



Published in final edited form as:

Methods. 2019 June 01; 162-163: 60–67. doi:10.1016/j.ymeth.2019.04.003.

How to Benchmark RNA Secondary Structure Prediction Accuracy

David H. Mathews

Center for RNA Biology, Department of Biochemistry & Biophysics, and Department of Biostatistics & Computational Biology, University of Rochester Medical Center, 601 Elmwood Avenue, Box 712, Rochester, NY 14642

Abstract

RNA secondary structure prediction is widely used. As new methods are developed, these are often benchmarked for accuracy against existing methods. This review discusses good practices for performing these benchmarks, including the choice of benchmarking structures, metrics to quantify accuracy, the importance of allowing flexibility for pairs in the accepted structure, and the importance of statistical testing for significance.

Keywords

Comparative Sequence Analysis; RNA Folding

1. Introduction

RNA sequences serve a variety of roles in Biology. It was appreciated early that RNA has core roles in the expression of proteins as outlined in the Central Dogma of Molecular Biology [1]. These roles include coding of genetic information, as served by messenger RNA, and noncoding roles, served by ribosomal RNA and transfer RNA. Since the discovery of these roles, we came to appreciate that RNA sequences serve a variety of non-coding roles [2], including catalysis [3], site-recognition [4, 5], and gene expression regulation [6–8].

RNA structure is hierarchical, from primary, to secondary, to tertiary, and to quaternary structure [9]. The primary structure is the covalent structure, i.e. the sequence of nucleotides. The secondary structure is the set of canonical base pairs, including A-U, G-C, and G-U pairs. These base pairs are organized in A-form helices, which flank nucleotides that are said to be in loops. The tertiary structure is the three-dimensional positions of the atoms in space, which demonstrate the additional intramolecular contacts beyond canonical base pairing. These contacts include non-canonical base pairs, stacks, and base-backbone interactions.

David_Mathews@urmc.rochester.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

The quaternary structure is the interactions with other RNA, macromolecules, or small molecules.

The secondary structure of RNA provides important information about the function of a sequence. It can characterize a family of RNA sequences that are homologous [10, 11]. It also provides a framework for understanding the mechanism by which an RNA sequence functions. Additionally, the secondary structure is informative for designing constructs for tertiary structure determination by NMR, x-ray crystallography, or cryogenic electron microscopy.

RNA secondary structure prediction is widely used. It is popular to predict secondary structure using a dynamic programming algorithm to find the lowest free energy structure, where nearest neighbor parameters are used to estimate folding stability [12–15]. Dynamic programming algorithms have also been developed to characterize the thermodynamic ensemble of structures, including estimating the probability of base pairs and sampling structures from the ensemble [16]. Alternative structure scoring methods have been developed using parameters learned from the database of RNA sequences with known structures [17–19]. It has been demonstrated that the thermodynamic methods and the scoring schemes are rooted in the same principles [17, 20].

The accuracy of RNA secondary structure prediction is improved by the use of additional information to inform the modeling. Structure probing data can guide structure prediction [21, 22]. Multiple homologous sequences can be used to predict a conserved structure [15, 23, 24]. It has also been shown that these two sources of information can be synergistic [25, 26]. In particular, developing improved methods for secondary structure prediction that use additional information is a field of active research.

The definition of secondary structure has also been broadened to include “extended” secondary structures. The extended secondary structure includes both canonical and non-canonical base pairs. Although non-canonical pairs are traditionally considered tertiary structure, recent advances have made it possible to predict non-canonical base pairs. First, a clear nomenclature was developed by Leontis and Westhof to describe these pairs [27, 28]. Next, software tools were developed to automate the identification of non-canonical pairs from atomic coordinates of structures [29, 30]. Finally, scoring functions and algorithms were developed to predict extended secondary structures from sequence [31–34].

This review discusses methods and good practices for benchmarking the accuracy of RNA secondary structure prediction. As new RNA secondary structure prediction methods are developed, they are typically benchmarked against existing methods by predicting structures for a set of sequences with accepted structures and then comparing the predictions to the accepted structures. This can highlight the strengths and weaknesses of the new method, informing potential users and providing direction for additional method development.

2. Choosing structures for benchmarks.

An informative benchmark requires a variety of high-quality structures. To have a sufficient number of tests, the structures are generally drawn from a set of structures determined by

comparative sequence analysis, which infers a conserved structure using a large set of homologs [35, 36]. RNA sequences drift across evolution, but the secondary structure is more highly conserved because structure dictates function. This results in the phenomenon of compensating base pair changes, by which a base pair is conserved by two changes in sequence, such as an A-U base pair observed in one species that is replaced by a G-C pair observed in the sequence homolog in the second species. Although some differences can be observed between comparative analysis structure predictions and subsequently solved crystal structures, the accuracy of comparative analysis can be quite excellent. For example, over 97% of base pairs modeled for rRNA structures were observed in subsequent crystal structures [37].

For our benchmarks, we use structures from RNA families that were determined by comparative analysis where the structures have been carefully studied by specific investigators and refined over time using a wide phylogeny of sequences. We currently use secondary structures from the 5S ribosomal RNA (rRNA), group I intron, group II intron, large subunit rRNA, RNase P RNA, signal recognition particle RNA (SRP RNA), small subunit rRNA, tRNA, tmRNA, and telomerase RNA families. We largely stayed away from using structures drawn from the Rfam database [11]. The Rfam database is an excellent resource, but the secondary structure models were deposited by the community and therefore the structures vary in quality as compared to structures that were refined as the specific focus of an investigator or group. A recent study [38], for example, noted that the Rfam-curated 5S rRNA structure contains errors as compared to the consensus structure from the 5S database [39].

In our benchmark dataset, the 5S rRNA structures are drawn from the 5S rRNA database maintained by Maciej Szymanski at Adam Mickiewicz University [39]. Group I introns, small subunit rRNA, and large subunit rRNA structures are studied by Robin Gutell and his lab. They are available from the Comparative RNA Web Site from the University of Texas at Austin [40]. The group II introns were studied and assembled by Francois Michel and published in a review [41]. RNase P RNA are studied by James Brown and his lab. They provide structures at the RNase P Database from North Carolina State University [42]. The SRP RNA and tmRNA structures were assembled and maintained by Christian Zwieb [43]. The databases are provided by a consortium of investigators and hosted at the Center for noncoding RNA in Technology and Health (RTH) at the University of Copenhagen (<https://rth.dk/resources/rnp/SRPDB/SRPDB.html>; <https://rth.dk/resources/rnp/tmRDB/>). The tRNA database was assembled by Mathias Sprinzl and coworkers, and is now hosted at the University of Leipzig as a collaboration between the Universities at Leipzig, Marburg, and Strasbourg [44]. The telomerase RNA structures were assembled and studied by Julian Chen and his group at Arizona State University. They provide a database of structures that are available as images [45], but we found it convenient to retrieve machine-readable files with the same information from the Rfam database [11]. The collection of benchmarking structures we collected, called archive II, is available for download from our lab website at <https://rna.urmc.rochester.edu/publications.html> [46].

Benchmarks can also use secondary structures derived from solved tertiary structures. When testing the accuracy of prediction for extended secondary structures that include non-

canonical base pairs, the use of these data is important because comparative analysis generally does not identify non-canonical base pairs. The RNA Strand database (<http://www.rnasoft.ca/strand/>) provides a combination of secondary structures determined by comparative sequence analysis and secondary structures extracted from tertiary structures [47]. The CompaRNA online benchmark uses the RNA Strand database and an additional database of structures that were extracted from tertiary structures. The tertiary structures for both these databases were derived from the protein databank [48] and the secondary structures determined using the RNAView program [29], which identifies base pairs from 3D coordinates.

3. Characterizing Accuracy.

Benchmarks of secondary structure prediction focus on the accuracy of predicted base pairs as compared to a known structure. This is a binary classification, in that a pair is either predicted or not and a pair is known to occur or not. It is therefore helpful to clarify the successes and failures of the prediction using a confusion matrix (Fig. 1). Pairs that are predicted and are also in the accepted structure are termed true positives. However, pairs that are predicted that are not in the accepted structure are termed false positives. Likewise, a pair in the accepted structure that is not predicted is a false negative, and a pair that is neither predicted nor in the accepted structure is a true negative.

It is clear that a prediction that contains only true positives and true negatives is a perfect prediction. To quantify the accuracy, we use terms from epidemiology. The sensitivity, also known as recall, is the fraction of pairs in the accepted structure that are predicted:

$$\text{sensitivity} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}} \quad (\text{eq. 1})$$

The positive predictive value (PPV), also known as precision, is the fraction of predicted pairs that are in the accepted the structure:

$$\text{PPV} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}} \quad (\text{eq. 2})$$

The sensitivity is therefore reduced from 1 (100%) when pairs in the known structure are missed by the prediction (false negatives) and the PPV is reduced from 1 (100%) when there are incorrectly predicted pairs (false positives). Both are needed to fully characterize the successes and failures of structure prediction. To summarize the sensitivity and PPV, the F_1 score, the harmonic mean of sensitivity and PPV, can be calculated:

$$\begin{aligned} F_1 &= \frac{2 \times \text{sensitivity} \times \text{PPV}}{\text{sensitivity} + \text{PPV}} \\ &= \frac{2 \times \text{true positives}}{2 \times \text{true positives} + \text{false positives} + \text{false negatives}} \end{aligned} \quad (\text{eq. 3})$$

The harmonic mean has the property of being conservative (i.e. pessimistic about the accuracy) as compared to the arithmetic mean. This is because, if either sensitivity or PPV is substantially lower than the other, the harmonic mean tends towards the lower value.

Alternatively, the geometric mean can also be used to summarize the overall prediction quality, and geometric mean is also conservative as compared to the arithmetic mean.

Another metric from epidemiology is specificity, but that is uninformative for RNA secondary structure prediction. Specificity is the fraction of pairs not in the accepted structure that are also not predicted:

$$\text{specificity} = \frac{\text{true negatives}}{\text{true negatives} + \text{false positives}} \quad (\text{eq. 4})$$

The number of true negatives, however, is quite large compared to the false positives because the number of possible pairs for a sequence tends to increase with the square of the length, but the number of predicted pairs can only increase linearly with length (a nucleotide can be in only one pair in a valid structure). Therefore, especially for long sequences, specificity will be close to 1.

The PPV for free energy minimization has been observed to be lower on average than sensitivity [18, 49, 50]. This is in part because accepted structures can be missing pairs that are experimentally known to occur. A good example of this is the fact that the Sprinzl database of tRNA structures does not annotate base pairs in the variable loop region of leucine, serine, and tyrosine tRNAs, for which some sequences have fifth arms in their cloverleaf structures [51]. A correct prediction of one of these pairs would lower the PPV. The other reason for lower PPV, and the larger contributor, is that free energy minimization tends to overestimate the number of canonical base pairs because the canonical pairs lower the folding free energy change [52].

When reporting a mean sensitivity, PPV, and F_1 score, we prefer to report the mean taken across the families, where each family contributes equally to the mean. This is important because the total number of base pairs and sequences in each family varies. The mean across families is an estimate of the accuracy that would be observed for a new RNA family on which the benchmarked program might be used.

The small and large subunit rRNA families contain sequences that are longer than the other families. RNA secondary structure prediction is a harder problem for longer sequences because the number of possible structures increases exponentially with sequence length [53]. Historically, when benchmarking single sequence secondary structure prediction accuracy, we divided these long families into folding domains based on the known structure to provide average structure prediction accuracy for sequences shorter than 800 nucleotides [54–57]. We also provided the accuracy for the full-length sequences, but did not include that accuracy in the mean across families. It is important that a benchmark clearly report details, such as this, so that users are well informed about the accuracy they might have on their sequences.

4. Accounting for Conformational Dynamics.

RNA structures are intrinsically dynamic, and this is important for function [58–61]. The dynamics are most pronounced for non-canonical pairs, i.e. the loop regions of secondary

structures [62, 63]. Secondary structure, however, is also dynamic. For example, it is well-established using NMR that terminal base pairs in nucleic acid helices are subject to opening, called fraying [64–67].

When assessing base pair prediction accuracy, it is important to account for conformational dynamics. We do this by considering a predicted pair to be consistent with a known pair even if it is displaced by up to one nucleotide on one side [54]. In other words, for base pair (i)-(j) in the accepted structure between nucleotides i and j, the following predicted base pairs would also be considered correct: (i+1)-(j), (i-1)-(j), (i)-(j+1), and (i)-(j-1).

There is considerable evidence supporting that base pairs sample multiple structures. First, measurements of folding free energy change at 37 °C for single-nucleotide bulges adjacent to one or more paired nucleotides of the same identity are consistent with the bulge migrating across multiple positions [55, 68]. The stability of these loops is similar to other single-nucleotide bulges when an entropic stability bonus is applied that reflects the multiple states available at equilibrium (Fig. 2).

Fluctuating pairs are also observed by NMR. Woodson & Crothers observed that the bulged C nucleotide in the DNA helix 5'GATGGGCAG3'–5'CTGccccATC3') fluctuated across the four possible positions on the NMR timescale [69]. Dethoff et al. observed a minor conformation for the HIV-1 RNA stem-loop 1 that exchanges with the major conformation on the NMR timescale. The major and minor conformations differ by two canonical base pairs (Fig. 3), and the minor conformation is approximately 1/10 as populated as the major conformation.

Given that base pairs can fluctuate, a secondary structure model from either structure prediction or comparative sequence analysis should be considered as a representation of these alternative states, not as a rigid structure. Likewise, crystal structures can also be considered as one model of possible multiple states. Careful analyses of ribosome crystal structures reveal the flexibility (especially in non-canonical base pairs) across the conformational states that occur as part of ribosome function [70–74].

The Gutell lab provides high quality models of rRNA secondary structures [40] based on the comparison of over 7,000 small subunit and 1,000 large subunit rRNA sequences [37]. When the *T. thermophilus* small subunit [75] and *H. marismortui* large subunit [76] ribosome crystal structures became available, Gutell et al. analyzed and reported discrepancies between their secondary structure models and the crystal structures [37]. As expected, their models were excellent, and, as noted above, the models had 97% PPV against the crystal structures. Interestingly, using the information they provided, there are two cases in the 16S rRNA and three cases in the 23S rRNA where one side of a canonical base pair was displaced by one nucleotide in the comparative analysis model as compared to the crystal structure. The number of these types of discrepancies increases to 19 when non-canonical base pairs are also considered. These discrepancies between the models and the crystal structures are likely inconsequential because they might simply reflect situations where a structure is flexible as noted above. The prevalence of these discrepancies for high

quality comparative analysis models also supports our choice to allow the one-nucleotide displacement when we score predicted structures against comparative models.

When we calculated the accuracy both allowing displacement and also requiring an exact match, we find the two scores generally differ by approximately 2%. For example, in a recent benchmark we reported sensitivity and PPV for *Dynalign II*, an algorithm that predicts the lowest free energy structure conserved for two sequences, and *Fold*, an algorithm that predicts the lowest free energy structure using a single sequence [77]. For 5S rRNA, the sensitivity was 91.8% and 73.1% for *Dynalign II* and *Fold*, respectively, when allowing the single-nucleotide displacement. The sensitivity was 90.6% and 70.9% for *Dynalign II* and *Fold*, respectively, when an exact match was required. PPV was also similar with the two scoring methods for 5S rRNA. It was 83.3% and 63.5% with single-nucleotide displacement allowed and 79.7% and 61.4% for exact match for *Dynalign II* and *Fold*, respectively.

5. The *scorer* program in RNAstructure.

The program *scorer* is available as a component of the RNAstructure package to compare a predicted secondary structure against an established structure [78]. It takes a connection table (ct) file of the predicted structure, which can contain multiple predicted structures, and a ct file of the known structure. It reports the sensitivity and PPV, along with the number of true positives, the number of predicted pairs, and the number of pairs in the known structure. By default, *scorer* allows pairs to be displaced by one nucleotide position on either the 5' or the 3' side of the pair. For comparison, this behavior can be changed to require exact match between predicted and accepted pairs.

A helpful adjunct to *scorer* is *CircleCompare*, which produces a postscript illustration of the predicted pairs and the accepted pairs (Fig. 4). The illustration shows the sequence in a circle and pairs as chords of the circle. The color of pairs provides information about whether pairs are predicted and/or in the accepted structure. Green pairs are those that are predicted and in the accepted structure. Black pairs are present in the accepted structure and not in the predicted structure. Red pairs are those in the predicted structure and not in the accepted structure. *CircleCompare* also reports the sensitivity and PPV for the prediction.

6. Statistical testing.

In addition to comparing mean performance in benchmarks, it is also essential to determine whether differences in performance are statistically significant. For example, we reported a small improvement in PPV when using maximum expected accuracy (MEA) structure prediction to assemble structures with pairs of high equilibrium pairing probability as compared to free energy minimization [56]. This followed previous work that developed MEA using parameters learned from known structures [18, 79]. This small improvement was found subsequently to not be statistically significant on the RNA Strand dataset using a bootstrap to assess significance [80]. Therefore, although MEA provides an ability to tune predictions to higher sensitivity or higher PPV (at the cost of the other) and MEA provides

helpful alternative hypotheses for the predicted structure, it is not reliably more accurate than free energy minimization.

For statistical testing, we use paired t-tests, which assess whether the mean difference in performance (either sensitivity or PPV) of two methods is significantly different than zero [81]. It uses the natural pairing of predictions made with two methods on the same sequence to compare the differences in performance. We also discussed ways to add structures to a benchmark set to test for significance if an initial set of structures does not have enough power to establish that a difference in performance is significant [81].

A second assessment for an improvement in one prediction method over another is whether the improvement is substantial. Generally, we believe that any statistically significant improvement is substantial because secondary structure prediction is well used.

Additionally, it can be helpful to assess whether an improvement is substantial by assessing the improvement on individual calculations rather than just the mean improvement. For example, we examined the accuracy performance for each sequence individually in a benchmark testing whether *TurboFold* [82], which predicts conserved structures using a set of homologous sequences, is improved with the addition of SHAPE mapping data, which reveals nucleotides that are unpaired [25]. We plotted the F_1 score of structure prediction including SHAPE mapping data [83] as a function of F_1 score without using SHAPE. For some families, such as tRNA [44] and group I introns [40], we observed that the structure prediction improvement was greatest for sequences that were poorly predicted without the benefit of SHAPE data. For these families, a modest improvement in mean performance was really a large improvement for sequences that needed improvement. Additionally, even for families where the improvement was modest, such as 5S rRNA [40], the accuracy of a vast majority of calculations were improved by adding SHAPE. Therefore, for these families, we could demonstrate modest improvements across the sequences, without observing a loss of accuracy for individual calculations.

7. Pseudoknots.

We consider base pairs in pseudoknots to be part of secondary structure [84]. These base pairs span from one loop segment to another loop segment. Formally, a pseudoknot occurs if there are two base pairs $(i)-(j)$ and $(i')-(j')$ with $i < i' < j < j'$. This distinction is important because many of the standard dynamic programming algorithms are not capable of predicting pseudoknotted base pairs; they can predict $(i)-(j)$ or $(i')-(j')$, but not both pairs in the same structure.

In our benchmarks, we keep pseudoknotted pairs present in the predicted structure. Therefore, methods that predict secondary structure without including pseudoknotted pairs are not capable of achieving 100% sensitivity.

When benchmarking methods that predict pseudoknots, it is also important to verify that pseudoknotted pairs are being predicted and to report the sensitivity and PPV of pseudoknotted base pairs. In our benchmark of programs that predict pseudoknotted secondary structures using a single input sequence, we found that the PPV of pseudoknotted

base pair prediction was less than 30% and the sensitivity was 3% for the programs with highest PPV and highest sensitivity, respectively [57]. A good mean performance across all base pairs disguised the fact that the accuracy pseudoknot prediction using a single sequence is quite poor. In a follow-up benchmark of methods that predict pseudoknotted secondary structures using multiple homologous sequences, the PPV of the best method was 30%, but the sensitivity was still 5% [85]. Although the accuracy is greatly improved with the help of multiple homologs, there is still a large need to improve the prediction of pseudoknots.

8. The importance of separate training and testing sets.

A number of secondary structure prediction methods were explicitly trained using known RNA secondary structures. The database of secondary structures can be used to train parameters that are free energy nearest neighbor parameters [19, 50] or are similar to free energy nearest neighbor parameters [18, 31, 86]. Alternatively, parameters can be estimated for probabilistic models, such as stochastic context-free grammars [17, 79, 87, 88]. These training methods have been discussed in detail [20]. Also, methods can have adjustable parameters that were optimized to sets of structures, such as the sequence alignment components of *Dynalign* [77, 89], *Foldalign* [90], *locARNA* [91], and *TurboFold* [82], which are programs that predict a conserved structure and sequence alignment for two or more homologous sequences.

To prevent the training from inflating the apparent accuracy, the test set of structures must be independent of the training structures. This extends to RNA families because the structures within families are similar [17, 56]. Therefore, the testing dataset should use different RNA families from any training dataset that was used.

9. The focus on base pairs and prospectus.

Since 1999, our benchmarks focused on the accuracy of predicting base pairs because it is clear and simple to convey [54]. At that time, at least two previous papers had started to focus on the accuracy of individual base pairs because they reported methods to identify the base pairs more likely to be correctly predicted using folding free energy change or base pair probabilities [92, 93]. Prior to this, benchmarks had focused on helices as the basic unit, but this required interpretation of what constitutes a helix [94, 95].

Going forward, one approach to expanding benchmarks might be approaches that focus more on the topology of the overall fold than on individual pairs. Abstract shapes, for example, provide a means for abstracting structure to the helix or to the branch level [96]. Benchmarks could be developed to use the branching representation to characterize topology.

10. Conclusions.

Benchmarks are commonly performed when new secondary structure prediction methods are developed. Here, good practices are outlined for performing and reporting these benchmarks. It is important to use a variety of high-quality secondary structures, it is

important to test for statistical significance, and it is important to use separate testing structure datasets from training datasets.

Acknowledgements.

Funding: This work was supported by the National Institutes of Health [grant number R01GM076485].

References.

- [1]. Crick F, Central dogma of molecular biology, *Nature* 227 (1970) 561–3. [PubMed: 4913914]
- [2]. Eddy SR, Non-coding RNA genes and the modern RNA world, *Nat. Rev* 2 (2001) 919–929.
- [3]. Doudna JA, Cech TR, The chemical repertoire of natural ribozymes, *Nature* 418 (2002) 222–228. [PubMed: 12110898]
- [4]. Bachellerie JP, Cavaille J, Huttenhofer A, The expanding snoRNA world, *Biochimie* 84 (2002) 775–90. [PubMed: 12457565]
- [5]. Karijolich J, Yi C, Yu YT, Transcriptome-wide dynamics of RNA pseudouridylation, *Nat. Rev. Mol. Cell Biol* 16 (2015) 581–5. [PubMed: 26285676]
- [6]. Wu L, Belasco JG, Let me count the ways: mechanisms of gene regulation by miRNAs and siRNAs, *Mol. Cell* 29 (2008) 1–7. [PubMed: 18206964]
- [7]. Storz G, Gottesman S, Versatile roles of small RNA regulators in bacteria, in: Gesteland RF, Cech TR, Atkins JF (Eds.), *The RNA World*, third edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 2006, pp. 567–594.
- [8]. Serganov A, Nudler E, A decade of riboswitches, *Cell* 152 (2013) 17–24. [PubMed: 23332744]
- [9]. Tinoco I Jr., Bustamante C, How RNA folds, *J. Mol. Biol* 293 (1999) 271–81. [PubMed: 10550208]
- [10]. Nawrocki EP, Eddy SR, Infernal 1.1: 100-fold faster RNA homology searches, *Bioinformatics* 29 (2013) 2933–5. [PubMed: 24008419]
- [11]. Kalvari I, Argasinska J, Quinones-Olvera N, Nawrocki EP, Rivas E, Eddy SR, Bateman A, Finn RD, Petrov AI, Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families, *Nucleic Acids Res* 46 (2018) D335–D342. [PubMed: 29112718]
- [12]. Hofacker IL, Energy-directed RNA structure prediction, *Methods Mol. Biol* 1097 (2014) 71–84. [PubMed: 24639155]
- [13]. Lorenz R, Wolfinger MT, Tanzer A, Hofacker IL, Predicting RNA secondary structures from sequence and probing data, *Methods* 103 (2016) 86–98. [PubMed: 27064083]
- [14]. Mathews DH, Moss WN, Turner DH, *Folding and Finding RNA Secondary Structure*, Cold Spring Harb. Perspect. Biol 2 (2010) a003665. [PubMed: 20685845]
- [15]. Seetin MG, Mathews DH, RNA structure prediction: an overview of methods, *Methods Mol. Biol* 905 (2012) 99–122. [PubMed: 22736001]
- [16]. Mathews DH, Revolutions in RNA secondary structure prediction, *J. Mol. Biol* 359 (2006) 526–532. [PubMed: 16500677]
- [17]. Rivas E, Lang R, Eddy SR, A range of complex probabilistic models for RNA secondary structure prediction that includes the nearest-neighbor model and more, *RNA* 18 (2012) 193–212. [PubMed: 22194308]
- [18]. Do CB, Woods DA, Batzoglou S, CONTRAfold: RNA secondary structure prediction without physics-based models, *Bioinformatics* 22 (2006) e90–8. [PubMed: 16873527]
- [19]. Andronescu M, Condon A, Hoos HH, Mathews DH, Murphy KP, Computational approaches for RNA energy parameter estimation, *RNA* 16 (2010) 2304–18. [PubMed: 20940338]
- [20]. Rivas E, The four ingredients of single-sequence RNA secondary structure prediction. A unifying perspective, *RNA Biol* 10 (2013) 1185–96. [PubMed: 23695796]
- [21]. Sloma MF, Mathews DH, Improving RNA secondary structure prediction with structure mapping data, *Methods Enzymol* 553 (2015) 91–114. [PubMed: 25726462]

- [22]. Eddy SR, Computational analysis of conserved RNA secondary structure in transcriptomes and genomes, *Annu. Rev. Biophys* 43 (2014) 433–56. [PubMed: 24895857]
- [23]. Havgaard JH, Gorodkin J, RNA structural alignments, part I: Sankoff-based approaches for structural alignments, *Methods Mol. Biol* 1097 (2014) 275–90. [PubMed: 24639164]
- [24]. Asai K, Hamada M, RNA structural alignments, part II: non-Sankoff approaches for structural alignments, *Methods Mol. Biol* 1097 (2014) 291–301. [PubMed: 24639165]
- [25]. Tan Z, Sharma G, Mathews DH, Modeling RNA Secondary Structure with Sequence Comparison and Experimental Mapping Data, *Biophys. J* 113 (2017) 330–338. [PubMed: 28735622]
- [26]. Lavender CA, Lorenz R, Zhang G, Tamayo R, Hofacker IL, Weeks KM, Model-Free RNA Sequence and Structure Alignment Informed by SHAPE Probing Reveals a Conserved Alternate Secondary Structure for 16S rRNA, *PLoS Comput. Biol* 11 (2015) e1004126. [PubMed: 25992778]
- [27]. Leontis NB, Stombaugh J, Westhof E, The non-Watson-Crick base pairs and their associated isostericity matrices, *Nucleic Acids Res* 30 (2002) 3497–3531. [PubMed: 12177293]
- [28]. Leontis NB, Westhof E, Geometric nomenclature and classification of RNA base pairs, *RNA* 7 (2001) 499–512. [PubMed: 11345429]
- [29]. Yang H, Jossinet F, Leontis N, Chen L, Westbrook J, Berman H, Westhof E, Tools for the automatic identification and classification of RNA base pairs, *Nucleic Acids Res* 31 (2003) 3450–60. [PubMed: 12824344]
- [30]. Lemieux S, Major F, RNA canonical and non-canonical base pairing types: a recognition method and complete repertoire, *Nucleic Acids Res* 30 (2002) 4250–63. [PubMed: 12364604]
- [31]. Parisien M, Major F, The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data, *Nature* 452 (2008) 51–5. [PubMed: 18322526]
- [32]. Honer zu Siederdisen C, Bernhart SH, Stadler PF, Hofacker IL, A folding algorithm for extended RNA secondary structures, *Bioinformatics* 27 (2011) i129–36. [PubMed: 21685061]
- [33]. Dallaire P, Major F, Exploring Alternative RNA Structure Sets Using MC-Flashfold and db2cm, *Methods Mol. Biol* 1490 (2016) 237–51. [PubMed: 27665603]
- [34]. Sloma MF, Mathews DH, Base pair probability estimates improve the prediction accuracy of RNA non-canonical base pairs, *PLoS Comput. Biol* 13 (2017) e1005827. [PubMed: 29107980]
- [35]. James BD, Olsen GJ, Pace NR, Phylogenetic comparative analysis of RNA secondary structure, *Methods Enzymol* 180 (1989) 227–239. [PubMed: 2482415]
- [36]. Pace NR, Thomas BC, Woese CR, Probing RNA structure, function, and history by comparative analysis, in: Gesteland RF, Cech TR, Atkins JF (Eds.), *The RNA World*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1999, pp. 113–141.
- [37]. Gutell RR, Lee JC, Cannone JJ, The accuracy of ribosomal RNA comparative structure models, *Curr. Opin. Struct. Biol* 12 (2002) 301–310. [PubMed: 12127448]
- [38]. Rivas E, Clements J, Eddy SR, A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs, *Nat. Methods* 14 (2017) 45–48. [PubMed: 27819659]
- [39]. Szymanski M, Zielezinski A, Barciszewski J, Erdmann VA, Karlowski WM, 5SRNadb: an information resource for 5S ribosomal RNAs, *Nucleic Acids Res* 44 (2016) D180–3. [PubMed: 26490961]
- [40]. Cannone JJ, Subramanian S, Schnare MN, Collett JR, D’Souza LM, Du Y, Feng B, Lin N, Madabusi LV, Muller KM, Pande N, Shang Z, Yu N, Gutell RR, The comparative RNA web (CRW) site: An online database of comparative sequence and structure information for ribosomal, intron, and other RNAs, *BMC Bioinformatics* 3 (2002) 2. [PubMed: 11869452]
- [41]. Michel F, Umesono K, Ozeki H, Comparative and functional anatomy of group II catalytic introns - a review, *Gene* 82 (1989) 5–30. [PubMed: 2684776]
- [42]. Brown JW, The ribonuclease P database, *Nucleic Acids Res* 27 (1999) 314. [PubMed: 9847214]
- [43]. Andersen ES, Rosenblad MA, Larsen N, Westergaard JC, Burks J, Wower IK, Wower J, Gorodkin J, Samuelsson T, Zwieb C, The tmRDB and SRPDB resources, *Nucleic Acids Res* 34 (2006) D163–8. [PubMed: 16381838]
- [44]. Juhling F, Morl M, Hartmann RK, Sprinzl M, Stadler PF, Putz J, tRNAdb 2009: compilation of tRNA sequences and tRNA genes, *Nucleic Acids Res* 37 (2009) D159–62. [PubMed: 18957446]

- [45]. Podlevsky JD, Bley CJ, Omana RV, Qi X, Chen JJ, The telomerase database, *Nucleic Acids Res* 36 (2008) D339–43. [PubMed: 18073191]
- [46]. Sloma MF, Mathews DH, Exact calculation of loop formation probability identifies folding motifs in RNA secondary structures, *RNA* 22 (2016) 1808–1818. [PubMed: 27852924]
- [47]. Andronescu M, Bereg V, Hoos HH, Condon A, RNA STRAND: the RNA secondary structure and statistical analysis database, *BMC Bioinformatics* 9 (2008) 340. [PubMed: 18700982]
- [48]. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE, The Protein Data Bank, *Nucleic Acids Res* 28 (2000) 235–42. [PubMed: 10592235]
- [49]. Mathews DH, Using an RNA secondary structure partition function to determine confidence in base pairs predicted by free energy minimization, *RNA* 10 (2004) 1178–1190. [PubMed: 15272118]
- [50]. Andronescu M, Condon A, Hoos HH, Mathews DH, Murphy KP, Efficient parameter estimation for RNA secondary structure prediction, *Bioinformatics* 23 (2007) i19–i28. [PubMed: 17646296]
- [51]. Westhof E, Auffinger P, *Transfer RNA Structure*, eLS (2012) doi: 10.1002/9780470015902.a0000527.pub2.
- [52]. Xia T, SantaLucia J Jr., Burkard ME, Kierzek R, Schroeder SJ, Jiao X, Cox C, Turner DH, Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick pairs, *Biochemistry* 37 (1998) 14719–14735. [PubMed: 9778347]
- [53]. Zuker M, Sankoff D, RNA secondary structures and their prediction, *Bull. Math. Biol* 46 (1984) 591–621.
- [54]. Mathews DH, Sabina J, Zuker M, Turner DH, Expanded sequence dependence of thermodynamic parameters provides improved prediction of RNA secondary structure, *J. Mol. Biol* 288 (1999) 911–940. [PubMed: 10329189]
- [55]. Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH, Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure, *Proc. Natl. Acad. Sci. U.S.A* 101 (2004) 7287–7292. [PubMed: 15123812]
- [56]. Lu ZJ, Gloor JW, Mathews DH, Improved RNA secondary structure prediction by maximizing expected pair accuracy, *RNA* 15 (2009) 1805–1813. [PubMed: 19703939]
- [57]. Bellaousov S, Mathews DH, ProbKnot: Fast prediction of RNA secondary structure including pseudoknots, *RNA* 16 (2010) 1870–1880. [PubMed: 20699301]
- [58]. Petrov A, Kornberg G, O’Leary S, Tsai A, Uemura S, Puglisi JD, Dynamics of the translational machinery, *Curr. Opin. Struct. Biol* 21 (2011) 137–45. [PubMed: 21256733]
- [59]. Korostelev A, Ermolenko DN, Noller HF, Structural dynamics of the ribosome, *Curr. Opin. Chem. Biol* 12 (2008) 674–83. [PubMed: 18848900]
- [60]. Whitford PC, The ribosome’s energy landscape: Recent insights from computation, *Biophys Rev* 7 (2015) 301–310. [PubMed: 28510226]
- [61]. Ray S, Chauvier A, Walter NG, Kinetics coming into focus: single-molecule microscopy of riboswitch dynamics, *RNA Biol* (2018) 1–9.
- [62]. Burkard ME, Turner DH, NMR structures of r(GCAGGCGUGC)₂ and determinants of stability for single guanosine-guanosine base pairs, *Biochemistry* 39 (2000) 11748–11762. [PubMed: 10995243]
- [63]. Chen G, Kennedy SD, Qiao J, Krugh TR, Turner DH, An alternating sheared AA pair and elements of stability for a single sheared purine-purine pair flanked by sheared GA pairs in RNA, *Biochemistry* 45 (2006) 6889–903. [PubMed: 16734425]
- [64]. Patel DJ, Hilbers CW, Proton nuclear magnetic resonance investigations of fraying in doublestranded d-ApTpGpCpApT in H₂O solution, *Biochemistry* 14 (1975) 2651–6. [PubMed: 238568]
- [65]. Andreatta D, Sen S, Perez Lustres JL, Kovalenko SA, Ernsting NP, Murphy CJ, Coleman RS, Berg MA, Ultrafast dynamics in DNA: “fraying” at the end of the helix, *J. Am. Chem. Soc* 128 (2006) 6885–92. [PubMed: 16719468]
- [66]. Nonin S, Leroy JL, Gueron M, Terminal base pairs of oligodeoxynucleotides: imino proton exchange and fraying, *Biochemistry* 34 (1995) 10652–9. [PubMed: 7654719]

- [67]. Kochoyan M, Lancelot G, Leroy JL, Study of structure, base-pair opening kinetics and proton exchange mechanism of the d-(AATTGCAATT) self-complementary oligodeoxynucleotide in solution, *Nucleic Acids Res* 16 (1988) 7685–702. [PubMed: 2842732]
- [68]. Znosko BM, Silvestri SB, Volkman H, Boswell B, Serra MJ, Thermodynamic parameters for an expanded nearest-neighbor model for the formation of RNA duplexes with single nucleotide bulges, *Biochemistry* 41 (2002) 10406–10417. [PubMed: 12173927]
- [69]. Woodson SA, Crothers DM, Proton nuclear magnetic resonance studies on bulge-containing DNA oligonucleotides from a mutational hot-spot sequence, *Biochemistry* 26 (1987) 904–912. [PubMed: 3567151]
- [70]. Loveland AB, Demo G, Grigorieff N, Korostelev AA, Ensemble cryo-EM elucidates the mechanism of translation fidelity, *Nature* 546 (2017) 113–117. [PubMed: 28538735]
- [71]. Mohan S, Noller HF, Recurring RNA structural motifs underlie the mechanics of L1 stalk movement, *Nat. Commun* 8 (2017) 14285. [PubMed: 28176782]
- [72]. Liu Q, Fredrick K, Intersubunit Bridges of the Bacterial Ribosome, *J. Mol. Biol* 428 (2016) 2146–64. [PubMed: 26880335]
- [73]. Dunkle JA, Wang L, Feldman MB, Pulk A, Chen VB, Kapral GJ, Noeske J, Richardson JS, Blanchard SC, Cate JH, Structures of the bacterial ribosome in classical and hybrid states of tRNA binding, *Science* 332 (2011) 981–4. [PubMed: 21596992]
- [74]. Schuwirth BS, Borovinskaya MA, Hau CW, Zhang W, Vila-Sanjurjo A, Holton JM, Cate JH, Structures of the bacterial ribosome at 3.5 Å resolution, *Science* 310 (2005) 827–34. [PubMed: 16272117]
- [75]. Wimberly BT, Brodersen DE, Clemons WM Jr., Morgan-Warren RJ, Carter AP, Vornrhein C, Hartsch T, Ramakrishnan V, Structure of the 30S ribosomal subunit, *Nature* 407 (2000) 327–339. [PubMed: 11014182]
- [76]. Ban N, Nissen P, Hansen J, Moore PB, Steitz TA, The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution, *Science* 289 (2000) 905–920. [PubMed: 10937989]
- [77]. Fu Y, Sharma G, Mathews DH, Dynalign II: common secondary structure prediction for RNA homologs with domain insertions, *Nucleic Acids Res* 42 (2014) 13939–48. [PubMed: 25416799]
- [78]. Reuter JS, Mathews DH, RNAstructure: software for RNA secondary structure prediction and analysis, *BMC Bioinformatics* 11 (2010) 129. [PubMed: 20230624]
- [79]. Knudsen B, Hein J, RNA secondary structure prediction using stochastic context-free grammars and evolutionary history, *Bioinformatics* 15 (1999) 446–454. [PubMed: 10383470]
- [80]. Hajiaghayi M, Condon A, Hoos HH, Analysis of energy-based algorithms for RNA secondary structure prediction, *BMC Bioinformatics* 13 (2012) 22. [PubMed: 22296803]
- [81]. Xu Z, Almudevar A, Mathews DH, Statistical evaluation of improvement in RNA secondary structure prediction, *Nucleic Acids Res* 40 (2011) e26. [PubMed: 22139940]
- [82]. Tan Z, Fu Y, Sharma G, Mathews DH, TurboFold II: RNA structural alignment and secondary structure prediction informed by multiple homologs, *Nucleic Acids Res* 45 (2017) 11570–11581. [PubMed: 29036420]
- [83]. Hajdin CE, Bellaousov S, Huggins W, Leonard CW, Mathews DH, Weeks KM, Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots, *Proc. Natl. Acad. Sci. U.S.A* 110 (2013) 5498–503. [PubMed: 23503844]
- [84]. Liu B, Mathews DH, Turner DH, RNA pseudoknots: folding and finding, *F1000 Biol. Rep* 2 (2010) 8. [PubMed: 20495679]
- [85]. Seetin MG, Mathews DH, TurboKnot: Rapid Prediction of Conserved RNA Secondary Structures Including Pseudoknots, *Bioinformatics* 28 (2012) 792–798. [PubMed: 22285566]
- [86]. Gardner DP, Ren P, Ozer S, Gutell RR, Statistical potentials for hairpin and internal loops improve the accuracy of the predicted RNA structure, *J. Mol. Biol* 413 (2011) 473–83. [PubMed: 21889515]
- [87]. Dowell RD, Eddy SR, Evaluation of several lightweight stochastic context-free grammars for RNA secondary structure prediction, *BMC Bioinformatics* 5 (2004) 71. [PubMed: 15180907]
- [88]. Knudsen B, Hein J, Pfold: RNA secondary structure prediction using stochastic context-free grammars, *Nucleic Acids Res* 31 (2003) 3423–8. [PubMed: 12824339]

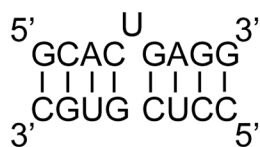
- [89]. Xu Z, Mathews DH, Multilign: an algorithm to predict secondary structures conserved in multiple RNA sequences, *Bioinformatics* 27 (2011) 626–32. [PubMed: 21193521]
- [90]. Havgaard JH, Lyngso RB, Stormo GD, Gorodkin J, Pairwise local structural alignment of RNA sequences with sequence similarity less than 40%, *Bioinformatics* 21 (2005) 1815–24. [PubMed: 15657094]
- [91]. Will S, Reiche K, Hofacker IL, Stadler PF, Backofen R, Inferring noncoding RNA families and classes by means of genome-scale structure-based clustering, *PLoS Comput. Biol* 3 (2007) e65. [PubMed: 17432929]
- [92]. Huynen M, Gutell R., D. Konings, Assessing the reliability of RNA folding using statistical mechanics, *J. Mol. Biol* 267 (1997) 1104–1112. [PubMed: 9150399]
- [93]. Zuker M, Jacobson AB, Using reliability information to annotate RNA secondary structures, *RNA* 4 (1998) 669–679. [PubMed: 9622126]
- [94]. Jaeger JA, Turner DH, Zuker M, Improved predictions of secondary structures for RNA, *Proc. Natl. Acad. Sci. U.S.A* 86 (1989) 7706–7710. [PubMed: 2479010]
- [95]. Walter AE, Turner DH, Kim J, Lyttle MH, Müller P, Mathews DH, Zuker M, Coaxial stacking of helices enhances binding of oligoribonucleotides and improves predictions of RNA folding, *Proc. Natl. Acad. Sci. U.S.A* 91 (1994) 9218–9222. [PubMed: 7524072]
- [96]. Giegerich R, Voss B, Rehmsmeier M, Abstract shapes of RNA, *Nucleic Acids Res* 32 (2004) 4843–51. [PubMed: 15371549]
- [97]. Zuber J, Sun H, Zhang X, McFadyen I, Mathews DH, A sensitivity analysis of RNA folding nearest neighbor parameters identifies a subset of free energy parameters with the greatest impact on RNA secondary structure prediction, *Nucleic Acids Res* 45 (2017) 6168–6176. [PubMed: 28334976]
- [98]. Longfellow CE, Kierzek R, Turner DH, Thermodynamic and spectroscopic study of bulge loops in oligoribonucleotides, *Biochemistry* 29 (1990) 278–285. [PubMed: 2322546]
- [99]. Strom S, Shiskova E, Hahm Y, Grover N, Thermodynamic examination of 1- to 5-nt purine bulge loops in RNA and DNA constructs, *RNA* 21 (2015) 1313–22. [PubMed: 26022248]
- [100]. Blose JM, Manni ML, Klapek KA, Stranger-Jones Y, Zyra AC, Sim V, Griffith CA, Long JD, Serra MJ, Non-nearest-neighbor dependence of the stability for RNA bulge loops based on the complete set of group I single-nucleotide bulge loops, *Biochemistry* 46 (2007) 15123–35. [PubMed: 18047298]
- [101]. McCann MD, Lim GF, Manni ML, Estes J, Klapek KA, Frattini GD, Knarr RJ, Gratton JL, Serra MJ, Non-nearest-neighbor dependence of the stability for RNA group II single-nucleotide bulge loops, *RNA* 17 (2011) 108–19. [PubMed: 21088109]
- [102]. Kent JL, McCann MD, Phillips D, Panaro BL, Lim GF, Serra MJ, Non-nearest-neighbor dependence of stability for group III RNA single nucleotide bulge loops, *RNA* 20 (2014) 825–34. [PubMed: 24742935]
- [103]. Lim GF, Merz GE, McCann MD, Gruskiewicz JM, Serra MJ, Stability of single-nucleotide bulge loops embedded in a GAAA RNA hairpin stem, *RNA* 18 (2012) 807–14. [PubMed: 22345128]
- [104]. Dethoff EA, Petzold K, Chugh J, Casiano-Negrone A, Al-Hashimi HM, Visualizing transient low-populated structures of RNA, *Nature* 491 (2012) 724–8. [PubMed: 23041928]
- [105]. Bellaousov S, Reuter JS, Seetin MG, Mathews DH, RNAstructure: Web servers for RNA secondary structure prediction and analysis, *Nucleic Acids Res* 41 (2013) W471–4. [PubMed: 23620284]

- Benchmarks of RNA secondary structure prediction should use a large set of well determined structures.
- Benchmarks must include a test of statistical significance.
- Benchmarks should consider a base pair correctly predicted even if it is displaced by up to one nucleotide on either the 5' or 3' side as compared to the accepted structure.
- Benchmarks should use entirely different RNA families than those used for training the methods being benchmarked.

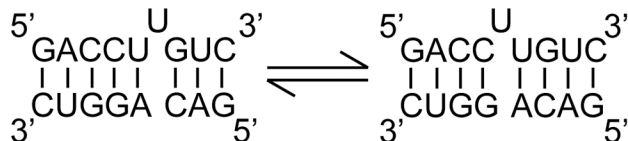
		Base Pairs in Accepted Structure	
		+	-
Base Pairs Predicted	+	True Positive	False Positive
	-	False Negative	True Negative

Figure 1.

The confusion matrix for binary classification. Base pairs that are in the accepted structure and are predicted are true positives. Base pairs that are not in the accepted structure and also not in the prediction are true negatives. Base pairs that are in the accepted structure, but not predicted, are false negatives. Base pairs that are not in the accepted structure, but are predicted, are false positives. The true positives and true negatives are therefore correct predictions (labeled in green). The false negatives and false positives are incorrect predictions (labeled in red).



$$\Delta G^{\circ}_{37 \text{ bulge}} = 4.1 \pm 0.45 \text{ kcal/mol}$$



$$\Delta G^{\circ}_{37 \text{ bulge}} = 3.3 \pm 0.45 \text{ kcal/mol}$$

Figure 2.

The comparison of folding free energy for two duplexes with a single U bulge. The top duplex has a single state. The second duplex has two states because the bulged U can migrate between two positions. The bulged U is assumed to not interrupt the helix, and therefore the two states in the second duplex have identical base pair stacks. In free energy, the entropic bonus for populating multiple states is [55, 69]:

$$\Delta G^{\circ}_{\text{bonus}} = -RT \ln(\text{number of states}).$$

For the lower duplex, the number of states populated by the bulge is 2, and the bonus is estimated as:

$$\Delta G^{\circ}_{\text{bonus}} = -\left(1.9859 \text{ cal K}^{-1} \text{ mol}^{-1}\right)(310.15 \text{ K}) \ln(2) = -0.43 \text{ kcal/mol},$$

which is consistent with the observed difference of stability of $-0.84 \pm 0.64 \text{ kcal/mol}$ [68, 97]. The average stability of all measured single nucleotide bulges in one state is $3.7 \pm 0.2 \text{ kcal/mol}$ at $37 \text{ }^{\circ}\text{C}$ [68, 98–103]. The average stability of all measured single nucleotide bulges in two states, excluding C bulges that are known to have additional stability [55], is $2.0 \pm 0.1 \text{ kcal/mol}$. Although this neglects sequence-specific effects, the average additional stability for a single nucleotide bulge that populates two states is therefore $-1.7 \pm 0.2 \text{ kcal/mol}$.

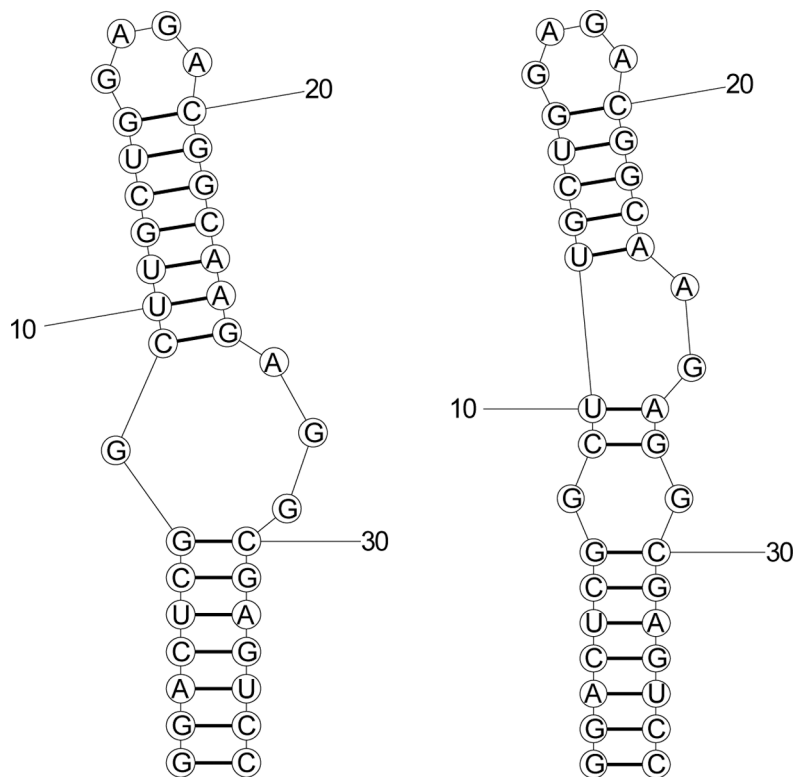


Figure 3.

The two NMR-observed conformations for the HIV-1 stem-loop 1 [104]. The left is the major conformation and the right is the minor conformation. Interestingly, RNAstructure predicts the left structure as the lowest free energy structure and the right structure as the maximum expected accuracy (MEA) structure [56, 78, 105]. In this case, the fluctuating pairs are displaced by two nucleotides on one side (i.e. pairs from nucleotides 9 and 10 to nucleotides 25 and 26 switch to pairs to nucleotides 27 and 28).

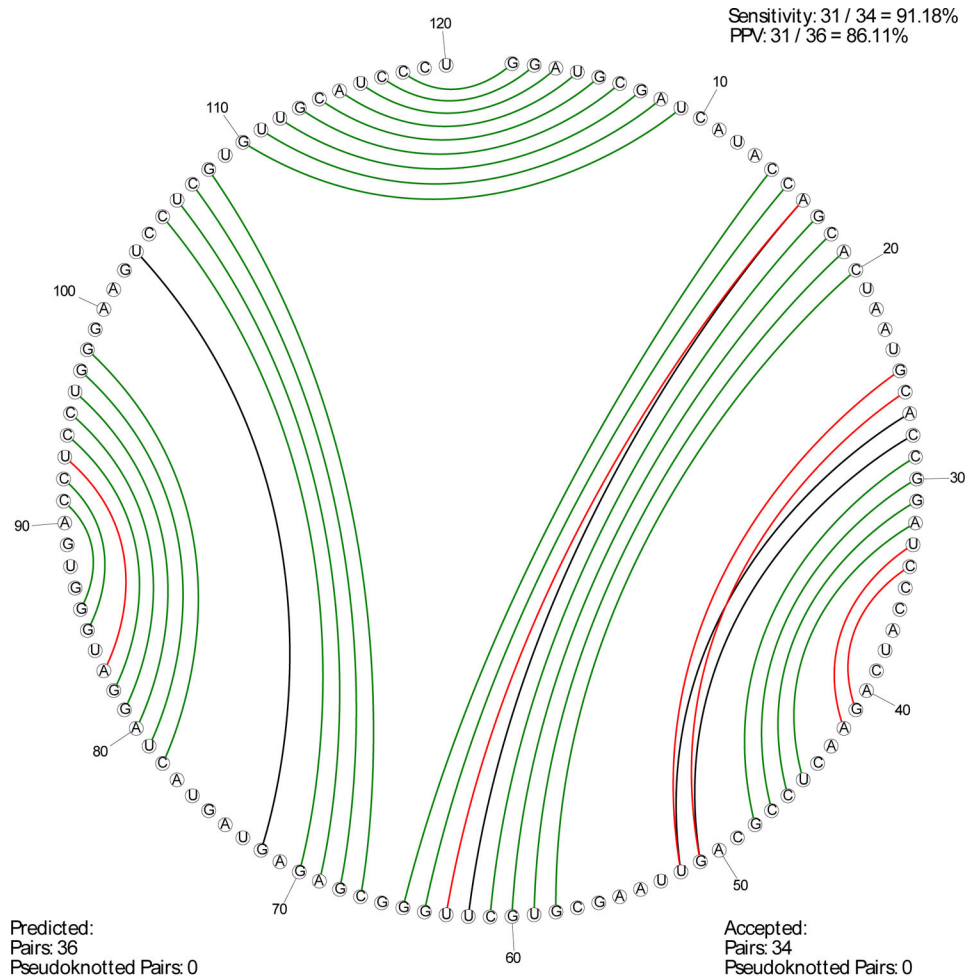


Figure 4.

Example output for *CircleCompare*. Here, the predicted lowest free energy structure [78] for the *Arabidopsis thaliana* 5S rRNA is compared against the accepted structure [39]. The sensitivity of the prediction is 91% and the PPV is 86%. A key in the lower left explains the color code for the pairs.