], which use adjacent atoms' van der Waals (vdW) radii. The 'overlap' between two atoms is defined as follows:

$$overlap = vdW\ radii\ of\ atom\ 1 + vdW\ radii\ of\ atom 2 - distance - hbond\ allowance$$

where an H-bond allowance of 0.4 Å is subtracted if the atoms form a donor–acceptor pair. Any pair with an overlap greater than 0.6 Å is flagged as a potential steric clash. These correspond to the default values used in ChimeraX[4]. Pairs of atoms within the same residue or directly connected along the backbone (i.e., in sequential residues) are excluded, as they are typically covalently bound.

Supplementary References

1. Kerpedjiev,P., Hammer,S., Hofacker,I.L. (2015) Forna (force-directed RNA): Simple and effective online RNA secondary structure diagrams. Bioinformatics 31, 3377–3379.

2. Mitra,R., Cohen,A.S., Rohs,R. (2024) RNAscape: geometric mapping and customizable visualization of RNA structure. Nucleic Acids Res. 52, W354–W361.

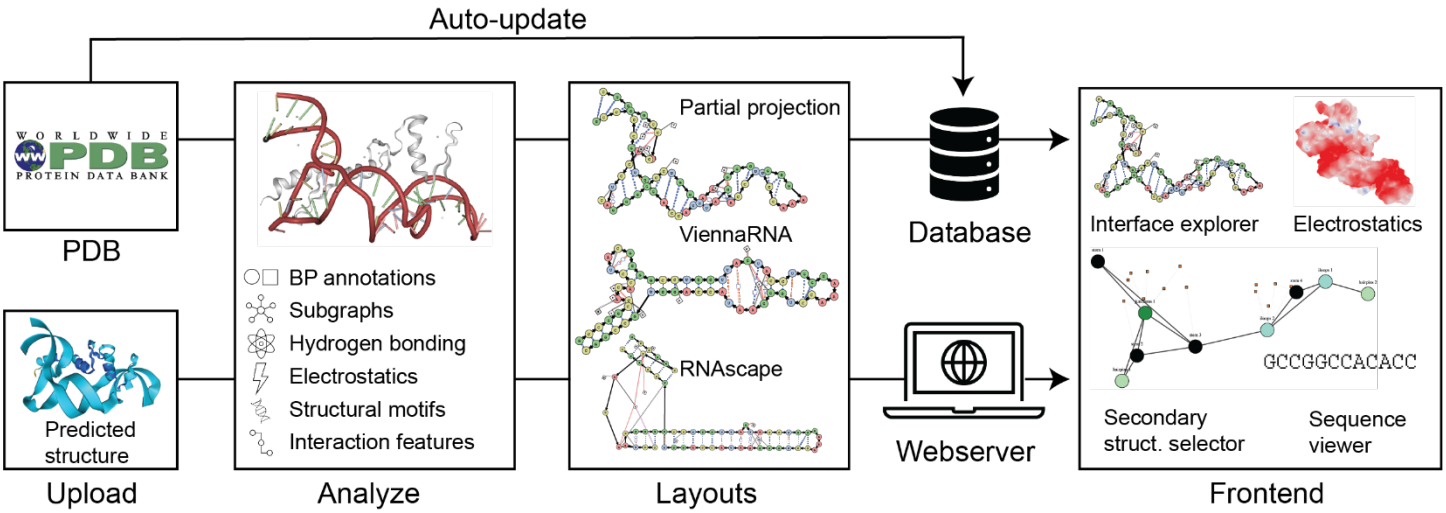
3. Lorenz,R., Bernhart,S.H., Höner zu Siederdissen,C., Tafer,H., Flamm,C., Stadler,P.F., Hofacker,I.L. (2011) ViennaRNA Package 2.0. Algorithms Mol. Biol. 6, 26.

4. Meng,E.C., Goddard,T.D., Pettersen,E.F., Couch,G.S., Pearson,Z.J., Morris,J.H., Ferrin,T.E. (2023). UCSF ChimeraX: Tools for structure building and analysis. Prot. Sci., 32, e4792.

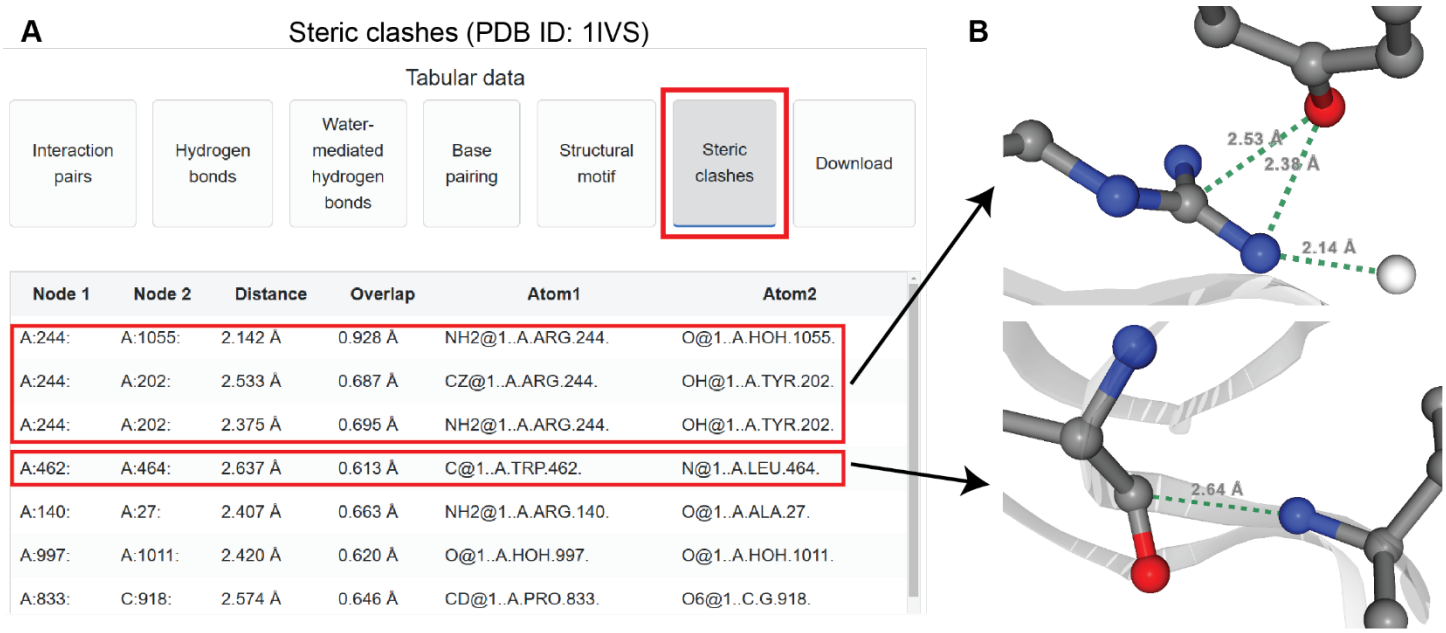
5. Schrodinger,L.L.C. (2015). The PyMOL molecular graphics system. Version, 1, 8.

6. Chen,V.B., Arendall,W.B., Headd,J.J., Keedy,D.A., Immormino,R.M., Kapral,G.J., Murray,L.W., Richardson,J.S., Richardson,D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D Biol. Crystallogr., 66, 12–21.

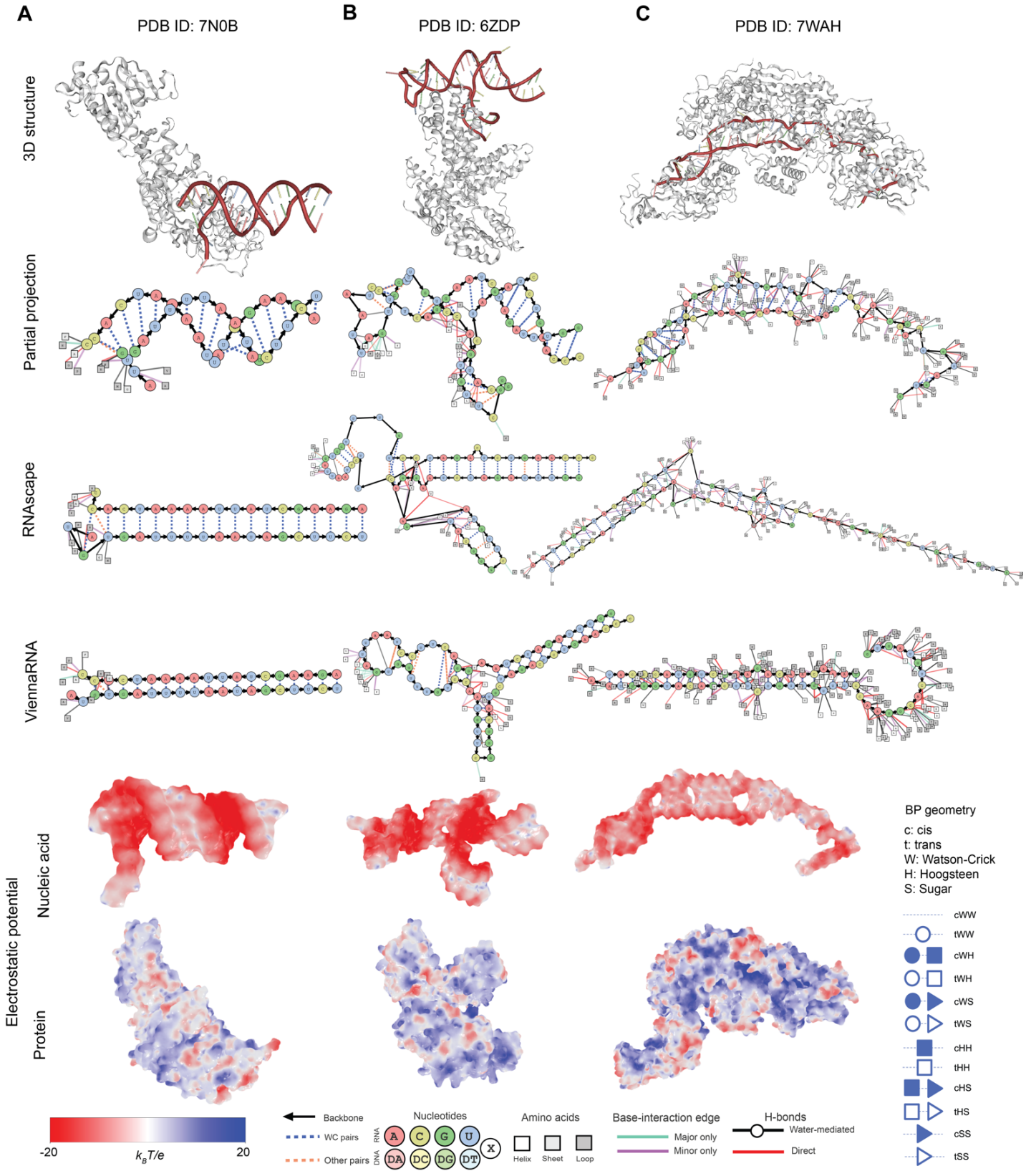
Supplementary Figures



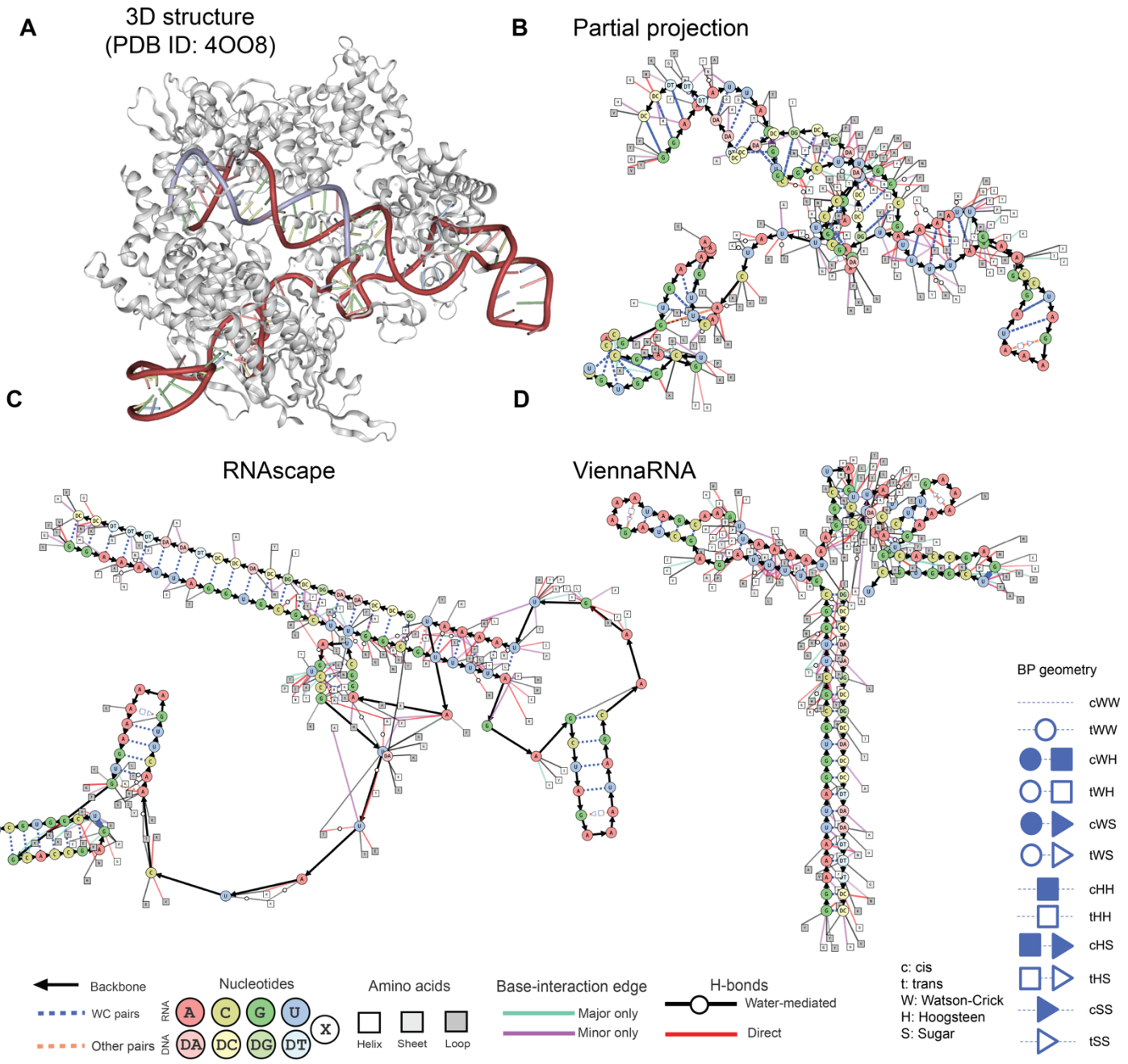
Supplementary Figure S1. Overview of RNAproDB processing pipeline. Structure files are sourced from either the PDB or user uploads. These structures are processed and analyzed for salient features. Several nucleic acid layouts are computed including a partial projection, a secondary structure, and a tertiary structure-aware RNAscape mapping. Analysis output from PDB structures is deposited within the RNAproDB database, while output from user structures is assigned a unique link for access. The preceding process is run weekly for newly released, applicable structures in the PDB to auto-update the RNAproDB database. Lastly, the RNAproDB frontend allows users to interact with visualizations and exploratory tools and enables searching for desired structures.



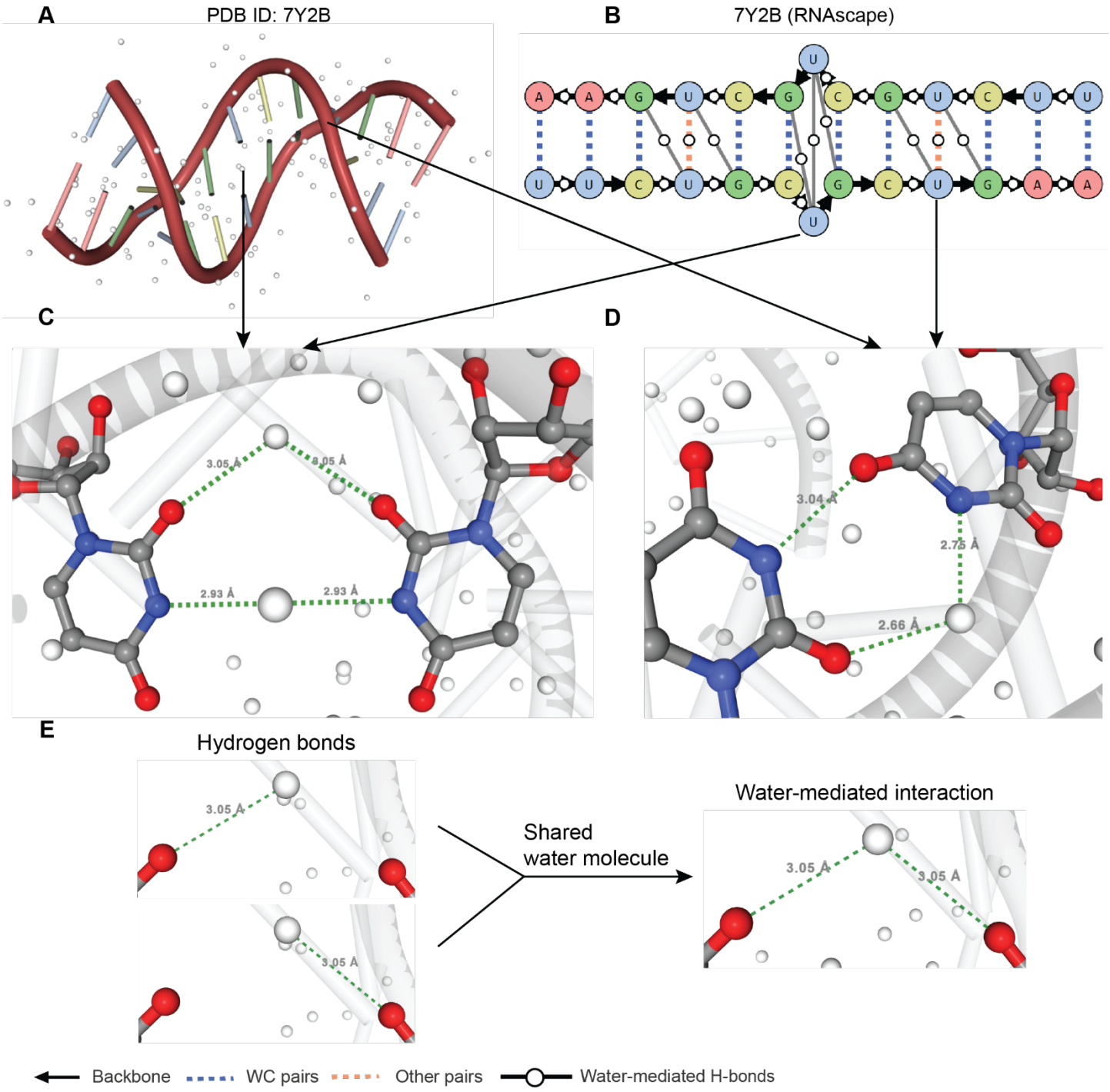
Supplementary Figure S2. Steric clash detection for a structure. A. Table of detected steric clashes in a t-RNA molecule (PDB ID 1IVS). Note that the parameters used[4] (i.e., overlap threshold of 0.6 Å and H-bond allowance of 0.4 Å) influence which atom pairs are classified as clashing. **B.** Example of steric clashes in '3D viewer', which are annotated by green dotted lines along with the distances.



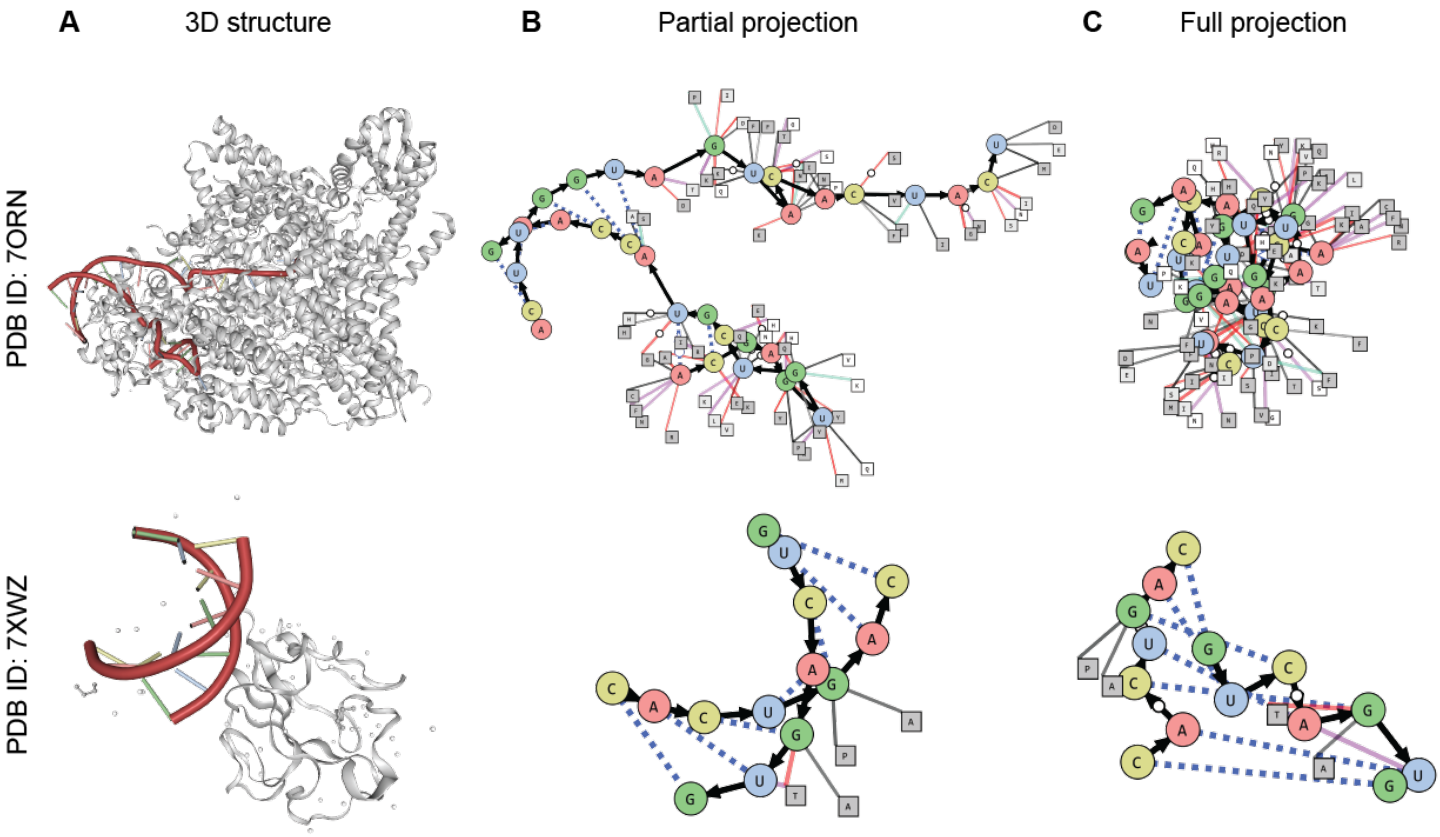
Supplementary Figure S3. RNAproDB layouts and surface electrostatics for selected protein–RNA structures. A. Structure of SARS-CoV-2 nsp10-nsp14 (WT)-RNA (PDB ID 7N0B) and **B.** Telomerase reverse transcriptase from the fungi *Candida tropicalis* (PDB ID 6ZDP). **C.** Type III-E CRISPR-Cas7-11 effector complex (PDB ID 7WAH).



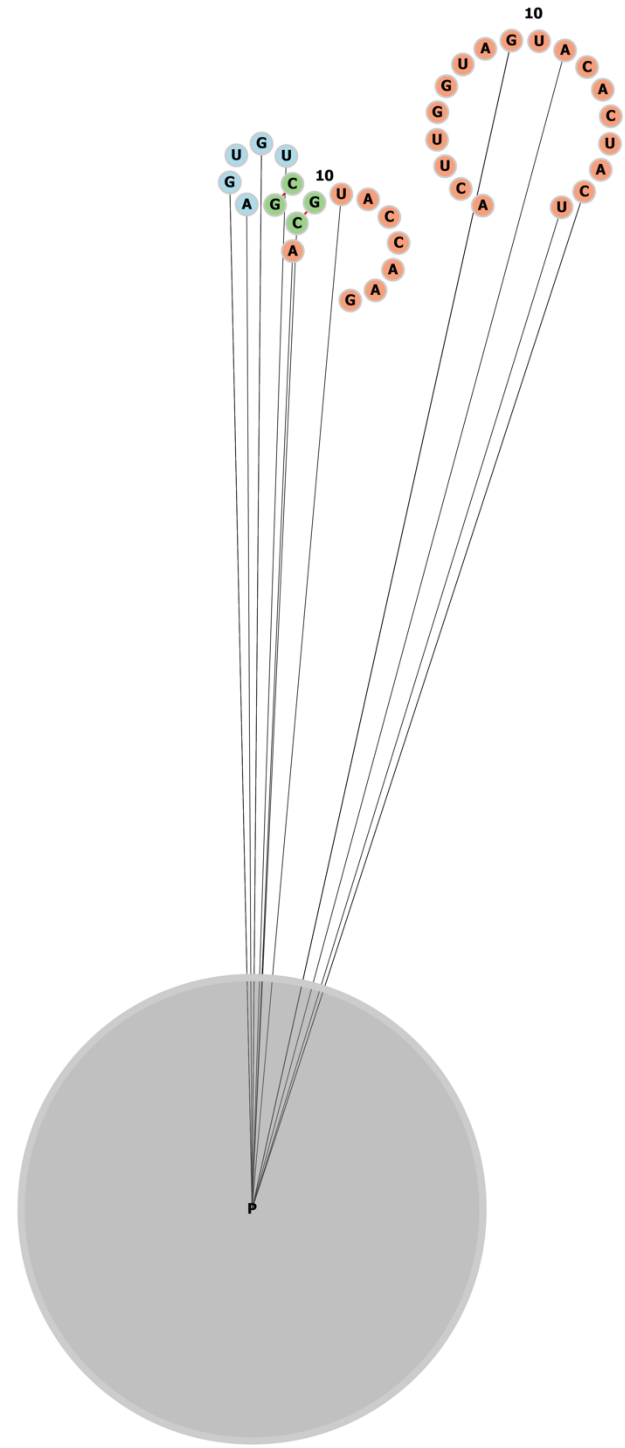
Supplementary Figure S4. 3D structure view juxtaposed with various structural mappings available in RNAproDB for a Cas9–DNA/RNA complex. **A.** 3D structure view in RNAproDB (PDB ID 4OO8). **B.** Partial projection mapping in RNAproDB which corresponds best with the 3D structure. **C.** RNAscape mapping in RNAproDB which displays helices as simplified ladders. **D.** ViennaRNA secondary structure mapping which places nucleotides without considering tertiary structure.



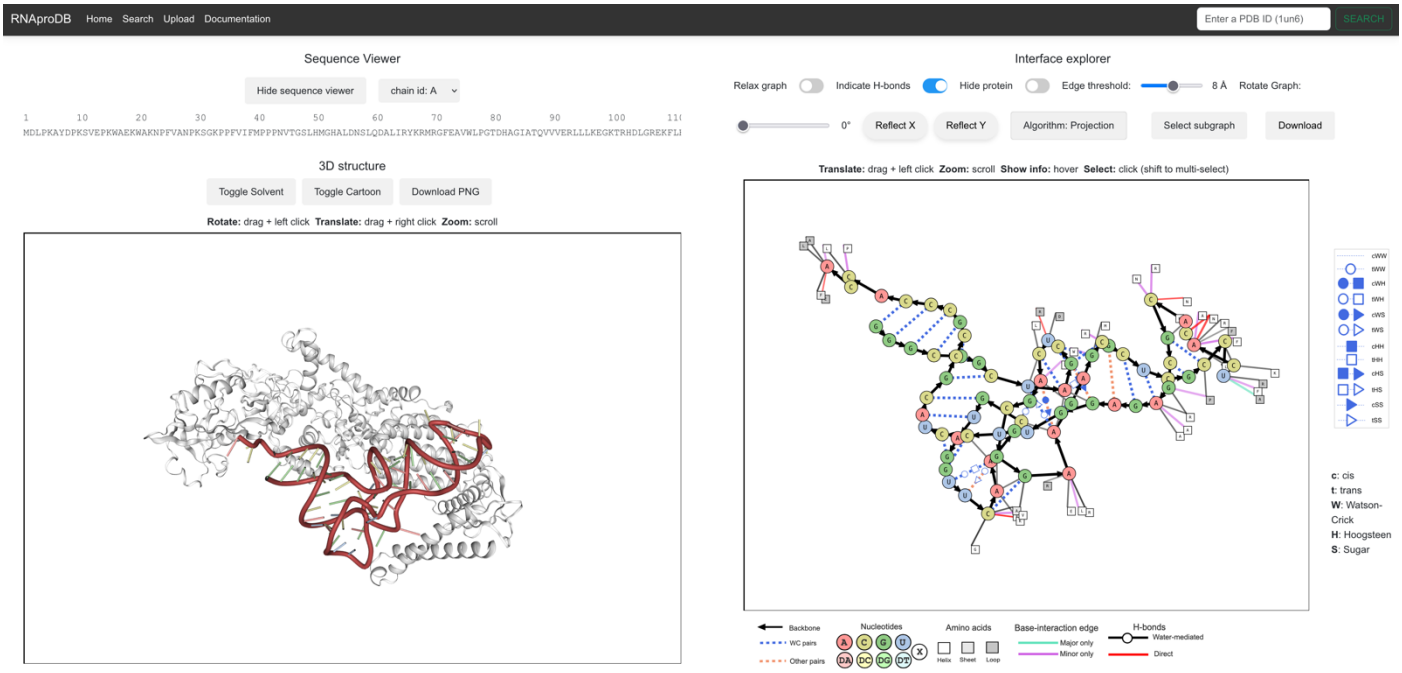
Supplementary Figure S5. Example illustration of RNA–RNA water-mediated hydrogen bonds facilitating non-Watson-Crick interactions. **A.** 3D structure of PDB ID 7Y2B. **B.** RNAproDB 'Interface explorer' layout (RNAscape algorithm) for PDB ID 7Y2B. The central U/U pair is considered un-paired by DSSR, but computing RNA–RNA water-mediated hydrogen bonds reveal interactions between them. **C.** Zoomed-in view of central U/U interactions with two water molecules facilitating water-mediated hydrogen bonds. **D.** Zoomed-in view of an off-center U/U interaction with one water molecule facilitating a water-mediated hydrogen bond in addition to a direct hydrogen bond. **E.** Identifying water-mediated interactions: if a single water molecule forms hydrogen bonds with both atom A and atom B, we consider atoms A and B to be connected through a water-mediated interaction.



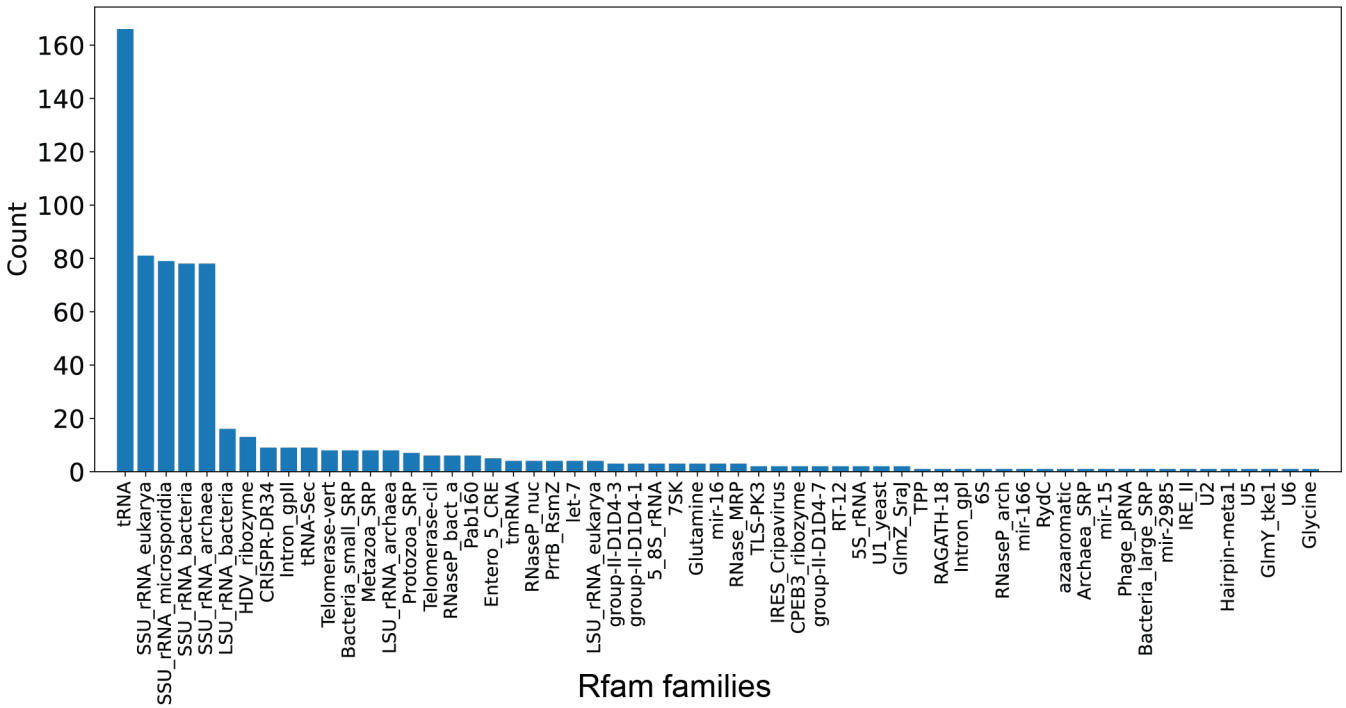
Supplementary Figure S6. Comparison of partial and full projections. A. 3D structure. **B.** Partial projection, generated by maximizing spatial variance of the nucleotide positions only. **C.** Full projection, generated by maximizing spatial variance across the entire structure (i.e., including protein residues).



Supplementary Figure S7. TBI-Forna visualization of PDB ID 7ORN. Protein residues are represented as a single node, and protein–RNA interactions are displayed as uniform edges with this node.



Supplementary Figure S8. RNAproDB report page for AlphaFold 3 predicted structure. The 'Sequence viewer' (seen at the top left) lists chains and corresponding sequences; upon clicking on a residue, the 'Interface explorer' and '3D structure viewer' will automatically highlight or orient around that residue. The 'Interface explorer' (seen at the top left) displays a highly interactive mapping of the structure and its protein interactions. Within the 'Interface explorer', a user can click 'Select subgraph' to generate a subgraph of the 'Interface explorer' based on first-order neighbors of selected residues.



Supplementary Figure S9. Counts of different Rfam families represented in the RNAproDB protein–RNA collection as of October 15, 2024. For each protein–RNA entry, we extracted the RNA sequence and matched it with families in the Rfam (RNA families) database. This matching process was conducted using the Rfam sequence search tool, where every query is compared with the Rfam library of RNA families. Entries for which the query produced a result were compiled to compute the distribution of RNA families.