



ANNUAL  
REVIEWS **Further**

Click [here](#) to view this article's online features:

- Download figures as PPT slides
- Navigate linked references
- Download citations
- Explore related articles
- Search keywords

# RNA Structure: Advances and Assessment of 3D Structure Prediction

Zhichao Miao and Eric Westhof

Architecture et Réactivité de l'ARN, Université de Strasbourg, Institut de Biologie Moléculaire et Cellulaire du CNRS, 67000 Strasbourg, France; email: z.miao@ibmc-cnrs.unistra.fr, e.westhof@ibmc-cnrs.unistra.fr

Annu. Rev. Biophys. 2017. 46:483–503

First published online as a Review in Advance on March 30, 2017

The *Annual Review of Biophysics* is online at [biophys.annualreviews.org](http://biophys.annualreviews.org)

<https://doi.org/10.1146/annurev-biophys-070816-034125>

Copyright © 2017 by Annual Reviews.  
All rights reserved

## Keywords

evaluation metrics, prediction, RNA-puzzles, RNA structure, structural module, three-dimensional

## Abstract

Biological functions of RNA molecules are dependent upon sustained specific three-dimensional (3D) structures of RNA, with or without the help of proteins. Understanding of RNA structure is frequently based on 2D structures, which describe only the Watson–Crick (WC) base pairs. Here, we hierarchically review the structural elements of RNA and how they contribute to RNA 3D structure. We focus our analysis on the non-WC base pairs and on RNA modules. Several computer programs have now been designed to predict RNA modules. We describe the RNA-Puzzles initiative, which is a community-wide, blind assessment of RNA 3D structure prediction programs to determine the capabilities and bottlenecks of current predictions. The assessment metrics used in RNA-Puzzles are briefly described. The detection of RNA 3D modules from sequence data and their automatic implementation belong to the current challenges in RNA 3D structure prediction.

## Contents

INTRODUCTION .....	484
RNA STRUCTURAL ELEMENTS .....	485
RNA Bases and Base Pairings .....	485
2D Structure and 3D Structural Modules .....	486
Example of an RNA Module: The G-Bulge .....	488
Pseudoknots .....	488
Loop–Loop Interactions .....	489
Ion and Ligand Binding .....	489
RECENTLY SOLVED RNA STRUCTURES .....	490
MODULE PREDICTION PROGRAMS .....	490
ASSESSMENT OF RNA 3D STRUCTURE PREDICTION .....	491
Evaluation Metrics .....	491
RNA-Puzzles .....	492
CHALLENGES AND PERSPECTIVES .....	496

## INTRODUCTION

RNA is central to numerous and various biological functions in coding, decoding, regulation, and gene expression (1, 6). RNAs must sustain stable and specific structures to perform their biological functions. Riboswitches and ribozymes often self-assemble to carry out biological functions, whereas most noncoding RNAs (ncRNAs), such as ribosomes, small nuclear ribonucleoproteins (snRNPs), small nucleolar ribonucleoproteins (snoRNPs), telomerase, microRNAs, and long ncRNAs, form RNA–protein complexes to operate (7). Essential understanding of the RNA molecules and their functions requires knowledge about their structures (103).

RNA secondary (2D) structure, which is a representation of the Watson–Crick (WC) base pairs present in the RNA structure, has long been used to describe RNA architectures. And much of RNA structure prediction work focuses on the 2D structure level. Detailed analyses of these structure prediction methods can be found in a number of reviews (19, 61, 89, 118).

Chemical probing of RNA structures dates back to the last century (24, 105, 115). Many different chemicals have been used to modify specific bases at certain sites to show their availability for specific modification type. RNA structure probing is now undergoing several trends: (a) the probing of RNA structures inside living cells (40, 85, 93) to illustrate the native state and flexibility of RNA structures; (b) the advent of high-throughput (53, 97) and genome-wide (23) probing of RNA structures; and (c) computational prediction (23, 52) to enhance the interpretation of experimental results.

Over the years, the probing of RNA structure has been extended to three-dimensional (3D) structure probing (11, 105). As 2D structure describes only the WC base pairs in RNA structure, knowledge of 3D structure is necessary for demonstrating molecular details and more insightful for interpreting molecular functions. Some new experiments have been developed to capture both WC and non-WC base-pair interactions (35, 89). With our accumulated knowledge on its structure, from RNA sequences to RNA tertiary structures, computational prediction of RNA 3D structure both with and without the aid of experimental data has been developed and improved over the years. Strategies range from physics-based (110) to experiment-aided (11, 16) to evolutionary

---

**Watson–Crick (WC) base pairs:** the base pairs between A–U and C = G in the *cis* orientation formed by H-bonds between the WC edges

**Base pair:** a unit consisting of two nucleobases bound to each other by hydrogen bonds

**RNA tertiary structure:** the three-dimensional structure of RNA, typically described by atomic coordinates

---

clue-derived models (18). Details of RNA 3D structure prediction methods are reviewed in the article “Theory and Modeling of RNA Structure” in this volume (95).

Due to the importance of RNA modules formed by non-WC base pairs (43, 48), computer programs have been designed to predict these RNA modules according to sequence information or sequence alignment. One such program has already proved useful for RNA 2D structure prediction on a genome-wide scale (99), suggesting the potential for further improvements in RNA 3D structure prediction.

A community effort of RNA 3D structure prediction, named RNA-Puzzles, was initiated in 2011 (14, 63). As a blind test of RNA 3D structure prediction, RNA-Puzzles employs various assessment metrics to describe different aspects of prediction accuracies. The challenges in RNA-Puzzles show a good diversity of RNA structures: double-stranded structures, riboswitches, ribozymes, RNA–RNA complexes, and RNA–protein complexes. According to the prediction results from many groups around the world, RNA-Puzzles have led to an improvement in prediction over the years. The value of the assessment for improving the computational methods depends critically on the validity and relevance of the criteria used during the evaluation.

An ultimate aim of RNA structure prediction is to help improve our understanding of the biological functions of the molecule and the interpretation of molecular data and processes. New RNA-Puzzles challenges have started to include RNA structures with ligand binding and ligand binding-induced conformational change. It is comforting that some predictions are topologically similar to their native structures and that remote homologs have been used to predict the ligand binding position. The right prediction of non-WC base pairs in RNA 3D modules as well as the production of optimized and stereochemically correct structures constitute two points of improvement.

## RNA STRUCTURAL ELEMENTS

A general understanding of RNA structure is similar in approach to that of protein structure, from sequence (1D) to secondary structure (2D) to tertiary structure (3D), with a hierarchical and modular integration of the various levels. However, the outstanding nature of pairwise interactions between RNA bases makes RNA structural elements different from the structural elements in proteins. The WC and non-WC base pairs occur through interactions between one of the three edges of each RNA base, leading to 12 pairing families (**Figure 1**). The *cis* WC/WC interaction is the basic element of RNA helical regions, whereas the other 11 non-WC interactions contribute to RNA structural modules and RNA 3D structural elements. Thus, RNA 3D structure is mainly a combination of helical regions (2D) and specific RNA 3D structural modules or elements.

### RNA Bases and Base Pairings

Hydrogen bonds between the bases form the basis of RNA recognition. Three interacting edges can be defined for both purine and pyrimidine bases: the WC edge (used in WC base pairs), Hoogsteen edge, and sugar edge (with the sugar ring linked) (**Figure 1**). All three edges may act as an interacting edge, and a given edge of a base can potentially form a base pair with a second base with any one of the three edges.

Considering the orientation of the glycosidic bonds, a base-pair interaction can be either *cis* or *trans*. Combining both the edge interaction types and the orientation, there are 12 different base-pairing types between any two bases, whereas the standard WC base pair is the *cis* WC/WC interaction. To facilitate the 2D representation of RNA 3D structure, visually accessible and informative 2D symbols have been designed: The circle, square, and triangle stand for WC edge,

---

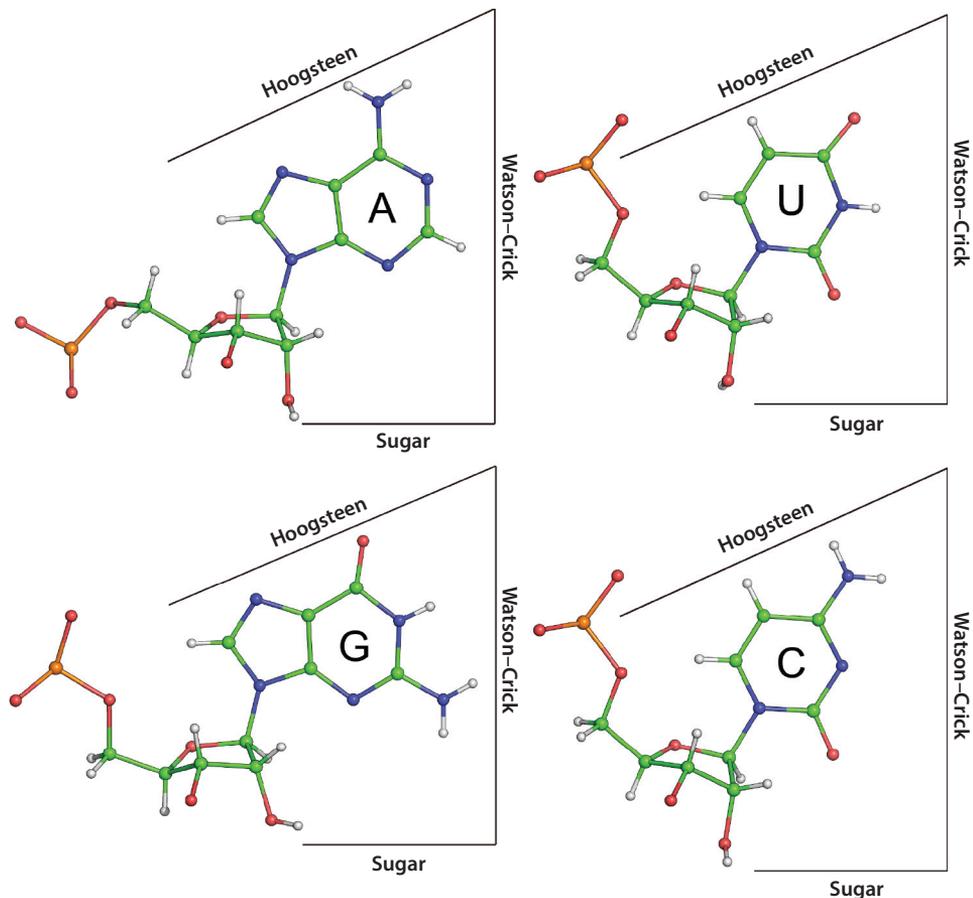
#### RNA module:

a set of ordered non-Watson–Crick (WC) base pairs embedded between WC pairs, which are recurrently observed in RNA families throughout the phylogeny

#### RNA-Puzzles:

a community-wide, blind experiment in RNA 3D structure prediction

---



**Figure 1**

The three edges around the four nucleic acid bases and from which H-bonded pairs can form (adapted from Reference 47).

Hoogsteen edge, and sugar edge, respectively, whereas black-filled shapes are *cis* conformations and unfilled shapes are *trans* conformations (44, 47). Symbolic representations of the 12 interaction types are shown in **Figure 2**, in which real structures solved in recent years are mapped.

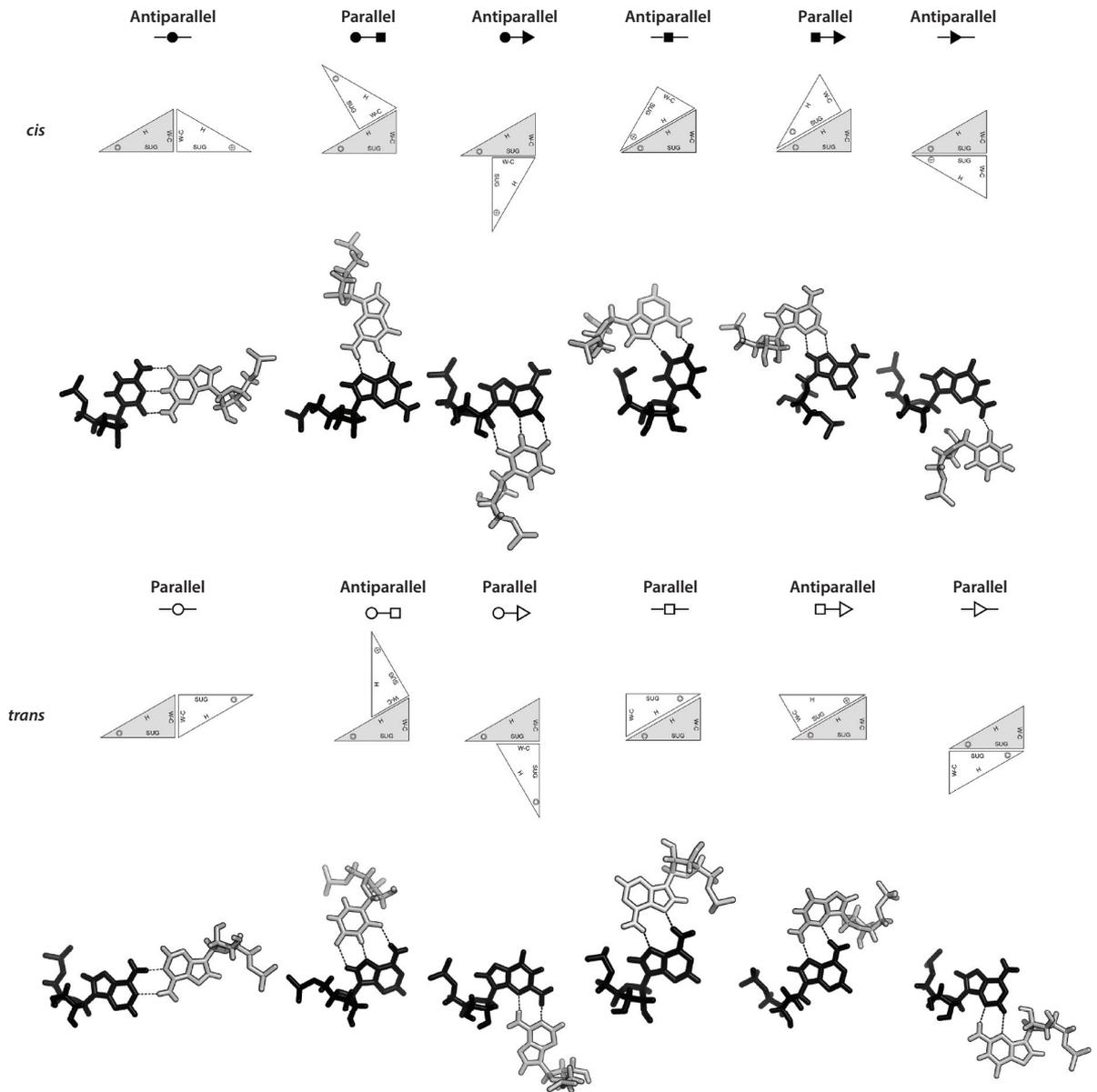
### RNA secondary structure:

an essentially two-dimensional representation of RNA in terms of its intramolecular, nested base-pairing interactions that form stems and loops

## 2D Structure and 3D Structural Modules

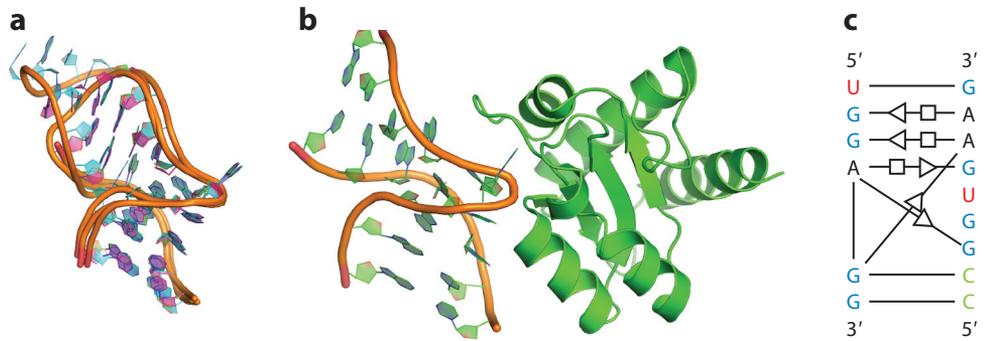
RNA secondary structures refer to the base-pairing interactions within a single molecule and are determined by the standard WC base pairs (*cis* WC/WC) of A-U and C = G base-base interactions (with possible G · U wobble pairs). The 2D structure is a good simplification of the important WC base pairings, which can be used to describe some of the structural elements, such as the double helix, hairpin loop, internal loop, and junction.

As the 11 cases of non-WC base-base interactions are not part of the secondary structure, a 2D structure generally displays internal loops. As in protein loops, most of the RNA loops are more structured than random conformation and can significantly contribute to their molecular functions (48). For example, an internal loop structure can actually be a kink-turn (37) structure



**Figure 2**

The twelve base-pair families that can form via two H-bonds between the three base edges. The circle, square, and triangle stand for the Watson–Crick, Hoogsteen, and sugar edges, respectively. Both *cis* (black-filled) and *trans* (gray, unfilled) conformations are illustrated (47). The relative orientations of the two nucleic acid strands at the base pairs can be either antiparallel (as in standard nucleic acid helices) or parallel, assuming that the conformation of the base with respect to the sugar is *anti* and the sugar-phosphate backbone is in the standard helical conformation (double *gauche*-minus for the phosphate bonds and *trans* for the C4'–C5' bond) (47). The name of the base pair gives immediately the types of H-bonds between the two bases. Other less precise names exist; for example, a sheared base pair between A and G is equivalent to a *trans* Hoogsteen/sugar edge between A and G.



**Figure 3**

The K-turn (37) as an example of an RNA module. (a) Superimposition of K-turns from *Homo sapiens* (green), *Haloarcula marismortui* (cyan), and *Archaeoglobus fulgidus* (magenta). (b) U4 K-turn interacting with the 15.5-kD protein (102). (c) Two-dimensional annotated diagram of a classic K-turn module (48).

and function as an L7Ae-like protein binding region; structural and schematic contacts are shown in **Figure 3**. Therefore, non-WC base pairs are crucial to RNA structure and may determine the RNA 3D topologies in hinge regions. This also explains why RNA 3D structures are difficult to build based solely on 2D structure. RNA 3D modules are recurring 3D building blocks, mainly constituted by non-WC base pairs, of similar structure and function occurring in various functional RNA molecules. As RNA secondary structure is insufficient to describe the unique structure of an RNA molecule, RNA 3D modules can greatly complement the deficiency of 2D structure in depicting molecular details. Here, we briefly review some crucial RNA structural modules.

### Example of an RNA Module: The G-Bulge

The sarcin/ricin module was first observed in the 5S ribosomal RNA (rRNA) of eukaryotes (12, 13, 45, 46, 96). The sarcin/ricin loop (SRL) includes a G-bulge module, in which a guanosine is extrahelical bulged and forms a non-WC interaction with its neighboring uridine. Nine bases and five non-WC base pairs that are surrounded by WC base pairs form the center of the sarcin/ricin module. As a recurrent structural module, the SRL has been found in different rRNAs, namely, 5S rRNA and 23S rRNA. The SRL is functionally essential as it is targeted by cytotoxins such as  $\alpha$ -sarcin and ricin that completely abolish translation. A recent study points out that SRL is critical for anchoring elongation factor G (EF-G) on the ribosome during the process of mRNA-tRNA translocation (91).

The bacterial loop E is another type of module, which is found in bacterial 5S rRNAs. The loop E region specifically binds ribosomal protein L25 in the conserved region. The loop E module is a symmetrical loop. According to the crystal structure of *Escherichia coli* 5S rRNA, the loop E module includes seven non-WC base pairs forming a severely distorted double helix. The structural consensus of this module is constituted by three key base pairs: *trans* Hoogsteen/sugar edge, *trans* WC/Hoogsteen edge or *trans* sugar/Hoogsteen edge, and *cis* bifurcated or *trans* sugar/Hoogsteen edge. In such local structures, the base edges face the grooves in different orientations than in standard WC pairs and thus form novel interaction interfaces with protein side chains.

### Pseudoknots

A pseudoknot (94, 107) is a structure that contains at least a hairpin loop and a single-stranded complementary sequence outside the loop that forms WC base pairs with the hairpin. A stricter

definition is, For two paired nucleotide positions  $(i, k)$ , there exists another base pair  $(m, n)$  that fulfills  $i < m < k, n < i, \text{ or } n > k$  (28). Pseudoknots, therefore, belong fundamentally to the 3D structure of RNA and are beyond typical 2D structure prediction. Pseudoknots play crucial roles in various types of biological processes. First recognized in the turnip yellow mosaic virus in 1982 (81), pseudoknots have been found in all classes of RNA. The H-type (hairpin loop, or classic) pseudoknot is the simplest type of pseudoknot. Either the leading or the leaving strand to the stem of a hairpin structure forms base pairs with the loop region of the hairpin to form the H-type pseudoknot. Among all known structures of pseudoknots, H-type pseudoknots are the most abundant and have been best characterized.

Due to the nonnested nature of the base pairs, it is difficult to computationally detect pseudoknot structures according to sequence information. The prediction of lowest free energy structures considering pseudoknots has proved to be NP complete (56, 57). In addition, metal ion binding also stabilizes RNA pseudoknots (109). In the scheme of popular RNA secondary structure prediction programs such as Mfold (118) and Pfold (38), pseudoknot structures cannot be predicted, but the most stable structure of the two pseudoknotted stems can be predicted. Using dynamic programming (82), it is now possible to identify a limited though not exhaustive class of pseudoknots, but the prediction is also sequence length dependent (20).

### Loop–Loop Interactions

In general, loop–loop interactions are formed by base pairing between the loop regions of two hairpin structures or between an internal loop (or single-stranded region) and a hairpin loop. Formally, they are pseudoknotted structures when they occur within a single RNA strand (i.e., intramolecularly). However, intermolecular loop–loop interactions constitute a frequent dimerization contact between two RNA molecules (26). The pairs are not necessarily all of the *cis* WC type. The overall topology of two interacting kissing hairpin loops is that of a coaxial stack of the three helices. Only a few intermolecular WC base pairs can be strong enough to stabilize the complex structure (9).

Various types of kissing loop complex structures have now been identified: the dimerization initiation site of human immunodeficiency virus type 1 (HIV-1) (41, 67, 71, 92), the Moloney murine leukemia virus H3 stem-loop (34), and the complex between RNA I and RNA II stem-loops of the ColE1 plasmid of *E. coli* (59), between yeast phenylalanine tRNA and RNA aptamers (88), between HIV *trans*-activating responsive (TAR) hairpin loop and its complement (8), between TAR and RNA hairpin aptamers (42), and between TAR and DNA aptamers (10). The number of residues in the linker fragments between two stems and the number of residues in the central duplex contribute to the main differences among these complexes. Such loop–loop interactions are of special importance in RNA structures because they are functionally critical or essential for RNA folding.

### Ion and Ligand Binding

Metal ions are necessary for efficient and biologically relevant RNA folding (2, 21, 22), stability, and various biological functions. Specifically, the positive charge of metal cations can be used to compensate for the negative charge of the RNA phosphate backbone. However, the functions of metal ions are not limited to the neutralization of charges, as metal ions also bind to very specific locations on RNA 3D structures. Further, metal ions have been found to directly mediate catalysis in some ribozymes (86). It is therefore essential to include hydrated ions, such as sodium, potassium, and magnesium, in any consideration of RNA 3D structure.

Aside from ions, many RNA molecules require ligands to function, especially the riboswitches and some ribozymes. Ligand binding may change the conformations of riboswitches and result in functional changes, that is, to regulate gene expression. Examples include the well-known Purine riboswitch (27), TPP riboswitch (100), SAM riboswitch (65), FMN riboswitch (31, 90), and *glmS* ribozyme (36). Riboswitches solved in recent years include the glutamine riboswitch (80), ZTP riboswitch (101), guanine riboswitch (30), and *ydaO* riboswitch (78).

## RECENTLY SOLVED RNA STRUCTURES

Our overall knowledge of RNA structure strongly depends on the number of solved RNA structures. Currently, >110,000 protein structures have been deposited in the Protein Data Bank (83). But only approximately 1,100 RNA structures and approximately 2,000 RNA-related complex structures are available. According to the latest release of the Rfam database (68), version 12.1, 2,474 RNA families have been reported but only 75 of them include 3D structural information. This ratio points to our limited knowledge of the RNA 3D structure sequence space. Still, of the 1,100 known RNA structures, >330 of them were solved after 2011. Some recently solved RNA structures have shed new light on RNA 3D structures.

The fluoride riboswitch structure (79), also known as the *crcB* RNA motif, and its encapsulation of fluoride ion with  $Mg^{2+}$  ions significantly contributed to our comprehension of RNA architecture, folding, and recognition. The twister ribozyme RNA (51) adopts a novel compact fold based on a double-pseudoknot structure forming an active site at its center. One WC and three non-WC base pairs stabilize the core of the ribozyme. This underscores the importance of non-WC base pairs and pseudoknot structures, as well as their functional importance. The guanine riboswitch (3, 30) shows a structural change during ligand binding. The *yypP-ykoY* orphan riboswitch (75) demonstrates a  $Mn^{2+}$ -sensing function with tertiary RNA–RNA interaction. The SRP Alu domain (33) includes a minor-saddle motif and an extended loop–loop pseudoknot. And the core of the Spinach RNA aptamer (104) is a three-tetrad quadruplex composed of two G-quartets stacked above a mixed-sequence tetrad and stabilized by two  $K^+$  ions.

## MODULE PREDICTION PROGRAMS

Computational efforts (15, 77, 98, 99, 114, 116) have been targeted at the prediction of 3D structural modules based on sequence or sequence alignment information. As discussed above, the prediction of RNA modules is beneficial for RNA secondary structure prediction (99) and, especially, tertiary structure prediction (5, 77).

RMDetect (15) was a pioneer program in RNA 3D module prediction and correctly identifies known 3D structural modules in single and multiple RNA sequences in the absence of any other information. RMDetect encodes a Bayesian probabilistic network for each structural module, integrating non-WC interaction network information derived from 3D structure and sequence alignment information. Then, it evaluates the probability of a new sequence or sequence alignment including the same module by threading the sequence into the Bayesian network using a slide-window approach. For example, RMDetect correctly identifies standard RNA modules like G-bulge, K-turn, C-loop, and tandem GA.

RMDetect offers the freedom to define the non-WC interaction network of an RNA module. It is possible to integrate prior experience to enforce certain important interactions. With a default interaction network derived from RNA 3D structures, RMDetect can be applied to large-scale tests. Recently, the metaRNAmodule (98) pipeline combined the RNA 3D Motif Atlas (74), a structure database of RNA modules generated by the FR3D (87) program; Rfam (68), an RNA

sequence alignment database; and RMDetect. The metaRNAmodule pipeline automatically builds an interaction network using the structural information from the RNA 3D Motif Atlas and maps to Rfam alignment information to derive Bayesian network models for several RNA modules. Based on large-scale tests, RMDetect identified 977 internal loops and 17 hairpin modules with clear discriminatory power in 35 unique locations in 11 different RNA families, demonstrating a successful result for RNA 3D module prediction.

JAR3D (116) is another automation of RNA module prediction that derives RNA modules from the RNA 3D Motif Atlas and Rfam, but JAR3D encodes RNA modules into a hybrid stochastic context-free grammar and Markov random field (SCFG/MRF) model for prediction. Considering the base-pairing nature of RNA structure, JAR3D represents RNA structure using SCFG. And MRF is used to model base triples and nonnested base pairs.

The prediction of RNA 2D structure typically considers these RNA modules as internal loops, because non-WC base pairs in RNA modules are not as a rule considered in 2D structure. Therefore, the identification of RNA modules can help determine the correct internal loops in 2D structure and thus improve 2D structure prediction. JAR3D and the metaRNAmodule were applied to genome-wide data (99), and the false discovery rate of RNA 2D structure prediction was significantly improved with the integration of RNA 3D module prediction information. RNA-MoIP (77) employed a similar idea in improving 2D structure prediction but extended the improvements to optimal 2D structure identification and RNA 3D structure prediction.

## ASSESSMENT OF RNA 3D STRUCTURE PREDICTION

With the efforts of computational biologists, we are now seeing a wealth of 3D RNA structure modeling approaches. Examples include FARNA (16), SimRNA (5, 58), ModeRNA (84), RNA-Composer (4, 76), MC-Fold/MC-Sym (73), MMB (25), NAST (32), DMD (19), Vfold (110), iFoldRNA (39), DCA (18), and EC\_RNA (106), which are discussed in detail in the article “Theory and Modeling of RNA Structure” in this volume (95). The evaluation of predicted models in the context of crystal structures as references to assess prediction methods is a nontrivial task in understanding how well a 3D prediction method can perform.

### Evaluation Metrics

As 3D RNA structure prediction is inspired by protein structure prediction, some routine structural comparison metrics for protein structures have been adopted. The root-mean-square deviation (RMSD) measures the distance deviation of each atom pair after rigid body superimposition. The CAD-score (69, 70) measures the structural similarity in a contact-area difference-based function, whereas the template modeling score (TM-score) (113) uses the Levitt–Gerstein (50) function to estimate similarity by superimposition based on maximum local similar fragment. And the global distance test (GDT) (111) adopted from the CASP (Critical Assessment of Methods of Protein Structure Prediction) (66) experiment calculates the largest set of amino acid residue  $\alpha$ -carbon atoms in the model structure that fall within a defined distance cut-off of their position in the experimental structure. Nevertheless, these rigid body scores ignore the fact that RNA structure is determined more by base stacking and base pairing than by backbone paths. The mean of circular quantities (MCQ) (117) is an attempt to overcome this issue by estimating structural similarity based on a torsion angle space representation of 3D structure.

But the essential base-pairing nature of RNA structures is still not considered by the above metrics. Therefore, the interaction network fidelity (INF) and deformation index (DI) were invented to assess the central characteristics of RNA architecture: the network of intramolecular

contacts, of WC and non-WC hydrogen bonds, and of stacking. The INF (72) is the Matthews correlation coefficient of interaction prediction, which can be formulated as

$$\text{INF} = \sqrt{\frac{\text{TP}}{\text{TP} + \text{FP}} \times \frac{\text{TP}}{\text{TP} + \text{FN}}},$$

where TP is the number of true positives, or correctly predicted contacts, and FP and FN are false positives and false negatives, respectively. The INF can either measure different interaction types (WC base pairing, non-WC base pairing, and base stacking) separately or combine all of the types, resulting in four metrics: INF<sub>WC</sub>, INF<sub>nWC</sub>, INF<sub>stacking</sub>, and INF<sub>all</sub>. The DI is the ratio between RMSD and INF. The deformation profile (DP) (72) measures pairwise local structure similarity at the nucleotide scale using a matrix rather than a single score for the whole structure. The DP can be presented as a 2D heat map, mapping to the RNA sequence, that illustrates which parts of the structure are more similar to the native structure or which long-distance interaction is incorrect (Figure 4). Finally, MolProbity (17) is a standard protocol to check atomic distances in the RNA structure to avoid unreasonable atomic clashes. The *P*-value (29) estimates the probability that a predicted model is better than that expected in prediction. A brief summary of these metrics is given in Table 1.

As these evaluation metrics compare different structural aspects, there is no single metric that can be used to make an overall judgment. For example, the TM-score uses one-atom coordinates to represent a residue and cannot describe whether the direction of the base is correct, whereas RMSD stands for an all-atom similarity, and the INF best represents the interaction network prediction. Consequently, it is better to evaluate a structure with several of these metrics to show different aspects. Additionally, RMSD is still a widely used metric to assess the general similarity between a predicted model and the native state. In RNA-Puzzles (14, 63), the RMSD, INF, DI, MCQ, MolProbity, and *P*-value are used and models are ranked by RMSD. In comparison with protein structures that involve only two flexible dihedral angles in the backbone, RNA has six dihedral angles, resulting in much larger degrees of freedom. Therefore, the RMSD of RNA structure is generally larger than that of protein models. For protein structures, an RMSD >15 Å would not be so meaningful, but for RNA structures, an RMSD <20 Å can still be meaningful in evaluation. Most of the abovementioned metrics have been separately implemented. RNAnalyzer (54) and RNAssess (55) are a program and a website, respectively, that in a practical way have implemented many of these metrics, simultaneously providing useful demonstrations and plots of the results.

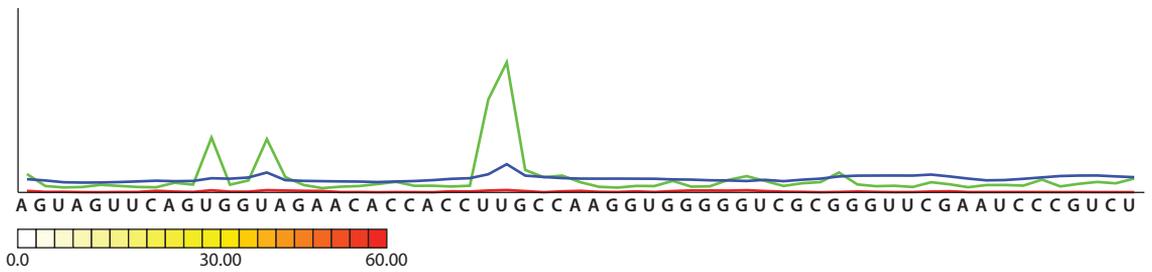
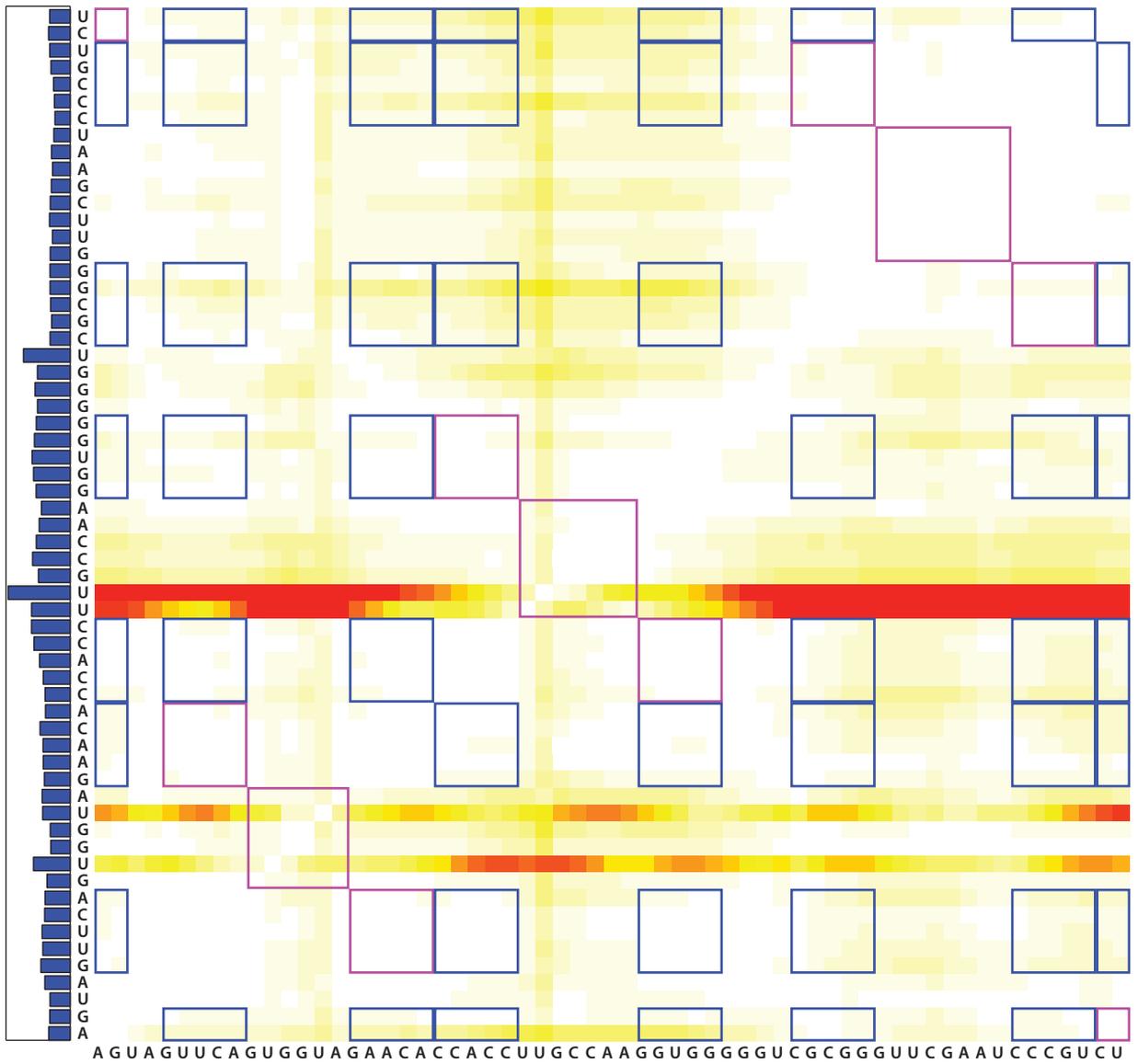
## RNA-Puzzles

Although the prediction of RNA structure dates back to the last century (49, 60, 64, 108), the problem is far from being completely solved. The development of novel prediction approaches is typically based on publicly available RNA structure data sets, and the results are sometimes presented in an optimistic but uncritical style. Without a blind data test, a prediction is likely to be postdiction. RNA-Puzzles (14, 63) is a community-wide series of blind trials in 3D RNA

---

**Figure 4**

Deformation profile (72) plotted together with B factors (from Reference 112). Blue and pink squares inside the matrix correspond to intra- and interdomain similarity relationships, respectively. Color scale goes from 0 Å (white) to (but not including) 20 Å (dark green) in ten equal steps and from 20 Å (yellow) to 80 Å (red) in five equal steps. Below the matrix, the average values of rows (green), columns (blue), and main diagonal (red) of the matrix are plotted.



**Table 1** Summary of the available evaluation metrics for RNA structures

Metrics	Structure	Summary	Annotation
RMSD	All atoms	Atom-wise deviation based on rigid superimposition	General metric
CADscore	All atoms	Assessment of the accuracy of interdomain or intersubunit interfaces	Adopted from protein
TMscore	Representative atom for each residue (P)	Deviation based on largest superimposed fragment	
GDTscore		Largest set of C $\alpha$ fall within a cut-off referred to as experimental structure	
MCQscore	Dihedrals	Dihedral angle deviations	RNA-oriented
INF (INFwc, INFnwc, INFstacking)	Residue interactions	Matthews correlation coefficient to measure accuracy in the prediction of key interaction networks	
DI, DP	All atoms	Average distance deviations based on residue pair superimpositions	
MolProbity	All atoms	Geometrical reasonableness of bonds, angles, and contact distances	General metric
<i>P</i> -value	All atoms	Prediction quality based on structure length and RMSD	RNA-oriented

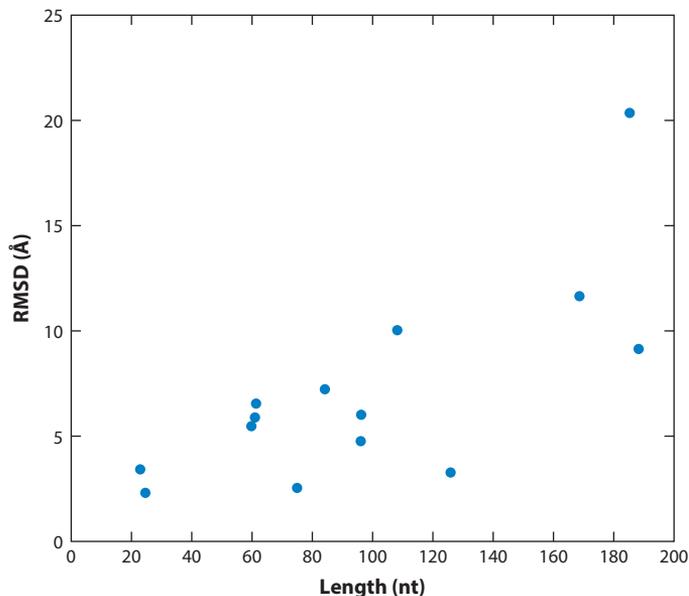
Abbreviations: DI, deformation index; DP, deformation profile; GDT, global distance test; INF, interaction network fidelity; MCQ, mean of circular quantities; nwc, non-Watson-Crick; RMSD, root-mean-square deviation; TM, template modeling; wc, Watson-Crick.

structure prediction employing a method similar to that of CASP (66). Experimentally determined RNA structures, especially crystal structures, are collected prior to the publication of the structure coordinates. Predictors around the world make predictions within a constrained time period based on the given sequence information. The predicted structures are compared with experimental structures after the publication of the RNA structures.

RNA-Puzzles aims to (a) determine the capabilities, limitations, and recent progress of current RNA 3D structure predictions; (b) identify the bottlenecks that hold back the field and how to solve them; (c) promote the application of RNA structure prediction in solving real-world biological problems; and (d) encourage the development and improvement of automated prediction tools.

Since 2011, 14 groups around the world have tackled 17 Puzzles. With the emergence of new prediction methods in recent years, more and more new groups are joining this effort. Further, additional groups developing assessment tools actively participate in the evaluation efforts. The targets in RNA-Puzzles are biologically significant, including riboswitches, ribozymes, and RNA complexes. According to the results, the lengths of the RNA sequences are <200 nt, whereas the best prediction RMSD ranges from 2.3 to 20 Å. Considering the assessment results, the majority of the best-predicted models are already quite similar to native structures. In general, approximately 90% of the WC interactions in native structures are predicted correctly in the best prediction models, and stacking prediction has also achieved a good level of success. But the non-WC interaction predictions vary greatly among the puzzles.

Comparing the sequence length with the best RMSD values (Figure 5), we find a good correlation (Pearson correlation coefficient, ~0.74), an observation that illustrates that longer RNA structures are still more difficult to predict. On the one hand, longer RNA structure can occupy



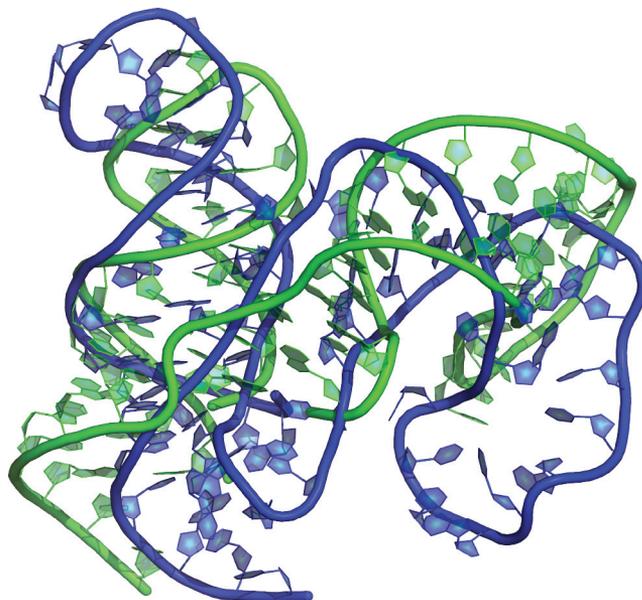
**Figure 5**

Distribution between sequence length and the best-predicted root-mean-square deviation (RMSD) for cases computed in RNA-Puzzles (14, 63).

a much larger volume and small variations close to the center of mass can have huge effects at the periphery. On the other hand, rigid comparison between native structure and predictions using RMSD introduces overlap errors in correct regions (because of the averaging over the whole molecule) and does not include consideration of structural flexibility.

Beyond the size of the RNA structure, prediction quality is most affected by our prior knowledge about the RNA structure. Human insight into some of the important base-pair interactions or ligand binding may greatly facilitate prediction. Various types of prior knowledge can be derived (for example, by sequence alignments or secondary structure homologies, or additional experimental data). Thus, when a homologous structure is already known, the prediction accuracy can be greatly improved by homology modeling. Chemical probing data and sequence alignment-derived evolutionary clues are also critical to directing the predictions. To assess the effectiveness of fast-track experiments and the capability of automated RNA 3D structure prediction web servers, RNA-Puzzles started three categories of predictions after Puzzle 14: (a) automated web servers that predict RNA structure without any human intervention (with a 48-hour time limit for web servers); (b) human expert prediction without experimental data, or pre-experiment predictions; and (c) human expert modeling with experimental data, or post-experiment predictions. In this way, RNA-Puzzles attempts to quantify the efficacy of human manipulation and of experimental data.

An interesting case is Puzzle 14 for which the free state modeling was provided by the Ding group (62) as a post-experiment prediction (**Figure 6**): The model has an INFwc score of 1.0, which means that all the WC interactions were predicted correctly. But an INF<sub>nwc</sub> score cannot be calculated as no non-WC interaction was predicted. This model is the best reflection of RNA 2D structure. It has a RMSD of 10 Å, and there are 15 other models that were better predictions than this one in terms of RMSD. This offers us a concrete example of how well RNA 2D structure



**Figure 6**

Predicted structure of a 100% correct two-dimensional structure (*blue*) compared with a native three-dimensional structure (*green*) for Puzzle 14 on the L-glutamine riboswitch (80).

can represent real RNA 3D structure and how important the contribution of non-WC interactions is to the final RNA 3D structure. Non-WC interactions are essential in determining RNA 3D structure, whereas RNA 2D structure, although forming inescapable building blocks of the final architecture, represent only isolated or disconnected helical regions when not many non-WC interactions or RNA modules exist in the structure.

According to **Table 2**, the clash scores of the best RMSD models range from 0 to 23, with many models having values greater than 10. This points out that atomic clashes are still very often observed in prediction models. Still, we also find atomic clashes in native structures; for instance, the crystal structure of Puzzle 5 has a clash score of 5.86. This indicates the necessity to update the refinement and prediction dictionaries of distances and angles for RNA structure toward common and accepted values.

To summarize, progress in RNA 3D structure prediction has been achieved during recent years. The prediction of WC base-pair interactions has already achieved nice accuracies with the help of sequence alignment and experimental probing. However, our knowledge of RNA structure is still limited by the number of known RNA structures compared with the number of known RNA sequences. The current bottlenecks in RNA 3D structure prediction lie at two points: the accurate recognition of non-WC interactions and RNA modules, and the optimization of atomic clashes within the constraints of the topology and compactness of RNA structures.

## CHALLENGES AND PERSPECTIVES

Prior to the prediction of any RNA structure, we need to evaluate the number of known structured RNAs. According to the Rfam database, only a very small number of RNA families have been solved for structure. Therefore, our knowledge of RNA structure in general is limited by what we know about the 3D structure of RNA.

**Table 2** A list of released RNA-Puzzles and the best predictions. Results for Puzzles 1, 2, 3 are from (14); results for Puzzles 5, 6, 10 are from (63); results for Puzzles 4, 8, 12, 13, 14 are from (62)

Puzzle	Name	Length (nt)	BestRMSD (Å)	DI all	INF all	INF wc	INF nwc	INF stacking	Clash Score
1	thymidylate synthase mRNA	23	3.4	3.657	0.934	0.953	-	0.924	0
2	Self-assembling RNA square	15+10	2.3	2.846	0.808	0.875	0	0.798	11.76
3	Glycine Riboswitch	84	7.241	9.819	0.737	0.857	0	0.731	0
4	SAM-I riboswitch aptamer	126	3.257	3.472	0.938	0.989	0.667	0.935	2.22
5	group I intron	188	9.152	12.019	0.761	0.906	0.334	0.751	6.79
6	adenosylcobalamin riboswitch	168	11.699	16.151	0.724	0.885	0.316	0.702	23.48
7	Varkud satellite ribozyme	185	20.370	27.245	0.748	0.895	0.105	0.726	10.68
8	SAM-I/IV riboswitch	96	4.801	5.651	0.850	0.985	0.577	0.816	13.8
10	T-box complex	96+75	6.803	8.365	0.813	0.946	0.700	0.786	11.09
	T-box	96	6.021	7.478	0.805	0.906	0.624	0.804	10.73
	tRNA	75	2.505	2.767	0.905	0.972	0.913	0.883	1.45
12	<i>ydaO</i> riboswitch	108	10.061	14.453	0.696	0.861	0.000	0.663	12.61
13	ZMP riboswitch	60	5.410	7.032	0.769	0.905	0.258	0.755	10.85
14	L-glutamine riboswitch (Free state)	61	6.513	8.276	0.787	0.884	0.354	0.805	10.66
	L-glutamine riboswitch (Bound state)	61	5.879	7.926	0.742	0.918	0.577	0.701	11.18

Abbreviations: see **Table 1**.

The structure space of RNA is much more intricate than the structure space limited to double-stranded helices, and more and more new RNA cases have been explored, including multiple ligand binding, ligand binding-induced conformational change, and ion-stabilized catalytic core. Many of the specific 3D structures of RNA are composed of non-WC base pairs. As a very important type of RNA structural element, non-WC base pairs greatly contribute to RNA modules; they are stable and recurrent in RNA structures. **Figure 6** highlights the fact that secondary structure is far from sufficient to describe real RNA structures without noting non-WC base pairs. Overlooking non-WC base pairs is a primary hindrance in current RNA structure prediction.

With the emergence of RNA module prediction programs, RNA secondary and 3D structure prediction have been promoted to an extent. However, there is still a gap between the prediction of RNA modules and their insertion within an overall RNA architecture, especially in the framework of automatic software prediction programs. Researchers have started to capture some of base-pair interactions with evolutionary coupling (106), but most of these are WC base pairs; non-WC base pairs are still rarely predicted. Nevertheless, direct coupling analysis is an attempt to capture some functional relevance from aligned homologous RNA sequences, for example, in RNA-protein or RNA-ligand interactions. Although protein sequences tend to be conserved at the RNA binding

interface, these protein binding residues, together with the interacting RNA nucleotides, form an interaction network to maintain functionality and demonstrate certain signals. A good example is the interaction between ribosomal 5S RNA and the ribosomal protein L25. The L25 protein is relatively conserved on its RNA binding interface, binding strictly to the loop E module of the rRNA. The loop E module of RNA can vary in sequence but sustains the same structure in binding.

Chemical probing of RNA structure has already been shown to improve RNA secondary structure determination. And, according to the results from RNA-Puzzles, these fast-track experiments continuously provide insightful information on RNA 3D structure prediction, thus promoting our understanding of RNA structure as a whole. In particular, some important 3D RNA–RNA interactions and non-WC base pairs, which are important to RNA 3D topology, may be determined through such experimental data.

Considering the structural differences between RNA and protein as well as the flexibility, degrees of freedom, and base pairing between side chains, it is a critical issue to compare RNA structures, and in particular, the non-WC base pairs. A number of assessment metrics that compare different aspects of RNA structure has been summarized herein. RNA-Puzzles has provided a good community-wide platform for RNA 3D structure prediction. The current results of RNA-Puzzles have highlighted some crucial facts and bottlenecks in RNA 3D structure prediction:

- Without non-WC base pairs, RNA 2D structure does not provide enough information for assembling 3D architecture.
- Non-WC base pairs and RNA loop–loop interaction constitute main bottlenecks of current predictions.
- The ranking of the set of predicted structures is not systematically in congruence with the metrics used, even the RMSDs.
- To consider ligand or ion binding in RNA structure prediction is a necessity in RNA structure prediction in order to understand its functions.
- Conformational changes in riboswitches and ribozymes constitute a future task for RNA 3D structure prediction.
- The prediction of stereochemically correct structures with reasonable atomic distances and angles is necessary for RNA structure optimization.

Several of the present-day deficiencies revealed by RNA-Puzzles in the RNA prediction programs (like more accurate ranking of the solutions or better clash scores) would require novel approaches and software implementations because they are at the crossroads between local and global methodologies and concepts. The same conflict between local and global similarities undermine the development of new criteria or metrics for assessing the comparisons between the predicted and experimental structures. A single overall assessment metric is still not available. Finally, even with an approximate 3D structure, we are still very far away from deducing the function of an RNA molecule, except by sequence alignments and homology deductions. The fundamental characteristics of RNA, the very same ones that make RNA an extraordinary vector and engine of selection and biological evolution, such as the great diversity of weak interactions, their frequent interchangeability or neutrality, and the ease of molecular adaptations and adjustments exacerbate the difficulties in predicting the foldings and molecular functions of RNA molecules. Systematic and detailed comparisons between sequences and homologous structures reveal empirical rules, albeit extremely complex to formalize—hence, the importance of continuing to develop and improve prediction programs because they are intended to integrate all of our knowledge on this fascinating RNA molecule.

## DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

## ACKNOWLEDGMENTS

Z.M. was supported by a French government grant (ANR-10-BINF-02-02 BACNET) to E.W.

## LITERATURE CITED

1. Atkins J, Gesteland RF, Cech TR, eds. 2011. *RNA Worlds: From Life's Origins to Diversity in Gene Regulation*. Cold Spring Harbor, NY: Cold Spring Harb. Lab. Press
2. Auffinger P, Grover N, Westhof E. 2011. Metal ion binding to RNA. *Met. Ions Life Sci.* 9:1–35
3. Batey RT, Gilbert SD, Montange RK. 2004. Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature* 432:411–15
4. Biesiada M, Pachulska-Wieczorek K, Adamiak RW, Purzycka KJ. 2016. RNAComposer and RNA 3D structure prediction for nanotechnology. *Methods* 103:120–27
5. Boniecki MJ, Lach G, Dawson WK, Tomala K, Lukasz P, et al. 2016. SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction. *Nucleic Acids Res.* 44:e63
6. Caprara MG, Nilsen TW. 2000. RNA: versatility in form and function. *Nat. Struct. Biol.* 7:831–33
7. Cech TR, Steitz JA. 2014. The noncoding RNA revolution—trashing old rules to forge new ones. *Cell* 157:77–94
8. Chang KY, Tinoco I Jr. 1997. The structure of an RNA “kissing” hairpin complex of the HIV TAR hairpin loop and its complement. *J. Mol. Biol.* 269:52–66
9. Chen Y, Varani G. 2001. RNA structure. *eLS*. doi: 10.1002/9780470015902.a0001339.pub2
10. Collin D, van Heijenoort C, Boiziau C, Toulme JJ, Guittet E. 2000. NMR characterization of a kissing complex formed between the TAR RNA element of HIV-1 and a DNA aptamer. *Nucleic Acids Res.* 28:3386–91
11. Cordero P, Das R. 2015. Rich RNA structure landscapes revealed by mutate-and-map analysis. *PLOS Comput. Biol.* 11:e1004473
12. Correll CC, Freeborn B, Moore PB, Steitz TA. 1997. Metals, motifs, and recognition in the crystal structure of a 5S rRNA domain. *Cell* 91:705–12
13. Correll CC, Munishkin A, Chan YL, Ren Z, Wool IG, Steitz TA. 1998. Crystal structure of the ribosomal RNA domain essential for binding elongation factors. *PNAS* 95:13436–41
14. Cruz JA, Blanchet MF, Boniecki M, Bujnicki JM, Chen SJ, et al. 2012. RNA-Puzzles: a CASP-like evaluation of RNA three-dimensional structure prediction. *RNA* 18:610–25
15. Cruz JA, Westhof E. 2011. Sequence-based identification of 3D structural modules in RNA with RMDetect. *Nat. Methods* 8:513–21
16. Das R, Baker D. 2007. Automated de novo prediction of native-like RNA tertiary structures. *PNAS* 104:14664–69
17. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, et al. 2007. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* 35:W375–83
18. De Leonardi E, Lutz B, Ratz S, Cocco S, Monasson R, et al. 2015. Direct-coupling analysis of nucleotide coevolution facilitates RNA secondary and tertiary structure prediction. *Nucleic Acids Res.* 43:10444–55
19. Ding F, Sharma S, Chalasani P, Demidov VV, Broude NE, Dokholyan NV. 2008. Ab initio RNA folding by discrete molecular dynamics: from structure prediction to folding mechanisms. *RNA* 14:1164–73
20. Dirks RM, Pierce NA. 2004. An algorithm for computing nucleic acid base-pairing probabilities including pseudoknots. *J. Comput. Chem.* 25:1295–304
21. Draper DE. 2008. RNA folding: thermodynamic and molecular descriptions of the roles of ions. *Biophys. J.* 95:5489–95
22. Draper DE, Grilley D, Soto AM. 2005. Ions and RNA folding. *Annu. Rev. Biophys. Biomol. Struct.* 34:221–43

23. Eddy SR. 2014. Computational analysis of conserved RNA secondary structure in transcriptomes and genomes. *Annu. Rev. Biophys.* 43:433–56
24. Ehresmann C, Baudin F, Mougél M, Romby P, Ebel JP, Ehresmann B. 1987. Probing the structure of RNAs in solution. *Nucleic Acids Res.* 15:9109–28
25. Flores SC, Sherman M, Bruns CM, Eastman P, Altman RB. 2011. Fast flexible modeling of RNA structure using internal coordinates. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 8:1247–57
26. Forsdyke DR. 1995. A stem-loop “kissing” model for the initiation of recombination and the origin of introns. *Mol. Biol. Evol.* 12:949–58
27. Gilbert SD, Stoddard CD, Wise SJ, Batey RT. 2006. Thermodynamic and kinetic characterization of ligand binding to the purine riboswitch aptamer domain. *J. Mol. Biol.* 359:754–68
28. Gulyaev AP, Olsthoorn RCL, Pleij CWA, Westhof E. 2001. RNA structure: pseudoknots. *eLS*. doi: 10.1002/9780470015902.a0003134.pub2
29. Hajdin CE, Ding F, Dokholyan NV, Weeks KM. 2010. On the significance of an RNA tertiary structure prediction. *RNA* 16:1340–49
30. Hernandez AR, Shao Y, Hoshika S, Yang Z, Shelke SA, et al. 2015. A crystal structure of a functional RNA molecule containing an artificial nucleobase pair. *Angew. Chem. Int. Ed.* 54:9853–56
31. Howe JA, Wang H, Fischmann TO, Balibar CJ, Xiao L, et al. 2015. Selective small-molecule inhibition of an RNA structural element. *Nature* 526:672–77
32. Jonikas MA, Radmer RJ, Laederach A, Das R, Pearlman S, et al. 2009. Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters. *RNA* 15:189–99
33. Kempf G, Wild K, Sinning I. 2014. Structure of the complete bacterial SRP Alu domain. *Nucleic Acids Res.* 42:12284–94
34. Kim CH, Tinoco I Jr. 2000. A retroviral RNA kissing complex containing only two G·C base pairs. *PNAS* 97:9396–401
35. Kladwang W, VanLang CC, Cordero P, Das R. 2011. A two-dimensional mutate-and-map strategy for non-coding RNA structure. *Nat. Chem.* 3:954–62
36. Klein DJ, Ferré-D’Amaré AR. 2006. Structural basis of *glmS* ribozyme activation by glucosamine-6-phosphate. *Science* 313:1752–56
37. Klein DJ, Schmeing TM, Moore PB, Steitz TA. 2001. The kink-turn: a new RNA secondary structure motif. *EMBO J.* 20:4214–21
38. Knudsen B, Hein J. 2003. Pfold: RNA secondary structure prediction using stochastic context-free grammars. *Nucleic Acids Res.* 31:3423–28
39. Krokhotin A, Houlihan K, Dokholyan NV. 2015. iFoldRNA v2: folding RNA with constraints. *Bioinformatics* 31:2891–93
40. Kubota M, Tran C, Spitale RC. 2015. Progress and challenges for chemical probing of RNA structure inside living cells. *Nat. Chem. Biol.* 11:933–41
41. Laughrea M, Jette L. 1994. A 19-nucleotide sequence upstream of the 5′ major splice donor is part of the dimerization domain of human immunodeficiency virus 1 genomic RNA. *Biochemistry* 33:13464–74
42. Lebars I, Richard T, Di Primo C, Toulme JJ. 2007. NMR structure of a kissing complex formed between the TAR RNA element of HIV-1 and a LNA-modified aptamer. *Nucleic Acids Res.* 35:6103–14
43. Leontis NB, Lescoute A, Westhof E. 2006. The building blocks and motifs of RNA architecture. *Curr. Opin. Struct. Biol.* 16:279–87
44. Leontis NB, Stombaugh J, Westhof E. 2002. The non-Watson–Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res.* 30:3497–531
45. Leontis NB, Westhof E. 1998. The 5S rRNA loop E: chemical probing and phylogenetic data versus crystal structure. *RNA* 4:1134–53
46. Leontis NB, Westhof E. 1998. A common motif organizes the structure of multi-helix loops in 16S and 23S ribosomal RNAs. *J. Mol. Biol.* 283:571–83
47. Leontis NB, Westhof E. 2001. Geometric nomenclature and classification of RNA base pairs. *RNA* 7:499–512
48. Lescoute A, Leontis NB, Massire C, Westhof E. 2005. Recurrent structural RNA motifs, Isostericity Matrices and sequence alignments. *Nucleic Acids Res.* 33:2395–409

49. Levitt M. 1969. Detailed molecular model for transfer ribonucleic acid. *Nature* 224:759–63
50. Levitt M, Gerstein M. 1998. A unified statistical framework for sequence comparison and structure comparison. *PNAS* 95:5913–20
51. Liu Y, Wilson TJ, McPhee SA, Lilley DM. 2014. Crystal structure and mechanistic investigation of the twister ribozyme. *Nat. Chem. Biol* 10:739–44
52. Lorenz R, Wolfinger MT, Tanzer A, Hofacker IL. 2016. Predicting RNA secondary structures from sequence and probing data. *Methods* 103:86–98
53. Loughrey D, Watters KE, Settle AH, Lucks JB. 2014. SHAPE-Seq 2.0: systematic optimization and extension of high-throughput chemical probing of RNA secondary structure with next generation sequencing. *Nucleic Acids Res.* 42:e165
54. Lukasiak P, Antczak M, Ratajczak T, Bujnicki JM, Szachniuk M, et al. 2013. RNAnalyzer—novel approach for quality analysis of RNA structural models. *Nucleic Acids Res.* 41:5978–90
55. Lukasiak P, Antczak M, Ratajczak T, Szachniuk M, Popena M, et al. 2015. RNAssess—a web server for quality assessment of RNA 3D structures. *Nucleic Acids Res.* 43:W502–6
56. Lyngso RB. 2004. Complexity of pseudoknot prediction in simple models. In *Automata, Languages and Programming: 31st International Colloquium, ICALP 2004, Turku, Finland, July 12–16, 2004, Proceedings*, ed. J Díaz, J Karhumäki, A Lepistö, D Sannella, pp. 919–31. Berlin: Springer-Verlag
57. Lyngso RB, Pedersen CN. 2000. RNA pseudoknot prediction in energy-based models. *J. Comput. Biol.* 7:409–27
58. Magnus M, Boniecki MJ, Dawson W, Bujnicki JM. 2016. SimRNAweb: a web server for RNA 3D structure modeling with optional restraints. *Nucleic Acids Res.* 44:W315–19
59. Marino JP, Gregorian RS Jr., Csankovszki G, Crothers DM. 1995. Bent helix formation between RNA hairpins with complementary loops. *Science* 268:1448–54
60. Massire C, Jaeger L, Westhof E. 1998. Derivation of the three-dimensional architecture of bacterial ribonuclease P RNAs from comparative sequence analysis. *J. Mol. Biol.* 279:773–93
61. Mathews DH, Moss WN, Turner DH. 2010. Folding and finding RNA secondary structure. *Cold Spring Harb. Perspect. Biol.* 2:a003665
62. Miao Z, Adamiak RW, Antczak M, Batey RT, Becka AJ, et al. 2017. RNA-Puzzles Round III: 3D RNA structure prediction of five riboswitches and one ribozyme. *RNA*. In press. <https://doi.org/10.1261/rna.060368.116>
63. Miao Z, Adamiak RW, Blanchet MF, Boniecki M, Bujnicki JM, et al. 2015. RNA-Puzzles Round II: assessment of RNA structure prediction programs applied to three large RNA structures. *RNA* 21:1066–84
64. Michel F, 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
65. Montange RK, Batey RT. 2006. Structure of the S-adenosylmethionine riboswitch regulatory mRNA element. *Nature* 441:1172–75
66. Moulton J, Fidelis K, Kryshtafovich A, Schwede T, Tramontano A. 2016. Critical assessment of methods of protein structure prediction (CASP)—progress and new directions in round XI. *Proteins* 84:4–14
67. Muriaux D, Girard PM, Bonnet-Mathoniere B, Paoletti J. 1995. Dimerization of HIV-1<sub>Lai</sub> RNA at low ionic strength. An autocomplementary sequence in the 5' leader region is evidenced by an antisense oligonucleotide. *J. Biol. Chem.* 270:8209–16
68. Nawrocki EP, Burge SW, Bateman A, Daub J, Eberhardt RY, et al. 2015. Rfam 12.0: updates to the RNA families database. *Nucleic Acids Res.* 43:D130–37
69. Olechnovic K, Kulberkyte E, Venclovas C. 2013. CAD-score: a new contact area difference-based function for evaluation of protein structural models. *Proteins* 81:149–62
70. Olechnovic K, Venclovas C. 2014. The CAD-score web server: contact area-based comparison of structures and interfaces of proteins, nucleic acids and their complexes. *Nucleic Acids Res.* 42:W259–63
71. Paillart JC, Skripkin E, Ehresmann B, Ehresmann C, Marquet R. 1996. A loop-loop “kissing” complex is the essential part of the dimer linkage of genomic HIV-1 RNA. *PNAS* 93:5572–77
72. Parisien M, Cruz JA, Westhof E, Major F. 2009. New metrics for comparing and assessing discrepancies between RNA 3D structures and models. *RNA* 15:1875–85

73. Parisien M, Major F. 2008. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* 452:51–55
74. Parlea LG, Sweeney BA, Hosseini-Asanjan M, Zirbel CL, Leontis NB. 2016. The RNA 3D Motif Atlas: computational methods for extraction, organization and evaluation of RNA motifs. *Methods* 103:99–119
75. Price IR, Gaballa A, Ding F, Helmann JD, Ke A. 2015. Mn<sup>2+</sup>-sensing mechanisms of *yypP-ykoY* orphan riboswitches. *Mol. Cell* 57:1110–23
76. Purzycka KJ, Popena M, Szachniuk M, Antczak M, Lukasiak P, et al. 2015. Automated 3D RNA structure prediction using the RNAComposer method for riboswitches. *Methods Enzymol.* 553:3–34
77. Reinharz V, Major F, Waldispuhl J. 2012. Towards 3D structure prediction of large RNA molecules: an integer programming framework to insert local 3D motifs in RNA secondary structure. *Bioinformatics* 28:i207–14
78. Ren A, Patel DJ. 2014. c-di-AMP binds the *ydaO* riboswitch in two pseudo-symmetry-related pockets. *Nat. Chem. Biol* 10:780–86
79. Ren A, Rajashankar KR, Patel DJ. 2012. Fluoride ion encapsulation by Mg<sup>2+</sup> ions and phosphates in a fluoride riboswitch. *Nature* 486:85–89
80. Ren A, Xue Y, Peselis A, Serganov A, Al-Hashimi HM, Patel DJ. 2015. Structural and dynamic basis for low-affinity, high-selectivity binding of L-glutamine by the glutamine riboswitch. *Cell Rep.* 13:1800–13
81. Rietveld K, van Poelgeest R, Pleij CW, van Boom JH, Bosch L. 1982. The tRNA-like structure at the 3' terminus of turnip yellow mosaic virus RNA. Differences and similarities with canonical tRNA. *Nucleic Acids Res.* 10:1929–46
82. Rivas E, Eddy SR. 1999. A dynamic programming algorithm for RNA structure prediction including pseudoknots. *J. Mol. Biol.* 285:2053–68
83. Rose PW, Prlic A, Bi C, Bluhm WF, Christie CH, et al. 2015. The RCSB Protein Data Bank: views of structural biology for basic and applied research and education. *Nucleic Acids Res.* 43:D345–56
84. Rother M, Rother K, Puton T, Bujnicki JM. 2011. ModeRNA: a tool for comparative modeling of RNA 3D structure. *Nucleic Acids Res.* 39:4007–22
85. Rouskin S, Zubradt M, Washietl S, Kellis M, Weissman JS. 2014. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* 505:701–5
86. Roychowdhury-Saha M, Burke DH. 2006. Extraordinary rates of transition metal ion-mediated ribozyme catalysis. *RNA* 12:1846–52
87. Sarver M, Zirbel CL, Stombaugh J, Mokdad A, Leontis NB. 2008. FR3D: finding local and composite recurrent structural motifs in RNA 3D structures. *J. Math. Biol.* 56:215–52
88. Scarabino D, Crisari A, Lorenzini S, Williams K, Tocchini-Valentini GP. 1999. tRNA prefers to kiss. *EMBO J.* 18:4571–78
89. Seetin MG, Kladwang W, Bida JP, Das R. 2014. Massively parallel RNA chemical mapping with a reduced bias MAP-seq protocol. *Methods Mol. Biol.* 1086:95–117
90. Serganov A, Huang L, Patel DJ. 2009. Coenzyme recognition and gene regulation by a flavin mononucleotide riboswitch. *Nature* 458:233–37
91. Shi X, Khade PK, Sanbonmatsu KY, Joseph S. 2012. Functional role of the sarcin-ricin loop of the 23S rRNA in the elongation cycle of protein synthesis. *J. Mol. Biol.* 419:125–38
92. Skripkin E, Paillart JC, Marquet R, Ehresmann B, Ehresmann C. 1994. Identification of the primary site of the human immunodeficiency virus type 1 RNA dimerization in vitro. *PNAS* 91:4945–49
93. Spitale RC, Crisalli P, Flynn RA, Torre EA, Kool ET, Chang HY. 2013. RNA SHAPE analysis in living cells. *Nat. Chem. Biol.* 9:18–20
94. Staple DW, Butcher SE. 2005. Pseudoknots: RNA structures with diverse functions. *PLOS Biol.* 3:e213
95. Sun L-Z, Zhang D, Chen S-J. 2017. Theory and modeling of RNA structure and interactions with metal ions and small molecules. *Annu. Rev. Biophys.* 46. In press
96. Szewczak AA, Moore PB, Chang YL, Wool IG. 1993. The conformation of the sarcin/ricin loop from 28S ribosomal RNA. *PNAS* 90:9581–85
97. Talkish J, May G, Lin Y, Woolford JL Jr., McManus CJ. 2014. Mod-seq: high-throughput sequencing for chemical probing of RNA structure. *RNA* 20:713–20
98. Theis C, Höner zu Siederdissen C, Hofacker IL, Gorodkin J. 2013. Automated identification of RNA 3D modules with discriminative power in RNA structural alignments. *Nucleic Acids Res.* 41:9999–10009

99. Theis C, Zirbel CL, Höner zu Siederdisen C, Anthon C, Hofacker IL, et al. 2015. RNA 3D modules in genome-wide predictions of RNA 2D structure. *PLOS ONE* 10:e0139900
100. Thore S, Leibundgut M, Ban N. 2006. Structure of the eukaryotic thiamine pyrophosphate riboswitch with its regulatory ligand. *Science* 312:1208–11
101. Trausch JJ, Marcano-Velazquez JG, Matyjasik MM, Batey RT. 2015. Metal ion-mediated nucleobase recognition by the ZTP riboswitch. *Chem. Biol.* 22:829–37
102. Vidovic I, Nottrott S, Hartmuth K, Lührmann R, Ficner R. 2000. Crystal structure of the spliceosomal 15.5 kD protein bound to a U4 snRNA fragment. *Mol. Cell* 6:1331–42
103. Wan Y, Kertesz M, Spitale RC, Segal E, Chang HY. 2011. Understanding the transcriptome through RNA structure. *Nat. Rev. Genet.* 12:641–55
104. Warner KD, Chen MC, Song W, Strack RL, Thorn A, et al. 2014. Structural basis for activity of highly efficient RNA mimics of green fluorescent protein. *Nat. Struct. Mol. Biol.* 21:658–63
105. Weeks KM. 2010. Advances in RNA secondary and tertiary structure analysis by chemical probing. *Curr. Opin. Struct. Biol.* 20:295–304
106. Weinreb C, Riesselman AJ, Ingraham JB, Gross T, Sander C, Marks DS. 2016. 3D RNA and functional interactions from evolutionary couplings. *Cell* 165:963–75
107. Westhof E, Jaeger L. 1992. RNA pseudoknots. *Curr. Opin. Struct. Biol.* 2:334–37
108. Westhof E, Romby P, Romaniuk PJ, Ebel JP, Ehresmann C, Ehresmann B. 1989. Computer modeling from solution data of spinach chloroplast and of *Xenopus laevis* somatic and oocyte 5 S rRNAs. *J. Mol. Biol.* 207:417–31
109. Wyatt JR, Puglisi JD, Tinoco I Jr. 1990. RNA pseudoknots: stability and loop size requirements. *J. Mol. Biol.* 214:455–70
110. Xu X, Chen SJ. 2015. A method to predict the 3D structure of an RNA scaffold. *Methods Mol. Biol.* 1316:1–11
111. Zemla A, Venclovas C, Moulton J, Fidelis K. 1999. Processing and analysis of CASP3 protein structure predictions. *Proteins* 3:22–29
112. Zhang JW, Ferré-D'Amaré AR. 2013. Co-crystal structure of a T-box riboswitch stem I domain in complex with its cognate tRNA. *Nature* 500:363–66
113. Zhang Y, Skolnick J. 2004. Scoring function for automated assessment of protein structure template quality. *Proteins* 57:702–10
114. Zhong C, Zhang S. 2012. Clustering RNA structural motifs in ribosomal RNAs using secondary structural alignment. *Nucleic Acids Res.* 40:1307–17
115. Ziehler WA, Engelke DR. 2001. Probing RNA structure with chemical reagents and enzymes. *Curr. Protoc. Nucleic Acid Chem.* 6:6.1.1–21
116. Zirbel CL, Roll J, Sweeney BA, Petrov AI, Pirrung M, Leontis NB. 2015. Identifying novel sequence variants of RNA 3D motifs. *Nucleic Acids Res.* 43:7504–20
117. Zok T, Popenda M, Szachniuk M. 2014. MCQ4Structures to compute similarity of molecule structures. *Cent. Eur. J. Oper. Res.* 22:457–73
118. Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31:3406–15