

Developing Community Resources for Nucleic Acid Structures

Helen M. Berman¹, Catherine L. Lawson² and Bohdan Schneider^{3,*}

¹ Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA; berman@rcsb.rutgers.edu

² Institute for Quantitative Biomedicine, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA; cathy.lawson@rutgers.edu

³ Institute of Biotechnology of the Czech Academy of Sciences, 252 50 Vestec, Czech Republic

* Correspondence: bohdan.schneider@ibt.cas.cz

Abstract: In this review, we describe the creation of the Nucleic Acid Database (NDB) at Rutgers University and how it became a testbed for the current infrastructure of the RCSB Protein Data Bank. We describe some of the special features of the NDB and how it has been used to enable research. Plans for the next phase as the Nucleic Acid Knowledgebase (NAKB) are summarized.

Keywords: nucleic acid structures; nucleic acid conformation; biological structure database; DNA; RNA; validation standards

1. Introduction

The first single crystal structures of nucleic acids were determined in the 1970s, almost twenty years after the model of the DNA double helix based on fiber data was published [1,2]. Short fragments of RNA yielded the first atomic-level views of the double helix and demonstrated conformational flexibility [3–5]. These structures were archived as small molecules in the Cambridge Crystallographic Database (CSD) [6]. The structure of tRNA, determined in 1974 [7–9], showed that RNA can fold into a compact structure and demonstrated the importance of tertiary interactions. As DNA synthesis became possible, structures of the DNA double helix with predefined sequences were determined. The first structures were left-handed Z-form DNA fragments [10], and in 1981, the first single crystal structure of a full turn of B-form DNA was published [11]. The tRNA structures and larger nucleic acid fragments were archived in the Protein Data Bank (PDB [12]). By 1990, there were nearly 100 publicly released nucleic acid structures, thus allowing analyses of sequence-dependent features, hydration patterns, and ligand interactions.

During the late 1970s and 1980s, several faculty members in the Chemistry Department at Rutgers University focused their research on nucleic acids. Ken Breslauer worked on the macroscopic properties of nucleic acids using calorimetric approaches [13–16]; these works, seminal for the understanding of thermodynamics of DNA, have continued to this day [17–20]. Roger Jones developed new methods to synthesize DNA [21]. Jerry Manning developed the counterion condensation theory to understand DNA folding [22], and continued this work in collaboration with the Breslauer group [23]. Wilma Olson performed detailed analyses of the structure of DNA [24]. During that period, Helen Berman carried out nucleic acid crystallography research at the Institute for Cancer Research in Philadelphia and had close interactions with the Rutgers group. In 1989, she joined the Chemistry faculty at Rutgers.

The setting at Rutgers was ideal for collaborative studies using both experimental and computational approaches to investigate nucleic acid structure. It was necessary to have a resource that contained the structural information which resided in the CSD, in the PDB, or in the laboratories of individual researchers to facilitate these efforts. In collaboration with David Beveridge, with whom Berman was collaborating on computational analyses of nucleic acid hydration, Olson and Berman proposed to create the Nucleic Acid Database



Citation: Berman, H.M.; Lawson, C.L.; Schneider, B. Developing Community Resources for Nucleic Acid Structures. *Life* **2022**, *12*, 540. <https://doi.org/10.3390/life12040540>

Academic Editors: Tigran Chalikian, Jens Völker and Bruce J. Nicholson

Received: 14 March 2022

Accepted: 31 March 2022

Published: 6 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

(NDB). In the early 1990s, funding was received from the National Science Foundation to establish “A Comprehensive Database of the Three-Dimensional Structures of Nucleic Acids”. The goal was to create a searchable database that would integrate information from several sources and make a variety of reports, thus enabling research on nucleic acid structure.

2. Development of the Nucleic Acid Database

The first step in the development of the NDB was to collect and curate the structural data [25]. Coordinates were accessed from the CSD and the PDB. Each structure and experiment were carefully reviewed to create appropriate annotations beyond what was available from each resource. Rather than working directly with the flat files maintained by the PDB, the NDB imported the parsed data files into a relational database management system (DBMS). Sybase [26] was chosen as the DBMS in large part because it was being used by Genbank [27,28]. A query system called NDBquery was put into place. In the early years, distribution was accomplished via FTP and a system called Gopher [29]. By 1995, a web server was set up, which generated a modest amount of activity to access and analyze the 350 structures represented in the NDB. The NDB was actively involved in the development of mmCIF, whose data model is compatible with a relational DBMS. By 1996, mmCIF [30] became the master format for the NDB. The software that was developed and the experience gained using this data representation set the stage for the management of the Protein Data Bank using mmCIF as the master format by the Research Collaboratory for Structural Bioinformatics (RCSB) beginning in 1998.

The NDB also became a driver for the creation of geometrical standards for nucleic acid structures. Careful analysis of high-resolution structures from CSD permitted the calculation of standard reference bond distances and angles for the bases, sugars, and phosphates of nucleic acids [31,32]. Using these values, Parkinson et al. [33] created new parameters that enabled improved refinement of nucleic acid-containing crystal structures against their experimental data. Those standards were widely used. In 1998, the NDB helped organize a conference whose outcome was the standard coordinate frame definition for nucleic acid bases [34]. This standard became widely adopted by researchers studying nucleic acid base morphology.

3. Features of the NDB

In addition to facilitating access to primary data for nucleic acid structures, the NDB provides tables of derived features, such as classifications of base pairing topologies [35], backbone torsion angles, and conformational and base pair classifications [36,37].

The NDB also offers different types of data visualization and presentation. The most important is the NDB Atlas page (Figure 1), which gives summary information about the structure, visualizations of the crystal asymmetric unit, the biological unit, unit cells, and for RNA structures; it provides a view that combines the secondary and tertiary structural features. Links to other resources are also provided.

The functionality of the NDB and its query engine was first and foremost driven by research projects on the nucleic acid structural and computational biologists. Careful attention was given to the quality and uniformity of the metadata so that it would be possible to use Boolean logic to create queries; individual questions could be made into logical constructs joined by logical AND, OR, and NOT. This requirement represented a challenge for building a robust system of precisely defined terms incorporated into a formal computer-readable language; mmCIF was that dictionary.

The NDB website was designed so that the user could select structures with features of interest and then use those structures for further analysis, e.g., through the creation of detailed tabular or graphical reports. Soon after the first functional version of the NDB was available, we started to use its potential to study the geometrical features of nucleic acids. The original NDB reporting capability allowed the user to obtain tabular reports of various properties of the selected nucleic acid structures from basic information about the

publication or refinement parameters and graphical reports of selected geometric features such as bond distances (Figure 2) or torsion angles (Figure 3). Once funding for the NDB became limited in the 2000s, it was not possible to maintain these reporting capabilities.

NDB ID: 7EQJ **PDB ID: 7EQJ**

Title:
CRYSTAL STRUCTURE OF E. COLI VALINE tRNA

Molecular Description:

Deposited:
2021-05-03

Released:
2022-01-19

Nucleic Acid Sequence:
[Click to show/hide 1 nucleic acid sequences](#)

Protein Sequence:
No Protein Sequence Found

Primary Citation:
Jeong, H., Kim, J.
Unique anticodon loop conformation with the flipped-out wobble nucleotide in the crystal structure of unbound tRNA Val. *Rna*, **27**, pp. 1330 - 1338, 2021.

Experimental Information:
X-RAY DIFFRACTION

Space Group:
P 1 2 1 1

Cell Constants:
a = 53.541 b = 33.059 c = 131.797 (Ångstroms)
α = 90.0 β = 98.15 γ = 90.0 (degrees)

Refinement:
The structure was refined using the PHENIX program. The R value is 0.2061 for 28771 reflections in the resolution range 29.488 to 2.043 Ångstroms with Fobs > 1.35 sigma(Fobs) and with I > 0.0 sigma(I)

Download Data:
[Asymmetric Unit coordinates \(pdb format, Unix compressed\(.gz\)\)](#)
[Asymmetric Unit coordinates \(cif format, Unix compressed\(.gz\)\)](#)
[Biological Assembly coordinates \(pdb format\) 1 2](#)
[Structure Factors \(cif format\)](#)
[XML | Complete with coordinates \(xml format, GNU compressed\(.gz\)\)](#)
[XML | Coordinates only \(xml format, GNU compressed\(.gz\)\)](#)
[XML | Header only \(xml format, GNU compressed\(.gz\)\)](#)

Structural Features

[RNA Base-Pairs, Stacking, etc.](#)

[Similar Structures](#)

[Interactive basepair map with 3D fragment visualization for: 7EQJ](#)

[RNAML](#)

[Base Pair Hydrogen Bonding Classification](#)

[Nucleic Acid Backbone Torsions](#)


[Base Pair Morphology Parameters](#)

[Base Pair Morphology Step Parameters](#)

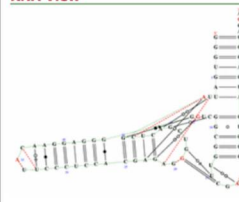
[Conformer Analysis \(DNATCO\)](#)

[RNA Pseudotorsions Analysis](#)

Biological Assembly 1



RNA View



[More Images...](#)

Figure 1. NDB Atlas page of a tRNA structure.

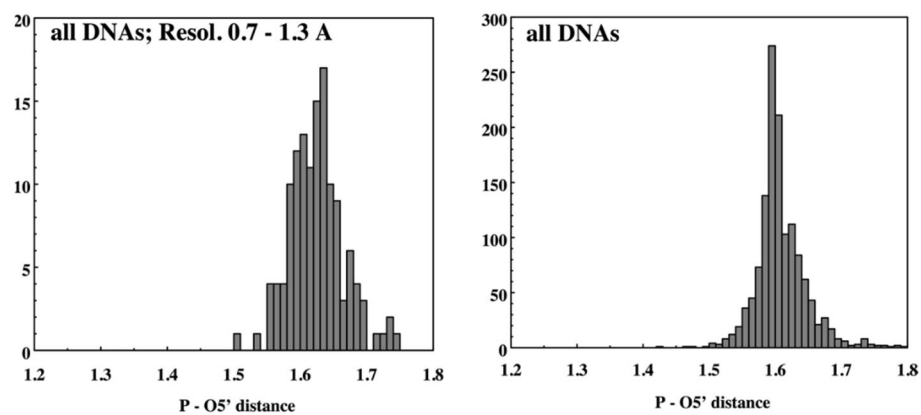


Figure 2. Example NDB report of geometric features of nucleic acids, based on structures available in the 1990s. Histograms show the P–O5' valence distances in high-resolution DNA, and in all DNA structures.

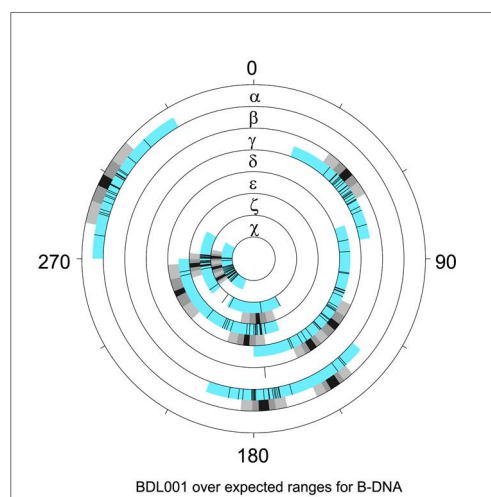


Figure 3. NDB graphical report of torsion angle distribution [38] for the Drew–Dickerson dodecamer, PDB ID 1BNA, NDB ID BDL001 [11]. Blue sectors indicate torsion angle limits for all structures annotated as B–DNA. Overlaid black tick marks are measured torsion values for BDL001. Adjacent black/grey sectors denote average values and spreads of 1 and 2 estimated standard deviations. Note that two averages are indicated for several torsions, e.g., for δ (two distinct sugar puckers) and ϵ (BI versus BII forms). Values reflect NDB data available in 1996.

4. Research Enabled by the NDB

The NDB has been used by many researchers to analyze the structures of nucleic acids. There are over 1100 citations to the original NDB article. The type of research enabled by the NDB includes DNA conformational analyses [39], DNA structure prediction [40], RNA structure prediction [41], analyses of protein–nucleic acid interactions [42,43], and the creation of new specialty databases [44]. In our research, we have used the NDB to study a variety of aspects of nucleic acids. For example, we surveyed A, B, and Z-form double helical DNA structures and used Fourier averaging to determine hydration patterns, e.g., for DNA nitrogenous bases [45]. Both base and later phosphate studies showed sequence and conformation-dependent water position preferences (Figure 4).

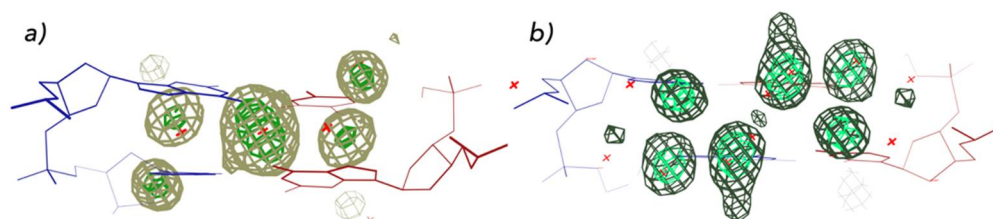


Figure 4. Sequence dependence of DNA hydration. Two distinct hydration patterns are shown in for the A-form major groove for (a) 5′–GC–3′ and (b) 5′–CG–3′, based on analyses of structures available in the mid–1990s NDB [44]. A more recent analysis of hydration using larger and functionally more relevant dinucleotide fragments is available at watlas.datmos.org/watna (accessed on 30 March 2022).

The growing volume of available crystal structures with ever growing sequence variability also led us to ask whether conformational properties of various DNA and RNA forms could be better characterized. This task posed new challenges to NDB querying and reporting capabilities. Specific subsets of structures were selected based on sequence, function, or structural features using SQL queries; their properties were reported as text or graphs (Figure 5). Ultimately, we were able to sharpen conformational definitions for established subtypes of A–B–Z forms (Figure 6) [46].

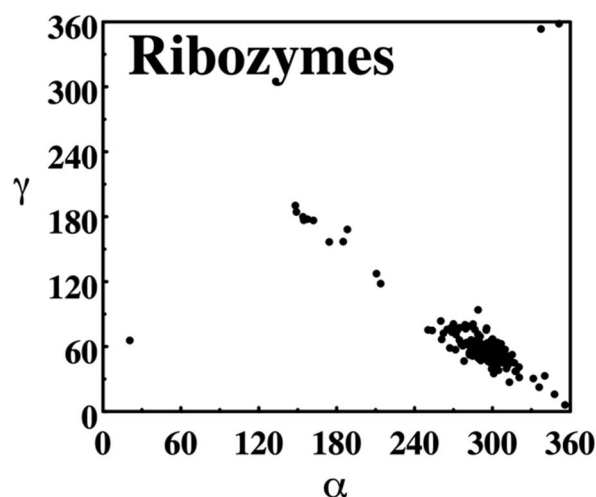


Figure 5. NDB torsion angle scatter plot. The distribution of backbone torsion angles α and γ observed in crystal structures of RNA annotated in the NDB as ribozymes in 1996 is shown. α describes rotations around the P–O5' phosphodiester bond, γ around C5'–C4' bond. Plots were created directly on the NDB website as PostScript formatted reports.

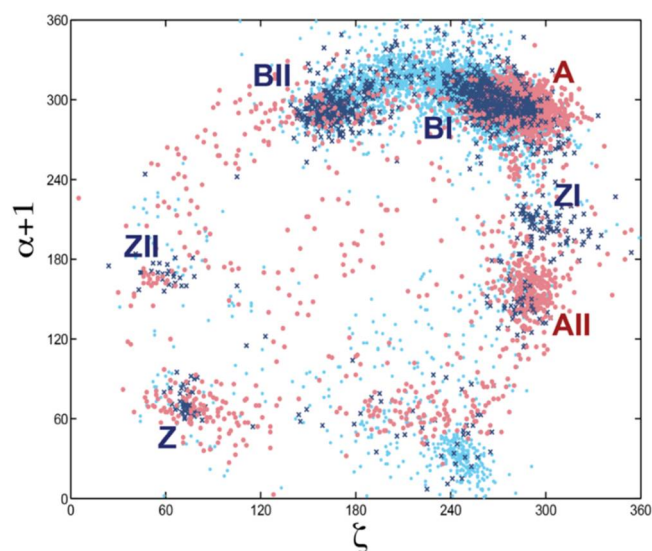


Figure 6. Scatter plot of backbone torsion angles ζ and $\alpha + 1$. ζ shows the variation in rotation around the O3'–P phosphodiester bond and $\alpha + 1$ around the P–O5' bond (labeled $\alpha + 1$ because this bond belongs to the sequentially following nucleotide). Data for the DNA alone is shown as dark blue crosses, for protein–DNA complexes: light blue dots, and for RNA: red dots. The scattergram shows data for all nucleotide residues in the 1998 NDB. Clusters of some major conformational types are labeled. This analysis revealed that no nucleic acid form can be unequivocally classified by torsion angle pairs; a more sophisticated multidimensional analysis was needed.

The growing number of nucleic acid structures and the appearance of new forms such as quadruplexes and large-folded RNAs demonstrated the plasticity of nucleic acid molecules. It became clear that the conformational space of nucleic acids is extremely complex and that capturing it would require a concerted understanding of base pairing motifs and the backbone structural variability.

Early analyses showed that backbone conformational variability was fundamentally influenced by flexibility around the O3'–P–O5' phosphodiester bonds that connected adjacent nucleotide residues, described by torsion angles ζ and α [47]. Our multidimensional statistical analysis, therefore, focused on dinucleotide fragments analyzed in torsion space, taking full advantage of the availability of the NDB and PDB.

In the 2000s, research conducted by several groups concentrated on analysis of RNA backbone flexibility culminated in an RNA Consortium consensus set of dinucleotide conformers [48]. The effort was later complemented by an analogous set of DNA conformers [49] and, ultimately, a comprehensive classification system for dinucleotide fragments covering both DNA and RNA [50]. This classification algorithm provides an automated structural ranking of dinucleotide fragments at two levels of detail: fully geometrical classification into dinucleotide conformational classes (NtC) and a more human-accessible structural alphabet (CANA). The assignment of the CANA and NtC classes makes it possible to study the structural propensities of dinucleotide sequences. For example, analysis of DNA in transcription factors and in histone core particle complexes showed important trends of protein interactions with specific bending associated NtC classes (Figure 7) [49].

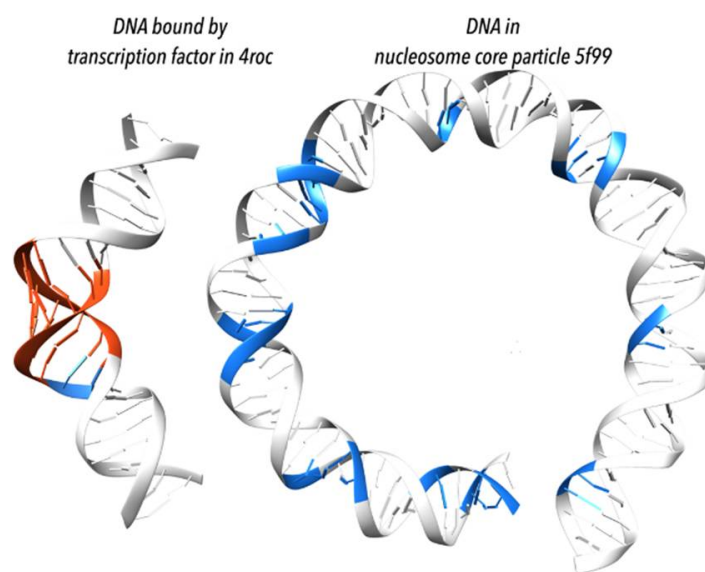


Figure 7. Transcription factors and proteins of the histone core particle bend DNA duplex differently [49]. **(Left)** bending by transcription factors is acquired mostly by local adaptation to the A form (highlighted in red); shown is DNA from complex with TFIIIB–Related Factor Brf2 (PDB id 4ROC [51]). **(Right)** bending by the histone core particle is associated with the BII form (highlighted in blue); shown are first 75 base pairs from a histone core particle (PDB id 5F99 [52]); when statistically measured over many structures, the BII form appears in histone-wrapped DNA every tenth step corresponding to one full turn of duplex; the periodicity of the BII form appearance explains the DNA bending.

NtC assignments have also inspired development of a new validation tool linking the global geometry criterion (closeness of fit to the nearest NtC class) and the quality of fit into electron density (Figure 8) [50]. It offers a simple information-rich graphical representation of the overall quality of nucleic acid structure in the form of a 2D graph.

In an additional effort to understand, classify, and validate nucleic acids, we have developed a procedure similar to Ramachandran analysis for proteins, making use of eta (η) and theta (θ) virtual torsion angles (pseudotorsions) [53,54]. Measured (η, θ) pairs define backbone conformations for each central residue within a trinucleotide. Plots are designed to quickly reveal rare conformations that may need extra checking (Figure 9). A web server was recently set up to investigate the utility of this approach for RNA structures determined using cryoEM (ptp.emdataresource.org) (accessed on 30 March 2022).

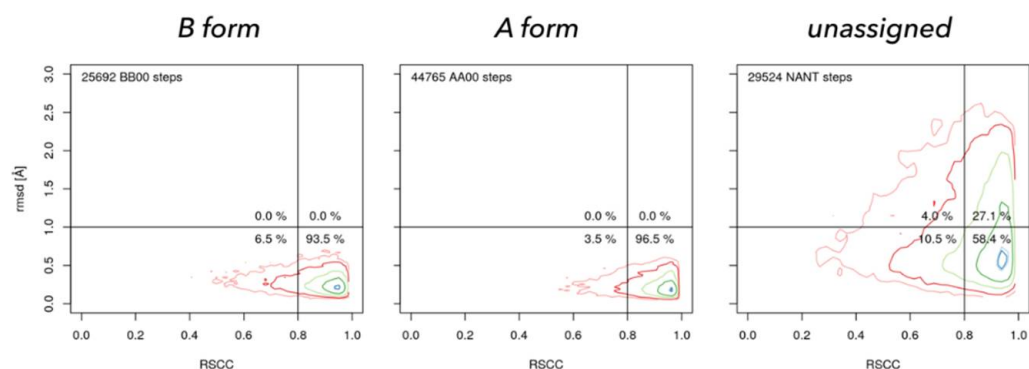


Figure 8. Geometrically well-defined dinucleotides fit well into their electron densities. Real Space Correlation Coefficient (RSCC, horizontal axis) measures how closely the model electron density resembles the experimental density and rmsd (vertical axis) measures how closely the geometry of the model resembles the closest NtC class in the so called golden set [50]. NtC class BB00 (**left**) characterizes the B form, AA00 (**center**) A form in both DNA and RNA, and NANT (**right**) are all unclassified dinucleotides. Geometrically unclassified dinucleotides fit significantly worse to the electron density.

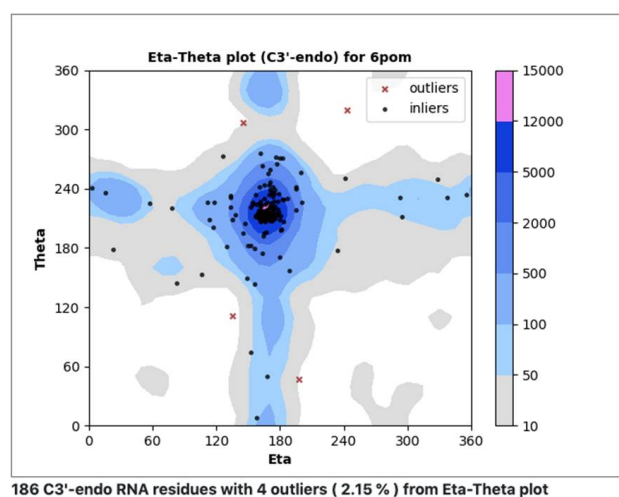


Figure 9. Pseudotorsion plot, a simple coarse-level RNA backbone conformation validation tool. Measured eta and theta (η, θ) pseudotorsion values for each trinucleotide (black dots, red x's) are plotted against a quality-filtered virtual torsion angle distribution derived from a large number of RNA structures (contours), analogous to the Ramachandran plot for proteins.

5. Current State of Nucleic Acid Structural Biology

When the NDB was established in the early 1990s, most of the nucleic acid structures were small fragments with the exception of tRNA. There were a few structures of protein-nucleic acid complexes, limited to virus capsids with viral genomic RNA or DNA and transcription factors bound to duplex DNA. Molecular machines, such as the ribosome, were yet to be determined. In contrast, there are now more than 14,000 nucleic acid-containing structures in the PDB and NDB (Figure 10). A notable trend is the recent increase in the use of electron microscopy (EM) for structure determination. Protein/DNA complexes are the most abundant, followed by protein/RNA, DNA-only, and RNA-only. In addition to the increase in the number of structures, the structures are very diverse, as shown in Figure 11.

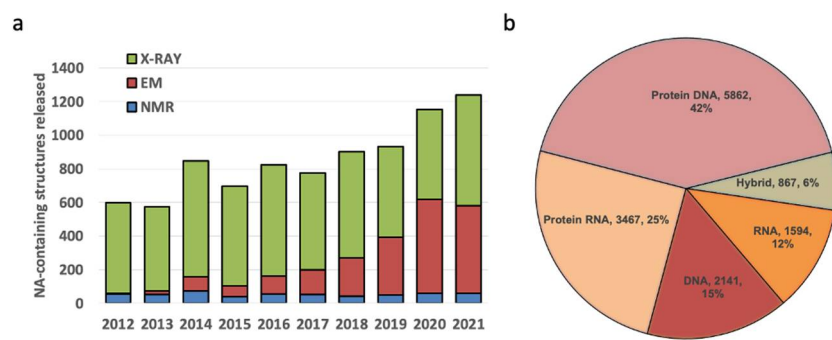


Figure 10. Current statistics for nucleic acid-containing structures. (a) New structures released into the PDB, by year and method; (b) Distribution of nucleic acid-containing structures. NDB archives and annotates structures determined using X-ray crystallography or NMR. Electron microscopy structures are not included in the NDB, but will be included in the NAKB, the planned successor to NDB.

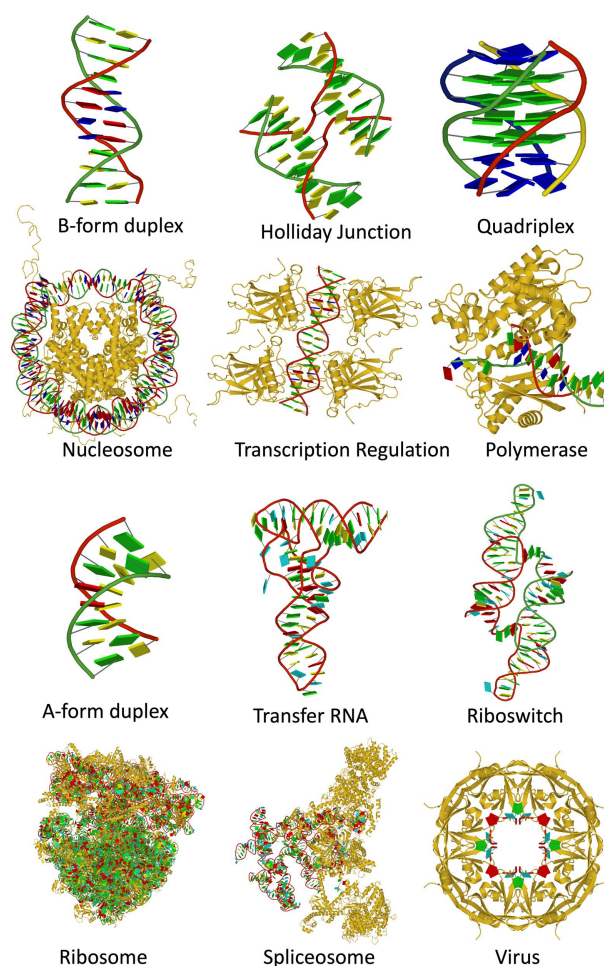


Figure 11. Diversity of structures containing DNA (top rows) and RNA (bottom rows). Nucleic acids are shown with ribbon backbones (random colors) and base blocks (A—red, C—yellow, G—green, T—blue, U—cyan). Proteins are shown as gold ribbons. From top left: B-form duplex DNA (1BNA [11]), Holliday junction (5DSB [55]), parallel-stranded DNA quadruplex (139D [56]), nucleosome core particle (1KX5 [57]), *trp* repressor/operator complex (1TRR [58]), DNA repair enzyme rev1 (6X6Z [59]), A-form duplex RNA (402D [60]), tRNA Asp (6UGG [61]), glutamine II riboswitch (6QN3 [62]), bacterial ribosome (4YBB [63]), spliceosomal E complex (6N7P [64]), Ebola virus matrix protein octamer (7K5L [65]). Images were generated using DSSR and PyMOL [66].

These structures have significantly expanded our knowledge of structure/function relationships and raised the potential of new knowledge from systematic analyses of structure collections. Many different databases and tools have been created to enable specialized analyses of nucleic acid structures. Some have focused on DNA [67], some on RNA [68–72], and some on the interactions between proteins and nucleic acids [73,74]. A systematic long-term analysis of dinucleotides led to a unified RNA + DNA automated classification system [50] available at DNATCO (dnatco.datmos.org) (accessed on 30 March 2022). The NDB (ndbserver.rutgers.edu) (accessed on 30 March 2022) is unique in that all nucleic acid structures and their complexes are contained in a single resource.

6. Going Forward

The NDB is maintained to the extent that new structures and manually curated annotations are added each week, but there is little significant development since its last full funding in 2003. Even so, thousands of users from the Americas, Asia, Europe, and other locations continue to make multiple visits to the NDB website each month. The most heavily visited pages are Advanced Search and DNA and RNA galleries.

In 2018, the collaborative group of scientists managing both the NDB (at Rutgers) and RNAhub services (at Bowling Green State University) proposed to create the Nucleic Acid Knowledge Base (NAKB), with the goal of integrating information already in the NDB with additional sequence, structure, function, and interaction-based annotations for all major classes of NA-containing 3D structures. This new service, which will ultimately replace the NDB, is currently under construction. The NAKB aims to enable users to quickly find and download all structures and metadata relevant to their search topic, whether broad or focused, based on the NDB's internal curation scheme, computationally generated annotations, and/or external database references for DNA, RNA, mixed NA, and for NA-binding enzymatic, regulatory, and structural proteins. All NA-containing structures in the PDB will be indexed, including structures obtained using Electron Microscopy. The NAKB will be updated weekly.

The NDB has employed manual expert curation collected over three decades to identify major NA secondary structure features (duplex, triplex, and quadruplex) and high-level classifications (e.g., ribosomal RNA or telomeric DNA), as well as interactions with ligands (e.g., minor groove binding) and protein classification [37]. Integrated computationally created annotations have included bond distance, angle, and torsion geometries, base and base-pair morphologies, as well as RNA 3D motifs, interactions (base pair types and parameters, base-to-backbone, and base stacking interactions), and RNA equivalence (3D structure similarity) classes [75,76].

New NAKB content will include equivalence class calculations for all nucleic acid molecule types (RNA, DNA, hybrid nucleic acids), enabling more accurate retrieval for closely related NA structures, analogous to the way that UniProt identifier mapping has improved search capabilities for related proteins in PDB [77]. Computationally derived annotations produced by DSSR software [78], including secondary structure features, sugar pucker type, and pseudo-torsion angles, will be added.

New search capabilities will be developed for specific classes of chemical modifications of nucleotides; nucleic acid 3D structure motifs by their common names, for example, G-Quadruplex, R-loop, Holliday Junction, Sarcin-Ricin, Kink-turn; ribosome functional states, e.g., full or single subunit, translational state, and numbers and positions of bound tRNAs; and deeper classification levels for selected proteins such as transcription factors.

The NAKB website will employ a modern web infrastructure with flexible data representation viewable on phones and tablets as well as desktop computers. For each NA-containing 3D structure, an atlas page will provide a summary overview of annotations as well as access to 1D, 2D, and 3D visualizations, external analysis tools, and file downloads. Mappings to external database links will initially include: PDB, Uniprot, RNACentral, Rfam. External analysis tools will include DNATCO and DNAProDB. Some of the report-

ing functions that were available in the original NDB so that the types of conformational analysis described earlier will be reenabled.

Author Contributions: Conceptualization, H.M.B.; writing—original draft, review, editing, H.M.B., C.L.L. and B.S. All authors have read and agreed to the published version of the manuscript.

Funding: The early funding of the NDB by the National Science Foundation is gratefully acknowledged. Funding for B.S. is from program Inter excellence of Ministry of Education, Youth, and Sports of the Czech Republic, grant number LTAUSA18197, and institutional support to the Institute of Biotechnology of the Czech Academy of Sciences, grant number RVO 86652036. Funding for C.L.L. is from the National Institutes of Health General Medical Sciences, grant numbers R01 GM079429 and R01 GM085238.

Acknowledgments: We are grateful to our collaborators for their many contributions over NDB's three decades of existence.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Watson, J.D.; Crick, F.H.C. A structure for deoxyribose nucleic acid. *Nature* **1953**, *171*, 737–738. [[CrossRef](#)] [[PubMed](#)]
2. Franklin, R.E.; Gosling, R.G. Molecular configuration in sodium thymonucleate. *Nature* **1953**, *171*, 740–741. [[CrossRef](#)] [[PubMed](#)]
3. Sussman, J.L.; Seeman, N.C.; Kim, S.-H.; Berman, H.M. The crystal structure of a naturally occurring dinucleotide phosphate uridylyl 3',5'-adenosine phosphate. models for RNA chain folding. *J. Mol. Biol.* **1972**, *66*, 403–421. [[CrossRef](#)]
4. Seeman, N.C.; Rosenberg, J.M.; Suddath, F.L.; Kim, J.J.P.; Rich, A. RNA double helical fragments at atomic resolution: I. The crystal and molecular structure of sodium adenylyl-3'-5'-uridine hexahydrate. *J. Mol. Biol.* **1976**, *104*, 109–144. [[CrossRef](#)]
5. Rosenberg, J.M.; Seeman, N.C.; Day, R.O.; Rich, A. RNA double helical fragments at atomic resolution: II. The structure of sodium guanylyl-3',5'-cytidine nonhydrate. *J. Mol. Biol.* **1976**, *104*, 145–167. [[CrossRef](#)]
6. Allen, F.H.; Bellard, S.; Brice, M.D.; Cartwright, B.A.; Doubleday, A.; Higgs, H.; Hummelink, T.; Hummelink-Peters, B.G.; Kennard, O.; Motherwell, W.D.S.; et al. The Cambridge Crystallographic Data Centre: Computer-based search, retrieval, analysis and display of information. *Acta Crystallogr.* **1979**, *B35*, 2331–2339. [[CrossRef](#)]
7. Robertus, J.D.; Ladner, J.E.; Finch, J.T.; Rhodes, D.; Brown, R.S.; Clark, B.F.C.; Klug, A. Structure of yeast phenylalanine tRNA at 3 Å resolution. *Nature* **1974**, *250*, 546–551. [[CrossRef](#)]
8. Suddath, F.; Quigley, G.; McPherson, A.; Sneden, D.; Kim, J.; Kim, S.; Rich, A. Three-dimensional structure of yeast phenylalanine transfer RNA at 3.0 Å resolution. *Nature* **1974**, *248*, 20–24. [[CrossRef](#)]
9. Quigley, G.J.; Seeman, N.C.; Wang, A.H.; Suddath, F.L.; Rich, A. Yeast phenylalanine transfer RNA: Atomic coordinates and torsion angles. *Nucleic Acids Res.* **1975**, *2*, 2329–2341. [[CrossRef](#)]
10. Wang, A.H.-J.; Quigley, G.J.; Kolpak, F.J.; Crawford, J.L.; van Boom, J.H.; van der Marel, G.A.; Rich, A. Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature* **1979**, *282*, 680–686. [[CrossRef](#)]
11. Drew, H.R.; Wing, R.M.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R.E. Structure of a B-DNA dodecamer: Conformation and dynamics. *Proc. Natl. Acad. Sci. USA* **1981**, *78*, 2179–2183. [[CrossRef](#)] [[PubMed](#)]
12. Bernstein, F.C.; Koetzle, T.F.; Williams, G.J.; Meyer, E.F., Jr.; Brice, M.D.; Rodgers, J.R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* **1977**, *112*, 535–542. [[CrossRef](#)]
13. Breslauer, K.J. A calorimetric determination of enthalpies and heat capacities of protonation. *J. Chem. Thermodyn.* **1979**, *11*, 527–530. [[CrossRef](#)]
14. Breslauer, K.J. Methods for Obtaining Thermodynamic Data on Oligonucleotide Transitions. In *Thermodynamic Data for Biochemistry and Biotechnology*; Hinz, H., Ed.; Springer: New York, NY, USA, 1986; pp. 402–427.
15. Breslauer, K.J. A thermodynamic perspective of DNA Bending. *Curr. Biol.* **1991**, *1*, 416–422. [[CrossRef](#)]
16. Breslauer, K.J. Extracting Thermodynamic Data From Equilibrium Melting Curves for Oligonucleotide Order-Disorder Transitions. In *Methods in Molecular Biology, Vol. 26: Protocols for Oligonucleotide Conjugates*; Agrawal, S., Ed.; Humana Press: Totowa, NJ, USA, 1994; Chapter 14; pp. 347–372.
17. Chalikian, T.V.; Breslauer, K.J. Thermodynamic analysis of biomolecules: A volumetric approach. *Curr. Opin. Struct. Biol.* **1998**, *8*, 657–664. [[CrossRef](#)]
18. Chalikian, T.V.; Volker, J.; Plum, G.E.; Breslauer, K.J. A more unified picture for the thermodynamics of nucleic acid duplex melting: A characterization by calorimetric and volumetric techniques. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 7853–7858. [[CrossRef](#)]
19. Klump, H.H.; Volker, J.; Breslauer, K.J. Energy mapping of the genetic code and genomic domains: Implications for code evolution and molecular Darwinism. *Q. Rev. Biophys.* **2020**, *53*, e11. [[CrossRef](#)]
20. Volker, J.; Klump, H.H.; Breslauer, K.J. DNA metastability and biological regulation: Conformational dynamics of metastable omega-DNA bulge loops. *J. Am. Chem. Soc.* **2007**, *129*, 5272–5280. [[CrossRef](#)]
21. Jones, R. Preparation of protected deoxyribonucleosides. In *Oligonucleotide Synthesis, a Practical Approach*; Gait, M.J., Ed.; IRL Press: Washington, DC, USA, 1984; pp. 22–34.

22. Manning, G. The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Q. Rev. Biophys.* **1978**, *11*, 179–246. [[CrossRef](#)]
23. Volker, J.; Klump, H.H.; Manning, G.S.; Breslauer, K.J. Counterion association with native and denatured nucleic acids: An experimental approach. *J. Mol. Biol.* **2001**, *310*, 1011–1025. [[CrossRef](#)]
24. Erie, D.; Sinha, N.; Olson, W.; Jones, R.; Breslauer, K. A dumbbell-shaped, double-hairpin structure of DNA: A thermodynamic investigation. *Biochemistry* **1987**, *26*, 7150–7159. [[CrossRef](#)] [[PubMed](#)]
25. Berman, H.M.; Olson, W.K.; Beveridge, D.L.; Westbrook, J.; Gelbin, A.; Demeny, T.; Hsieh, S.H.; Srinivasan, A.R.; Schneider, B. The nucleic acid database. A comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.* **1992**, *63*, 751–759. [[CrossRef](#)]
26. Kitakami, H.; Tateno, Y.; Gojobori, T. Toward unification of taxonomy databases in a distributed computer environment. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **1994**, *2*, 227–235. [[PubMed](#)]
27. Bilofsky, H.S.; Burks, C.; Fickett, J.W.; Goad, W.B.; Lewitter, F.I.; Rindone, W.P.; Swindell, C.D.; Tung, C.S. The GenBank genetic sequence databank. *Nucleic Acids Res.* **1986**, *14*, 1861–1863. [[CrossRef](#)]
28. Benson, D.A.; Karsch-Mizrachi, I.; Lipman, D.J.; Ostell, J.; Rapp, B.A.; Wheeler, D.L. GenBank. *Nucleic Acids Res.* **2000**, *28*, 15–18. [[CrossRef](#)]
29. Parker, M. Biological data access through Gopher. *Trends Biochem. Sci.* **1993**, *18*, 485–486. [[CrossRef](#)]
30. Fitzgerald, P.M.D.; Westbrook, J.D.; Bourne, P.E.; McMahon, B.; Watenpaugh, K.D.; Berman, H.M. 4.5 Macromolecular dictionary (mmCIF). In *International Tables for Crystallography G. Definition and Exchange of Crystallographic Data*; Hall, S.R., McMahon, B., Eds.; Springer: Dordrecht, The Netherlands, 2005; pp. 295–443.
31. Clowney, L.; Jain, S.C.; Srinivasan, A.R.; Westbrook, J.; Olson, W.K.; Berman, H.M. Geometric Parameters in Nucleic Acids: Nitrogenous Bases. *J. Am. Chem. Soc.* **1996**, *118*, 509–518. [[CrossRef](#)]
32. Gelbin, A.; Schneider, B.; Clowney, L.; Hsieh, S.-H.; Olson, W.K.; Berman, H.M. Geometric parameters in nucleic acids: Sugar and phosphate constituents. *J. Am. Chem. Soc.* **1996**, *118*, 519–528. [[CrossRef](#)]
33. Parkinson, G.; Vojtechovsky, J.; Clowney, L.; Brunger, A.T.; Berman, H.M. New parameters for the refinement of nucleic acid-containing structures. *Acta Cryst. D Biol. Cryst.* **1996**, *52 Pt 1*, 57–64. [[CrossRef](#)]
34. Olson, W.K.; Bansal, M.; Burley, S.K.; Dickerson, R.E.; Gerstein, M.; Harvey, S.C.; Heinemann, U.; Lu, X.J.; Neidle, S.; Shakked, Z.; et al. A standard reference frame for the description of nucleic acid base-pair geometry. *J. Mol. Biol.* **2001**, *313*, 229–237. [[CrossRef](#)]
35. Leontis, N.B.; Westhof, E. Geometric nomenclature and classification of RNA base pairs. *RNA* **2001**, *7*, 499–512. [[CrossRef](#)] [[PubMed](#)]
36. Schneider, B.; de la Cruz, J.; Feng, Z.; Chen, L.; Dutta, S.; Persikova, I.; Westbrook, J.; Yang, H.; Young, J.; Zardecki, C.; et al. The Nucleic Acid Database. In *Structural Bioinformatics*, 2nd ed.; Gu, J., Bourne, P.E., Eds.; Wiley-Blackwell: Hoboken, NJ, USA, 2009; pp. 305–319.
37. Coimbatore Narayanan, B.; Westbrook, J.; Ghosh, S.; Petrov, A.I.; Sweeney, B.; Zirbel, C.L.; Leontis, N.B.; Berman, H.M. The Nucleic Acid Database: New features and capabilities. *Nucleic Acids Res.* **2014**, *42*, D114–D122. [[CrossRef](#)]
38. Srinivasan, A.R.; Olson, W.K. Yeast tRNA^{Phe} conformation wheels: A novel probe of the monoclinic and orthorhombic models. *Nucleic Acids Res.* **1980**, *8*, 2307–2329. [[CrossRef](#)]
39. Dans, P.D.; Perez, A.; Faustino, I.; Lavery, R.; Orozco, M. Exploring polymorphisms in B-DNA helical conformations. *Nucleic Acids Res.* **2012**, *40*, 10668–10678. [[CrossRef](#)]
40. Gupta, A.; Kulkarni, M.; Mukherjee, A. Accurate prediction of B-form/A-form DNA conformation propensity from primary sequence: A machine learning and free energy handshake. *Patterns* **2021**, *2*, 100329. [[CrossRef](#)] [[PubMed](#)]
41. Bayrak, C.S.; Kim, N.; Schlick, T. Using sequence signatures and kink-turn motifs in knowledge-based statistical potentials for RNA structure prediction. *Nucleic Acids Res.* **2017**, *45*, 5414–5422. [[CrossRef](#)]
42. Corsi, F.; Lavery, R.; Laine, E.; Carbone, A. Multiple protein-DNA interfaces unravelled by evolutionary information, physico-chemical and geometrical properties. *PLoS Comput. Biol.* **2020**, *16*, e1007624. [[CrossRef](#)]
43. Srivastava, A.; Ahmad, S.; Gromiha, M.M. Deciphering RNA-Recognition Patterns of Intrinsically Disordered Proteins. *Int. J. Mol. Sci.* **2018**, *19*, 1595. [[CrossRef](#)] [[PubMed](#)]
44. Sagendorf, J.M.; Markarian, N.; Berman, H.M.; Rohs, R. DNAProDB: An expanded database and web-based tool for structural analysis of DNA-protein complexes. *Nucleic Acids Res.* **2020**, *48*, D277–D287. [[CrossRef](#)]
45. Schneider, B.; Berman, H.M. Hydration of the DNA bases is local. *Biophys. J.* **1995**, *69*, 2661–2669. [[CrossRef](#)]
46. Schneider, B.; Neidle, S.; Berman, H.M. Conformations of the sugar-phosphate backbone in helical DNA crystal structures. *Biopolymers* **1997**, *42*, 113–124. [[CrossRef](#)]
47. Kim, S.-H.; Berman, H.M.; Seeman, N.C.; Newton, M.D. Seven basic conformations of nucleic acid structural units. *Acta Crystallogr. Sect. B* **1973**, *29*, 703–710. [[CrossRef](#)]
48. Richardson, J.S.; Schneider, B.; Murray, L.W.; Kapral, G.J.; Immormino, R.M.; Headd, J.J.; Richardson, D.C.; Ham, D.; Hershkovits, E.; Williams, L.D.; et al. RNA backbone: Consensus all-angle conformers and modular string nomenclature (an RNA Ontology Consortium contribution). *RNA* **2008**, *14*, 465–481. [[CrossRef](#)] [[PubMed](#)]
49. Schneider, B.; Bozikova, P.; Cech, P.; Svozil, D.; Cerny, J. A DNA Structural Alphabet Distinguishes Structural Features of DNA Bound to Regulatory Proteins and in the Nucleosome Core Particle. *Genes* **2017**, *8*, 278. [[CrossRef](#)]

50. Cerny, J.; Bozikova, P.; Svoboda, J.; Schneider, B. A unified dinucleotide alphabet describing both RNA and DNA structures. *Nucleic Acids Res.* **2020**, *48*, 6367–6381. [[CrossRef](#)]
51. Gouge, J.; Satia, K.; Guthertz, N.; Widya, M.; Thompson, A.J.; Cousin, P.; Dergai, O.; Hernandez, N.; Vannini, A. Redox Signaling by the RNA Polymerase III TFIIB-Related Factor Brf2. *Cell* **2015**, *163*, 1375–1387. [[CrossRef](#)]
52. Frouws, T.D.; Duda, S.C.; Richmond, T.J. X-ray structure of the MMTV-A nucleosome core. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 1214–1219. [[CrossRef](#)]
53. Olson, W.K. Configurational statistics of polynucleotide chains. An updated virtual bond model to treat effects of base stacking. *Macromolecules* **1980**, *13*, 721–728. [[CrossRef](#)]
54. Wadley, L.M.; Keating, K.S.; Duarte, C.M.; Pyle, A.M. Evaluating and learning from RNA pseudotorsional space: Quantitative validation of a reduced representation for RNA structure. *J. Mol. Biol.* **2007**, *372*, 942–957. [[CrossRef](#)]
55. Vander Zanden, C.M.; Rowe, R.K.; Broad, A.J.; Robertson, A.B.; Ho, P.S. Effect of Hydroxymethylcytosine on the Structure and Stability of Holliday Junctions. *Biochemistry* **2016**, *55*, 5781–5789. [[CrossRef](#)]
56. Wang, Y.; Patel, D.J. Solution structure of a parallel-stranded G-quadruplex DNA. *J. Mol. Biol.* **1993**, *234*, 1171–1183. [[CrossRef](#)] [[PubMed](#)]
57. Davey, C.A.; Sargent, D.F.; Luger, K.; Maeder, A.W.; Richmond, T.J. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *J. Mol. Biol.* **2002**, *319*, 1097–1113. [[CrossRef](#)]
58. Lawson, C.L.; Carey, J. Tandem binding in crystals of a trp repressor/operator half-site complex. *Nature* **1993**, *366*, 178–182. [[CrossRef](#)]
59. Weaver, T.M.; Cortez, L.M.; Khoang, T.H.; Washington, M.T.; Agarwal, P.K.; Freudenthal, B.D. Visualizing Rev1 catalyze protein-template DNA synthesis. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 25494–25504. [[CrossRef](#)]
60. Jang, S.B.; Hung, L.W.; Chi, Y.I.; Holbrook, E.L.; Carter, R.J.; Holbrook, S.R. Structure of an RNA internal loop consisting of tandem C-A+ base pairs. *Biochemistry* **1998**, *37*, 11726–11731. [[CrossRef](#)] [[PubMed](#)]
61. Chan, C.W.; Badong, D.; Rajan, R.; Mondragon, A. Crystal structures of an unmodified bacterial tRNA reveal intrinsic structural flexibility and plasticity as general properties of unbound tRNAs. *RNA* **2020**, *26*, 278–289. [[CrossRef](#)]
62. Huang, L.; Wang, J.; Watkins, A.M.; Das, R.; Lilley, D.M.J. Structure and ligand binding of the glutamine-II riboswitch. *Nucleic Acids Res.* **2019**, *47*, 7666–7675. [[CrossRef](#)] [[PubMed](#)]
63. Noeske, J.; Wasserman, M.R.; Terry, D.S.; Altman, R.B.; Blanchard, S.C.; Cate, J.H. High-resolution structure of the Escherichia coli ribosome. *Nat. Struct. Mol. Biol.* **2015**, *22*, 336–341. [[CrossRef](#)]
64. Li, X.; Liu, S.; Zhang, L.; Issaian, A.; Hill, R.C.; Espinosa, S.; Shi, S.; Cui, Y.; Kappel, K.; Das, R.; et al. A unified mechanism for intron and exon definition and back-splicing. *Nature* **2019**, *573*, 375–380. [[CrossRef](#)]
65. Landeras-Bueno, S.; Wasserman, H.; Oliveira, G.; VanAernum, Z.L.; Busch, F.; Salie, Z.L.; Wysocki, V.H.; Andersen, K.; Sapphire, E.O. Cellular mRNA triggers structural transformation of Ebola virus matrix protein VP40 to its essential regulatory form. *Cell Rep.* **2021**, *35*, 108986. [[CrossRef](#)]
66. Lu, X.J. DSSR-enabled innovative schematics of 3D nucleic acid structures with PyMOL. *Nucleic Acids Res.* **2020**, *48*, e74. [[CrossRef](#)]
67. Schneider, B.; Bozikova, P.; Necasova, I.; Cech, P.; Svozil, D.; Cerny, J. A DNA structural alphabet provides new insight into DNA flexibility. *Acta Cryst. D Struct. Biol.* **2018**, *74 Pt 1*, 52–64. [[CrossRef](#)] [[PubMed](#)]
68. Appasamy, S.D.; Hamdani, H.Y.; Ramlan, E.I.; Firdaus-Raih, M. InterRNA: A database of base interactions in RNA structures. *Nucleic Acids Res.* **2016**, *44*, D266–D471. [[CrossRef](#)] [[PubMed](#)]
69. Boccaletto, P.; Machnicka, M.A.; Purta, E.; Piatkowski, P.; Baginski, B.; Wirecki, T.K.; de Crecy-Lagard, V.; Ross, R.; Limbach, P.A.; Kotter, A.; et al. MODOMICS: A database of RNA modification pathways. 2017 update. *Nucleic Acids Res.* **2018**, *46*, D303–D307. [[CrossRef](#)] [[PubMed](#)]
70. Chojnowski, G.; Walen, T.; Bujnicki, J.M. RNA Bricks—A database of RNA 3D motifs and their interactions. *Nucleic Acids Res.* **2014**, *42*, D123–D131. [[CrossRef](#)]
71. Zok, T.; Antczak, M.; Zurkowski, M.; Popena, M.; Blazewicz, J.; Adamiak, R.W.; Szachniuk, M. RNAPdb 2.0: Multifunctional tool for RNA structure annotation. *Nucleic Acids Res.* **2018**, *46*, W30–W35. [[CrossRef](#)]
72. The RNAcentral Consortium; Petrov, A.I.; Kay, S.J.E.; Kalvari, I.; Howe, K.L.; Gray, K.A.; Bruford, E.A.; Kersey, P.J.; Cochrane, G.; Finn, R.D. RNAcentral: A comprehensive database of non-coding RNA sequences. *Nucleic Acids Res.* **2017**, *45*, D128–D134.
73. Bernier, C.R.; Petrov, A.S.; Waterbury, C.C.; Jett, J.; Li, F.; Freil, L.E.; Xiong, X.; Wang, L.; Migliozi, B.L.; Hershkovits, E.; et al. RiboVision suite for visualization and analysis of ribosomes. *Faraday Discuss.* **2014**, *169*, 195–207. [[CrossRef](#)]
74. Sagendorf, J.M.; Berman, H.M.; Rohs, R. DNAProDB: An interactive tool for structural analysis of DNA-protein complexes. *Nucleic Acids Res.* **2017**, *45*, W89–W97. [[CrossRef](#)]
75. Petrov, A.I.; Zirbel, C.L.; Leontis, N.B. Automated classification of RNA 3D motifs and the RNA 3D Motif Atlas. *RNA* **2013**, *19*, 1327–1340. [[CrossRef](#)]
76. Zirbel, C.; Leontis, N. Nonredundant 3D Structure Datasets for RNA Knowledge Extraction and Benchmarking. In *RNA 3D Structure Analysis and Prediction*; Leontis, N., Westhof, E., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; Volume 27, pp. 281–298.

-
77. Dana, J.M.; Gutmanas, A.; Tyagi, N.; Qi, G.; O'Donovan, C.; Martin, M.; Velankar, S. SIFTS: Updated Structure Integration with Function, Taxonomy and Sequences resource allows 40-fold increase in coverage of structure-based annotations for proteins. *Nucleic Acids Res.* **2019**, *47*, D482–D489. [[CrossRef](#)] [[PubMed](#)]
 78. Lu, X.J.; Bussemaker, H.J.; Olson, W.K. DSSR: An integrated software tool for dissecting the spatial structure of RNA. *Nucleic Acids Res.* **2015**, *43*, e142. [[CrossRef](#)] [[PubMed](#)]