



# Evaluation of RNA secondary structure prediction for both base-pairing and topology

Yunjie Zhao<sup>1,2</sup>, Jun Wang<sup>1</sup>, Chen Zeng<sup>3</sup>, Yi Xiao<sup>1</sup>✉

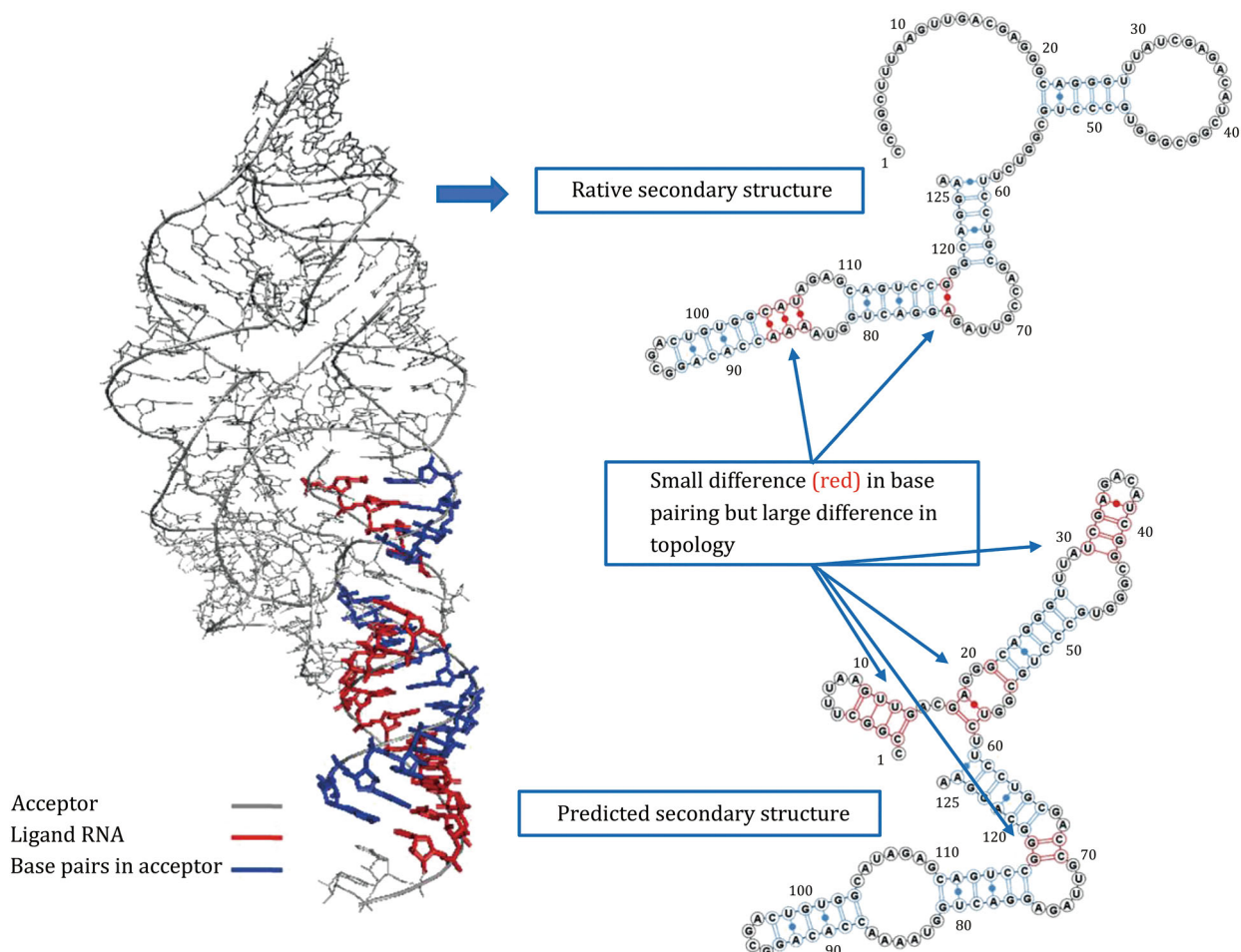
<sup>1</sup> Institute of Biophysics, School of Physics and Key Laboratory of Molecular Biophysics of the Ministry of Education, Huazhong University of Science and Technology, Wuhan 430074, China

<sup>2</sup> Institute of Biophysics and Department of Physics, Central China Normal University, Wuhan 430079, China

<sup>3</sup> Department of Physics, The George Washington University, Washington, DC 20052, USA

Received: 22 March 2018 / Accepted: 8 April 2018 / Published online: 20 June 2018

## Graphical abstract



✉ Correspondence: yxiao@hust.edu.cn (Y. Xiao)

**Abstract** Secondary structures of RNAs are crucial to the understanding of their tertiary structures and functions. At present, many theoretical methods are widely used to predict RNA secondary structures. The performance of these methods has been evaluated but only for their ability of base-pairing prediction. However, the topology of a RNA secondary structure is more important for understanding its tertiary structure and function, especially for long RNAs. In this paper, we constructed a new non-redundant RNA database containing 73 RNA with lengths of 50–300 nucleotides and benchmarked four popular algorithms for both base pairing and topology. The results show that the prediction accuracy of secondary structure topology is only 38%, in contrast to 70% for that of base pairing. Furthermore, the topological consistency is not strongly correlated to the base-pairing consistency. Our results will be helpful to understand the limitations of RNA secondary structure prediction methods from a different point of view and also to their improvements in future.

**Keywords** RNA secondary structure prediction, Base pairs, Structural topology, Prediction accuracy

## INTRODUCTION

Now it is recognized that RNA plays more important roles in the life process than expected (Li *et al.* 2017; Zhao *et al.* 2016). Besides the messenger RNA, transfer RNA, and ribosomal RNA in the genetic central dogma, many new non-coding RNAs have been discovered to have important roles in various biological processes. Among them, there are large RNAs, like ribonuclease, ribozyme, signal recognition particle RNA (SRP RNA), and riboswitches. The tertiary structures of these RNA molecules are very important for their biological functions (Zhao *et al.* 2013). For example, the riboswitches are capable of binding the metabolites to regulate gene expressions using a variety of secondary and tertiary structures (Gong *et al.* 2014; Weinberg *et al.* 2017). Due to the technology limitations, it is still a challenge to determine RNA tertiary structures experimentally. So, many groups developed computational strategies to predict RNA tertiary structures. Since RNA folding is considered as a hierarchical process (Tinoco and Bustamante 1999), most successful approaches of predicting RNA tertiary structures are based on secondary structures (Cao and Chen 2011; Popenda *et al.* 2012; Wang *et al.* 2015, 2017; Xu and Chen 2015; Zhao *et al.* 2012) and this can improve the accuracy of RNA tertiary structure predictions significantly.

## REVIEW

The methods of RNA secondary structure prediction have been well established (Mathews *et al.* 2016). The

most accurate secondary structure prediction method is to use the multiple sequence analysis or shape-directed to find the conserved motifs (Tan *et al.* 2017). However, this is not always practical due to some unique sequences and limited information at present. Therefore, the physics-based free energy minimization approach of RNA secondary structure prediction is still widely used by biologists. The performances of this type of secondary structure predictions have been evaluated previously and were found to be about 70% when comparing predicted and native base pairs on an RNA database without riboswitches and ribozymes (Mathews *et al.* 2004; Xu *et al.* 2012). Under such an accuracy of base-pair prediction, if using the predicted secondary structure information in the tertiary structure prediction, the topology of the former is critical to the performance of the latter. Therefore, it is also important to know how accurately the current algorithms predict the topologies of RNA secondary structures besides base pairing. Here, we introduce a topology consistency metrics to evaluate the consistency of the topology of a predicted secondary structure with that of the corresponding experimental one. We shall use a database consisting of 73 RNAs with length of 50–300 nt extracted from the RCSB Protein Data Bank (Westbrook *et al.* 2003) and including all types of RNAs known by now, and predict their secondary structures using four popular methods based on free energy minimization: Mfold (Zuker 2003), RNAfold (Gruber *et al.* 2015), RNAstructure (Bellaousov *et al.* 2013), and RNAshapes (Janssen and Giegerich 2015). We also analyze some possible reasons for the difficulties of the prediction of correct topologies of RNA secondary structures.

## RESULTS AND DISCUSSION

### Features of native secondary structures

There are several types of RNAs in the non-redundant RNA database, including riboswitch, tRNA (transfer RNA), HCV-IRES (hepatitis C viral-internal ribosome entry site) RNA, Ribonuclease, Ribozyme, Ribosomal RNA, SRP (signal recognition particle) RNAs, Group introns, and others (listed in Table 1). RNA secondary structure can be divided into helical stems and various kinds of loops, such as internal loops, bulge loops, and multi-way junction loops. The helical stems are formed by complementary canonical Watson-Crick and non-canonical Watson-Crick base pairs. The internal loops or bulge loops are unpaired nucleotides between helical stems. The multi-way junction loops are the connections between three or more helical stems. Table 2 provides the features of the secondary structures of RNAs in the database. It shows that about 58% of the total 6936 nucleotides are involved in forming base pairs, and, therefore, about half of the nucleotides are involved in base pairs and other half in different types of loops. The correct formations of both helices and loops are crucial to the overall topology of the secondary structures.

Current algorithms prefer to consider the canonical base pairs (A-U, C-G, and G-U base pairs) in RNA secondary structure predictions. However, there are nearly 10% of the 2028 native base pairs that are non-canonical base pairs (A-G, A-C, A-A, U-U, C-U, C-C, and G-G base pairs). This implies that the prediction accuracy of base pairs cannot be higher than 90% by using current secondary structure prediction methods.

### Base-pairing consistency of native and predicted secondary structures

Table 3 shows the base-pairing consistency of native and predicted secondary structures by the four methods stated above. It is shown that the mean values of the sensitivity and positive predictive value (PPV) are similar for the four algorithms and both are around 0.70. This is in agreement with previous results (Mathews *et al.* 2004) where they used a RNA database without riboswitches and ribozymes. Among the nine types of RNA, the values of the sensitivity (0.56–0.60) and PPV (0.44–0.49) for ribosomal RNA are significantly lower than other types for all the four algorithms. One of the reasons for this may be due to the base-pairing or tertiary interactions with proteins or other RNAs because of compact assembly of ribosomal RNAs and proteins in ribosomes. For HCV-IRES RNA, RNAsHapes, Mfold, and

**Table 1** PDB ID of RNAs in the database

PDB	Type	Length	PDB	Type	Length
Unbound RNA					
3E5C	Riboswitch	54	1KH6	HCV-IRES	53
1U8D	Riboswitch	68	1P5O	HCV-IRES	77
3IVN	Riboswitch	69	1U9S	Ribonuclease	161
1Y26	Riboswitch	71	2QUS	Ribozyme	69
2GIS	Riboswitch	94	2OIU	Ribozyme	71
3F2Q	Riboswitch	112	1C2X	Ribosomal	120
2QBZ	Riboswitch	161	2IL9	Ribosomal	142
3DIG	Riboswitch	174	1Z43	SRP RNA	101
1YFG	tRNA	75	1KXK	Group intron	70
1EHZ	tRNA	76	1GRZ	Group intron	247
2 K4C	tRNA	76	1S9S	Others	101
2TRA	tRNA	79	1FOQ	Others	109
3A3A	tRNA	90			
Bound RNA in RNA-protein and RNA-RNA complexes					
3EGZ	Riboswitch	65	3A2K	tRNA	77
3K0J	Riboswitch	87	2ZUE	tRNA	78
3IWN	Riboswitch	93	1H3E	tRNA	86
2DLC	tRNA	69	1WZ2	tRNA	88
3EPH	tRNA	69	3ADB	tRNA	92
2DU3	tRNA	71	1M5 K	Ribozyme	92
2ZNI	tRNA	72	2GCS	Ribozyme	125
1B23	tRNA	74	2NZ4	Ribozyme	141
1GIX	tRNA	74	1DK1	Ribosomal	60
2D6F	tRNA	74	1UN6	Ribosomal	61
3AKZ	tRNA	74	3BBO_c	Ribosomal	103
1GAX	tRNA	75	3JYX_3	Ribosomal	113
2AZX	tRNA	75	3BBO_b	Ribosomal	117
1EIY	tRNA	76	1NKW	Ribosomal	124
1F7U	tRNA	76	1S1I	Ribosomal	125
2DER	tRNA	76	3JYX_4	Ribosomal	157
2WRN	tRNA	76	3KTW	SRP RNA	96
3KFU	tRNA	76	1L9A	SRP RNA	128
1C0A	tRNA	77	1U6B	Group intron	197
1H4Q	tRNA	77	2RKJ	Group intron	246
1J1U	tRNA	77	3HJW	Others	58
1J2B	tRNA	77	3HAY	Others	71
2FMT	tRNA	77	2IHX	Others	75
2WWL	tRNA	77	3CUL	Others	92

RNAfold have much lower sensitivity (0.54–0.57) and PPV (0.60) except RNAstructure. For other types of RNA, the performances of the four algorithms are similar.

To further analyze the base-pairing consistency of the native and predicted secondary structures, the data in Tables 2 and 3 are also visualized in different ways in Fig. 1. Figure 1A shows the number of base pairs versus the lengths for the native secondary structures in the

**Table 2** The summary of native secondary structures

RNA type	Sequence number	Nucleotides	Base pairs
Riboswitch	11	1048	327
tRNA	31	2386	679
HCV-IRES	2	130	46
Ribonuclease	1	161	48
Ribozyme	5	498	140
Ribosomal RNA	10	1122	270
SRP RNA	3	325	122
Group intron	4	760	218
Others	6	506	178
Total	73	6936	2028

database. As expected, the base-pair number increases as the sequence length and has a high linear correlation coefficient of 0.86. The slope of the fitting line describes the probability of the base pairs formation, which is about one base pair for every four nucleotides as mentioned above. Figure 1B shows the consistent base-pair number (the number of true positives) of predicted base pairs versus that of native base pairs in the database. The two numbers also have a high linear correlation with a correlation coefficient of 0.83. It also shows that the base-pairing consistency does not decrease quickly with the number of the native base pairs or length of RNA because many short RNAs also have low base-pairing consistency. This is shown more clearly in Fig. 1C and D, which gives the base-pair sensitivity versus RNA length and the number of native base pairs, respectively. They show that about 20% short RNAs (<100 nt) has sensitivity less than 0.5. These results indicate that even for short RNAs (<100 nt) the base pairs cannot be correctly predicted by using current algorithms completely. For long RNAs (>100 nt), since the number of long RNA chains in the database is small, more long RNA sequences are needed to give a reliable conclusion about the base-pairing consistency.

### Topological consistency of native and predicted secondary structures

Table 3 also shows the topological consistency of the native and predicted secondary structures. It shows that on average only about 38% of the predicted secondary structures have identical topologies with those of the native ones (Fig. 2A). This consistency rises up to 60%–70% if we include those predicted secondary structures that have similar topologies with the native ones (Fig. 2B). The topological consistency levels of the

predicted secondary structures of the group intron and ribosomal RNA are the lowest ones for all the four algorithms besides ribonuclease which has only one sequence in the database. The reason for this low topological consistency is that the PPV, *i.e.*, the percentage of predicted base pairs occurs in the native secondary structure, is only 69%. In other words, in the predicted base pairs there are about 31% that are inconsistent with the native ones. This inconsistent base pairing may form different internal bulge and multiple loops from the native ones and lead to different overall topologies of the secondary structures. Figure 2 gives three examples of topological consistency of the native and predicted secondary structures by RNashapes. Figure 2A is an example (the ribozyme RNA (PDB ID: 2QUS)) that the topologies of the native and predicted secondary structures are identical and there are only two missing non-canonical base pairs C12–A53 and A38–A51 in the predicted secondary structure in comparison with the native one (marked in red color). In this case, the sensitivity of the predicted base pairs is about 0.91. Figure 2B shows an example (ribosomal RNA (PDB ID: 1DK1)) that the topologies of the native and predicted secondary structures are similar but there are base-pairing shifts (the region in red color). In this case, the sensitivity of the predicted base pairs is about 0.72. There is a bulge C50 in the native secondary structure. However, in the predicted secondary structure, C50 forms a base pair with G22 and there is an internal loop (U23, G24, A48, and C49) formed with shifted base pairs G25–C47, G26–C46, and A27–U45. Figure 2C is an example (tRNA (PDB ID: 3BBV)) that the topologies of the native and predicted secondary structures are completely different. The native secondary structure has a four-way junction but the predicted result is a long helical stem only with internal loops and bulge loops. In this case, the sensitivity of the predicted base pairs is about 0.52.

### Some discussions

The results above show that it is still a challenge to predict the topology of RNA secondary structures correctly and the identical topology consistency of predicted secondary structures with native ones is only about 38% on average. Therefore, the RNA tertiary structure prediction based on pure predicted secondary structures is greatly limited. Even the similar topology consistency of the predicted secondary structures with the native ones can reach 60%–70%, but in this case the predicted secondary structures usually introduce different or additional bulge and internal loops that may also decrease the accuracy of

**Table 3** The base-pairing and topological consistencies by different RNA secondary structure prediction algorithms

RNA type	RNASHAPES			Mfold		
	Sensitivity	PPV	Identical(similar) topology consistency	Sensitivity	PPV	Identical(similar) topology consistency
Riboswitch	0.73	0.76	0.45(0.73)	0.71	0.74	0.45(0.64)
tRNA	0.69	0.67	0.39(0.48)	0.68	0.66	0.39(0.55)
HCV-IRES	0.54	0.60	0.50(0.50)	0.54	0.60	0.50(0.50)
Ribonuclease	0.88	0.84	0.00(1.00)	0.88	0.84	0.00(1.00)
Ribozyme	0.81	0.76	0.40(0.60)	0.81	0.75	0.40(0.60)
Ribosomal RNA	0.58	0.47	0.10(0.70)	0.60	0.49	0.10(0.70)
SRP RNA	0.75	0.85	0.33(0.67)	0.76	0.89	0.33(0.67)
Group intron	0.69	0.68	0.00(0.25)	0.70	0.69	0.00(0.25)
Others	0.89	0.92	1.00(1.00)	0.89	0.92	1.00(1.00)
Total	0.71	0.69	0.38(0.60)	0.71	0.69	0.38(0.62)

RNA type	RNAfold			RNAstructure		
	Sensitivity	PPV	Identical(similar) topology consistency	Sensitivity	PPV	Identical(similar) topology consistency
Riboswitch	0.69	0.71	0.45(0.73)	0.69	0.73	0.36(0.73)
tRNA	0.69	0.65	0.35(0.48)	0.78	0.75	0.48(0.81)
HCV-IRES	0.57	0.60	0.50(0.50)	0.85	0.91	0.50(1.00)
Ribonuclease	0.88	0.84	0.00(1.00)	0.88	0.82	0.00(1.00)
Ribozyme	0.72	0.67	0.20(0.40)	0.80	0.74	0.20(0.60)
Ribosomal RNA	0.56	0.44	0.10(0.70)	0.57	0.44	0.00(0.50)
SRP RNA	0.70	0.81	0.00(0.67)	0.66	0.77	0.00(0.33)
Group intron	0.70	0.67	0.00(0.25)	0.69	0.66	0.00(0.25)
Others	0.83	0.86	0.83(0.83)	0.83	0.86	0.67(1.00)
Total	0.69	0.66	0.33(0.58)	0.73	0.70	0.34(0.71)

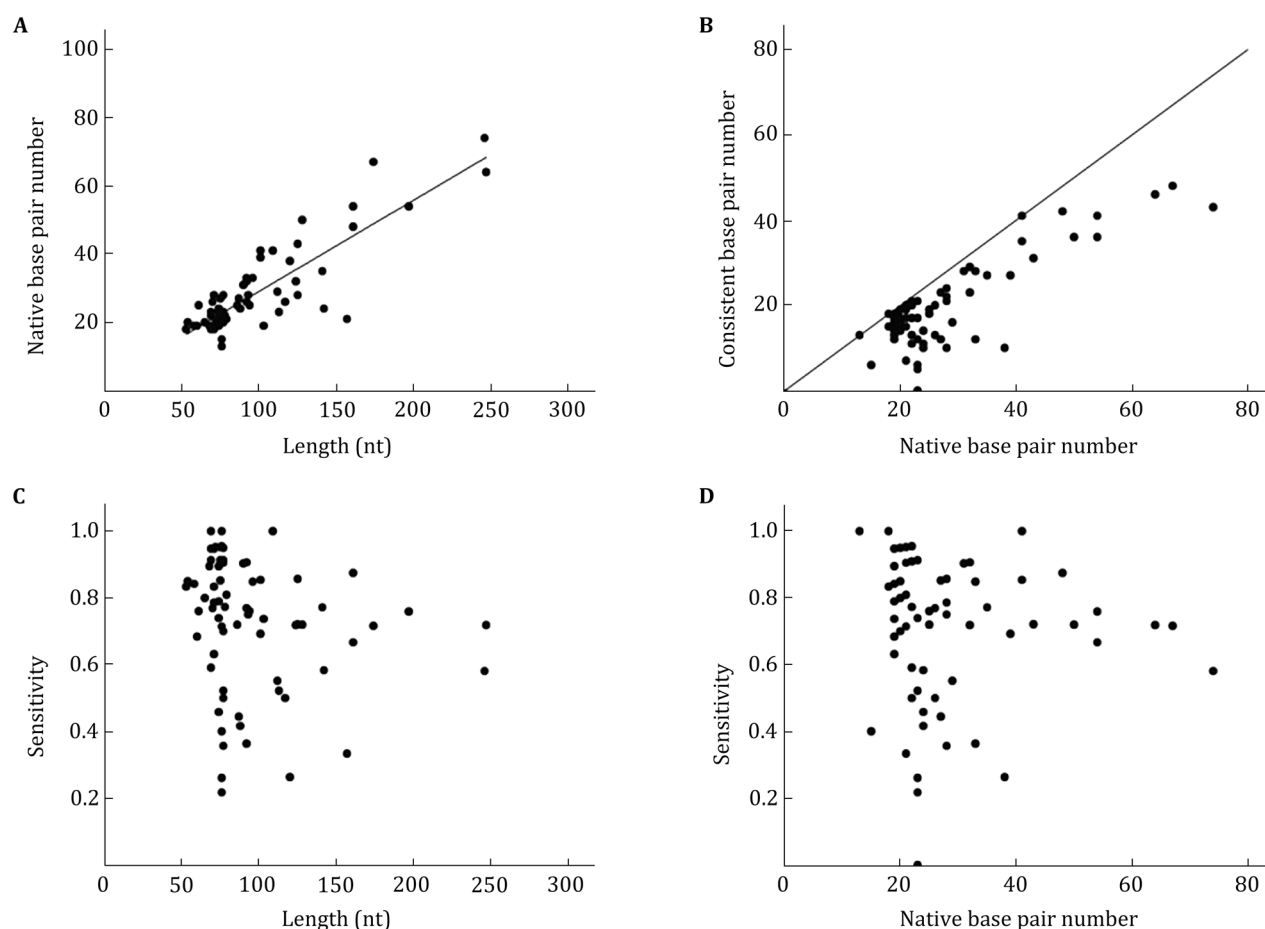
the RNA tertiary structure prediction based on them. Therefore, from a topological point of view, the success rate of RNA tertiary structure prediction based on pure predicted secondary structures can reach more than 38% or 60%–70% if wrong internal and bulge loops are ignored.

The results above (Fig. 2) also show that the topological consistency is unnecessarily always correlated to the base-pair consistency, as indicated in Fig. 3. For example (Fig. 4), the sensitivity and PPV of the predicted base pairs of a ribozyme (PDB ID: 2GCS) are about 0.86 and 0.65, respectively, but the small number of inconsistent base pairs makes the topology of the predicted secondary structure significantly different from the native one (Figs. 4B, C). Furthermore, we analyzed the statistical correlation of sensitivity/specificity and topological consistency (including both identical and similar topologies) for RNASHAPES (0.64/0.54), Mfold (0.65/0.52), RNAfold (0.50/0.45), and RNAstructure (0.78/0.63), respectively (Fig. 3). The low

correlation values indicate that the topological consistency is not strongly correlated to the base-pairing consistency and small number of wrong-paired bases can change the topologies significantly.

One of the reasons that affects the base-pairing and topological predictions may be due to the base-pairing or tertiary interactions with other molecules (Perederina *et al.* 2002). Figure 4A is the native complex structure of a ribozyme (PDB ID: 2GCS) and its amino RNA inhibitor. It shows that G3 to U13 and G14 form canonical and non-canonical base pairs with the amino RNA inhibitor in the native secondary structure and cannot form the hairpin structure in the predicted structure (Fig. 4C). Similarly, C16–A18 form tertiary interactions with U31–G33 and cannot form the internal loop G17–C58 and A18–U57 in the predicted structure. These indicate that accurate prediction of the topology of secondary structures of RNA molecules also needs to consider their interactions with other molecules, although the number of these interactions is





**Fig. 1** Statistics of the secondary structure prediction results. The line in **A** is linearly fitting and the line in **B** is diagonal. The unit of length and base-pair number is nucleotide. Sensitivity describes what percentage of native base pairs occurs in the predicted secondary structure

small in comparison with that of the total native base pairs.

## CONCLUSION

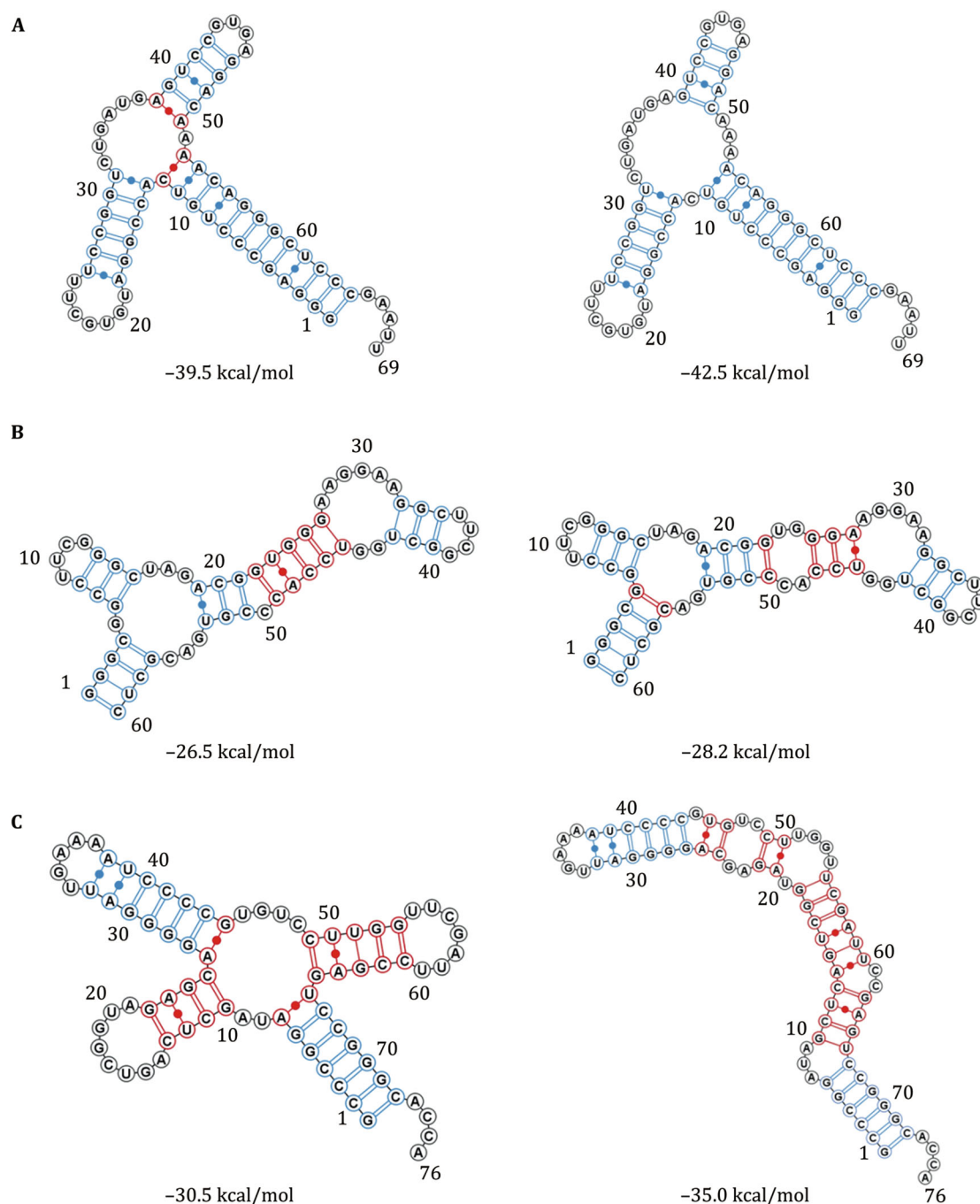
In this paper, we evaluate the physics-based free energy minimization secondary structure prediction methods by comparing the base-pairing as well as topological consistencies. We built a non-redundant RNA tertiary structure database consisting main types of RNAs to construct our statistical analysis. The benchmark tests show that the percentages of correct predictions of the base-pair predictions and topology are about 70% and 38% on average, respectively. Furthermore, the topological consistency is not strongly correlated to the base-pairing consistency under current accuracy of base-pair prediction. Relatively high accuracy of base-pair prediction does not mean correct topology of secondary structure. This suggests that experimental

information about secondary structure is usually needed to build accurate tertiary structures of RNAs. Our results will be helpful to understand the limitations of RNA secondary structure prediction methods and their applications in RNA tertiary structure prediction.

## MATERIALS AND METHODS

### Database of experimental RNA tertiary structures

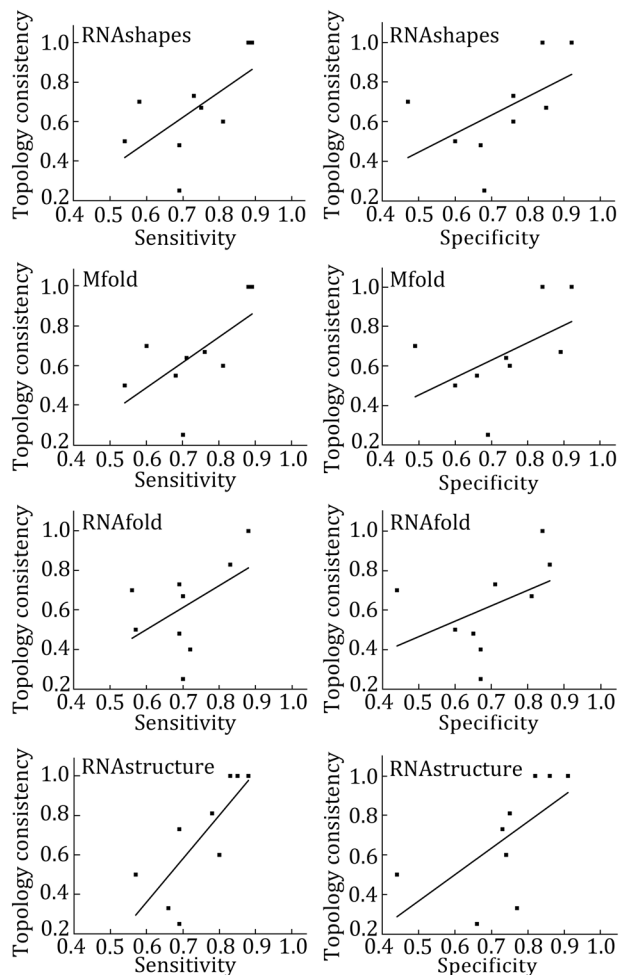
In order to evaluate the performance of different RNA secondary structure prediction methods fairly, we built a non-redundant RNA tertiary structure database from the experimental RNA tertiary structures in the RCSB Protein Data Bank (PDB) (Westbrook *et al.* 2003). The RNAs with size of 50–300 nt are practical for RNA tertiary structure prediction (Zhao *et al.* 2012). Therefore, we first collected all the RNA structures with lengths



**Fig. 2** Three examples of native (left) and predicted (right) secondary structures. Their overall topologies are identical **(A)**, similar **(B)**, and different **(C)**. The consistent and inconsistent base pairs are colored in *blue* and *red*, respectively. Their free energies are also listed

between 50 and 300 nucleotides in the PDB database. The current non-redundant RNA tertiary structure databases of statistical potential used the sequence identity of 95% and 80% to reduce redundancy (Bernauer *et al.* 2011; Capriotti *et al.* 2011; Wang *et al.* 2015). Here, we used a lower sequence identity of 75% to remove possible homology structures in the selected

RNAs. Although a lot of efforts, it is still a challenge to predict the pseudoknots. And so the RNAs with pseudoknots are not included in the non-redundant RNA tertiary structure database. The database totally contains 73 RNA tertiary structures (Table 1), including 25 in unbound state and 48 in RNA-protein and RNA-RNA complexes.



**Fig. 3** Statistical linear fitting of sensitivity/specificity and topological consistency for RNAshapes, Mfold, RNAfold, and RNAstructure, respectively

### Secondary structure prediction and analysis

The free energy minimization approach of RNA secondary structure prediction is the most popular method when RNA homologous information is limited (Mathews and Turner 2006), *e.g.*, Mfold (Zuker 2003), RNAfold (Gruber *et al.* 2015), RNAstructure (Bellaousov *et al.* 2013), and RNAshapes (Janssen and Giegerich 2015). Mfold uses a dynamic programming algorithm to generate a set of candidates of secondary structure for an RNA, and then calculates their free energies by adding up those of independent subunits using the nearest neighborhood approximation. The free energies of the subunits were determined experimentally. RNAstructure is similar to Mfold but uses alternative set of thermodynamic parameters. RNAfold is also based on the minimum free energy model but it can compute the equilibrium partition functions and base-pairing probabilities. RNAshapes is different from the three

algorithms above. It first clusters potential secondary structures of an RNA into different abstract shapes and finds a minimum free energy structure as the representative of each shape. Then, it can calculate the shape probability or use other biological information to identify the possible native secondary structure from the representatives. Here, we use these four algorithms to evaluate the performance of secondary structure prediction. The experimental secondary structures (called as “native secondary structures”) were generated from the experimental tertiary structures in the RNA database using the sequence to structure (S2S) algorithm (Jossinet and Westhof 2005). S2S can display the RNA data, like sequences, secondary structures, and tertiary structures. In addition, we analyzed the Watson–Crick base pairs (Leontis and Westhof 2001) (A–U, C–G, and G–U base pairs) and non-Watson–Crick base pairs (Leontis *et al.* 2002) (A–G, A–C, A–A, U–U, C–U, C–C, and G–G base pairs). In order to observe the consequence of RNA–protein or RNA–RNA tertiary interactions on predicted RNA secondary structures, we also calculated the hydrogen bonds between RNA and protein or RNAs in the complexes by the hydrogen bond calculation algorithm HBPLUS (McDonald and Thornton 1994). The free energy of RNA secondary structure was calculated by using the RNAeval algorithm in Vienna RNA package (Gruber *et al.* 2015).

We analyzed base-pairing consistencies between native and predicted secondary structures of the RNA structures in the database. The base-pairing consistency was measured by both sensitivity (*STY*) and positive predictive value of precision (*PPV*) (Parisien *et al.* 2009).

$$STY(sensitivity) = \frac{TP}{TP + FN}$$

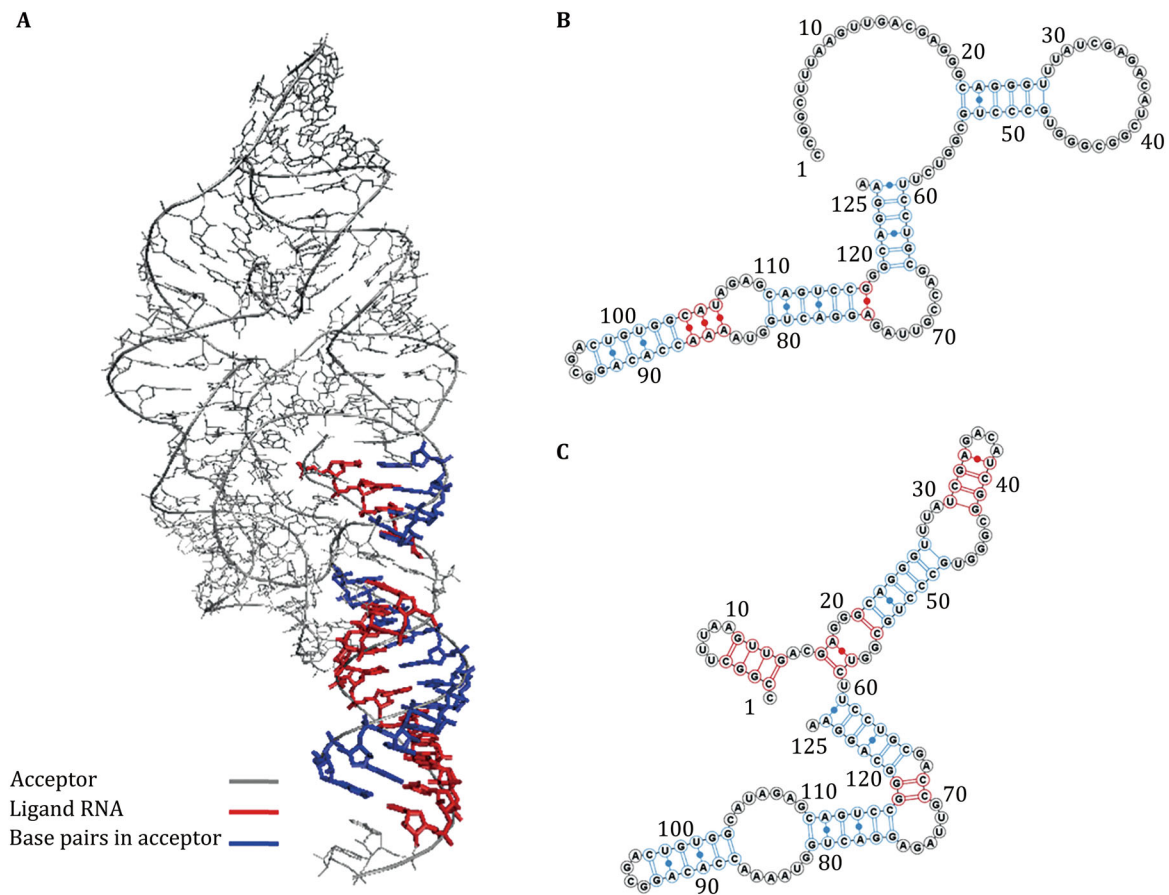
$$PPV(precision) = \frac{TP}{TP + FP}$$

The base pairs found in both native and prediction sets are true positives (*TP*). The base pairs found in native sets but not in prediction sets are false negatives (*FN*). The base pairs found in prediction sets but not in native sets are false positives (*FP*). Sensitivity (*STY*) describes what percentage of native base pairs occurs in the predicted secondary structure. Specificity (*PPV*) denotes what percentage of predicted base pairs occurs in the native secondary structure.

### Topological consistency analysis

The topological consistency measures the topological similarity of the native and predicted secondary structures. Since it is difficult to find a simple index to clearly





**Fig. 4** **A** Tertiary interactions between a ribozyme (PDB ID: 2GCS) and ligand RNAs. **B** and **C** are the native and predicted secondary structures of 2GCS, respectively. The consistent and inconsistent base pairs between them are colored in *blue* and *red*, respectively

distinguish the difference of RNA secondary structures, we divide the topological consistency into three types: identical, similar, and different (Fig. 2). The topology of a predicted secondary structure is considered to be identical to that of the native one if the former has the same loops with the latter (Fig. 2A). In this case, the predicted secondary structures may have a few additional base pairs that do not occur in the native ones or a few of the native base pairs that do not form. The topology of a predicted secondary structure is considered to be similar to that of the native one if the former has the same multi-way junction loops with the latter but different internal and bulge loops (Fig. 2B). Finally, the topology of a predicted secondary structure is considered to be different from that of the native one if the former has different multi-way junction loops with the latter (Fig. 2C).

**Acknowledgements** This work is supported by the NSFC under Grant Nos. 31570722 and 11374113.

### Compliance with Ethical Standards

**Conflict of interest** Yunjie Zhao, Jun Wang, Chen Zeng, and Yi Xiao declare that they have no conflict of interest.

**Human and animal rights and informed consent** This article does not contain any studies with human or animal subjects performed by any of the authors.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

### References

- Bellaousov S, Reuter JS, Seetin MG, Mathews DH (2013) RNAstructure: web servers for RNA secondary structure prediction and analysis. *Nucleic Acids Res* 41:W471–W474

- Bernauer J, Huang X, Sim AY, Levitt M (2011) Fully differentiable coarse-grained and all-atom knowledge-based potentials for RNA structure evaluation. *RNA* 17:1066–1075
- Cao S, Chen SJ (2011) Physics-based *de novo* prediction of RNA 3D structures. *J Phys Chem B* 115:4216–4226
- Capriotti E, Norambuena T, Marti-Renom MA, Melo F (2011) All-atom knowledge-based potential for RNA structure prediction and assessment. *Bioinformatics* 27:1086–1093
- Gong Z, Zhao Y, Chen C, Duan Y, Xiao Y (2014) Insights into ligand binding to PreQ1 Riboswitch Aptamer from molecular dynamics simulations. *PLoS ONE* 9:e92247
- Gruber AR, Bernhart SH, Lorenz R (2015) The ViennaRNA web services. *Methods Mol Biol* 1269:307–326
- Janssen S, Giegerich R (2015) The RNA shapes studio. *Bioinformatics* 31:423–425
- Jossinet F, Westhof E (2005) Sequence to structure (S2S): display, manipulate and interconnect RNA data from sequence to structure. *Bioinformatics* 21:3320–3321
- Leontis NB, Westhof E (2001) Geometric nomenclature and classification of RNA base pairs. *RNA* 7:499–512
- Leontis NB, Stombaugh J, Westhof E (2002) The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res* 30:3497–3531
- Li X, Bu D, Sun L, Wu Y, Fang S, Li H, Luo H, Luo C, Fang W, Chen R, Zhao Y (2017) Using the NONCODE Database Resource. *Curr Protoc Bioinform* 58:12.16.1–12.16.19
- Mathews DH, Turner DH (2006) Prediction of RNA secondary structure by free energy minimization. *Curr Opin Struct Biol* 16:270–278
- Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci USA* 101:7287–7292
- Mathews DH, Turner DH, Watson RM (2016) RNA secondary structure prediction. *Curr Protoc Nucleic Acid Chem* 67:1–19
- McDonald IK, Thornton JM (1994) Satisfying hydrogen bonding potential in proteins. *J Mol Biol* 238:777–793
- Parisien M, Cruz JA, Westhof E, Major F (2009) New metrics for comparing and assessing discrepancies between RNA 3D structures and models. *RNA* 15:1875–1885
- Perederina A, Nevskaya N, Nikonov O, Nikulin A, Dumas P, Yao M, Tanaka I, Garber M, Gongadze G, Nikonov S (2002) Detailed analysis of RNA–protein interactions within the bacterial ribosomal protein L5/5S rRNA complex. *RNA* 8:1548–1557
- Popenda M, Szachniuk M, Antczak M, Purzycka KJ, Lukasiak P, Bartol N, Blazewicz J, Adamiak RW (2012) Automated 3D structure composition for large RNAs. *Nucleic Acids Res* 40:e112
- Tan Z, Fu Y, Sharma G, Mathews DH (2017) TurboFold II: RNA structural alignment and secondary structure prediction informed by multiple homologs. *Nucleic Acids Res* 45:11570–11581
- Tinoco I Jr, Bustamante C (1999) How RNA folds. *J Mol Biol* 293:271–281
- Wang J, Zhao Y, Zhu C, Xiao Y (2015) 3dRNAscore: a distance and torsion angle dependent evaluation function of 3D RNA structures. *Nucleic Acids Res* 43:e63
- Wang J, Mao K, Zhao Y, Zeng C, Xiang J, Zhang Y, Xiao Y (2017) Optimization of RNA 3D structure prediction using evolutionary restraints of nucleotide-nucleotide interactions from direct coupling analysis. *Nucleic Acids Res* 45:6299–6309
- Weinberg Z, Nelson JW, Lunse CE, Sherlock ME, Breaker RR (2017) Bioinformatic analysis of riboswitch structures uncovers variant classes with altered ligand specificity. *Proc Natl Acad Sci USA* 114:E2077–E2085
- Westbrook J, Feng Z, Chen L, Yang H, Berman HM (2003) The protein data bank and structural genomics. *Nucleic Acids Res* 31:489–491
- Xu X, Chen SJ (2015) Physics-based RNA structure prediction. *Biophys Rep* 1:2–13
- Xu Z, Almudevar A, Mathews DH (2012) Statistical evaluation of improvement in RNA secondary structure prediction. *Nucleic Acids Res* 40:e26
- Zhao Y, Huang Y, Gong Z, Wang Y, Man J, Xiao Y (2012) Automated and fast building of three-dimensional RNA structures. *Sci Rep* 2:734
- Zhao Y, Wang J, Chen X, Luo H, Zhao Y, Xiao Y, Chen R (2013) Large-scale study of long non-coding RNA functions based on structure and expression features. *Sci China Life Sci* 56:953–959
- Zhao Y, Li H, Fang S, Kang Y, Wu W, Hao Y, Li Z, Bu D, Sun N, Zhang MQ, Chen R (2016) NONCODE 2016: an informative and valuable data source of long non-coding RNAs. *Nucleic Acids Res* 44:D203–D208
- Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31:3406–3415