

# An RNA-centric historical narrative around the Protein Data Bank

Received for publication, November 5, 2020, and in revised form, February 17, 2021. Published, Papers in Press, March 18, 2021, <https://doi.org/10.1016/j.jbc.2021.100555>

Eric Westhof<sup>1,\*</sup> and Neocles B. Leontis<sup>2</sup>

From the <sup>1</sup>Institut de Biologie Moléculaire et Cellulaire du CNRS, Architecture et Réactivité de l'ARN, Université de Strasbourg, Strasbourg, France; <sup>2</sup>Department of Chemistry, Bowling Green State University, Bowling Green, Ohio, USA

Edited by Karin Musier-Forsyth

Some of the amazing contributions brought to the scientific community by the Protein Data Bank (PDB) are described. The focus is on nucleic acid structures with a bias toward RNA. The evolution and key roles in science of the PDB and other structural databases for nucleic acids illustrate how small initial ideas can become huge and indispensable resources with the unflinching willingness of scientists to cooperate globally. The progress in the understanding of the molecular interactions driving RNA architectures followed the rapid increase in RNA structures in the PDB. That increase was consecutive to improvements in chemical synthesis and purification of RNA molecules, as well as in biophysical methods for structure determination and computer technology. The RNA modeling efforts from the early beginnings are also described together with their links to the state of structural knowledge and technological development. Structures of RNA and of its assemblies are physical objects, which, together with genomic data, allow us to integrate present-day biological functions and the historical evolution in all living species on earth.

The *Protein Data Bank* (PDB) is an icon for structural biologists. In its virtual vaults, the PDB stores the structures of biological macromolecules obtained painstakingly by thousands of crystallographers, electron microscopists, and nuclear magnetic resonance spectroscopists over the last 50 years or so. This note is primarily written to honor and acknowledge those who set up the PDB and who have strived throughout all these years to improve and maintain the database. We thank the *Journal of Biological Chemistry* for organizing the publication of such pieces that provide a mixture of personal recollections, historical records about nucleic acids structure, and some general comments about databases and their future.

## The evolution of structural databases

When one of us was a postdoctoral fellow with M. Sundaralingam at the Department of Biochemistry of the University of Wisconsin in Madison, the only way to work with a published molecular structure was to first type in the atomic coordinates at the console of a PDP computer (Programmed Data Processor from Digital Equipment Corporation,

Massachusetts). Only then could one deduce or check atomic distances, angles, stereochemistry, contacts and visualize the 3D structure using programs developed in-house. This was a tedious job, but very instructive to students and postdoctoral fellows. In those days, atomic coordinates were part of the paper and presented with temperature factors in tables (for an example, see (1) and for general guidelines (2)). The number and diversity of errors that escaped detection before publication, while rarely casting doubt on the crystallography itself, led generally to inaccurately described molecular structures with, for example, some wrong chiralities, or atoms of the molecular unit in different unit cells. This manual handling of data constituted for sure an excellent training ground. Happily, the nucleic acid structures that could be tackled at that time were restricted to the bases (often modified), nucleosides, and nucleotides, with the first structure of a dinucleotide appearing in 1971 (3, 4). In 1965, Olga Kennard, inspired by J.D. Bernal, had started to work on the *Cambridge Structural Database*, CSD (5) and, in 1971, the PDB was established jointly by the *Crystallographic Data Centre*, Cambridge and the *Brookhaven National Laboratory* (Protein Data Bank). PDB originally served as a repository system where crystallographers could mail in their data on punch cards, then computer tapes, for archiving and distribution. The announcement of the founding of the PDB in *Nature New Biology* was a simple little insert of 256 words (6)! Full publications appeared later (6, 7). The CSD archived structures of individual nucleotides, as well as di- and trinucleotides. These structures are not in PDB. In fact, the first release of PDB did not include any nucleic acid structures, just seven protein structures (7). Later, in 1992, the Nucleic Acid Database (NDB) was founded by Berman *et al.* (8) and structures of RNA units were included in the NDB.

In the early days of the development of structural databases, the main objectives were to prevent the loss of data by archiving solved structures and making them freely available upon request. The deposited data were checked for atom names, numberings, and geometry by database curators, but not as systematically and thoroughly as it is done today. Dictionaries and defined formats were developed over the years and applied to standardize nomenclature, file formats, and metadata. As regards nucleic acids, the relational NDB archived data accompanied by structural information and descriptions (8). NDB served as a testbed for new data formats

\* For correspondence: Eric Westhof, [E.Westhof@ibmc-cnrs.unistra.fr](mailto:E.Westhof@ibmc-cnrs.unistra.fr).



**Eric Westhof** is Professor Emeritus of Structural Biochemistry at the University of Strasbourg at the Institute of Molecular and Cellular Biology of the CNRS. Starting with his early work with M. Sundaralingam, he has made seminal contributions to the structural biology of nucleic acids.

that could accommodate much larger structures and much more metadata, which were eventually adopted by PDB. Nowadays, these databases are rich in metadata and tools for assessing and describing structures. Very valuable tools for structural validations (9) and corrections (10) of nucleic acid structures were introduced. The PDB now offers various validation tools with compelling metrics for all types of structures (11). While the PDB remains, understandably, protein-centric, other databases and tools are available for nucleic acids (8, 12, 13)—see also <https://www.bgsu.edu/research/rna/>.

### Quality and accuracy of structural data

All of the precise structural data regarding RNA comes ultimately from atomic-resolution X-ray structures of nucleotides, oligonucleotides, and various biologically relevant structures, ranging in size from individual helical elements to the full ribosome. The early work was carried out by pioneers such as A. Rich, O. Kennard, and M. Sundaralingam, who from the mid-1960s and into the 1980s carried out precise crystallographic studies of nucleosides, nucleotides, and dinucleotides. These data comprise all our basic knowledge of bond lengths, angles, and stereochemistry, as well as interaction preferences, including all types of base pairs, and most stacking and base–backbone interactions. The new data led early on to useful concepts including the conformational wheel, describing ribose conformations (14), and better understanding of the conformational preferences of nucleic acids, the analysis of which was quite daunting to theoreticians and modelers alike (15–17). Unlike proteins, which effectively have just two degrees of freedom per monomer unit, nucleotides present seven, six dihedrals along the backbone, and the one around the glycosidic bond. In 1965, based on the small number of protein structures known, as well as better resolved peptide structures, Ramakrishnan and Ramachandran (18) worked out the sterically favored backbone configurations and summarized them graphically in two dimensions, the dihedral angles phi and psi. Such Ramachandran plots continue to be used as easily visualized metrics for protein structures. Shortly afterward, Sundaralingam (19) carried out a series of landmark studies to accurately determine molecular structures of nucleic acid constituents, culminating in a highly cited 1969 paper in *Biopolymers* that laid the foundation for conformational analysis of nucleotides and polynucleotides, and gave the first stereochemical rules for the sugar-phosphate backbone. This

work revealed conclusively that backbone dihedrals are restricted and that correlations exist between them and supported theoretical work that proposed various schemes to simplify nucleic acid conformational analysis using virtual bonds (15, 20).

High-resolution data from small molecules are also used to build force fields and to infer rules for assembly of molecular moieties. These force fields and energetic rules are then used for producing and optimizing structures, sampling the conformational space, or simulating molecular dynamics. The quality and general value of the deduced force fields strongly depend on the number and variety of structures available. The importance of efficient and accurate force fields is paramount in modern biochemical research, since these force fields are used not only for computing trajectories of molecular dynamics simulations, but also in determination of new structure using NMR and cryo-electron microscopy.

It cannot be overemphasized that the quality of the deposited structures is of primary importance; it is directly related to the crystallographic resolution of the X-ray data and on the refinement process, since only a minor fraction of X-ray structures is obtained at true atomic resolution (better than 1 Å). For example, can one really be confident in RNA structures at resolutions above 3.3 Å with average B-factors around 200 Å<sup>2</sup>, bad clash scores, and poor PDB validation metrics? Only a systematic analysis of the various regions of the RNA molecule in electron density maps would allow a knowledgeable and interested person to reach an informed opinion. One key parameter for compiling reference databases for knowledge extraction is the nonredundancy of the structures that are included. Nonredundant structure databases, by reducing bias in the parameters deduced from structures, are extremely valuable for extracting knowledge about RNA as well as for benchmarking modeling strategies (21). In this respect, it is worth noting that less than 100 nonredundant RNA structures have been solved at 2 Å resolution or better. Well-defined metrics for 3D structures determined from crystallography (X-ray and NMR) have been established and are reported at the nucleotide level by PDB (22–24). Metrics have recently also been defined for structures determined by cryo-electron microscopy, which is largely replacing X-ray diffraction for structure determination of large macromolecular machines such as ribosomes, viruses, and spliceosomes (25).

### The first 20 years of RNA structures

The first nucleic acid structures were of short synthetic DNA helices and were deposited in 1981 (Z-DNA, 2DCG (26) and B-DNA, 1BNA (27)), with the molecular descriptions published respectively in 1979 (28) and 1980 (29). These high-resolution DNA structures only became possible once solid-state synthesis could provide sufficient quantities of pure oligonucleotides for crystallization and structure determination (30, 31). The early RNA structures, of biological origin, were those of various tRNAs isolated from *Escherichia coli* or yeast (32–34), an accomplishment that required novel purification and crystallization protocols. It took roughly 10 years to produce refined structures of yeast tRNA<sup>Phe</sup> (deposited in 1978,

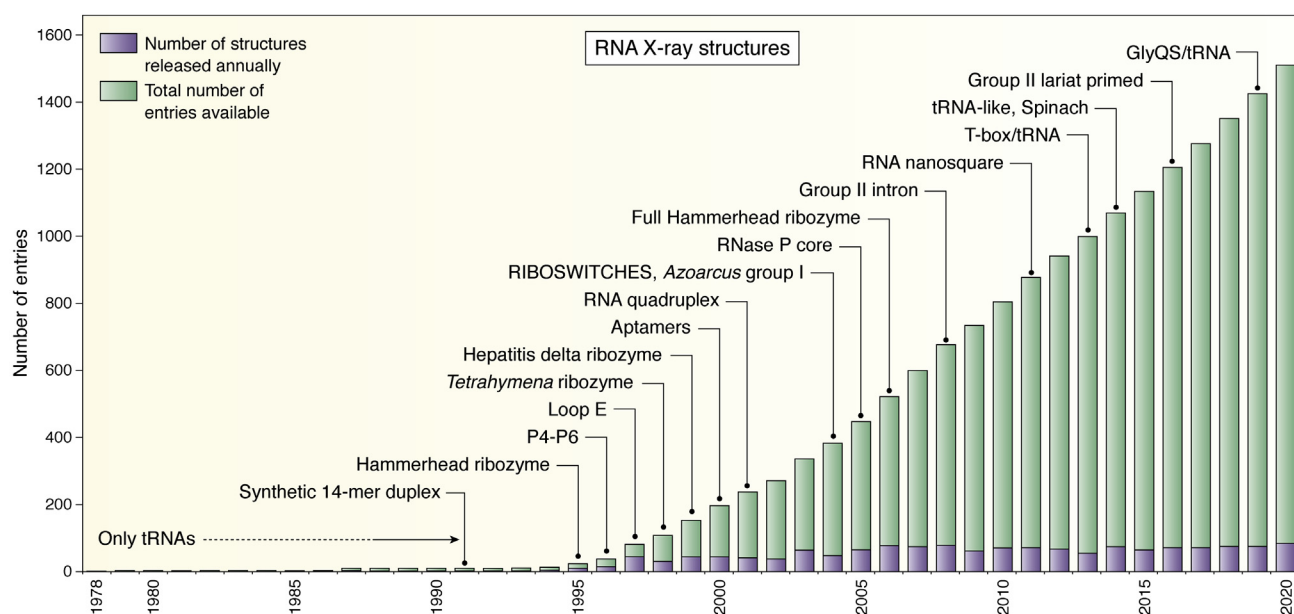
4TNA (35) and 6TNA (36)). The choice of tRNA for study was obvious because of the amount present in cells and the availability of some purification protocols. Its primary and secondary structures had been determined in the 1960s, by chemical and biochemical means (37, 38). In the 1980s, additional tRNA structures were solved as well as the first RNA viruses (1BMV (39) and 2TMV (40)), containing short fragments of the viral genome bound to the coat proteins. In those years, at crystallography meetings and conferences, talks on nucleic acids, and RNA in particular, were generally relegated to the last day, either after the meeting dinner (always very lively and joyous events) in the early morning or just before the departure of the bus.

Following novelty and progress in chemical synthesis and purification of RNA oligonucleotides (41), the first X-ray structures of synthetic RNAs appeared in 1988 (42) and were deposited in 1991. That structure showed first examples of intermolecular interactions involving the ribose hydroxyl groups between RNA helices (43). The next advances came from NMR structures of small recurrent RNA motifs, such as the frequent GNRA tetraloop (44) - identified by sequence comparisons in large RNAs (autocatalytic introns, ribosomal RNAs, RNase P). These included most revealingly the loop E of 5S rRNA (45, 46). Even though they lacked atomic resolution, the NMR structures were sufficiently detailed to provide models that expanded our understanding on how the RNA structures could form, in particular the variety of non-Watson-Crick base

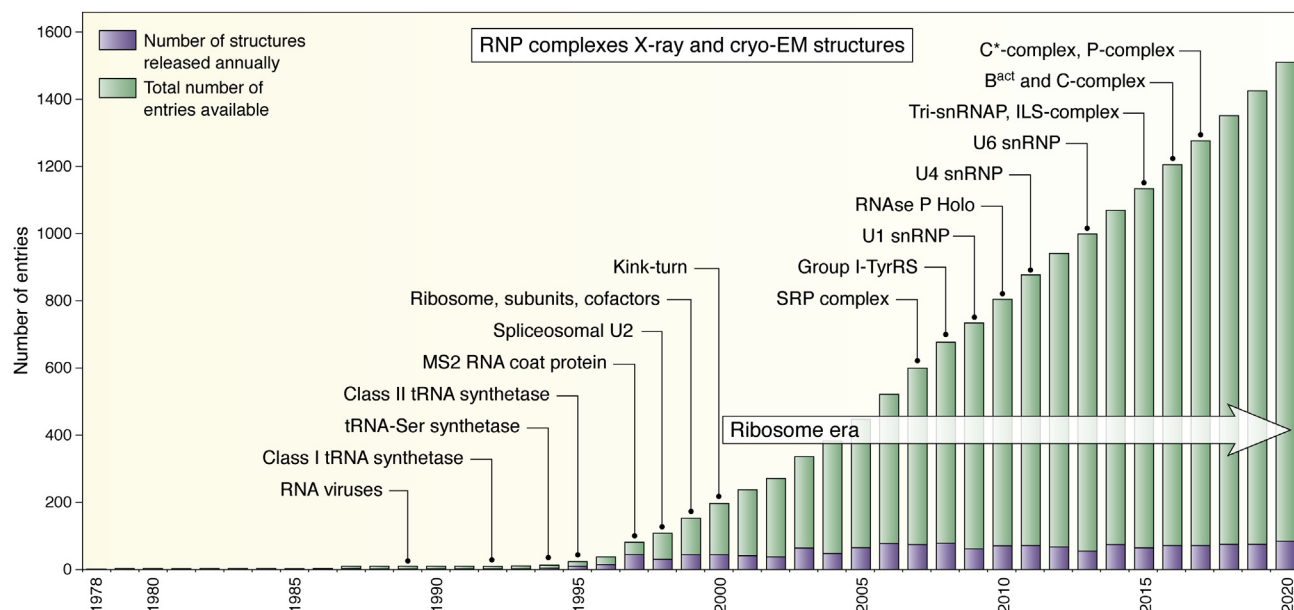
pairs, and they provided enough information to infer sequence signatures to identify recurrences of the same motifs in other structures (44, 47). The willingness of its directors to expand PDB to include NMR structures proved to be a wise decision.

By 1995, the number of X-ray RNA structures (alone or in complex with protein) in the PDB amounted to only about 1% of the present content of total RNA and RNA-protein complex (RNP) structures. Figures 1 and 2 show the evolution in the number of structures of RNA alone and RNP with time. Some key structures are indicated to illustrate the increase in complexity of the solved structures. Again progress in chemical synthesis (48, 49), biochemical RNA production (50) and purification (51), as well as in X-ray technology, including the spreading use of synchrotron radiation at cryogenic temperatures (52, 53) were central to the rapid increase in the resolution of RNA and RNP structures. Many advances in crystallography (54, 55) and in biochemical preparation for crystallization of RNA (56–58) or RNA complexes (59–61) appeared in those years and the whole community benefited greatly from those novel approaches. Amazing breakthroughs in cryo-electron microscopy are now accelerating the pace at which highly complex particles can be observed (62–64).

In 1996, a breakthrough in RNA crystallography was achieved in the laboratories of Tom Cech and Jennifer Doudna with the crystal structure of the P4–P6 domain of the *Tetrahymena* group I intron, a structure twice the size of tRNA, featuring striking compact folds and novel types of RNA



**Figure 1. The evolution of the number of RNA structures in the PDB.** The figure is downloaded from the option “Analyze PDB statistics.” All RNA structures are included (from X-ray, NMR, and cryo-EM). Some key X-ray structures are indicated. Up to 1991, only tRNA structures were present. Following time, these structures are highlighted: (1) A synthetic 14-mer duplex, 1RNA (43); (2) the core hammerhead ribozyme, 1MME (189); (3) the P4-P6 domain of the *Tetrahymena* ribozyme, 1GID (65); (4) the eukaryotic loop E structure, 354D (191); (5) the core of the *Tetrahymena* ribozyme, 1GRZ (200); (6) the hepatitis delta ribozyme, 1DRZ (192); (7) Aptamers binding to malachite green, 1F1T (201), and vitamin B12, 1DDY (202); (8) RNA quadruplex, 1J8G (203); an earlier NMR structure was solved before, 1RAU (204); (9) the purine riboswitch, 1Y27 (119), since then the structures of a great variety of riboswitches have appeared (117, 120, 205); (10) the group I intron from *Azoarcus*, 1U6B (206); (11) the core of a RNase P ribozyme, 2A2E (207); (12) the full hammerhead ribozyme with long-range loop-loop contacts stabilizing the core, 3ZD5 (208); (13) a complete group II intron, 3EOH (101); (14) the structure of a RNA nanosquare, 3P59 (209); (15) the complex between the T-box riboswitch and its tRNA target, 4MGN (210); (16) the TYMV tRNA-like, 4P5J (211); (17) the Spinach fluorescent aptamer, 4TSO (113); (18) a group II intron with a lariat primed for transposition, 5J01 (102); (19) the full structure of the T-box between GlyQS and its tRNA, 6POM (212).



**Figure 2. The evolution of the number of structures of RNA-protein complexes (RNPs) in the PDB.** The figure is downloaded from the option “Analyze PDB statistics.” All RNP structures are included (from X-ray, NMR, and cryo-EM). The number of structures related to ribosomes and its cofactors is much too large to show them on such a figure. We preferred to emphasize the complexes formed in the spliceosome (for detailed reviews, see (99, 213)). Most of the large RNP structures after 2015 are based on cryo-EM data. Following time, these structures are highlighted: (1) RNA viruses, 1BMV (39); (2) class I tRNA synthetase complex, 1GSG (214); (3) tRNA<sup>Ser</sup>, a class II tRNA with a long variable loop, complexed with its specific synthetase, 1SER (215); (4) class II tRNA synthetase complex, 1ASY (216); (5) MS2 RNA coat protein, 1AQ3 (217); (6) spliceosomal U2 complex, 1A9N (218); (7) the kink-turn was first observed in the complex of U4 sRNA fragment, 1E7K (219), before being recurrently observed in the ribosome structure (193); (8) the Signal Recognition Particle complex, 2V3C (220); (9) complex between a tyrosyl tRNA synthetase and a group I intron, 2RKJ (221); (10) U1 snRNP, 3CW1 (222); (11) an RNase P holoenzyme, 3Q1Q (223); (12) in the U4 snRNP, 4WZJ (224); (13) Lsm/U6 snRNP complex, 4M7A (225); (14) the tri-snRNP structure, 3JCM (226); (15) Intron-lariat complex, 3JB9 (227); (16) B<sup>act</sup> complex, 5GM6 (228), C-complex, 5GMK (229), 5LJ3 (230); (17) C\*-complex, 5W5G (231), 5MPS (232); P-complex, 5YLZ (233), 6EXN (234), 6BK8 (235).

tertiary contacts (65, 66). With the amazing structure of P4–P6, many interested scientists learned that large RNAs could also be crystallized starting from *in vitro* synthesis and production. This breakthrough spurred fruitful efforts that quickly expanded our knowledge of the repertoire of surprising and beautiful RNA architectures, culminating in the structures of the ribosomal subunits themselves (67–70) and the Nobel Prize in Chemistry to Venki Ramakrishnan, Tom Steitz, Ada Yonath in 2009 (<https://www.nobelprize.org/prizes/chemistry/>). And, after the Nobel Prize in Physiology or Medicine for RNA interference (RNAi) in 2006 to Andrew Fire and Craig Mello, the RNA community is celebrating the 2020 Nobel Prize in Chemistry awarded to Jennifer Doudna and Emmanuelle Charpentier for the structure-based design of the splendidly efficient RNA-programmed Crispr-cas9 system (71, 72).

### The unfolding of RNA structural folding rules

The accumulation of RNA structures, each one bringing either a new insight key to folding or additional confirmation and sequence variants, led to deep understanding of the main physicochemical features underlying RNA architecture. Some of these features, gained from structures of Figure 1, are gathered in Table 1. Up to the 1970s, structures were restricted to the bases, nucleosides, or nucleotides and analyzed the stereochemistry of nucleic acids (19), the orientation of the base with respect to the ribose, stacking (73), sugar pucker (14), protonation states, tautomeric forms,

effects of modifications (74). These fundamental studies are still very relevant as shown by the roles of (de)protonation in nucleolytic ribozymes (75–77) or of tautomeric forms in ribosomal translation (78–81). Despite its relatively small size (<80 nts), tRNAs have a tightly folded structure and provided many surprising and eye-opening insights into the logic of RNA folding (see also Table 1): short Watson–Crick paired helices coaxially stacked (Fig. 3, A and B), short-range and long-range interactions (Fig. 3C), the central roles of non-Watson–Crick base pairs in mediating tertiary contacts (Fig. 4), base–phosphate or sugar–phosphate contacts (Fig. 4, A and B), structured hairpin loops (with the famous U-turn) and their propensities to intimately interact, and base–base intercalative stacking interactions (Fig. 4C) that contribute to long-range loop–loop interactions.

The quest for the structure of the hammerhead ribozyme took more than 10 years before a coherent picture emerged and overall, about 20 years since its discovery (82, 83) (for short overviews see (84, 85)). In any case, it displays a three-way junction maintained in the proper relative orientations of the helices by non-Watson–Crick pairs (Fig. 5). As mentioned above, the structure of P4–P6, a fragment of the group I intron from *Tetrahymena thermophila* (65), unveiled several key recurrent RNA folding rules. As alluded to above, that structure, beyond the breathtaking views on RNA folds, initiated a strong impetus on the RNA crystallography community. An illustration is shown in Figure 6 and short descriptions in Table 1.

**Table 1**  
A short historical overview of folding rules derived from crystal structures

Date	RNA	Key advances in revealing interactions within RNA structures
50s, 60s, 70s	Bases, nucleo-sides/-tides	Base pairing (74), stacking (73), sugar puckers (14), protonation, tautomers, modifications, metal binding (187)
1976	tRNAs	- coaxial stacking between helices (continuous strand) - loop-loop interactions - non-Watson-Crick base pairs - U-turn structuration of loops - base triple contacts between a single-strand and the deep major groove of a helix - intercalation of unpaired bases - polyamines, Mg ions, lead cleavage
1991	RNA helices	O2'H...O2' H-bonds between adjacent helices (43), 4-way junctions (188)
1995	Hammerhead ribozyme	3-way junction structured by non-Watson-Crick pairs (189, 190), GNRA/helix minor groove contacts (190)
1996	P4-P6	A-minor contacts, ribose zippers, platform triples, packing of helices, loop-loop interactions (65, 66)
1997	Loop E	Continuous stack of non-Watson-Crick pairs, Mg ions (191)
1999	Hepatitis delta virus ribozyme	Highly constrained pseudoknot, continuous A-minor contacts (192)
2000	K-turn	Recurrent RNA modular unit based on non-WC pairs (193, 194)
2000	Ribosome	A goldmine of molecular contacts. Prevalence of A-minor contacts (193); A-minor recognition of codon/anticodon triplet (194), base-phosphate contacts (195), conservation of GoU pairs (196), multiple junctions (197)

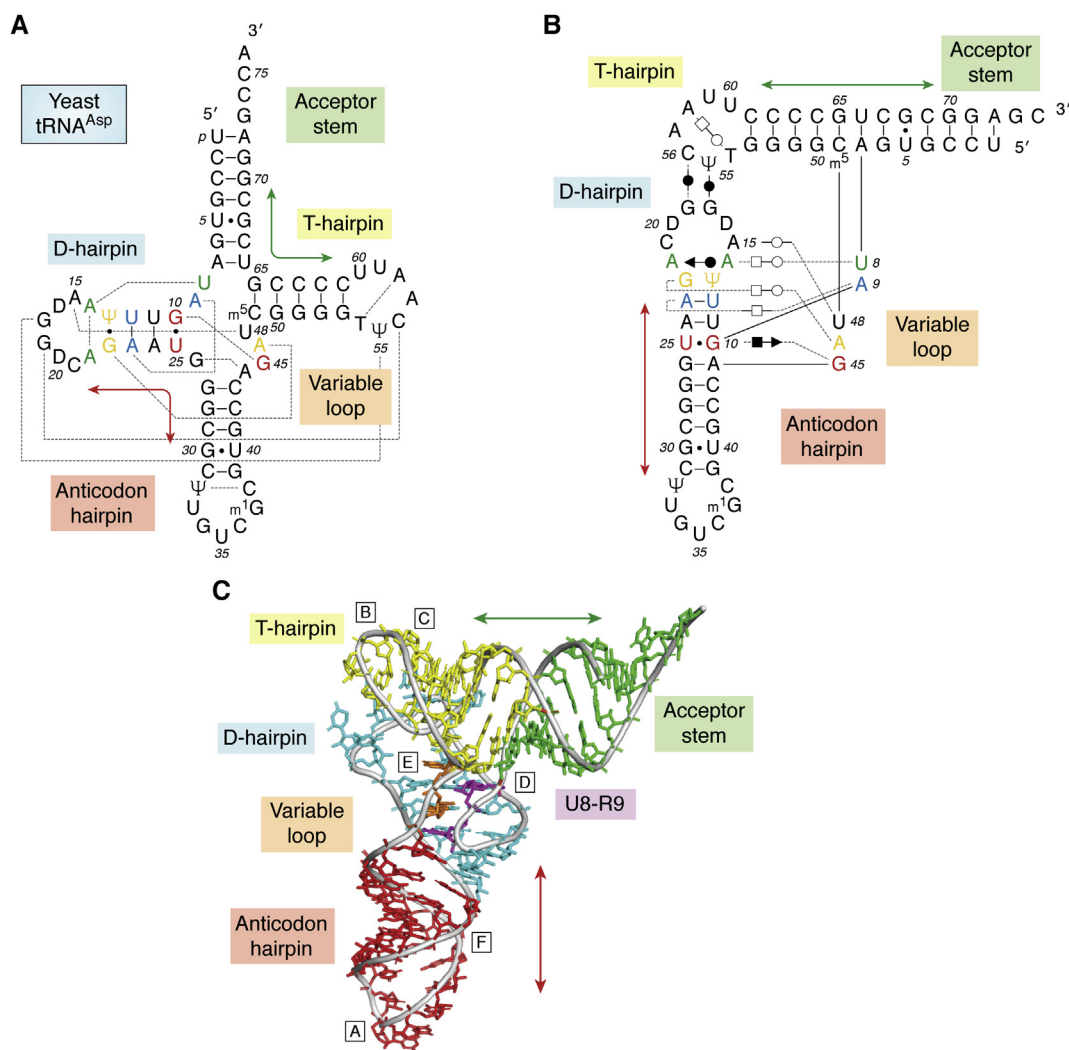
A GNRA tetraloop is a stretch of four nucleotides capping a hairpin and starting with a guanine (G), any of the four nucleotides (A, G, C, U, or N), a purine (A or G, R) and ending with an adenine (A). The GoU pairs were first suggested by Francis Crick (198) for explaining the degeneracy of the genetic code. In such a pair, a U pairs with a G instead of a C. GoU pairs have definite characteristics and play key roles in RNA structures and biology (for a recent overview, see (199)).

## Form and function

The quote from Wainwright (86) captures aptly the conundrum: “Structure without function is a corpse; function without structure is a ghost” (86). And it would take a lot more pages to do justice to what RNA structures have taught us about biology. Many RNA structures have been milestones and fueled our advances in understanding biochemical function and biological evolution. Forty years after the establishment by Tom Cech (87) and Sidney Altman (88) that all cells exploit RNA catalysis, we now have high-resolution structures of different catalytic RNAs in various stages (see Fig. 1) (89–93). We have also learned from ribosome structures that peptide bond catalysis during translation is performed in an RNA only environment (94–96). And the recent structures of the active complexes of the spliceosome (see Fig. 2) (97–100), together with the structural similarities with the autocatalytic group II introns (101, 102), indicate clearly also that the chemistry of splicing is performed by RNA elements. In those large RNP complexes, such as the ribosomes and the spliceosomes, the RNA elements are positioned precisely for function by the concomitant actions of RNA folding and protein complex formation. In the ribosome field, technological development in cryo-electron microscopy has led to stunning increases in resolution; for example, we now can delve inside ribosomal structures at 2 Å resolution (103, 104). Further, active ribosomes are assembled through long and convoluted maturation processes. Structures are now appearing showing various steps in the maturation of ribosomes with changes in RNA base pairing and structures and with binding of maturation cofactors absent in the final assembled ribosomes (105–108). While the ribosome performs the programmed steps in translation (initiation, translocation, termination) essentially as a dynamical single complex entity, this is not the case for the spliceosome. Driven by several splicing or regulatory factors and RNA-dependent ATPase/helicases, the RNA and protein composition as well as the structures of the spliceosomal

complex change along the catalytic steps, from the recognition of the 5' and 3' splice junctions and the branching point leading to the B\* complex ready for the first step of splicing (cleavage at 5' splice site) to the C\* complex ready for the second step of splicing, the ligation of the exons, and the disassembly of the spliceosome for initiating a new cycle (98, 99). As shown in Figure 2, we now can visualize many of the states along the spliceosomal cycle and derive illuminating movies after decades of sustained efforts by several groups around the world.

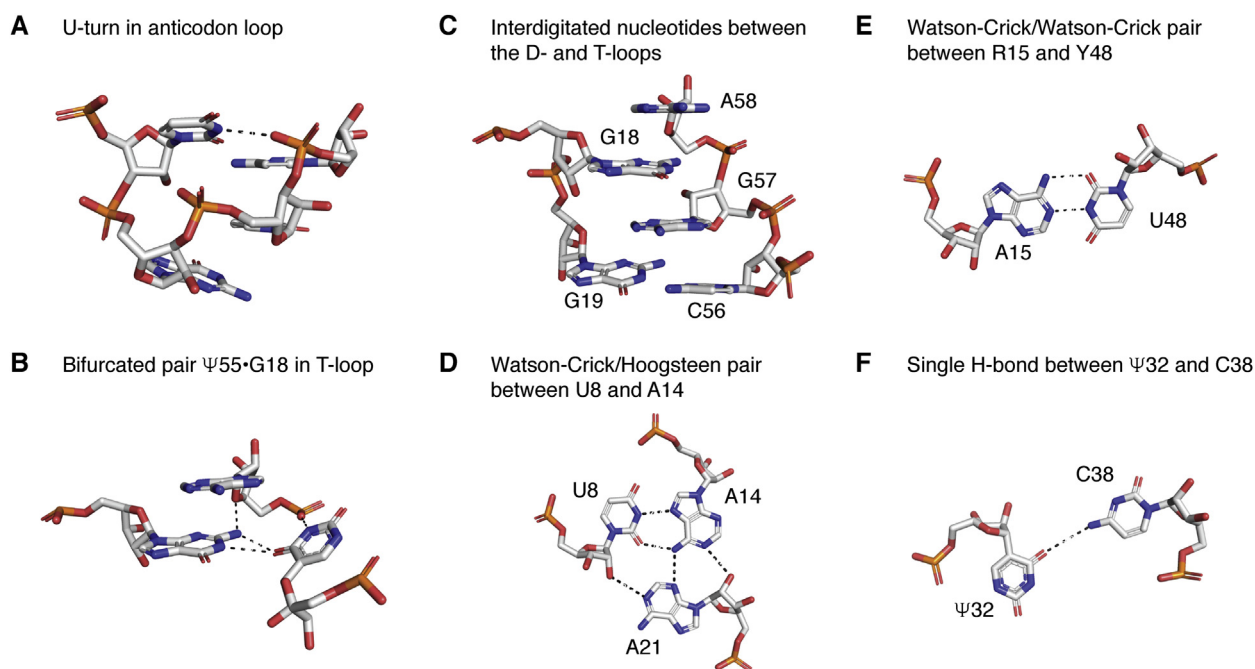
One should always keep in mind the famous quote by T. Dobzhansky: “Nothing in Biology Makes Sense except in the Light of Evolution” (109). The accumulated RNA structures offer us amazing insights into the principles of molecular evolution underlying biological evolution. Powerful techniques in molecular evolution have demonstrated that starting from a random sequence one can isolate sequences specific for binding a given ligand or with a defined function (110–112). The structures of many aptamers are in the PDB now and new ones are still being selected and crystallized, such as the RNA fluorophores (113–115). The SELEX experiments led to the discovery of riboswitches in bacteria (116). Riboswitches are RNA sequences that, during transcription in the presence of a given ligand, fold into a native structure different from the one obtained after transcription in the absence of the same ligand (117–119). The alternative folds exert different action on the production of metabolic enzymes through transcription termination (or anti-termination) or inhibition (or not) of ribosome binding for initiation of translation. There are now many structures of riboswitches with and without ligand bound (117, 120). Amazingly, the ligand can be as small as a fluoride ion (121) or as large as vitamin B12 (122, 123). Furthermore, there are several families of sequences for the same or similar ligand; up to four classes of guanidine and SAM riboswitches have been identified (120, 124).



**Figure 3. Three representations of the interactions present between nucleotides in transfer RNA with increasing levels of structural complexity.** *A*, standard cloverleaf structure of yeast tRNA<sup>Asp</sup> (236). *B*, a two-dimensional view of the tertiary structure of yeast tRNA<sup>Asp</sup>, it follows the representation proposed by Kim (237) that stresses the two main arms made of helical stems, the acceptor-stem with the Thymine (T)-stem and of the Dihydrouridine (D)-stem with the anticodon stem. A stem capped by a loop is called a hairpin. The numbering follows that of yeast tRNA<sup>Phe</sup> and because the numbers of nucleotides are not the same in the D- and variable-loops, residues 17 and 47 are skipped and the residue following D20 is C20a. The representation clearly shows the contacts linking the T- and D-loops and the tertiary base pairs and triples between the single-stranded segments and the D-hairpin. The contacts represented correspond to those observed in the yeast tRNA<sup>Asp</sup> structure (238, 239). For characterizing the tertiary pairs, the following nomenclature is used (240). Nucleic acid bases can interact through three possible edges: the Watson-Crick edge, the Hoogsteen edge (the edge with N7 in purines or C5 in pyrimidines), and the sugar edge (O2 in pyridines or N3 and N2 in purines, with often the hydroxyl O2' of the ribose). The nucleotides can interact with the sugars on the same side of the H-bonds (like in normal Watson-Crick pairs) and the pair is called *cis*; or on opposite sides and the pair is called *trans*. The three symbols, *circle, square, triangle*, represent respectively the Watson-Crick, the Hoogsteen, and the sugar edges. When the pair is *cis*, the symbols are *dark* and, when in *trans*, they are *white*. This nomenclature applies to the large number of specific base-base interactions. Pairs form through single H-bond (see Fig. 4F) or bifurcated H-bonds (see Fig. 4B) are not easily annotated. *C*, the tertiary structure of yeast tRNA<sup>Asp</sup> with the four domains colored (*green*: acceptor stem; *yellow*: T-hairpin; *blue*: D-hairpin; *red*: anticodon hairpin; the two nucleotides U8 and R9 (generally a purine) linking the 5'-end acceptor strand to the D-strand is *magenta*; the variable loop linking the 3'-end of the anticodon hairpin to the 5'-end T-strand is *orange*). The *double arrows* (*green* and *red*) indicate the sets of helices that stack upon each other in a coaxial manner in the three-dimensional fold. Capital letters indicate the position of the contacts shown in Figure 4.

Beyond the comforting insights on the RNA world at the origins of life about 2.8 billion years ago, all these RNA structures, coupled with structural alignments of homologous sequences, offer us multiple “Rosetta stones” for deciphering the RNA molecular evolution principles (125). Prior to the visualization of three-dimensional RNA architectures, RNA sequence alignments played major roles (and they still continue to do so). Through the analysis of nucleotide positions covarying according to the Watson-Crick rules, one can deduce the secondary structure (*i.e.*, the ensemble of base-

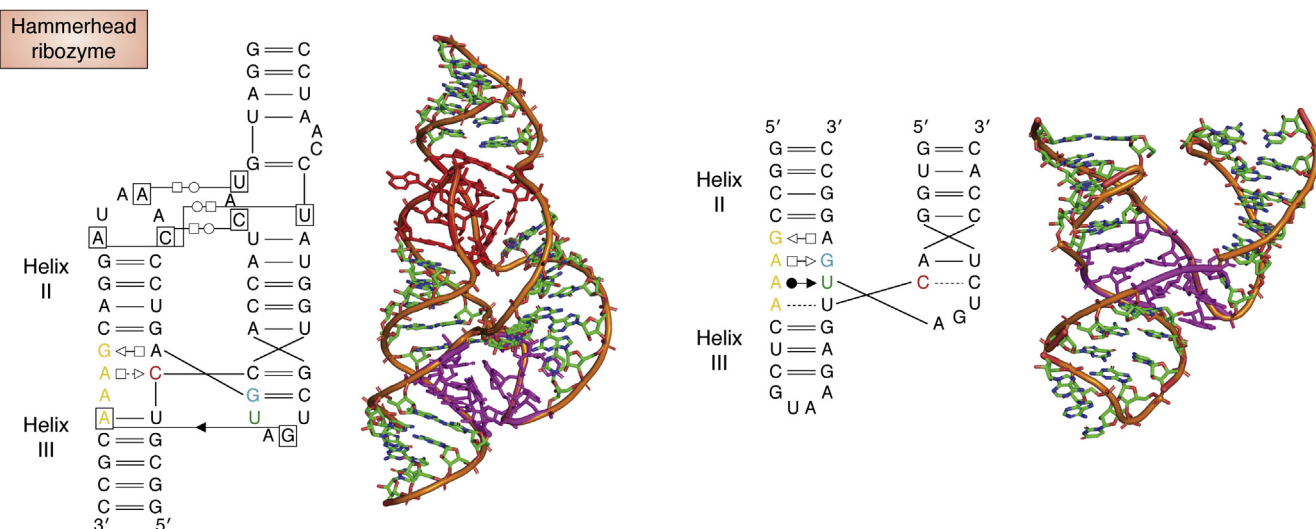
paired RNA helices) of a set of homologous RNA molecules. The third kingdom of life, the *Archaea*, were discovered by Carl Woese by sequencing ribosomal RNA and aligning them (126, 127). The bacteria are still identified according to the 16S rRNA sequences (128). The secondary structures of many functional structured RNAs were obtained by sequence alignments (129–134). The iconic cloverleaf structure of all cellular tRNAs was deduced with the first two sequences determined (37, 38). The availability of both RNA sequences and RNA structures allows structural alignment patterned on



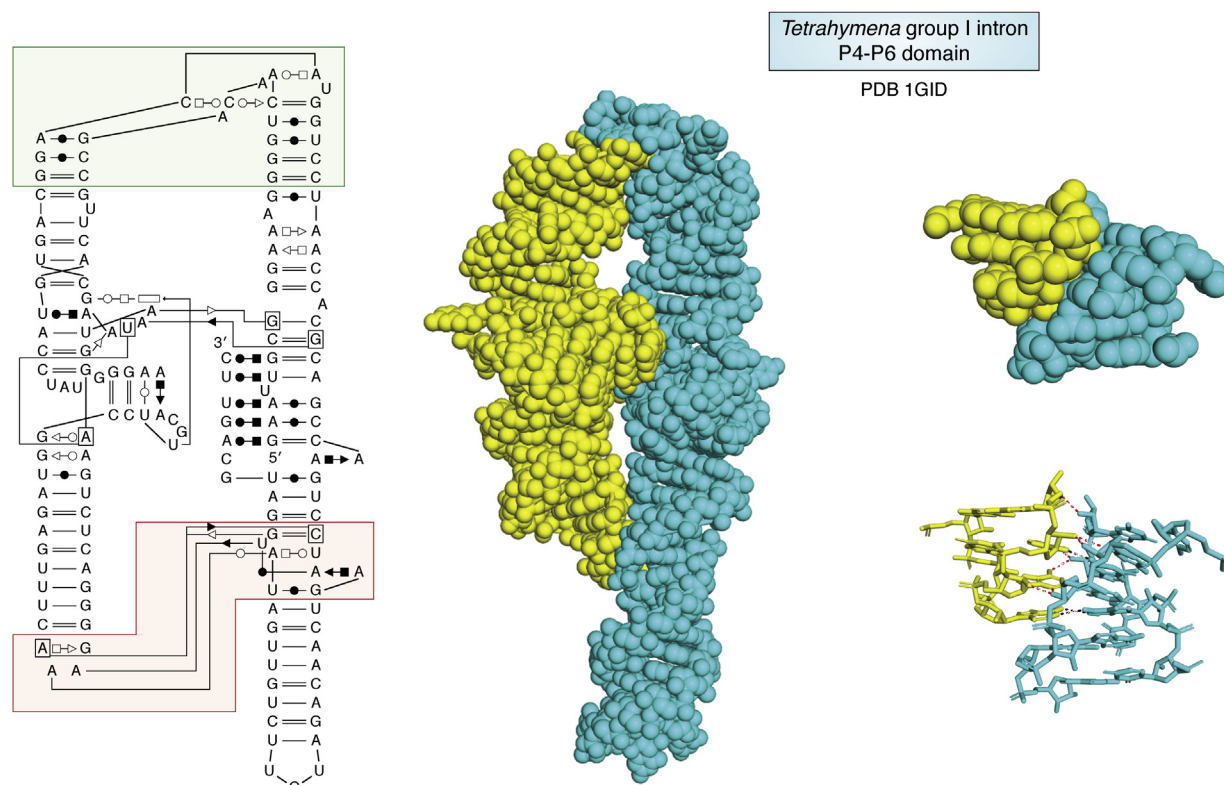
**Figure 4. Illustrations of contacts discussed in the text.** *A*, the U-turn after U33 in the anticodon loop: the torsion angle about P-O5' of the 5'-phosphate of residue 34 is *trans* (180°) instead of the usual *gauche-minus* (-60°); the 5'-phosphate of residue 35 stacks below U33; there is a H-bond between N3-H of U33 and an anionic phosphate oxygen of the 5' phosphate of residue 36. Thus, all three residues of the anticodon triplet have some interaction with the highly conserved U33 (in mammalian initiator tRNA<sup>Met</sup>, C33 occurs). *B*, the equivalent U-turn in the T-loop where the U is a pseudouridine (noted Ψ or Psi). The bifurcated pair Ψ55•G18 in which the O4(Ψ) interacts with both N1-H and N2-H of the guanine is also shown. The residue A58 stacks above G18. *C*, interdigitated nucleotides between the D- and T-loops. *D*, the highly conserved *trans* Watson-Crick/Hoogsteen pair between U8 and A14 forms three H-bonds with A21, also highly conserved. *E*, the famous *trans* Watson-Crick/Watson-Crick pair between R15 and Y48. Levitt (141) noted that residue 15 is always a purine and residue always a pyrimidine and modeled that pair as a regular *cis* Watson-Crick/Watson-Crick. Notice that, in standard nucleotide conformations, *trans* base pairs lead to parallel strands and not antiparallel strands as in usual helices (241). *F*, nucleotides 32 and 38 immediately adjacent to the last base pair of the anticodon stem generally present a single H-bond (for details (242)).

RNA architectures. These allow us to learn about how RNAs evolve, in other words how changes or mutations in homologous sequences maintain the RNA folded architecture. One can learn which variations are neutral and which molecular interactions are conserved through species. When several

homologous three-dimensional structures are available, one can learn also how molecular accommodations occur and how ions or water molecules compensate for the loss of some interaction. By analysis of sequence alignments coupled with structures, one can also visualize how molecular units can



**Figure 5. A native state of the hammerhead ribozyme (left drawing) that promotes cleavage at low magnesium concentrations (243) that is reached through intricate contacts (and variable depending on the type of ribozymes) between an apical loop and an internal loop (drawn in red color).** Without these tertiary contacts, a different core of the three-way junction (drawn in magenta color) is observed (compare the regions in magenta color). The nomenclature described in Figure 3B is used for the non-Watson-Crick pairs (240).



**Figure 6.** From left to right, a two-dimensional view of the P4–P6 domain with the nomenclature described in Figure 3B (240) and, with next to it, the whole molecule shown in space-filling mode highlighting the coaxial stacking of helices and their parallel packing. In the left two-dimensional view, the region boxed in green shows how the RNA is able to bend 180° and the regions boxed in red are shown in space-filling and atomic views on the right of the figure; these views show the precise and tight contact with the tetraloop GAAA and its receptor, called the 11-nt motif; notice that there are twice as many H-bonds between the hydroxyl groups (red dotted lines) than between the bases (black dotted lines). That type of RNA–RNA contacts were discovered through sequence analysis and SELEX experiments (244).

swap or interchange and determine which interactions are opportunistic and not critical for form and function. Integrated databases with sophisticated techniques for guaranteeing the interoperability of data will need to be developed to exploit fully the one-dimensional and the three-dimensional data. In complexes between RNA and proteins, the diversity and multiplicity of intermolecular contacts are enormously potentiated. The understanding of RNP formation and structure is a major challenge for years to come.

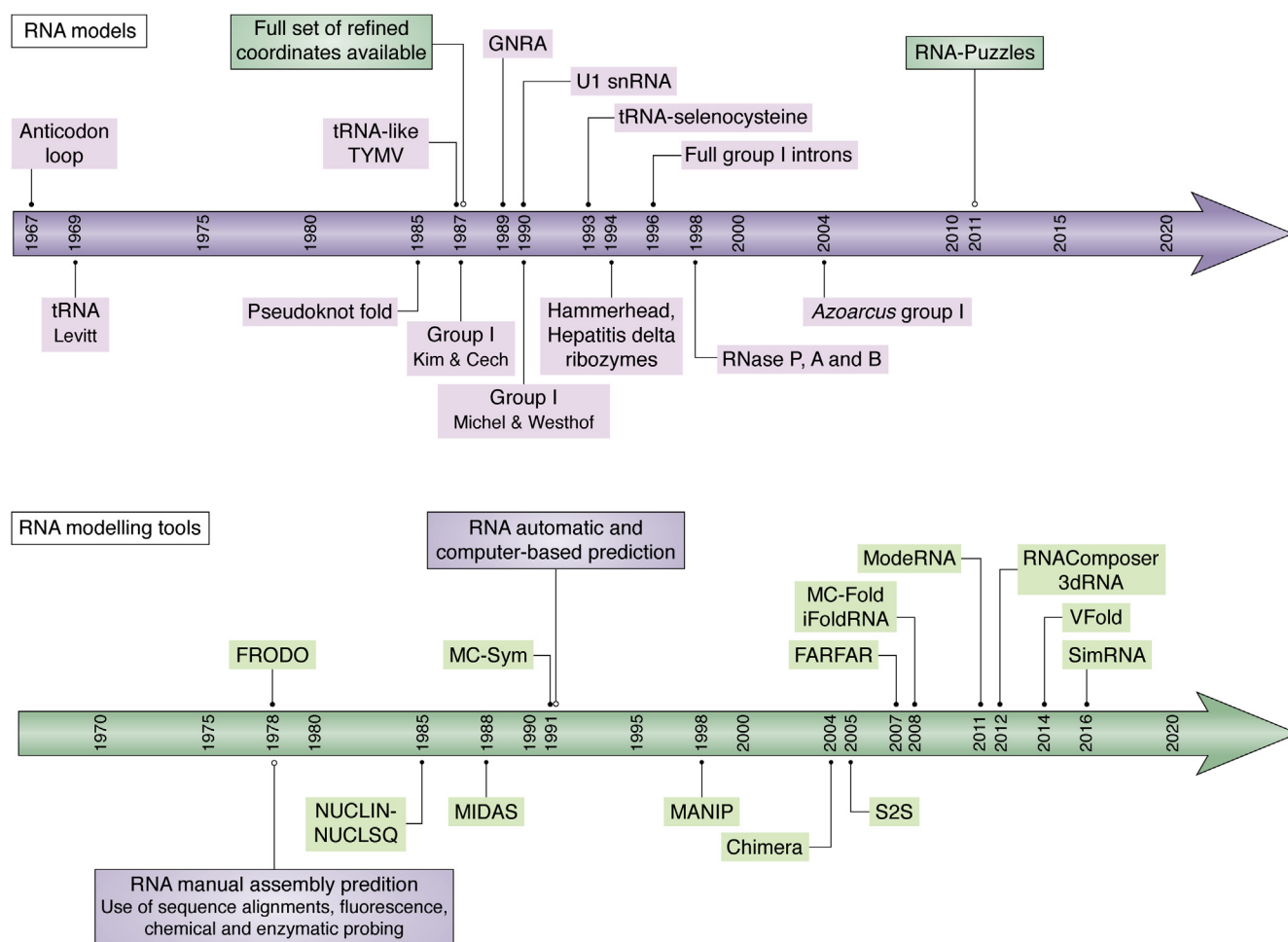
### A short history of RNA modeling and RNA assembly computational tools

In the 1980s, the PDB also accepted structures derived by computational modeling. One of us did deposit several structures of RNA modeled on the basis of sequence alignments and/or chemical and enzymatic probing in solution. While PDB may have accepted these submissions somewhat reluctantly, many of these modeled structures were of intense interest to the growing crowd of RNA aficionados, and their accessibility stimulated experimentalists to put them to the test. After presenting the modeling of the core of group I introns (135) at *The 49th Pittsburgh Diffraction Conference*, Columbus, in November 1991, there was almost a riot in the room with half of the participants arguing for and the other half against the validity of the procedure. These modeled structures can no longer be found on the PDB but can be

retrieved from the following website <https://eric-westhof.ibmc.cnrs.fr/>. Interestingly, wwPDB recently held a workshop on data standards for integrative structural models (136), and *Protein Data Bank Japan* (PDBj) recently launched the *Biological Structure Model Archive* (BSM-Arc) (137).

Figure 7 (top) shows a timeline of selected RNA models that have been published. It starts with the modeling work of Fuller and Hodgson (138) on the anticodon loop of tRNA. They manually assembled space filling (139) and Kendrew (140) models. They stated very clearly the objectives that are still of value today: “We do not necessarily believe that our models describe the actual molecular conformations to an accuracy of a few hundredths of an angstrom, but the analysis shows that a model with the general characteristics we propose can be built with acceptable stereochemistry. Only if model building is treated as a rigid discipline with strict attention paid to detailed stereochemistry can the results of a study such as this be considered reliable and meaningful.” They did identify and rationalize the correct stacking of the five nucleotides on the 3'-end of the loop with the 3'-strand of the anticodon helix (called 3'-stack). Michael Levitt assembled in 1969 a whole tRNA (141). For that model, the available sequences were used and analyzed. The choice of the stacked arms followed previous small-angle X-ray scattering (142) (Fig. 3, A–C). Interestingly, contacts between the D- and T-loops are proposed as well as base pairs between residues 8 and 14 as well as 15 and





**Figure 7. Top, a timeline of RNA models.** Before 1987, no sets of coordinates for the suggested models were made available. With the development of modeling tools based on computer graphics, one could derive coordinates (without manually building physical models) and refine them (see bottom). Along time, the following RNA models are highlighted: (1) the anticodon loop (138); (2) the Levitt tRNA (141); (3) the pseudoknot fold (143); (4) the Kim & Cech core of group I intron (144); (5) the tRNA-like in TYMV (245); (6) the GNRA loop in 5S rRNA (246); (7) the Michel & Westhof core of group I intron (135); (7) the 4-way junction of U1 snRNA (188); (8) the tRNA selenocysteine (247); (9) the hammerhead (248) and the hepatitis delta (249) ribozymes; (10) full group I introns (250); (11) A and B families of RNase P (251); (12) the *Azoarcus* group I intron (252). Many RNA structures were also modeled afterward, especially within the RNA-Puzzles Consortium (155–158). **Bottom**, a timeline of some RNA assembly and computing tools. The most recent ones are regularly used and actively improved. (1) FRODO developed by Alwyn Jones was a pioneering tool in computer molecular graphics (151); (2) NUCLIN-NUCLSQ (159), an inclusive refinement program dedicated to nucleic acids and based on Hendrickson–Konnert PROLSQ (253, 254); (3) MIDAS (255, 256); (4) MC-Sym (257); (5) MANIP (258); (6) Chimera (259); (7) S2S (260, 261); (8) FARFAR (262); (9) MC-Fold (263), iFoldRNA (264); (10) ModeRNA (265); (11) RNAComposer 3dRNA (267); (12) VFold (268); (13) SimRNA (269).

48 (now sometimes called the Levitt pair), but both in the usual Watson–Crick configurations instead of the observed *trans* configuration (Fig. 4, D and E). Sequence analysis coupled with model building led to the suggestion that some RNAs can form pseudoknots, for example, between a hairpin loop and a single strand (meaning that the single strand does not go through the loop) (143). This fold, now on the logo of the International RNA Society (<https://www.rnasociety.org/>), is present in a large number of functional structured RNAs. In 1987, Kim and Cech (144) published a model of the core of group I intron on the basis of the available experimental data and with clear model building principles that are still valid. They are worth restating: “(i) RNA duplexes were assumed to have A-form RNA helix conformation. (ii) If two duplex stems were separated by fewer than three unpaired nucleotides, they were stacked collinearly. (iii) If one helix competed with two others for collinear stacking, the two helices separated by the

least number of unpaired nucleotides were chosen to form a stacked helix. (iv) Non-Watson-Crick base-pairing was allowed at the junction of two helices. (v) Single “bulged” bases were stacked within a helix. (vi) If two conserved bases in a single-stranded region were in proximity, base-pairing was attempted subject to the constraints of the chemical-modification data. (vii) Loop conformations were taken from those in the tRNA structure.” Kim had built the tRNA structure in electron density in Alex Rich’s laboratory (145, 146) some years before. The coaxial stacks P3–P7–P8 and P4–P6 were identified (144). Afterward, with the development of molecular graphics and refinement programs, full molecular models with coordinates could be generated (Fig. 7).

Figure 7 (bottom) shows the evolution of the computer tools necessary for the manipulation and refinement of RNA structures. The advent of computer graphics and the development of systems allowing the display and the manipulation

of molecular objects in real time, pioneered by Levinthal (147), Feldmann (148), Langridge (149, 150), and Jones (151), were breakthroughs in structural biology. Several of these tools were developed for crystallography and modeling purposes as well. Model building is the process in crystallography (or cryo-EM) in which one actually assembles a molecular model by fitting molecular objects into electron density (manually either with physical objects in a Richard's box or with computer graphics, a process now almost automatically done). The model building is guided and constrained by the electron density. The molecular models are then refined geometrically and stereochemically so as to satisfy best the crystallographic data or the electron density. In *ab initio* modeling, the molecular objects are built without the constraints of crystallographic data. The model building can be driven by geometric and physico-chemical energetic terms in (semi-) automatic fashion by various computing tools (for RNA, see the recent reviews (152, 153)). Such tools should integrate the knowledge, accumulated at a given time, on the molecular systems to be modeled. This knowledge integration is still a formidable challenge owing to the diversity of molecular interactions to consider and treat simultaneously (weak forces, water molecules, ions, etc.). When performed manually with physical objects, model building is still very powerful to assimilate and understand the rules underlying the folding of macromolecules. Richard Hamming said, "The purpose of computing is insight, not numbers." Similarly, the purpose of modeling is insight, not models. One can also recall the famous quote from Richard Feynman: "What I cannot create, I do not understand."

The point of archiving 3D models is to provide a record of previous modeling efforts that can help to benchmark methodological progress in computational modeling and to develop metrics to assess contemporary modeling efforts (see (154)). Toward that aim, the *RNA-Puzzles* computational contests were created in 2011 (155–158). The *RNA-Puzzles* consortium is a community-wide assessment of RNA 3D structure prediction that aims to expose the bottlenecks in current RNA 3D structure prediction and promote the improvement of these prediction methods. *RNA-Puzzles* attempts to provide the RNA modeling community what CASP (*Critical Assessment of Methods of Protein Structure Prediction*) has been doing for protein modeling (159). All the data sets for the modeled structures and codes for assessment are now available as open source on GitHub (<https://github.com/RNA-Puzzles>) (160).

### Databases play central roles in science

In 2003, three organizations, the *Research Collaboratory for Structural Bioinformatics* (RCSB) in the United States, the *Macromolecular Structure Database* (MSD) at EBI, and the PDBj in Osaka, created the worldwide PDB (wwPDB; <http://www.wwpdb.org/>) with the goal of maintaining a single archive of structure data of macromolecules that is freely and publicly available to the global community (161). The Biological Magnetic Resonance Data Bank is now also part of wwPDB, and soon the *China Protein Data Bank* will join. This structure also facilitates unified deposition, curation, and

distribution of structures to the growing global community of structural biologists and scientists that use 3D data (161). At the same time, each independent center is free to innovate new visualization and analysis tools to promote the reuse of the data in creative and useful ways for nonspecialists. Each center is also engaged in creating educational materials to make structural biology accessible to students at all levels, for example, PDB101's *Molecule of the Month* and the PDBj's *Encyclopedia of Protein Structures* (162). A number of the *Molecule of the Month* entries are RNA or DNA structures, and several others are proteins that interact with nucleic acids.

We would like to recall here a sentence written, in 1990, by Brändén and Jones (163): "It is the crystallographer's responsibility to make sure that incorrect protein structures do not reach the literature." They could as well have written: "It is the structural biologist's responsibility to make sure that incorrect macromolecular structures do not reach the literature," which would encompass also NMR and cryo-EM technologies as well as nucleic acids and their complexes. PDB continues to adapt in the new century by adopting new data formats to accommodate the much larger structures being solved. The development of metrics for assessing structure validity both globally and locally at the nucleotide level is proving key to understanding the significance of structural variation and mapping out the dynamic behavior of nucleic acid structures in response to interactions with small molecules, proteins, and other nucleic acids. Cryo-EM is rapidly displacing X-ray crystallography for very large complexes and molecular machines that cycle through multiple functional states and are hard to crystallize in unique states. New metrics for cryo-EM are in progress and should soon be available (164). In this context, referees and journal editors have also a major role to play, to prevent incorrect structures from reaching the literature and worse, the archival databases. Referees are encouraged to request (and in many cases are actually provided with) complete and detailed statistics tables, validation reports and quality indicators, coordinates, and electron-density maps. Journal editors, as well as authors, should comply with such requests despite fierce competition for publication. The PDB has played an important role by providing sophisticated metrics and automation that provides rapid feedback to contributors of data when they submit structures.

Databases constitute an absolute necessity for advancing science, to facilitate the work of scientists in the same and related disciplines. They are repositories of observational data organized with carefully curated dictionaries and controlled vocabularies upon which future science can build and develop. In the immediate future, there are at least two main challenges to which we would like to draw attention. Owing to the sizes and complexities of modern databases, these two challenges are interconnected and interdependent.

First, the sustainability and maintenance of databases for the use of a now internationalized community of scientists raise the issue of how the associated costs shall be equitably apportioned; these are recurrent costs to support trained and competent personnel, together with the maintenance and

constant upgrading of the extensive infrastructure required to keep up with the growth in the size and complexity of the data. Short-term grants with a dedicated focus are not adapted to maintain large international resources (164). New business models are being explored and promoted (165). The *Global Biodata Coalition* is an initiative started by the *International Human Frontier Science Program Organization* that according to their website (<https://globalbiodata.org>) aims “to stabilize and ensure sustainable financial support for the global biodata infrastructure and in particular to identify for prioritized long-term support a set of *Global Core Data Resources* that are crucial for sustaining the broader biodata infrastructure.”

And, secondly, databases fundamentally comprise well-organized archives of chosen sets of objects. In structural biology, the objects are mainly molecular sequences and structures. However, as these two types of data attach to the same natural entities, they need to be integrated to grasp biological function and evolution more holistically. New developments in cryo-electron tomography (166), which allows visualization of macromolecules *in situ*, will open up visualization of macromolecules in their cellular environment. Maybe one day we will also include digital descriptions of phenotypes, including bones, skeletons, or wings in order to relate their forms and colors to the underlying genes and genetic networks. Launched by the *RNAcentral Consortium* in 2014 at the *EMBL-EBI, Wellcome Genome Campus* (Hinxton, UK), *RNAcentral* (<https://rnacentral.org>) integrates and unifies access to all types of noncoding RNA sequences from all organisms. This is a major enterprise that goes in the right direction for the integration and interoperability of data. A bridge between sequences and standard secondary structures has been achieved recently (167). In the future this bridging integration should be extended to three-dimensional structures and thus the PDB and the NDB.

### In the light of evolution

It would take a whole book with several dozens of figures to convey the amazing knowledge accumulated on RNA structures and architectures since the beginning of RNA structural biology and the advent of the PDB. Only glimpses of RNA structural biology are presented here, and we would like to apologize for not mentioning the contributions of many structural biologists, especially those using NMR spectroscopy, a field that contributed several seminal structures, for example, on RNPs and telomerase (168–172), with unique critical data on RNA dynamics (173–177). Almost each RNA structure still brings some surprising and unexpected features, and these unveil the constraints of the physicochemical interactions and the congruent accommodations due to the historical contingencies of biological evolution. If we can cite again T. Dobzhansky (109): “Seen in the light of evolution, biology is, perhaps, intellectually the most satisfying and inspiring science. Without that light it becomes a pile of sundry facts, some of them interesting or curious but making no meaningful picture as a whole.” A quote from Ernest Rutherford states “All science is either physics or stamp collecting,” which is in sharp

contrast to the lifelong accumulation of observations and data by the biologist Charles Darwin. The RNA structures in the PDB are absolutely not a pile of miscellaneous items, “sundry facts”; they are physical objects, which, together with genomic data, allow us to integrate present-day biological functions and the historical evolution in all living species on earth.

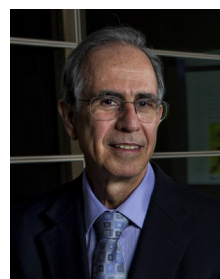
Technological improvements and breakthroughs are at the heart of discovery and scientific progress. One may recall the famous quote from Sydney Brenner “progress in science depends on new techniques, new discoveries and new ideas, probably in that order” (178). As alluded above, the recent and constant progress in cryo-EM techniques is totally changing structural biology (179). The protein prediction field is now overwhelmed by the achievements of artificial intelligence tools (180–182). For those raised with the warning of the Levinthal’s paradox (183), we admire in awe the progress and breakthroughs. Could such tools be applied to RNA structure prediction? Since they are based on the analysis of hundreds of structures, do we have enough RNA structures sufficiently nonredundant? As remarked by Jane Richardson (184), among the “1491 RNA-only structures, just 373 of which have a chain  $\geq 60$  nucleotides” and of those a large number consists of ribosomal structures with various ligands or from various biological sources. As appropriately reminded by John Helliwell (185), in the end, these breakthroughs would not have been feasible without the PDB.

*Acknowledgments*—We wish to thank the numerous scientists, engineers, technicians who have been involved since the origins in the development and maintenance of the PDB. We gratefully acknowledge also the funding bodies that offered to the scientific community worldwide this essential component of its everyday life.

*Author contributions*—E. W. and N. B. L. wrote the article, did the bibliographic research, and made first drafts of the drawings and figures.

*Conflict of interest*—The authors declare that they have no conflicts of interest with the contents of this article.

### Dedication



Neocles B. Leontis  
(1955–2020)

During the writing of this work, on December 8, 2020, **Neocles B. Leontis** died in a car accident (186). Together with the editors, we would like to dedicate this article to his memory. I have been fortunate to exchange and collaborate with Neocles for more than 20 years. Beyond the many scientific papers published together, always occasions of in-depth, dynamic, and always fair discussions, Neocles was a friend intensely human, joyous, and positive with whom it was recomforting to be. I admired him deeply for his move toward local politics (in 2019, he was elected to the City Council of Bowling Green, Ohio) and his involvement in many social and climate issues. He applied his intellectual power to these issues as thoroughly and seriously as he did to the various scientific problems he tackled. His untimely and brutal death is a loss to the local community around his University and town, but also to all the RNA scientists around the world who use the

concepts, tools database he contributed to develop. Neocles earned his PhD from Yale with Peter Moore as mentor. He became professor of Chemistry at Bowling Green State University in 1987. With Peter Moore, he worked using NMR on the 5S rRNA and showed the key role of Mg ions and the presence of non-Watson–Crick pairs. Later using sequence comparisons, he could identify loop E modules in RNA. Together we published a nomenclature for classifying logically and nonambiguously non-Watson–Crick. Neocles Leontis and his group at Bowling Green developed many websites and tools for analyzing and compiling RNA structures.

**Abbreviations**—The abbreviations used are: PDB, Protein Data Bank; PDBj, PDB Japan; PDP, Programmed Data Processor; NDB, Nucleic Acid Database; RNP, RNA–protein complex; wwPDB, worldwide PDB.

## References

- Sundaralingam, M., and Jensen, L. H. (1965) Stereochemistry of nucleic acid constituents: I. Refinement of the structure of cytidylic acid b. *J. Mol. Biol.* **13**, 914–929
- Kennard, O., Speakman, J. C., and Donnay, J. D. H. (1967) Primary crystallographic data. *Acta Cryst.* **22**, 445–449
- Rubin, J., Brennan, T., and Sundaralingam, M. (1971) Crystal structure of a naturally occurring dinucleoside monophosphate: Uridyl (3',5') adenosine hemihydrate. *Science* **174**, 1020–1022
- Seeman, N. C., Sussman, J. L., Berman, H. N., and Kim, S. H. (1971) Nucleic acid conformation: Crystal structure of a naturally occurring dinucleoside phosphate (UpA). *Nat. New Biol.* **233**, 90–92
- Kennard, O. A., Brice, F. H., Hummelink, M. D., Motherwell, T. W. A., Roidgers, W. D. S., Watson, J. R., and D.G.. (1977) Computer based systems for the retrieval of data: Crystallography. *Pure Appl. Chem.* **49**, 1807–1816
- Crystallography: Protein Data Bank. *Nat. New Biol.* **233**, (1971), 223
- Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977) The Protein Data Bank. A computer-based archival file for macromolecular structures. *Eur. J. Biochem.* **80**, 319–324
- Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S. H., Srinivasan, A. R., and Schneider, B. (1992) The nucleic acid database. A comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.* **63**, 751–759
- Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., 3rd, Snoeyink, J., Richardson, J. S., and Richardson, D. C. (2007) MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383
- Chou, F. C., Sripakdeevong, P., Dibrov, S. M., Hermann, T., and Das, R. (2013) Correcting pervasive errors in RNA crystallography through enumerative structure prediction. *Nat. Methods* **10**, 74–76
- Read, R. J., Adams, P. D., Arendall, W. B., 3rd, Brunger, A. T., Emsley, P., Joosten, R. P., Kleywegt, G. J., Krissinel, E. B., Lutheke, T., Otwinowski, Z., Perrakis, A., Richardson, J. S., Sheffler, W. H., Smith, J. L., Tickle, I. J., et al. (2011) A new generation of crystallographic validation tools for the protein data bank. *Structure* **19**, 1395–1412
- Petrov, A. I., Zirbel, C. L., and Leontis, N. B. (2011) WebFR3D—a server for finding, aligning and analyzing recurrent RNA 3D motifs. *Nucleic Acids Res.* **39**, W50–W55
- Coimbatore Narayanan, B., Westbrook, J., Ghosh, S., Petrov, A. I., Sweeney, B., Zirbel, C. L., Leontis, N. B., and Berman, H. M. (2014) The nucleic acid database: New features and capabilities. *Nucleic Acids Res.* **42**, D114–D122
- Altona, C., and Sundaralingam, M. (1972) Conformational analysis of the sugar ring in nucleosides and nucleotides. A new description using the concept of pseudorotation. *J. Am. Chem. Soc.* **94**, 8205–8212
- Olson, W. K., and Flory, P. J. (1972) Spatial configurations of polynucleotide chains. 3. Polydeoxyribonucleotides. *Biopolymers* **11**, 57–66
- Olson, W. K., and Flory, P. J. (1972) Spatial configurations of polynucleotide chains. I. Steric interactions in polyribonucleotides: A virtual bond model. *Biopolymers* **11**, 1–23
- Olson, W. K., and Flory, P. J. (1972) Spatial configuration of polynucleotide chains. II. Conformational energies and the average dimensions of polyribonucleotides. *Biopolymers* **11**, 25–56
- Ramakrishnan, C., and Ramachandran, G. N. (1965) Stereochemical criteria for polypeptide and protein chain conformations. II. Allowed conformations for a pair of peptide units. *Biophys. J.* **5**, 909–933
- Sundaralingam, M. (1969) Stereochemistry of nucleic acids and their constituents.† IV. Allowed and preferred conformations of nucleosides, nucleoside mono-, di-, tri-, tetraphosphates, nucleic acids and polynucleotides. *Biopolymers* **7**, 821–860
- Yathindra, N., and Sundaralingam, M. (1976) Analysis of the possible helical structures of nucleic acids and polynucleotides. Application of (n-h) plots. *Nucleic Acids Res.* **3**, 729–747
- Petrov, A. I., Zirbel, C. L., and Leontis, N. B. (2013) Automated classification of RNA 3D motifs and the RNA 3D motif atlas. *RNA* **19**, 1327–1340
- Gore, S., Sanz Garcia, E., Hendrickx, P. M. S., Gutmanas, A., Westbrook, J. D., Yang, H., Feng, Z., Baskaran, K., Berrisford, J. M., Hudson, B. P., Ikegawa, Y., Kobayashi, N., Lawson, C. L., Mading, S., Mak, L., et al. (2017) Validation of structures in the Protein Data Bank. *Structure* **25**, 1916–1927
- Wuthrich, K. (1991) *NMR of Proteins and Nucleic Acids*, Wiley, New York, NY
- Allain, F. H., and Varani, G. (1997) How accurately and precisely can RNA structure be determined by NMR? *J. Mol. Biol.* **267**, 338–351
- Lawson, C. L., Berman, H. M., and Chiu, W. (2020) Evolving data standards for cryo-EM structures. *Struct. Dyn.* **7**, 014701
- Wang, A. J., Quigley, G. J., Kolpak, F. J., van der Marel, G., van Boom, J. H., and Rich, A. (1981) Left-handed double helical DNA: Variations in the backbone conformation. *Science* **211**, 171–176
- Drew, H. R., Wing, R. M., Takano, T., Broka, C., Tanaka, S., Itakura, K., and Dickerson, R. E. (1981) Structure of a B-DNA dodecamer: Conformation and dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **78**, 2179–2183
- Wang, A. H., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G., and Rich, A. (1979) Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature* **282**, 680–686
- Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K., and Dickerson, R. E. (1980) Crystal structure analysis of a complete turn of B-DNA. *Nature* **287**, 755–758
- Clerici, L., Campagnari, F., de Rooij, J. F., and van Boom, J. H. (1979) Preparation of polydeoxynucleotides linked to a solid support by coupling CNBr-activated cellulose with 5'-NH<sub>2</sub>-terminated oligo and poly(pdT)'s. *Nucleic Acids Res.* **6**, 247–258
- Itakura, K., Katagiri, N., Bahl, C. P., Wightman, R. H., and Narang, S. A. (1975) Improved triester approach for the synthesis of pentadecathymidylic acid. *J. Am. Chem. Soc.* **97**, 7327–7332
- Clark, B. F. C., Doctor, B. P., Holmes, K. C., Klug, A., Marcker, K. A., Morris, S. J., and Paradies, H. H. (1968) Crystallization of transfer RNA. *Nature* **219**, 1222–1224
- Kim, S. H., and Rich, A. (1968) Single crystals of transfer RNA: An x-ray diffraction study. *Science* **162**, 1381–1384
- Hampel, A., Labanauskas, M., Connors, P. G., Kirkegard, L., RajBhandary, U. L., Sigler, P. B., and Bock, R. M. (1968) Single crystals of transfer RNA from formylmethionine and phenylalanine transfer RNA's. *Science* **162**, 1384–1387
- Hingerty, B., Brown, R. S., and Jack, A. (1978) Further refinement of the structure of yeast tRNA<sup>Phe</sup>. *J. Mol. Biol.* **124**, 523–534
- Sussman, J. L., Holbrook, S. R., Warrant, R. W., Church, G. M., and Kim, S. H. (1978) Crystal structure of yeast phenylalanine transfer RNA. I. Crystallographic refinement. *J. Mol. Biol.* **123**, 607–630
- Holley, R. W., Appgar, J., Everett, G. A., Madison, J. T., Marquisee, M., Merrill, S. H., Penswick, J. R., and Zamir, A. (1965) Structure of a ribonucleic acid. *Science* **147**, 1462–1465
- Madison, J. T., Everett, G. A., and Kung, H. (1966) Nucleotide sequence of a yeast tyrosine transfer RNA. *Science* **153**, 531–534

39. Chen, Z. G., Stauffacher, C., Li, Y., Schmidt, T., Bomu, W., Kamer, G., Shanks, M., Lomonosoff, G., and Johnson, J. E. (1989) Protein-RNA interactions in an icosahedral virus at 3.0 Å resolution. *Science* **245**, 154–159
40. Namba, K., Pattanayek, R., and Stubbs, G. (1989) Visualization of protein-nucleic acid interactions in a virus. Refined structure of intact tobacco mosaic virus at 2.9 Å resolution by X-ray fiber diffraction. *J. Mol. Biol.* **208**, 307–325
41. Hayes, J. A., Brunden, M. J., Gilham, P. T., and Gough, G. R. (1985) High-yield synthesis of oligoribonucleotides using o-nitrobenzyl protection of 2'-hydroxyls. *Tetrahedron Lett.* **26**, 2407–2410
42. Dock-Bregeon, A. C., Chevrier, B., Podjarny, A., Moras, D., deBear, J. S., Gough, G. R., Gilham, P. T., and Johnson, J. E. (1988) High resolution structure of the RNA duplex [U(U-A)6A]<sub>2</sub>. *Nature* **335**, 375–378
43. Dock-Bregeon, A. C., Chevrier, B., Podjarny, A., Johnson, J., de Bear, J. S., Gough, G. R., Gilham, P. T., and Moras, D. (1989) Crystallographic structure of an RNA helix: [U(UA)6A]<sub>2</sub>. *J. Mol. Biol.* **209**, 459–474
44. Heus, H. A., and Pardi, A. (1991) Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science* **253**, 191–194
45. Gewirth, D. T., Abo, S. R., Leontis, N. B., and Moore, P. B. (1987) Secondary structure of 5S RNA: NMR experiments on RNA molecules partially labeled with nitrogen-15. *Biochemistry* **26**, 5213–5220
46. Leontis, N. B., and Moore, P. B. (1986) Imino proton exchange in the 5S RNA of *Escherichia coli* and its complex with protein L25 at 490 MHz. *Biochemistry* **25**, 5736–5744
47. Woese, C. R., Winker, S., and Gutell, R. R. (1990) Architecture of ribosomal RNA: Constraints on the sequence of “tetra-loops”. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 8467–8471
48. Caruthers, M. H. (2013) The chemical synthesis of DNA/RNA: Our gift to science. *J. Biol. Chem.* **288**, 1420–1427
49. Scaringe, S. A. (2000) Advanced 5'-silyl-2'-orthoester approach to RNA oligonucleotide synthesis. *Methods Enzymol.* **317**, 3–18
50. Milligan, J. F., Groebe, D. R., Witherell, G. W., and Uhlenbeck, O. C. (1987) Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res.* **15**, 8783–8798
51. Anderson, A. C., Scaringe, S. A., Earp, B. E., and Frederick, C. A. (1996) HPLC purification of RNA for crystallography and NMR. *RNA* **2**, 110–117
52. Greenhough, T. J., and Helliwell, J. R. (1983) The uses of synchrotron X-radiation in the crystallography of molecular biology. *Prog. Biophys. Mol. Biol.* **41**, 67–123
53. Hope, H., Frolow, F., von Bohlen, K., Makowski, I., Kratky, C., Halfon, Y., Danz, H., Webster, P., Bartels, K. S., Wittmann, H. G., and Yonath, A. (1989) Cryocrystallography of ribosomal particles. *Acta Crystallogr. B* **45**(Pt 2), 190–199
54. Giege, R., Lorber, B., and Theobald-Dietrich, A. (1994) Crystallography of biological macromolecules: Facts and perspectives. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 339–350
55. Giege, R. (2017) What macromolecular crystallography tells us - what is needed in the future. *IUCr* **4**, 340–349
56. Meyer, M., and Masquida, B. (2014) cis-Acting 5' hammerhead ribozyme optimization for *in vitro* transcription of highly structured RNAs. *Methods Mol. Biol.* **1086**, 21–40
57. Ferre-D'Amare, A. R., and Doudna, J. A. (1996) Use of cis- and trans-ribozymes to remove 5' and 3' heterogeneities from milligrams of *in vitro* transcribed RNA. *Nucleic Acids Res.* **24**, 977–978
58. Ferre-D'Amare, A. R., Zhou, K., and Doudna, J. A. (1998) A general module for RNA crystallization. *J. Mol. Biol.* **279**, 621–631
59. Price, S. R., Ito, N., Oubridge, C., Avis, J. M., and Nagai, K. (1995) Crystallization of RNA-protein complexes. I. Methods for the large-scale preparation of RNA suitable for crystallographic studies. *J. Mol. Biol.* **249**, 398–408
60. Oubridge, C., Ito, N., Teo, C. H., Fearnley, I., and Nagai, K. (1995) Crystallization of RNA-protein complexes. II. The application of protein engineering for crystallization of the U1A protein-RNA complex. *J. Mol. Biol.* **249**, 409–423
61. Ferre-D'Amare, A. R. (2010) Use of the spliceosomal protein U1A to facilitate crystallization and structure determination of complex RNAs. *Methods* **52**, 159–167
62. Nakane, T., Kotecha, A., Sente, A., McMullan, G., Masiulis, S., Brown, P., Grigoras, I. T., Malinauskaitė, L., Malinauskas, T., Miehling, J., Uchanski, T., Yu, L., Karia, D., Pechnikova, E. V., de Jong, E., et al. (2020) Single-particle cryo-EM at atomic resolution. *Nature* **587**, 152–156
63. Cheng, Y. (2015) Single-particle cryo-EM at crystallographic resolution. *Cell* **161**, 450–457
64. Subramaniam, S., Earl, L. A., Falconieri, V., Milne, J. L., and Egelman, E. H. (2016) Resolution advances in cryo-EM enable application to drug discovery. *Curr. Opin. Struct. Biol.* **41**, 194–202
65. Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., Cech, T. R., and Doudna, J. A. (1996) Crystal structure of a group I ribozyme domain: Principles of RNA packing. *Science* **273**, 1678–1685
66. Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Szewczak, A. A., Kundrot, C. E., Cech, T. R., and Doudna, J. A. (1996) RNA tertiary structure mediation by adenosine platforms. *Science* **273**, 1696–1699
67. Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289**, 905–920
68. Cate, J. H., Yusupov, M. M., Yusupova, G. Z., Earnest, T. N., and Noller, H. F. (1999) X-ray crystal structures of 70S ribosome functional complexes. *Science* **285**, 2095–2104
69. Schluenzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., Bashan, A., Bartels, H., Agmon, I., Franceschi, F., and Yonath, A. (2000) Structure of functionally activated small ribosomal subunit at 3.3 Å resolution. *Cell* **102**, 615–623
70. Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Jr., Morgan-Warren, R. J., Carter, A. P., Vornheim, C., Hartsch, T., and Ramakrishnan, V. (2000) Structure of the 30S ribosomal subunit. *Nature* **407**, 327–339
71. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821
72. Jinek, M., Jiang, F., Taylor, D. W., Sternberg, S. H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S., Kaplan, M., Iavarone, A. T., Charpentier, E., Nogales, E., and Doudna, J. A. (2014) Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343**, 1247997
73. Bugg, C. E., Thomas, J. M., Sundaralingam, M., and Rao, S. T. (1971) Stereochemistry of nucleic acids and their constituents. X. Solid-state base-stacking patterns in nucleic acid constituents and polynucleotides. *Biopolymers* **10**, 175–219
74. Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer Verlag, Berlin
75. Lilley, D. M. (2011) Catalysis by the nucleolytic ribozymes. *Biochem. Soc. Trans.* **39**, 641–646
76. Seith, D. D., Bingaman, J. L., Veenis, A. J., Button, A. C., and Bevilacqua, P. C. (2018) Elucidation of catalytic strategies of small nucleolytic ribozymes from comparative analysis of active sites. *ACS Catal.* **8**, 314–327
77. Wilson, T. J., Liu, Y., Li, N. S., Dai, Q., Piccirilli, J. A., and Lilley, D. M. J. (2019) Comparison of the structures and mechanisms of the pistol and hammerhead ribozymes. *J. Am. Chem. Soc.* **141**, 7865–7875
78. Weixlbaumer, A., Murphy, F. V. T., Dziergowska, A., Malkiewicz, A., Vendeix, F. A., Agris, P. F., and Ramakrishnan, V. (2007) Mechanism for expanding the decoding capacity of transfer RNAs by modification of uridines. *Nat. Struct. Mol. Biol.* **14**, 498–502
79. Kurata, S., Weixlbaumer, A., Ohtsuki, T., Shimazaki, T., Wada, T., Kirino, Y., Takai, K., Watanabe, K., Ramakrishnan, V., and Suzuki, T. (2008) Modified uridines with C5-methylene substituents at the first position of the tRNA anticodon stabilize U.G wobble pairing during decoding. *J. Biol. Chem.* **283**, 18801–18811
80. Demeshkina, N., Jenner, L., Westhof, E., Yusupov, M., and Yusupova, G. (2012) A new understanding of the decoding principle on the ribosome. *Nature* **484**, 256–259
81. Rozov, A., Demeshkina, N., Khusainov, I., Westhof, E., Yusupov, M., and Yusupova, G. (2016) Novel base-pairing interactions at the tRNA wobble position crucial for accurate reading of the genetic code. *Nat. Commun.* **7**, 10457

82. Hutchins, C. J., Rathjen, P. D., Forster, A. C., and Symons, R. H. (1986) Self-cleavage of plus and minus RNA transcripts of avocado sunblotch viroid. *Nucleic Acids Res.* **14**, 3627–3640
83. Prody, G. A., Bakos, J. T., Buzayan, J. M., Schneider, I. R., and Bruening, G. (1986) Autolytic processing of dimeric plant virus satellite RNA. *Science* **231**, 1577–1580
84. Westhof, E. (2007) A tale in molecular recognition: The hammerhead ribozyme. *J. Mol. Recognit.* **20**, 1–3
85. Uhlenbeck, O. C. (2003) Less isn't always more. *RNA* **9**, 1415–1417
86. Wainwright, S. A. (1988) Form and function in organisms. *Amer. Zool.* **28**, 671–680
87. Cech, T. R., Zaug, A. J., and Grabowski, P. J. (1981) *In vitro* splicing of the ribosomal RNA precursor of tetrahymena: Involvement of a guanosine nucleotide in the excision of the intervening sequence. *Cell* **27**, 487–496
88. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., and Altman, S. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**, 849–857
89. Lilley, D. M. J. (2019) Classification of the nucleolytic ribozymes based upon catalytic mechanism. *F1000Res.* **8**, F1000 Faculty Rev-1462
90. Wilson, T. J., Liu, Y., Domnick, C., Kath-Schorr, S., and Lilley, D. M. (2016) The novel chemical mechanism of the twister ribozyme. *J. Am. Chem. Soc.* **138**, 6151–6162
91. Strobel, S. A., and Cochrane, J. C. (2007) RNA catalysis: Ribozymes, ribosomes, and riboswitches. *Curr. Opin. Chem. Biol.* **11**, 636–643
92. Marcia, M., and Pyle, A. M. (2012) Visualizing group II intron catalysis through the stages of splicing. *Cell* **151**, 497–507
93. Doudna, J. A., and Lorsch, J. R. (2005) Ribozyme catalysis: Not different, just worse. *Nat. Struct. Mol. Biol.* **12**, 395–402
94. Cech, T. R. (2000) Structural biology. The ribosome is a ribozyme. *Science* **289**, 878–879
95. Hiller, D. A., Singh, V., Zhong, M., and Strobel, S. A. (2011) A two-step chemical mechanism for ribosome-catalysed peptide bond formation. *Nature* **476**, 236–239
96. Schmeing, T. M., Huang, K. S., Strobel, S. A., and Steitz, T. A. (2005) An induced-fit mechanism to promote peptide bond formation and exclude hydrolysis of peptidyl-tRNA. *Nature* **438**, 520–524
97. Wan, R., Bai, R., Yan, C., Lei, J., and Shi, Y. (2019) Structures of the catalytically activated yeast spliceosome reveal the mechanism of branching. *Cell* **177**, 339–351.e313
98. Wan, R., Bai, R., and Shi, Y. (2019) Molecular choreography of pre-mRNA splicing by the spliceosome. *Curr. Opin. Struct. Biol.* **59**, 124–133
99. Wilkinson, M. E., Charenton, C., and Nagai, K. (2020) RNA splicing by the spliceosome. *Annu. Rev. Biochem.* **89**, 359–388
100. Kastner, B., Will, C. L., Stark, H., and Luhrmann, R. (2019) Structural insights into nuclear pre-mRNA splicing in higher eukaryotes. *Cold Spring Harb. Perspect. Biol.* **11**, a032417
101. Toor, N., Keating, K. S., Taylor, S. D., and Pyle, A. M. (2008) Crystal structure of a self-spliced group II intron. *Science* **320**, 77–82
102. Costa, M., Walbott, H., Monachello, D., Westhof, E., and Michel, F. (2016) Crystal structures of a group II intron lariat primed for reverse splicing. *Science* **354**, aaf9258
103. Watson, Z. L., Ward, F. R., Meheust, R., Ad, O., Schepartz, A., Banfield, J. F., and Cate, J. H. (2020) Structure of the bacterial ribosome at 2 Å resolution. *Elife* **9**, e60482
104. Bai, X. C., Fernandez, I. S., McMullan, G., and Scheres, S. H. (2013) Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. *Elife* **2**, e00461
105. Jomaa, A., Fu, Y. H., Boehringer, D., Leibundgut, M., Shan, S. O., and Ban, N. (2017) Structure of the quaternary complex between SRP, SR, and translocon bound to the translating ribosome. *Nat. Commun.* **8**, 15470
106. Lau, B., Cheng, J., Flemming, D., La Venuta, G., Berninghausen, O., Beckmann, R., and Hurt, E. (2021) Structure of the maturing 90S pre-ribosome in association with the RNA exosome. *Mol. Cell* **81**, 293–303.e294
107. Ameisemeier, M., Zemp, I., van den Heuvel, J., Thoms, M., Berninghausen, O., Kutay, U., and Beckmann, R. (2020) Structural basis for the final steps of human 40S ribosome maturation. *Nature* **587**, 683–687
108. Cheng, J., Lau, B., La Venuta, G., Ameisemeier, M., Berninghausen, O., Hurt, E., and Beckmann, R. (2020) 90S Pre-ribosome transformation into the primordial 40S subunit. *Science* **369**, 1470–1476
109. Dobzhansky, T. (1973) Nothing in biology makes sense except in the light of evolution. *Am. Biol. Teach.* **3**, 125–129
110. Ellington, A. D., and Szostak, J. W. (1990) *In vitro* selection of RNA molecules that bind specific ligands. *Nature* **346**, 818–822
111. Tuerk, C., and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505–510
112. Beaudry, A. A., and Joyce, G. F. (1992) Directed evolution of an RNA enzyme. *Science* **257**, 635–641
113. Warner, K. D., Chen, M. C., Song, W., Strack, R. L., Thorn, A., Jaffrey, S. R., and Ferre-D'Amare, A. R. (2014) Structural basis for activity of highly efficient RNA mimics of green fluorescent protein. *Nat. Struct. Mol. Biol.* **21**, 658–663
114. Trachman, R. J., 3rd, Abdolazadeh, A., Andreoni, A., Cojocar, R., Knutson, J. R., Ryckelynck, M., Unrau, P. J., and Ferre-D'Amare, A. R. (2018) Crystal structures of the mango-II RNA aptamer reveal heterogeneous fluorophore binding and guide engineering of variants with improved selectivity and brightness. *Biochemistry* **57**, 3544–3548
115. Trachman, R. J., 3rd, Autour, A., Jeng, S. C. Y., Abdolazadeh, A., Andreoni, A., Cojocar, R., Garipov, R., Dolgosheina, E. V., Knutson, J. R., Ryckelynck, M., Unrau, P. J., and Ferre-D'Amare, A. R. (2019) Structure and functional reselection of the mango-III fluorogenic RNA aptamer. *Nat. Chem. Biol.* **15**, 472–479
116. Mandal, M., Lee, M., Barrick, J. E., Weinberg, Z., Emilsson, G. M., Ruzzo, W. L., and Breaker, R. R. (2004) A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science* **306**, 275–279
117. McCown, P. J., Corbino, K. A., Stav, S., Sherlock, M. E., and Breaker, R. R. (2017) Riboswitch diversity and distribution. *RNA* **23**, 995–1011
118. Salvail, H., Balaji, A., Yu, D., Roth, A., and Breaker, R. R. (2020) Biochemical validation of a fourth guanidine riboswitch class in bacteria. *Biochemistry* **59**, 4654–4662
119. Serganov, A., Yuan, Y. R., Pikovskaya, O., Polonskaia, A., Malinina, L., Phan, A. T., Hobartner, C., Micura, R., Breaker, R. R., and Patel, D. J. (2004) Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chem. Biol.* **11**, 1729–1741
120. Breaker, R. R. (2020) Imaginary ribozymes. *ACS Chem. Biol.* **15**, 2020–2030
121. Ren, A., Rajashankar, K. R., and Patel, D. J. (2012) Fluoride ion encapsulation by Mg<sup>2+</sup> ions and phosphates in a fluoride riboswitch. *Nature* **486**, 85–89
122. Johnson, J. E., Jr., Reyes, F. E., Polaski, J. T., and Batey, R. T. (2012) B12 cofactors directly stabilize an mRNA regulatory switch. *Nature* **492**, 133–137
123. Peselis, A., and Serganov, A. (2012) Structural insights into ligand binding and gene expression control by an adenosylcobalamin riboswitch. *Nat. Struct. Mol. Biol.* **19**, 1182–1184
124. Sun, A., Gasser, C., Li, F., Chen, H., Mair, S., Krasheninina, O., Micura, R., and Ren, A. (2019) SAM-VI riboswitch structure and signature for ligand discrimination. *Nat. Commun.* **10**, 5728
125. Zuckerkandl, E., and Pauling, L. (1965) Molecules as documents of evolutionary history. *J. Theor. Biol.* **8**, 357–366
126. Woese, C. R., and Fox, G. E. (1977) Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5088–5090
127. Woese, C. R., Kandler, O., and Wheelis, M. L. (1990) Towards a natural system of organisms: Proposal for the domains archaea, bacteria, and eucarya. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 4576–4579
128. Fox, G. E., Stackebrandt, E., Hespell, R. B., Gibson, J., Maniloff, J., Dyer, T. A., Wolfe, R. S., Balch, W. E., Tanner, R. S., Magrum, L. J., Zablen, L. B., Blakemore, R., Gupta, R., Bonen, L., Lewis, B. J., et al. (1980) The phylogeny of prokaryotes. *Science* **209**, 457–463
129. Noller, H. F., and Woese, C. R. (1981) Secondary structure of 16S ribosomal RNA. *Science* **212**, 403–411

130. Michel, F., Jacquier, A., and Dujon, B. (1982) Comparison of fungal mitochondrial introns reveals extensive homologies in RNA secondary structure. *Biochimie* **64**, 867–881
131. James, B. D., Olsen, G. J., Liu, J. S., and Pace, N. R. (1988) The secondary structure of ribonuclease P RNA, the catalytic element of a ribonucleoprotein enzyme. *Cell* **52**, 19–26
132. Michel, F., Costa, M., Massire, C., and Westhof, E. (2000) Modeling RNA tertiary structure from patterns of sequence variation. *Methods Enzymol.* **317**, 491–510
133. Michel, F., Umesono, K., and Ozeki, H. (1989) Comparative and functional anatomy of group II catalytic introns—a review. *Gene* **82**, 5–30
134. Michel, F., and Ferat, J. L. (1995) Structure and activities of group II introns. *Annu. Rev. Biochem.* **64**, 435–461
135. Michel, F., and Westhof, E. (1990) Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J. Mol. Biol.* **216**, 585–610
136. Berman, H. M., Adams, P. D., Bonvin, A. A., Burley, S. K., Carragher, B., Chiu, W., DiMaio, F., Ferrin, T. E., Gabanyi, M. J., Goddard, T. D., Griffin, P. R., Haas, J., Hanke, C. A., Hoch, J. C., Hummer, G., *et al.* (2019) Federating structural models and data: Outcomes from a workshop on archiving integrative structures. *Structure* **27**, 1745–1759
137. Bekker, G. J., Kawabata, T., and Kurisu, G. (2020) The biological structure model archive (BSM-Arc): An archive for in silico models and simulations. *Biophys. Rev.* **12**, 371–375
138. Fuller, W., and Hodgson, A. (1967) Conformation of the anticodon loop intRNA. *Nature* **215**, 817–821
139. Koltun, W. L. (1965) Precision space-filling atomic models. *Biopolymers* **3**, 665–679
140. Langridge, R., Marvin, D. A., Seeds, W. E., Wilson, H. R., Hooper, C. W., Wilkins, M. H. F., and Hamilton, L. D. (1960) The molecular configuration of deoxyribonucleic acid: II. Molecular models and their Fourier transforms. *J. Mol. Biol.* **2**, 38–62
141. Levitt, M. (1969) Detailed molecular model for transfer ribonucleic acid. *Nature* **224**, 759–763
142. Lake, J. A., and Beeman, W. W. (1968) On the conformation of yeast transfer RNA. *J. Mol. Biol.* **31**, 115–125
143. Pleij, C. W., Rietveld, K., and Bosch, L. (1985) A new principle of RNA folding based on pseudoknotting. *Nucleic Acids Res.* **13**, 1717–1731
144. Kim, S. H., and Cech, T. R. (1987) Three-dimensional model of the active site of the self-splicing rRNA precursor of tetrahymena. *Proc. Natl. Acad. Sci. U. S. A.* **84**, 8788–8792
145. Kim, S. H., Quigley, G., Suddath, F. L., McPherson, A., Sneden, D., Kim, J. J., Weinzierl, J., Blattmann, P., and Rich, A. (1972) The three-dimensional structure of yeast phenylalanine transfer RNA: Shape of the molecule at 5.5-Å resolution. *Proc. Natl. Acad. Sci. U. S. A.* **69**, 3746–3750
146. Kim, S. H., Sussman, J. L., Suddath, F. L., Quigley, G. J., McPherson, A., Wang, A. H., Seeman, N. C., and Rich, A. (1974) The general structure of transfer RNA molecules. *Proc. Natl. Acad. Sci. U. S. A.* **71**, 4970–4974
147. Katz, L., and Levinthal, C. (1972) Interactive computer graphics and representation of complex biological structures. *Annu. Rev. Biophys. Bioeng.* **1**, 465–504
148. Feldmann, R. J. (1976) The design of computing systems for molecular modeling. *Annu. Rev. Biophys. Bioeng.* **5**, 477–510
149. Langridge, R. (1974) Interactive three-dimensional computer graphics in molecular biology. *Fed. Proc.* **33**, 2332–2335
150. Langridge, R., Ferrin, T. E., Kuntz, I. D., and Connolly, M. L. (1981) Real-time color graphics in studies of molecular interactions. *Science* **211**, 661–666
151. Jones, T. A. (1978) A graphics model building and refinement system for macromolecules. *J. Appl. Cryst.* **11**, 268–272
152. Li, B., Cao, Y., Westhof, E., and Miao, Z. (2020) Advances in RNA 3D structure modeling using experimental data. *Front. Genet.* **11**, 574485
153. Miao, Z., and Westhof, E. (2017) RNA structure: Advances and assessment of 3D structure prediction. *Annu. Rev. Biophys.* **46**, 483–503
154. DellaVigna, S., Pope, D., and Vivaldi, E. (2019) Predict science to improve science. *Science* **366**, 428–429
155. Cruz, J. A., Blanchet, M. F., Boniecki, M., Bujnicki, J. M., Chen, S. J., Cao, S., Das, R., Ding, F., Dokholyan, N. V., Flores, S. C., Huang, L., Lavender, C. A., Lisi, V., Major, F., Mikolajczak, K., *et al.* (2012) RNA-puzzles: A CASP-like evaluation of RNA three-dimensional structure prediction. *RNA* **18**, 610–625
156. Miao, Z., Adamiak, R. W., Antczak, M., Batey, R. T., Becka, A. J., Biesiada, M., Boniecki, M. J., Bujnicki, J. M., Chen, S. J., Cheng, C. Y., Chou, F. C., Ferre-D'Amare, A. R., Das, R., Dawson, W. K., Ding, F., *et al.* (2017) RNA-puzzles round III: 3D RNA structure prediction of five riboswitches and one ribozyme. *RNA* **23**, 655–672
157. Miao, Z., Adamiak, R. W., Antczak, M., Boniecki, M. J., Bujnicki, J., Chen, S. J., Cheng, C. Y., Cheng, Y., Chou, F. C., Das, R., Dokholyan, N. V., Ding, F., Geniesse, C., Jiang, Y., Joshi, A., *et al.* (2020) RNA-puzzles round IV: 3D structure predictions of four ribozymes and two aptamers. *RNA* **26**, 982–995
158. Miao, Z., Adamiak, R. W., Blanchet, M. F., Boniecki, M., Bujnicki, J. M., Chen, S. J., Cheng, C., Chojnowski, G., Chou, F. C., Cordero, P., Cruz, J. A., Ferre-D'Amare, A. R., Das, R., Ding, F., Dokholyan, N. V., *et al.* (2015) RNA-puzzles round II: Assessment of RNA structure prediction programs applied to three large RNA structures. *RNA* **21**, 1066–1084
159. Moulton, J. (1996) The current state of the art in protein structure prediction. *Curr. Opin. Biotechnol.* **7**, 422–427
160. Magnus, M., Antczak, M., Zok, T., Wiedemann, J., Lukasiak, P., Cao, Y., Bujnicki, J. M., Westhof, E., Szachniuk, M., and Miao, Z. (2020) RNA-puzzles toolkit: A computational resource of RNA 3D structure benchmark datasets, structure manipulation, and evaluation tools. *Nucleic Acids Res.* **48**, 576–588
161. Young, J. Y., Westbrook, J. D., Feng, Z., Sala, R., Peisach, E., Oldfield, T. J., Sen, S., Gutmanas, A., Armstrong, D. R., Berrisford, J. M., Chen, L., Chen, M., Di Costanzo, L., Dimitropoulos, D., Gao, G., *et al.* (2017) OneDep: Unified wwPDB system for deposition, biocuration, and validation of macromolecular structures in the PDB archive. *Structure* **25**, 536–545
162. Goodsell, D. S., Zardecki, C., Berman, H. M., and Burley, S. K. (2020) Insights from 20 years of the molecule of the month. *Biochem. Mol. Biol. Educ.* **48**, 350–355
163. Branden, C. I., and Jones, T. A. (1990) Between objectivity and subjectivity. *Nature* **343**, 687–689
164. Anderson, W. P., and Global Life Science Data Resources Working Group. (2017) Data management: A global coalition to sustain core data. *Nature* **543**, 179
165. Bourne, P. E., Lorsch, J. R., and Green, E. D. (2015) Perspective: Sustaining the big-data ecosystem. *Nature* **527**, S16–S17
166. Turk, M., and Baumeister, W. (2020) The promise and the challenges of cryo-electron tomography. *FEBS Lett.* **594**, 3243–3261
167. RNAcentral Consortium (2021) RNAcentral 2021: Secondary structure integration, improved sequence search and new member databases. *Nucleic Acids Res* **49**, D212–D220
168. Daubner, G. M., Clery, A., and Allain, F. H. (2013) RRM-RNA recognition: NMR or crystallography...and new findings. *Curr. Opin. Struct. Biol.* **23**, 100–108
169. Dominguez, C., and Allain, F. H. (2006) NMR structure of the three quasi RNA recognition motifs (qRRMs) of human hnRNP F and interaction studies with Bcl-x G-tract RNA: A novel mode of RNA recognition. *Nucleic Acids Res.* **34**, 3634–3645
170. Wu, H., Finger, L. D., and Feigon, J. (2005) Structure determination of protein/RNA complexes by NMR. *Methods Enzymol.* **394**, 525–545
171. Zhang, Q., Kim, N. K., and Feigon, J. (2011) Architecture of human telomerase RNA. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 20325–20332
172. Wang, Y., Susac, L., and Feigon, J. (2019) Structural biology of telomerase. *Cold Spring Harb. Perspect. Biol.* **11**, a032383
173. Dethoff, E. A., Petzold, K., Chugh, J., Casiano-Negroni, A., and Al-Hashimi, H. M. (2012) Visualizing transient low-populated structures of RNA. *Nature* **491**, 724–728
174. Dethoff, E. A., Chugh, J., Mustoe, A. M., and Al-Hashimi, H. M. (2012) Functional complexity and regulation through RNA dynamics. *Nature* **482**, 322–330
175. Kimsey, I. J., Petzold, K., Sathyamoorthy, B., Stein, Z. W., and Al-Hashimi, H. M. (2015) Visualizing transient Watson-Crick-like mispairs in DNA and RNA duplexes. *Nature* **519**, 315–320

176. Rinnenthal, J., Buck, J., Ferner, J., Wacker, A., Furtig, B., and Schwalbe, H. (2011) Mapping the landscape of RNA dynamics with NMR spectroscopy. *Acc. Chem. Res.* **44**, 1292–1301
177. Reining, A., Nozinovic, S., Schlepckow, K., Buhr, F., Furtig, B., and Schwalbe, H. (2013) Three-state mechanism couples ligand and temperature sensing in riboswitches. *Nature* **499**, 355–359
178. Robertson, M. (1980) Biology in the 1980s, plus or minus a decade. *Nature* **285**, 358–359
179. Kappel, K., Zhang, K., Su, Z., Watkins, A. M., Kladwang, W., Li, S., Pintilie, G., Topkar, V. V., Rangan, R., Zheludev, I. N., Yesselman, J. D., Chiu, W., and Das, R. (2020) Accelerated cryo-EM-guided determination of three-dimensional RNA-only structures. *Nat. Methods* **17**, 699–707
180. Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., Qin, C., Zidek, A., Nelson, A. W. R., Bridgland, A., Penedones, H., Petersen, S., Simonyan, K., Crossan, S., Kohli, P., *et al.* (2020) Improved protein structure prediction using potentials from deep learning. *Nature* **577**, 706–710
181. Gao, W., Mahajan, S. P., Sulam, J., and Gray, J. J. (2020) Deep learning in protein structural modeling and design. *Patterns (N. Y.)* **1**, 100142
182. Callaway, E. (2020) 'It will change everything': DeepMind's AI makes gigantic leap in solving protein structures. *Nature* **588**, 203–204
183. Levintal, C. (1968) Are there pathways for protein folding? *J. Chim. Phys.* **65**, 44–45
184. Richardson, J. S. (2020) A new way to see RNAs. *Nat. Methods* **17**, 663–664
185. Helliwell, J. R. (2020) DeepMind and CASP14. *IUCr Newsl.* **28**, 6
186. Moore, P. B., Petrov, A., Westhof, E., and Zirbel, C. L. (2021) Neocles B. Leontis (1955 - 2020). *RNA*. <https://doi.org/10.1261/rna.078673.121>
187. Swaminathan, V., and Sundaralingam, M. (1979) The crystal structures of metal complexes of nucleic acids and their constituents. *CRC Crit. Rev. Biochem.* **6**, 245–336
188. Krol, A., Westhof, E., Bach, M., Luhrmann, R., Ebel, J. P., and Carbon, P. (1990) Solution structure of human U1 snRNA. Derivation of a possible three-dimensional model. *Nucleic Acids Res.* **18**, 3803–3811
189. Scott, W. G., Finch, J. T., and Klug, A. (1995) The crystal structure of an all-RNA hammerhead ribozyme: A proposed mechanism for RNA catalytic cleavage. *Cell* **81**, 991–1002
190. Pley, H. W., Flaherty, K. M., and McKay, D. B. (1994) Model for an RNA tertiary interaction from the structure of an intermolecular complex between a GAAA tetraloop and an RNA helix. *Nature* **372**, 111–113
191. Correll, C. C., Freeborn, B., Moore, P. B., and Steitz, T. A. (1997) Metals, motifs, and recognition in the crystal structure of a 5S rRNA domain. *Cell* **91**, 705–712
192. Ferre-D'Amare, A. R., Zhou, K., and Doudna, J. A. (1998) Crystal structure of a hepatitis delta virus ribozyme. *Nature* **395**, 567–574
193. Nissen, P., Ippolito, J. A., Ban, N., Moore, P. B., and Steitz, T. A. (2001) RNA tertiary interactions in the large ribosomal subunit: The A-minor motif. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 4899–4903
194. Ogle, J. M., Brodersen, D. E., Clemons, W. M., Jr., Tarry, M. J., Carter, A. P., and Ramakrishnan, V. (2001) Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* **292**, 897–902
195. Zirbel, C. L., Sponer, J. E., Sponer, J., Stombaugh, J., and Leontis, N. B. (2009) Classification and energetics of the base-phosphate interactions in RNA. *Nucleic Acids Res.* **37**, 4898–4918
196. Wu, J., Zhou, C., Li, J., Li, C., Tao, X., Leontis, N. B., Zirbel, C. L., Bisaro, D. M., and Ding, B. (2020) Functional analysis reveals G/U pairs critical for replication and trafficking of an infectious non-coding viroid RNA. *Nucleic Acids Res.* **48**, 3134–3155
197. Noller, H. F. (2005) RNA structure: Reading the ribosome. *Science* **309**, 1508–1514
198. Crick, F. H. (1966) Codon–anticodon pairing: The wobble hypothesis. *J. Mol. Biol.* **19**, 548–555
199. Westhof, E., Yusupov, M., and Yusupova, G. (2019) The multiple flavors of GoU pairs in RNA. *J. Mol. Recognit.* **32**, e2782
200. Golden, B. L., Gooding, A. R., Podell, E. R., and Cech, T. R. (1998) A preorganized active site in the crystal structure of the tetrahymena ribozyme. *Science* **282**, 259–264
201. Baugh, C., Grate, D., and Wilson, C. (2000) 2.8 Å crystal structure of the malachite green aptamer. *J. Mol. Biol.* **301**, 117–128
202. Sussman, D., Nix, J. C., and Wilson, C. (2000) The structural basis for molecular recognition by the vitamin B 12 RNA aptamer. *Nat. Struct. Biol.* **7**, 53–57
203. Vasudevan, S. S., and Sundaralingam, M. (2001) The occurrence of the syn-C3' endo conformation and the distorted backbone conformations for C4'-C5' and P-O5' in oligo and polynucleotides. *J. Biomol. Struct. Dyn.* **18**, 824–831
204. Cheong, C., and Moore, P. B. (1992) Solution structure of an unusually stable RNA tetraplex containing G- and U-quartet structures. *Biochemistry* **31**, 8406–8414
205. Garst, A. D., Edwards, A. L., and Batey, R. T. (2011) Riboswitches: Structures and mechanisms. *Cold Spring Harb. Perspect. Biol.* **3**, a003533
206. Adams, P. L., Stahley, M. R., Gill, M. L., Kosek, A. B., Wang, J., and Strobel, S. A. (2004) Crystal structure of a group I intron splicing intermediate. *RNA* **10**, 1867–1887
207. Torres-Larios, A., Swinger, K. K., Krasilnikov, A. S., Pan, T., and Mondragon, A. (2005) Crystal structure of the RNA component of bacterial ribonuclease P. *Nature* **437**, 584–587
208. Martick, M., and Scott, W. G. (2006) Tertiary contacts distant from the active site prime a ribozyme for catalysis. *Cell* **126**, 309–320
209. Dibrov, S. M., McLean, J., Parsons, J., and Hermann, T. (2011) Self-assembling RNA square. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 6405–6408
210. Grigg, J. C., and Ke, A. (2013) Structural determinants for geometry and information decoding of tRNA by T box leader RNA. *Structure* **21**, 2025–2032
211. Colussi, T. M., Costantino, D. A., Hammond, J. A., Ruehle, G. M., Nix, J. C., and Kieft, J. S. (2014) The structural basis of transfer RNA mimicry and conformational plasticity by a viral RNA. *Nature* **511**, 366–369
212. Li, S., Su, Z., Lehmann, J., Stamatopoulou, V., Giaramoglou, N., Henderson, F. E., Fan, L., Pintilie, G. D., Zhang, K., Chen, M., Ludtke, S. J., Wang, Y. X., Stathopoulos, C., Chiu, W., and Zhang, J. (2019) Structural basis of amino acid surveillance by higher-order tRNA-mRNA interactions. *Nat. Struct. Mol. Biol.* **26**, 1094–1105
213. Yan, C., Wan, R., and Shi, Y. (2019) Molecular mechanisms of pre-mRNA splicing through structural biology of the spliceosome. *Cold Spring Harb. Perspect. Biol.* **11**, a032409
214. Rould, M. A., Perona, J. J., Soll, D., and Steitz, T. A. (1989) Structure of E. coli glutamyl-tRNA synthetase complexed with tRNA(Gln) and ATP at 2.8 Å resolution. *Science* **246**, 1135–1142
215. Belrhali, H., Yaremchuk, A., Tukalo, M., Larsen, K., Berthet-Colominas, C., Leberman, R., Beijer, B., Sproat, B., Als-Nielsen, J., Grubel, G., Legrand, J.-F., Lehmann, M., and Cusack, S. (1994) Crystal structures at 2.5 angstrom resolution of seryl-tRNA synthetase complexed with two analogs of seryl adenylate. *Science* **263**, 1432–1436
216. Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J. C., and Moras, D. (1991) Class II aminoacyl transfer RNA synthetases: Crystal structure of yeast aspartyl-tRNA synthetase complexed with tRNA(Asp). *Science* **252**, 1682–1689
217. van den Worm, S. H., Stonehouse, N. J., Valegard, K., Murray, J. B., Walton, C., Fridborg, K., Stockley, P. G., and Liljas, L. (1998) Crystal structures of MS2 coat protein mutants in complex with wild-type RNA operator fragments. *Nucleic Acids Res.* **26**, 1345–1351
218. Price, S. R., Evans, P. R., and Nagai, K. (1998) Crystal structure of the spliceosomal U2B'-U2A' protein complex bound to a fragment of U2 small nuclear RNA. *Nature* **394**, 645–650
219. Vidovic, I., Nottrott, S., Hartmuth, K., Luhrmann, R., and Ficner, R. (2000) Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. *Mol. Cell* **6**, 1331–1342
220. Hainzl, T., Huang, S., and Sauer-Eriksson, A. E. (2007) Interaction of signal-recognition particle 54 GTPase domain and signal-recognition particle RNA in the free signal-recognition particle. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 14911–14916
221. Paukstelis, P. J., Chen, J. H., Chase, E., Lambowitz, A. M., and Golden, B. L. (2008) Structure of a tyrosyl-tRNA synthetase splicing factor bound to a group I intron RNA. *Nature* **451**, 94–97



222. Pomeranz Krummel, D. A., Oubridge, C., Leung, A. K., Li, J., and Nagai, K. (2009) Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution. *Nature* **458**, 475–480
223. Reiter, N. J., Osterman, A., Torres-Larios, A., Swinger, K. K., Pan, T., and Mondragon, A. (2010) Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. *Nature* **468**, 784–789
224. Leung, A. K., Nagai, K., and Li, J. (2011) Structure of the spliceosomal U4 snRNP core domain and its implication for snRNP biogenesis. *Nature* **473**, 536–539
225. Zhou, L., Hang, J., Zhou, Y., Wan, R., Lu, G., Yin, P., Yan, C., and Shi, Y. (2014) Crystal structures of the Lsm complex bound to the 3' end sequence of U6 small nuclear RNA. *Nature* **506**, 116–120
226. Wan, R., Yan, C., Bai, R., Wang, L., Huang, M., Wong, C. C., and Shi, Y. (2016) The 3.8 Å structure of the U4/U6.U5 tri-snRNP: Insights into spliceosome assembly and catalysis. *Science* **351**, 466–475
227. Yan, C., Hang, J., Wan, R., Huang, M., Wong, C. C., and Shi, Y. (2015) Structure of a yeast spliceosome at 3.6-angstrom resolution. *Science* **349**, 1182–1191
228. Yan, C., Wan, R., Bai, R., Huang, G., and Shi, Y. (2016) Structure of a yeast activated spliceosome at 3.5 Å resolution. *Science* **353**, 904–911
229. Wan, R., Yan, C., Bai, R., Huang, G., and Shi, Y. (2016) Structure of a yeast catalytic step I spliceosome at 3.4 Å resolution. *Science* **353**, 895–904
230. Galej, W. P., Wilkinson, M. E., Fica, S. M., Oubridge, C., Newman, A. J., and Nagai, K. (2016) Cryo-EM structure of the spliceosome immediately after branching. *Nature* **537**, 197–201
231. Yan, C., Wan, R., Bai, R., Huang, G., and Shi, Y. (2017) Structure of a yeast step II catalytically activated spliceosome. *Science* **355**, 149–155
232. Fica, S. M., Oubridge, C., Galej, W. P., Wilkinson, M. E., Bai, X. C., Newman, A. J., and Nagai, K. (2017) Structure of a spliceosome remodelled for exon ligation. *Nature* **542**, 377–380
233. Bai, R., Yan, C., Wan, R., Lei, J., and Shi, Y. (2017) Structure of the post-catalytic spliceosome from *Saccharomyces cerevisiae*. *Cell* **171**, 1589–1598.e1588
234. Wilkinson, M. E., Fica, S. M., Galej, W. P., Norman, C. M., Newman, A. J., and Nagai, K. (2017) Postcatalytic spliceosome structure reveals mechanism of 3'-splice site selection. *Science* **358**, 1283–1288
235. Liu, S., Li, X., Zhang, L., Jiang, J., Hill, R. C., Cui, Y., Hansen, K. C., Zhou, Z. H., and Zhao, R. (2017) Structure of the yeast spliceosomal postcatalytic P complex. *Science* **358**, 1278–1283
236. Gangloff, J., Keith, G., Ebel, J. P., and Dirheimer, G. (1971) Structure of aspartate-tRNA from brewer's yeast. *Nat. New Biol.* **230**, 125–126
237. Kim, S. H. (1978) Three-dimensional structure of transfer RNA and its functional implications. *Adv. Enzymol. Relat. Areas Mol. Biol.* **46**, 279–315
238. Moras, D., Comarmond, M. B., Fischer, J., Weiss, R., Thierry, J. C., Ebel, J. P., and Giege, R. (1980) Crystal structure of yeast tRNA<sup>Asp</sup>. *Nature* **288**, 669–674
239. Westhof, E., Dumas, P., and Moras, D. (1985) Crystallographic refinement of yeast aspartic acid transfer RNA. *J. Mol. Biol.* **184**, 119–145
240. Leontis, N. B., and Westhof, E. (2001) Geometric nomenclature and classification of RNA base pairs. *RNA* **7**, 499–512
241. Westhof, E. (1992) Westhof's rule. *Nature* **358**, 459–460
242. Auffinger, P., and Westhof, E. (2001) An extended structural signature for the tRNA anticodon loop. *RNA* **7**, 334–341
243. Khvorova, A., Lescoute, A., Westhof, E., and Jayasena, S. D. (2003) Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity. *Nat. Struct. Biol.* **10**, 708–712
244. Costa, M., and Michel, F. (1995) Frequent use of the same tertiary motif by self-folding RNAs. *EMBO J.* **14**, 1276–1285
245. Dumas, P., Moras, D., Florentz, C., Giege, R., Verlaan, P., Van Belkum, A., and Pleij, C. W. (1987) 3-D graphics modelling of the tRNA-like 3'-end of turnip yellow mosaic virus RNA: Structural and functional implications. *J. Biomol. Struct. Dyn.* **4**, 707–728
246. Westhof, E., Romby, P., Romaniuk, P. J., Ebel, J. P., Ehresmann, C., and Ehresmann, B. (1989) Computer modeling from solution data of spinach chloroplast and of *Xenopus laevis* somatic and oocyte 5 S rRNAs. *J. Mol. Biol.* **207**, 417–431
247. Sturchler, C., Westhof, E., Carbon, P., and Krol, A. (1993) Unique secondary and tertiary structural features of the eucaryotic selenocysteine tRNA(Sec). *Nucleic Acids Res.* **21**, 1073–1079
248. Tuschl, T., Gohlke, C., Jovin, T. M., Westhof, E., and Eckstein, F. (1994) A three-dimensional model for the hammerhead ribozyme based on fluorescence measurements. *Science* **266**, 785–789
249. Tanner, N. K., Schaff, S., Thill, G., Petit-Koskas, E., Crain-Denoyelle, A. M., and Westhof, E. (1994) A three-dimensional model of hepatitis delta virus ribozyme based on biochemical and mutational analyses. *Curr. Biol.* **4**, 488–498
250. Serganov, A. A., Masquida, B., Westhof, E., Cachia, C., Portier, C., Garber, M., Ehresmann, B., and Ehresmann, C. (1996) The 16S rRNA binding site of *Thermus thermophilus* ribosomal protein S15: Comparison with *Escherichia coli* S15, minimum site and structure. *RNA* **2**, 1124–1138
251. Massire, C., Jaeger, L., and Westhof, E. (1998) Derivation of the three-dimensional architecture of bacterial ribonuclease P RNAs from comparative sequence analysis. *J. Mol. Biol.* **279**, 773–793
252. Rangan, P., Masquida, B., Westhof, E., and Woodson, S. A. (2004) Architecture and folding mechanism of the *Azoarcus* group I pre-tRNA. *J. Mol. Biol.* **339**, 41–51
253. Konnert, J. H., and Hendrickson, W. A. (1980) A restrained-parameter thermal-factor refinement procedure. *Acta Cryst.* **A36**, 344–349
254. Hendrickson, W. A., and Konnert, J. H. (1981). In Srinivasan, R., ed., *Biomolecular Structure, Conformation, Function and Evolution* (Vol. 1, Pergamon Press, Oxford: 43–57
255. Ferrin, T. E., Couch, G. S., Huang, C. C., Pettersen, E. F., and Langridge, R. (1991) An affordable approach to interactive desktop molecular modeling. *J. Mol. Graph.* **9**, 27–32, 37–8
256. Ferrin, T. E., Huang, C. C., Jarvis, L. E., and Langridge, R. (1988) The MIDAS display system. *J. Mol. Graph.* **6**, 13–27
257. Major, F., Turcotte, M., Gautheret, D., Lapalme, G., Fillion, E., and Cedergren, R. (1991) The combination of symbolic and numerical computation for three-dimensional modeling of RNA. *Science* **253**, 1255–1260
258. Massire, C., and Westhof, E. (1998) MANIP: An interactive tool for modelling RNA. *J. Mol. Graph. Model.* **16**, 197–205, 255–7
259. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612
260. Jossinet, F., Ludwig, T. E., and Westhof, E. (2010) Assemble: An interactive graphical tool to analyze and build RNA architectures at the 2D and 3D levels. *Bioinformatics* **26**, 2057–2059
261. Jossinet, F., and Westhof, E. (2005) Sequence to structure (S2S): Display, manipulate and interconnect RNA data from sequence to structure. *Bioinformatics* **21**, 3320–3321
262. Das, R., and Baker, D. (2007) Automated de novo prediction of native-like RNA tertiary structures. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 14664–14669
263. Parisien, M., and Major, F. (2008) The MC-fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* **452**, 51–55
264. Ding, F., Sharma, S., Chalasani, P., Demidov, V. V., Broude, N. E., and Dokholyan, N. V. (2008) *Ab initio* RNA folding by discrete molecular dynamics: From structure prediction to folding mechanisms. *RNA* **14**, 1164–1173
265. Rother, M., Rother, K., Puton, T., and Bujnicki, J. M. (2011) ModeRNA: A tool for comparative modeling of RNA 3D structure. *Nucleic Acids Res.* **39**, 4007–4022
266. Popenda, M., Szachniuk, M., Antczak, M., Purzycka, K. J., Lukasiak, P., Bartol, N., Blazewicz, J., and Adamiak, R. W. (2012) Automated 3D structure composition for large RNAs. *Nucleic Acids Res.* **40**, e112
267. Zhao, Y., Huang, Y., Gong, Z., Wang, Y., Man, J., and Xiao, Y. (2012) Automated and fast building of three-dimensional RNA structures. *Sci. Rep.* **2**, 734
268. Xu, X., Zhao, P., and Chen, S. J. (2014) Vfold: A web server for RNA structure and folding thermodynamics prediction. *PLoS One* **9**, e107504
269. Boniecki, M. J., Lach, G., Dawson, W. K., Tomala, K., Lukasz, P., Soltynski, T., Rother, K. M., and Bujnicki, J. M. (2016) SimRNA: A coarse-grained method for RNA folding simulations and 3D structure prediction. *Nucleic Acids Res.* **44**, e63