

# VfoldMCPX: Predicting multistrand RNA complexes

Sicheng ZHANG<sup>1+</sup>, Yi CHENG<sup>1+</sup>, Peixuan GUO<sup>2</sup> and Shi-Jie CHEN<sup>1,\*</sup>

<sup>1</sup>Department of Physics, Department of Biochemistry, and Institute for Data Science and Informatics, University of Missouri, Columbia, Missouri 65211, United States

<sup>2</sup>Center for RNA Nanobiotechnology and Nanomedicine, College of Pharmacy, Dorothy M. Davis Heart and Lung Research Institute, James Comprehensive Cancer Center, The Ohio State University, Columbus, Ohio 43210, United States

## Abstract

Multistrand RNA complexes play a critical role in RNA-related biological processes. The understanding of RNA functions and the rational design of RNA nanostructures require accurate prediction of the structure and folding stability of the complexes, including those containing pseudoknots. Here, we present VfoldMCPX, a new model for predicting two-dimensional (2D) structures and folding stabilities of multistrand RNA complexes. Based on a partition function-based algorithm combined with physical loop free energy parameters, the VfoldMCPX model predicts not only the native structure but also the folding stability of the complex. An important advantage of the model is the ability to treat pseudoknotted structures. Extensive tests on structure predictions show the VfoldMCPX model provides improved accuracy for multistranded RNA complexes, especially for RNA complexes with three or more strands and/or containing pseudoknots. We have developed a freely accessible VfoldMCPX web server at "<http://rna.physics.missouri.edu/vfoldMCPX2>".

**Keywords:** multistrand RNA; folding thermodynamics; secondary structure; RNA pseudoknot

---

\*Author to whom correspondence should be addressed; E-mail: [chenshi@missouri.edu](mailto:chenshi@missouri.edu)

<sup>+</sup>These authors contributed equally to this work.

## Introduction

Multistrand RNAs can assemble to form complexes through RNA-RNA interactions. RNA-RNA interactions play crucial roles in many biological processes, such as mRNA splicing (Griffith and Swartz 2006), RNA dimerization (Paillart et al. 2004), and prokaryotic/eukaryotic gene regulation (Dutta and Srivastava 2018; Guil and Esteller 2015). Because of the hierarchical characteristics of RNA folding and the versatility/variety of RNA structures, various multistrand RNA complexes have become promising elementary building blocks for RNA nanostructure design (Sharma et al. 2016; Afonin et al. 2014a; Shu et al. 2011; Li et al. 2015). Recent advances in structural biology have led to accurate RNA structure determination for various systems (Chen 2008; Sun et al. 2017; Zhang et al. 2013) and rapid development of RNA nanotechnology (Guo 2010; Shukla et al. 2011; Jasinski et al. 2017; Afonin et al. 2014b; Yingling and Shapiro 2007; Bindewald et al. 2008). For example, Chworos et al. generated controllable jigsaw puzzle units by predicting and designing artificial RNA building blocks (Chworos et al. 2004). Dibrov et al. solved the crystal structure of the square-shaped nanoobject self-assembled from double-stranded RNA (Dibrov et al. 2011). Furthermore, the crystal structures of the pRNA 3WJ (Zhang et al. 2013) and the hexamer structure (Saeed et al. 2021; Zhang et al. 2013) of phi29 DNA packing motor have been demonstrated to serve as powerful building blocks for RNA nanotechnology.

Motivated by the demand for the accurate prediction of (two-dimensional, 2D) structures of RNA complexes, many computational models have been developed using thermodynamic approaches (Bernhart et al. 2006; Busch et al. 2008; Andronescu et al. 2005; Rehmsmeier et al. 2004; Dimitrov and Zuker 2004; Bindewald et al. 2011, 2016; Dirks et al. 2007; Zadeh et al. 2011; Montaseri et al. 2014; Legendre et al. 2019; Kato et al. 2017). RNAcofold (Bernhart et al. 2006; Lorenz et al. 2011; Hofacker et al. 1994; Bindewald et al. 2016), bifold (Reuter and Mathews 2010), IntaRNA (Busch et al. 2008), and PairFold (Andronescu et al. 2005) enable predictions of pseudoknot-free 2D structures with intra- and inter-strand base pairs, while RNAhybrid (Rehmsmeier et al. 2004) and UNAFold (Dimitrov and Zuker 2004) mainly focus on the inter-strand base pairs. To treat multistrand RNA complexes, MultiRNAFold (Andronescu et al. 2005, 2007), NUPACK (Dirks et al. 2007; Zadeh et al. 2011), and NanoFolder (Bindewald et al. 2011) programs were developed to predict multistrand structures by linking the input strands in a specific order. RCPred (Legendre et al. 2019) provides a novel method for predicting RNA complexes by modeling the predicted secondary structures and interactions as a graph, and finding the optimal solution to the constrained maximum weight clique problem. However, most of those models, NUPACK for example, can only predict multistrand RNA complexes without crossing base pairs. Because of the huge cost of computational resources and the lack of experimentally determined thermodynamic parameters, very few models can treat multistrand RNA complexes with crossing base pairs or pseudoknots. For the models mentioned above, RCPred can handle the prediction of multiple strands and complex motifs such as pseudoknots. NanoFolder, which does not explicitly account for the

energetic contributions from pseudoknotted loops, can also successfully treat pseudoknots.

On the other hand, the previously developed VfoldCPX model (Xu and Chen 2016b, 2016a) employs the Vfold-derived loop entropic parameters (Cao and Chen 2005, 2006b, 2009, 2006a, 2011, 2012; Xu et al. 2014; Cao et al. 2014; Zhao et al. 2017) to calculate the folding stabilities, and can predict RNA-RNA complex structures with tertiary motifs, such as pseudoknotted and hairpin loop-loop kissed motifs. However, the VfoldCPX model is limited to two-strand RNA complex systems, which hampers its applications in RNA design and predictions of multistrand complexes. In the multistrand system, the intermediate states formed by any two of the strands should be accounted for because the final multistrand native structure may involve dimeric intermediate states for the global optimization of the interactions in the multistrand system (Binzel et al. 2016, 2021). Here, we develop a new model, VfoldMCPX, to treat RNA systems with multiple strands. The VfoldMCPX model is based on a partition function-based algorithm, which includes secondary structural motifs, such as hairpins, bulge/internal loops and multi-way junctions (see Fig. 1a), as well as tertiary motifs, such as H-type pseudoknotted motif (see Fig. 1b). We have developed a VfoldMCPX web server to offer a user-friendly online interface for fully automated predictions of 2D structures and base pairing probabilities of RNA multistrand complexes. The web server outputs a set of energetically stable structures for RNA complexes with up to ten strands. This new model, VfoldMCPX, as well as the web server, can provide extensive physical insights into the role of RNA-RNA interactions in RNA complexes, and can aid in the design of RNA nanostructures.

## Method

VfoldMCPX transforms the input multistrand RNA to an effective single-strand RNA sequence using the phantom linkers. The phantom linker is a three-nucleotide linker marked as "XXX" in the connected sequence and depicted as red dots in Fig. 1. From the input  $N$  RNA strands, the model determines the permutation of the strands through the strand number (from strand 1 to strand  $N$ ) and inserts  $N-1$  phantom linkers to connect the strands. The nucleotides "X" in the phantom linkers are not allowed to participate in any base pairing. The strand concentration-dependent free energy for the initiation of strand association is assumed to be independent of the RNA sequence, therefore, all the predicted multistrand RNA complexes have the same constant initiation energy. As a result, the different structures for the complexation of a given number of strands would involve the same initiation, thus we do not include the constant initiation energy term in the free energy calculation. Employing the Vfold model-derived (Xu et al. 2014; Cao et al. 2010, 2014; Cao and Chen 2006a, 2011, 2012) loop entropy parameters to calculate the free energies for each sampled structure, VfoldMCPX can predict 2D structures for multistrand RNA complexes without crossing base pairs and with H-type pseudoknots. The main features and the algorithms of the model are described below.

## Structures without crossing base pairs (secondary structures)

After linking the input RNA strands to a single-strand sequence, VfoldMCPX enumerates all the possible base pair arrangements (2D structures). Non-pseudoknotted structural motifs enumerated by the model including helices, hairpin loops, bulge/internal loops, and multi-way junctions (see Fig. 1a). VfoldMCPX computes the enthalpy and entropy for the helices based on the Turner's experimental data (Mathews et al. 2004) for the canonical and mismatched base stacks, and employs the Vfold-derived entropies for the loops. Because the nucleotides in the phantom linkers are not allowed to form base pairs, they can only appear in loops. The phantom linkers are not physical linkers, they break the chain connectivity, hence we treat linker-containing loops as loops with zero entropic contributions. This treatment allow us to perform structural predictions following the procedure below:

1. Enumerate all the possible base pair arrangements for the connected single-strand sequence.
2. Calculate the free energy for each helix by adding the free energies of all the constituent base stacks:  $\Delta G_{\text{helix}} = \sum_{\text{stack}} \Delta G_{\text{stack}}$ , where  $\Delta G_{\text{stack}}$  are given by the Turner's experimental data (Mathews et al. 2004).
3. Enumerate all the possible intra-loop (mismatched) base pairs for the loops and calculate the loop free energy by

$$\Delta G_{\text{loop}} = -k_B T \ln \sum_{\text{mismatch}} e^{-(\Delta G_{\text{mismatch}} - T \Delta S_{\text{loop}})/(k_B T)},$$

where the loop partition function is computed from the Boltzmann sum over all the possible intra-loop mismatched base stacks.  $\Delta G_{\text{mismatch}}$  denotes the free energy for mismatched base stacks,  $\Delta S_{\text{loop}}$  denotes the loop entropy and is given by the pre-tabulated Vfold-derived parameters if the loop does not contain any phantom linkers, and  $\Delta S_{\text{loop}} = 0$  otherwise.

4. Compute the total partition function from all the possible 2D structures by

$$Q_{\text{total}} = \sum_{\text{all structures}} e^{-(\Delta G_{\text{helix}} + \Delta G_{\text{loop}})/(k_B T)}$$

5. Compute the base-pairing probability  $P_{ij}$  for nucleotides  $i$  and  $j$  from the conditional partition function  $Q_{ij}$  and the total partition function  $Q_{\text{total}}$ :  $P_{ij} = Q_{ij}/Q_{\text{total}}$ , where  $Q_{ij}$  is computed for all the possible structures that contain base pair  $(i, j)$  between nucleotides  $i$  and  $j$ . We predict the most probable and the alternative structures from  $P_{ij}$  for all the possible  $(i, j)$  pairs.

## Structures containing H-type pseudoknots

A multistrand RNA can fold into structures with pseudoknotted base pairs. VfoldMCPX separates out pseudoknot-containing structures from pseudoknot-free structures when computing the partition functions. VfoldMCPX can predict 2D structures with H-type pseudoknots using the available thermodynamic parameters. A schematic figure for the H-type pseudoknot motif is given in Fig. 2a. The H-type pseudoknot motif consists of three single-stranded loops ( $L_1$ ,  $L_2$ , and  $L_3$ ) and two helix stems ( $S_1$  and  $S_2$ ). The helix stem may contain a short internal/bulge loop with loop lengths of  $(l_{11}, l_{12})$  and  $(l_{21}, l_{22})$ . We consider loop lengths  $\leq 3$  nts and assume that with such short loops, stems  $S_1$  and  $S_2$  remain quasi-linear in the form of A-form helix. Therefore, considering the 3D helical shape of the strands, we use effective lengths  $S_1^{\text{eff}}$  and  $S_2^{\text{eff}}$  for the two stems (Cao and Chen 2009)

$$S_1^{\text{eff}} = l_{12} + h_{12} + h_{11}; S_2^{\text{eff}} = l_{21} + h_{21} + h_{22}.$$

We use  $l_{12}$  and  $l_{21}$  for calculating the effective lengths because they are closely positioned upstream and downstream from the center loop  $L_2$ , which is the core loop of the motif that directly determines the relative orientation of the two stems. Therefore, the pseudoknot motif shown in Fig. 2a can be described using 5 parameters:  $(L_1, L_2, L_3, S_1^{\text{eff}}, S_2^{\text{eff}})$ . The  $\Delta S(S_1^{\text{eff}}, S_2^{\text{eff}}, L_1, L_2, L_3)$  parameters for various pseudoknot structures as characterized by  $(S_1^{\text{eff}}, S_2^{\text{eff}}, L_1, L_2, L_3)$  have been computed and tabulated by the Vfold model-based computations (Cao and Chen 2009, 2012, 2011; Cao et al. 2010). However, as shown in Fig. 2b-i, depending on the positions in the structure, the phantom linkers can alter/break a pseudoknot, therefore we treat the different scenarios separately. Specifically, VfoldMCPX predicts 2D structures with H-type pseudoknots following the procedure below:

1. Generate a helix pool for all the possible helices. Enumerate all the possible pseudoknot structures assembled from the helices.
2. Assign entropy  $\Delta S_{\text{pseudoknot}}$  to the H-type pseudoknot motif according to the numbers and positions of the phantom linkers:
  - If no phantom linkers (Figs. 2a and b) are present in the loops, we treat the structure as an intact H-type pseudoknot motif:  $\Delta S_{\text{pseudoknot}} = \Delta S(S_1, S_2, L_1, L_2, L_3)$ , where  $\Delta S(S_1, S_2, L_1, L_2, L_3)$  is the Vfold-derived entropy for a single-strand H-type pseudoknot.
  - If a phantom linker is located in loop  $L_1$  or  $L_3$  (Figs. 2c and d), we treat the structure as a hairpin with an effective reduced loop size by counting only the unpaired nucleotides for the loop. Due to lack of entropy parameters, the current version of the model does not consider structures where the phantom linkers are located in loop  $L_2$  (Fig. 2e).
  - If there exist two or more phantom linkers (Fig. 2f-i) in the pseudoknot loops, the structure is effectively decomposed to form a secondary motif without crossing base pairs.

3. Compute the pseudoknot free energy:

$$\Delta G_{\text{pseudoknot}} = \Delta G_{\text{helix stem}}(S_1) + \Delta G_{\text{helix stem}}(S_2) - T\Delta S_{\text{pseudoknot}}$$

where  $\Delta G_{\text{helix stem}}(S_1)$  ( $\Delta G_{\text{helix stem}}(S_2)$ ) is the free energy of the helix stem  $S_1$  ( $S_2$ ).

4. Calculate the total partition function  $Q_{\text{total}}$  from all the possible structures that contain pseudoknots. Compute the base-pairing probability  $P_{ij}$  from the conditional partition function  $Q_{ij}$ . Predict the most probable structures and alternative structures from  $P_{ij}$ .

Due to the availability of thermodynamic parameters, currently, we use the tabulated parameters for  $S_1$  and  $S_2 \leq 10$  base pairs,  $L_1$  and  $L_3 \leq 7$  nts, and  $L_2 \leq 6$  nts. For large  $L_1$  and  $L_3$  loops ( $> 7$  nts), we use the formula  $\Delta S = a \log(l) + b$  to compute the loop entropy, where  $l$  is the loop length,  $a$  and  $b$  are fitted parameters (Serra and Turner 1995). Furthermore, we allow nested helices in three loops ( $L_1$ ,  $L_2$ , and  $L_3$ ) and calculate the effective loop length as the sum of the number of unpaired nucleotides and the number of nested helices.

## Results and Discussions

### Test of the VfoldMCPX model for structure prediction.

To evaluate the prediction accuracy of VfoldMCPX, we construct a test set by collecting experimentally determined RNA structures that consist of two or more unique strands. We use VfoldMCPX and other available models to predict the structure for each test case and compare the performance of the different models. The structures in the test data set are from the PDB database (Berman et al. 2000), and for each 3D structure, its 2D structure is extracted by RNApdbee (Antczak et al. 2014; Zok et al. 2018). We only consider multistrand RNAs with the total nucleotide length larger than 25 nt. After removing redundant structures (structures with sequence identity larger than 90%), we have 29 pseudoknot-free structures with two unique strands, 4 H-type pseudoknot-containing structures with two unique strands, and 5 structures with three or more unique strands. We then use these structures to test VfoldMCPX and other four models that are available through web servers or source codes: RNAcofold (Bernhart et al. 2006; Lorenz et al. 2011; Hofacker et al. 1994), bifold (from the RNAstructure suite) (Reuter and Mathews 2010), MultiRNAFold (Andronescu et al. 2005, 2007) and NUPACK (Dirks et al. 2007; Zadeh et al. 2011). The results from RNAcofold and bifold are obtained from the respective web servers <http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAcofold.cgi> for RNAcofold and <https://rna.urmc.rochester.edu/RNAstructureWeb/Servers/bifold/bifold.html> for bifold, respectively. The results from MultiRNAFold and NUPACK are obtained by downloading the source code, version 3.2 for NUPACK and version 2.1 for MultiRNAFold, and running the predictions locally. We note that RNAcofold and bifold are limited to pseudoknot-free structures with two RNA strands and MultiRNAFold and NUPACK can

treat RNA structures with three or more strands, however, pseudoknots are not allowed in these predictions.

The prediction accuracy is evaluated by the positive predictive value (PPV), sensitivity (STY) values, and the  $F_1$  score:

$$STY = TP/(TP + FN); PPV = TP/(TP + FP)$$

$$F_1 \text{ score} = 2 \times (STY \times PPV)/(STY + PPV)$$

where  $TP$ ,  $FN$ ,  $FP$  denote the true positive, false negative, and false positive base pairs, respectively. PPV and STY measure the precision (positive predictive value) and sensitivity (recall) of the model, respectively, and  $F_1$  score is a balanced measurement between PPV and STY, defined as the harmonic mean of them.

We first compare the performance of VfoldMCPX and other four models on the 29 pseudoknot-free structures with two unique strands. As shown in Fig. 3, five models are ranked from top to bottom according to the mean STY, PPV and  $F_1$  score, respectively. The mean values are labeled and depicted as dots in the figure. The error bars are depicted using confidence intervals at 95% confidence level. The upper and lower bounds of the error bar reflect the statistical difference, hence a narrow error bar indicates less random errors and better accuracy. As shown in Fig. 3, all the models show nearly the same performance with VfoldMCPX giving slightly higher mean STY, PPV, and  $F_1$  score, as well as narrower error bars. The result suggests that the VfoldMCPX model is on par with the other models and provides slightly improved predictions for structures containing two strands.

To evaluate the performance of VfoldMCPX on more complicated RNA complex systems, we test VfoldMCPX on five structures with three or more unique strands, and four H-type pseudoknot-containing structures (PDB ID shown in Table 1). Two models that can treat three or more RNA strands are also tested (MultiRNAFold and NUPACK). The result is reported in Table 1, boldface numbers indicate the models that give better or the same results. The test result suggests that VfoldMCPX outperforms NUPACK and performs on par with MultiRNAFold on structures with three or more unique strands. We note that with STY, PPV, and  $F_1$  score all equal to 1.00, VfoldMCPX makes the exact prediction for PDB structure 4KZ2. For the prediction of structures with H-type pseudoknots, as shown in Table 1, VfoldMCPX gives high STY, PPV, and  $F_1$  score in all four cases. We note that among the tested models, VfoldMCPX is the only one that can treat pseudoknots, and all the test scores for VfoldMCPX are higher than 0.80 except for the STY in PDB 6YML. The predicted structures from VfoldMCPX and the native structures extracted from the 3D crystal structures are drawn in Supplemental Figure S3 for structures with three or more strands, and in Supplemental Figure S4 for structures containing H-type pseudoknots. We note that VfoldMCPX gives low scores for structure PDB-4G6R. The native structure of PDB-4G6R contains a  $7 \times 9$  internal loop and a  $4 \times 4$  internal loop, the incorrect prediction for the two large internal loops causes the low scores for VfoldMCPX. The crystal structure of 4G6R indicates strong intra- and inter-loop interactions for the two internal loops. Currently, VfoldMCPX cannot explicitly treat

the loop sequence-dependence and cannot consider tertiary structural motifs such as intra loop kissing motifs, which may cause inaccuracy in the predictions.

We then compare VfoldMCPX with NanoFolder. In the comparison, we use the data set in the published work of NanoFolder (Bindewald et al. 2011). The data set contains 10 pseudoknot-free structures with two strands (PDB ID shown in Supplemental Table S1). The results are shown as the Box-Whisker plot of the Matthews correlation coefficients (Supplemental Figure S1), where the results of NanoFolder are from the published work (Bindewald et al. 2011). The Matthews correlation coefficient (MCC) is defined from the true positive (TP), false negative (FN), false positive (FP), and true negative (TN) base pairs:  $MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$ . MCC measures the quality of binary classifications. The result suggests that VfoldMCPX can make predictions on this data set with MCC for all the tested structures higher than 0.80. We note that VfoldMCPX gives higher median MCC (0.94) and a better minimum score (0.80).

We further test VfoldMCPX against RCPred. In the test, we use the structure data set in the published work of RCPred (Legendre et al. 2019). The data set consists of 23 PDB structures (Berman et al. 2000) (PDB ID shown in Supplemental Table S2) each containing two or more unique strands with a chain length longer than 25 nt. The structures predicted by RCPred are obtained by the RCPred home page at [https://evryrna.ibisc.univ-evry.fr/evryrna/RCPred/RCPred\\_home](https://evryrna.ibisc.univ-evry.fr/evryrna/RCPred/RCPred_home). The native 2D structures are extracted from the 3D crystal structures using RNApdbee (Zok et al. 2018; Antczak et al. 2014). The prediction accuracy is evaluated by the PPV, STY, and  $F_1$  score. As shown in Supplemental Figure S2, the mean values are depicted as dots while the error bars are shown with confidence intervals at a 95% confidence level. The result suggests that VfoldMCPX can give reliable predictions with all the mean values larger than 0.88. The two models show similar performances with VfoldMCPX giving slightly higher mean PPV, STY, and  $F_1$  score. We note that RCPred relies on available thermodynamic models for generating the input secondary structures and RNA-RNA interactions, while VfoldMCPX can directly predict folding thermodynamics and thermodynamic properties such as base-pairing probabilities.



**Table 1:** Comparisons between the structure prediction results from VfoldMCPX, MultiRNAFold (Andronescu et al. 2005, 2007) and NUPACK (Dirks et al. 2007; Zadeh et al. 2011) on five structures with three or more unique strands and structure prediction results from VfoldMCPX on 4 H-type pseudoknot-containing structures. All the structures in the data set have available PDB structures (Berman et al. 2000). We measure the prediction accuracy using the Positive Predictive Value (PPV), Sensitivity (STY), and  $F_1$  score. Boldface numbers indicate the model that gives better or the same values.

Structures with three or more strands											Structures containing H-type pseudoknots				
PDB-id	Strands	VfoldMCPX			NUPACK			MultiRNAFold			PDB-id	Strands	VfoldMCPX		
		STY	PPV	$F_1$ score	STY	PPV	$F_1$ score	STY	PPV	$F_1$ score			STY	PPV	$F_1$ score
6UFJ	3	<b>1.00</b>	0.77	0.87	<b>1.00</b>	0.77	0.87	0.94	<b>0.84</b>	<b>0.89</b>	3MJA	2	0.86	0.92	0.89
3BBM	3	<b>1.00</b>	<b>0.75</b>	<b>0.86</b>	0.94	0.74	0.83	<b>1.00</b>	<b>0.75</b>	<b>0.86</b>	2M23	2	0.82	1.00	0.90
364D	3	<b>1.00</b>	<b>0.91</b>	<b>0.95</b>	<b>1.00</b>	<b>0.91</b>	<b>0.95</b>	<b>1.00</b>	<b>0.91</b>	<b>0.95</b>	6YML	2	0.75	0.90	0.82
4KZ2	3	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	0.92	<b>1.00</b>	0.96	0.96	<b>1.00</b>	0.98	1F27	2	0.91	0.83	0.87
4G6R	4	0.56	0.67	0.61	0.50	0.64	0.56	<b>1.00</b>	<b>0.75</b>	<b>0.86</b>					

## Predicting melting thermodynamics for two-strand systems

For two-strand systems, we are able to predict heat capacity melting curves and melting temperatures using VfoldMCPX. We compute the heat capacity melting curve  $C(T)$  from the partition function  $Q(T)$ :  $C(T) = \frac{\partial}{\partial T} \left[ k_B T^2 \frac{\partial}{\partial T} \ln Q(T) \right]$ , where the partition function  $Q(T)$  is calculated from the sum over all the possible 2D structures  $Q_{\text{total}}$ . We include the concentration-dependent free energy term  $\Delta G_{\text{concentration}} = -k_B T \ln(C_T/a)$  in the free energy calculation. Here  $C_T$  denotes the total concentration of RNA strands,  $k_B$  is the Boltzmann constant,  $a$  is equal to 1 for self-complementary strands and 4 for non-self-complementary strands (Xia et al. 1998).

As an example, we investigate the folding thermodynamics for the 3WJ-forming two-strand system (Liu et al. 2011). The thermodynamic measurements are available for the system by optical melting and fluorescence competition experiments. When predicting the melting curves, we use the sequences, the order of strands, and the solution condition (23.5  $\mu\text{M}$  of total RNA concentration) obtained from the experiment (Liu et al. 2011). The predicted melting curve is reported in Fig. 4a, which shows two peaks at  $T = 59.5^\circ\text{C}$  and  $T = 80.5^\circ\text{C}$ , respectively. We note that the heat capacity melting temperature  $T = 59.5^\circ\text{C}$  predicted by VfoldMCPX is close to the experiment-determined melting temperature  $T_m = 57.6^\circ\text{C}$ . The second peak at  $T = 80.5^\circ\text{C}$  is out of the experiment range. As shown in Fig. 4b, structures are inferred from the base pairing probabilities at four different temperatures,  $T = 37^\circ\text{C}$ ,  $T = 55^\circ\text{C}$ ,  $T = 70^\circ\text{C}$ , and  $T = 85^\circ\text{C}$ . The structure predicted at  $T = 37^\circ\text{C}$  agrees with the experimental determined structure, which shows a 3WJ structure with Stem 1, Stem 2, and Stem 3 formed. Our

calculation further indicates that the unfolding of Stem 2 occurs between  $T = 55^{\circ}\text{C}$  and  $T = 70^{\circ}\text{C}$ . The unfolding of Stem 1 occurs between  $T = 70^{\circ}\text{C}$  and  $T = 85^{\circ}\text{C}$ .

## Multistrand RNA structure prediction

VfoldMCPX provides a new approach for predicting multistrand RNA complexes that may contain pseudoknots. In this section, we show the applications of the model by predicting 2D structures for two experimentally studied systems: the core domain of pRNA (Sharma et al. 2016) and the body region of snRNA (Nguyen et al. 2016). To perform structural predictions, we use the sequences and the order of strands from the nanostructure assembly experiments (Sharma et al. 2016) for the pRNA system and from the 3D crystal structure for the snRNA system.

The multistrand pRNA structure motif can be utilized as the core scaffold for constructing RNA dendrimers (Sharma et al. 2016; Shu et al. 2011). The pRNA-3WJ core is a three-way junction (3WJ) system that consists of three unique strands: 3WJ-a, 3WJ-b and 3WJ-c (see Fig. 5a). Experimental observations have suggested that the pRNA-3WJ can be reversed by switching the 5'- and 3'-ends. The reversed system is denoted as pRNA-3WJ\_reverse, which contains three reversed strands: 3WJ-a\_reverse, 3WJ-b\_reverse and 3WJ-c\_reverse (see Fig. 5b). We use VfoldMCPX to compute the base pairing probabilities and determine the most probable structures for the pRNA-3WJ and the pRNA-3WJ\_reverse. The results are reported in Fig. 5a-b. We find that the predicted 2D structures are consistent with the 3WJ structures observed in the experiments. Furthermore, the base pairing probabilities for the most probable structure (Fig. 5a) are around 99%. The high probability is consistent with the experimental finding, which suggests that pRNA-3WJ motif is ultra-stable.

To further illustrate the application of the VfoldMCPX model, we investigate the "Module-2" (Sharma et al. 2016), a functional module that is assembled from three different units with pRNA-3WJ core harboring 3WJ-a\_reverse, 3WJ-b\_reverse, and 3WJ-c\_reverse, respectively. The large number of RNA strands in "Module-2" makes the structure prediction and experimental structure assembly quite challenging. Experimentally, a step-wise assembly is used to ensure the formation of "Module-2" and to reduce the probability of misfolding. Here, we treat the "Module-2" as a RNA complex with nine strands, and predict the most probable structure using VfoldMCPX. The result is reported in Fig. 5c, with nine strands colored by blue and red. The predicted "Module-2" structure is consistent with the structure designed and observed in the experiment. Despite the large number of strands, VfoldMCPX well predicts the designed structure, which suggests that this model is promising for predicting large RNA complexes.

Multistrand RNA complexes with tertiary contacts frequently occur in spliceosome (Will and Lührmann 2011). Here, we investigate the body region of the yeast U4/U6.U5 tri-snRNP, which is an essential part of the spliceosome. The complex consists of three small nuclear RNAs (snRNAs): U4, U6 and U5. Fig. 6a shows the 3D and 2D structures extracted from the cryo-electron microscopy structure (PDB 5GAP) (Nguyen et al. 2016). The 2D structure and the tertiary interactions of the three snRNAs are critical for revealing the activation mechanism of the spliceosome (Will and Lührmann 2011).

Here, we treat the system as an RNA complex with three strands and predict the most probable pseudoknot-containing structure using VfoldMCPX. Three snRNAs are connected using two phantom linkers in the strand order U5-U6-U4. As shown in Fig. 6b, the predicted structure is in overall agreement with the experimentally determined native structure: both the theoretically predicted and the experimentally determined structures contain the H-type pseudoknot with crossing base pairs between U5 and U6 RNAs and two stems with crossing base pairs between U4 and U6 RNAs. We note that in the crystal structure, the body region of the yeast U4/U6.U5 tri-snRNP contains multiple proteins interacting with the three snRNAs. For example, in the cryo-electron microscopy structure, the conserved hairpin loop in U5 RNA points towards the conserved surface of the Prp8 protein due to RNA-protein interactions. Such RNA-protein interactions may affect the inaccuracy in our predictions.

In VfoldMCPX, the input RNA strands are connected as an effective single-strand sequence connected by the phantom linkers. However, the permutation of the input strands directly affects the effective single-strand RNA sequence and hence impacts the the result of structure prediction. Therefore, we develop an algorithm to test all the possible strand orders and find the one that gives the most stable structure: (1) enumerate all the possible permutations of the  $N$  input strands; (2) for each permutation, predict a set of energetically stable structures, including the structures with and without H-type pseudoknots; (3) combine all the predicted structures and rank the structures according to the predicted free energies. The structure (with the corresponding strand order) of the lowest free energy is predicted to be the most stable structure.

We have tested the above algorithm on six RNA structures with three or more strands obtained from the PDB database (Berman et al. 2000). We use RNApdb (Antczak et al. 2014; Zok et al. 2018) to extract the 2D structures, identify the strand IDs and the native strand orders from the 3D crystal structures. The test results and the structures are reported in Table 2. Out of the 6 tested structures, the VfoldMCPX predicted the correct strand order for the most stable structure for 5 cases. Furthermore, VfoldMCPX predicts that the most stable structures do not contain H-type pseudoknots, which also agrees with the experimentally determined structures. However, the large number of strand permutations may lead to exceedingly long computational time for the structure prediction for all the possible strand orders ( $N!$  for  $N$  strands).

**Table 2:** Comparisons of the different strand orders between the predicted most stable structures and experimentally determined structures. The test data set consist of six RNA complex structures with three or more strands. We use RNAPdb (Antczak et al. 2014; Zok et al. 2018) to extract the 2D structures and identify the strand IDs. A most stable structure is predicted using VfoldMCPX for all the possible strand permutations.

VfoldMCPX: rank structures from all the possible strand orders				
PDB-id	Strands	Native strand order	Most stable structure	
			H-type pseudoknots	strand order
6UFJ	3	A-B-E	no	A-B-E
3BBM	3	A-B-C	no	A-B-C
364D	3	A-B-C	no	A-B-C
4KZ2	3	A-B-C	no	A-B-C
4G6R	4	A-B-C-D	no	B-A-D-C

## Web server

We have developed a freely accessible web server to provide a user-friendly interface for using VfoldMCPX to predict multistrand 2D structures and base pairing probabilities. The input of the web server is a set of RNA strands. The user can specify the number of strands  $N$ , the folding temperature, and the base stacking energy parameters either from Turner's parameters (Turner and Mathews 2010) or from the MFOLD parameter set (Zuker 2003). Furthermore, the user has the option to specify the type of the structure included in the prediction: (1) No pseudoknots are allowed to form; (2) H-type pseudoknots are allowed to form; (3) The (virtual) connection of the RNA strands follows a user-defined sequential order (from RNA 1 to RNA  $N$ ); (4) All the different sequential orders of the strands are enumerated.

After submitting the job, the user receives a page displaying the job information. Each submitted job has a unique job ID, containing the job name provided by the user and a 4-letter unique job code randomly generated by the server. The user can bookmark the page to check the job status. The page is automatically updated and displays the results once the computation is finished. Fig. 7 shows a screenshot of the result page for the prediction of phi29 pRNA 3WJ core (PDB 4KZ2) (Zhang et al. 2013). The VfoldMCPX web server outputs two types of files: (1) a set of energetically stable structures, ranked by the free energies and visualized by the VARNA applet (Darty et al. 2009); (2) base pairing probabilities. Note that the results for structures with H-type pseudoknots are listed separately. Because multistrand RNA complexes are prone to the formation alternative folds with similar stabilities, it is important for the user to identify possible alternative structures from the predicted base pairing probabilities. The URL link for the web server is "<http://rna.physics.missouri.edu/vfoldMCPX2>".

## Computational efficiency and restrictions

The computational time for the prediction without pseudoknots ranges from seconds to minutes for a multistrand RNA system with the total nucleotide length less than 300 nt.

Predictions for structures with H-type pseudoknots are computational time-consuming, and could take several hours for RNA complexes with the total nucleotide length around 150 nt.

To avoid extremely long computational time, we restrict the total number of strands to 9 and the total length of the strands to  $L_{tot} \leq 300$  nucleotides for predictions without pseudoknots and  $L_{tot} \leq 150$  nucleotides for predictions with H-type pseudoknots.

## Conclusions

We have developed VfoldMCPX, a new, physics-based model for predicting 2D structures and folding thermodynamics of multistrand RNA complexes, with and without pseudoknots. Theory-experiment comparisons indicate that VfoldMCPX can provide accurate predictions for pseudoknot-forming RNA complexes with multiple strands. Comparisons between different models show that VfoldMCPX gives improved structure predictions for multistrand RNA complexes. VfoldMCPX and the web server may aid in the design of RNA nanostructures and provide novel physical insights into RNA-RNA interactions. Currently, VfoldMCPX can treat RNA complex systems with H-type pseudoknots, future development of the model should consider more complicated tertiary structural motifs such as intramolecular kissing motifs.

## Acknowledgments

This work was supported by the National Institutes of Health under Grants R35-GM134919 to S.-J. C. , U01CA207946 and R01EB019036 to P.G.

## References

- Afonin KA, Kasprzak WK, Bindewald E, Kireeva M, Viard M, Kashlev M, Shapiro BA. 2014a. In silico design and enzymatic synthesis of functional RNA nanoparticles. *Accounts of chemical research* **47**: 1731–1741.
- Afonin KA, Viard M, Koyfman AY, Martins AN, Kasprzak WK, Panigaj M, Desai R, Santhanam A, Grabow WW, Jaeger L, et al. 2014b. Multifunctional RNA nanoparticles. *Nano letters* **14**: 5662–5671.
- Andronescu M, Condon A, Hoos HH, Mathews DH, Murphy KP. 2007. Efficient parameter estimation for RNA secondary structure prediction. *Bioinformatics* **23**: i19–i28.
- Andronescu M, Zhang ZC, Condon A. 2005. Secondary structure prediction of interacting RNA molecules. *Journal of molecular biology* **345**: 987–1001.
- Antczak M, Zok T, Popena M, Lukasiak P, Adamiak RW, Blazewicz J, Szachniuk M. 2014. RNAPdb— a webserver to derive secondary structures from pdb files of knotted and unknotted RNAs. *Nucleic acids research* **42**: W368–W372.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. 2000. The protein data bank. *Nucleic acids research* **28**: 235–242.

Bernhart SH, Tafer H, Mückstein U, Flamm C, Stadler PF, Hofacker IL. 2006. Partition function and base pairing probabilities of RNA heterodimers. *Algorithms for Molecular Biology* **1**: 1–10.

Bindewald E, Afonin KA, Viard M, Zakrevsky P, Kim T, Shapiro BA. 2016. Multistrand structure prediction of nucleic acid assemblies and design of RNA switches. *Nano letters* **16**: 1726–1735.

Bindewald E, Afonin K, Jaeger L, Shapiro BA. 2011. Multistrand RNA secondary structure prediction and nanostructure design including pseudoknots. *ACS nano* **5**: 9542–9551.

Bindewald E, Grunewald C, Boyle B, O'Connor M, Shapiro BA. 2008. Computational strategies for the automated design of RNA nanoscale structures from building blocks using NanoTiler. *Journal of Molecular Graphics and Modelling* **27**: 299–308.

Binzel DW, Khisamutdinov E, Vieweger M, Ortega J, Li J, Guo P. 2016. Mechanism of three-component collision to produce ultrastable pRNA three-way junction of Phi29 DNA-packaging motor by kinetic assessment. *RNA* **22**: 1710–1718.

Binzel DW, Li X, Burns N, Khan E, Lee W-J, Chen L-C, Ellipilli S, Miles W, Ho YS, Guo P. 2021. Thermostability, tunability, and tenacity of RNA as rubbery anionic polymeric materials in nanotechnology and nanomedicine—specific cancer targeting with undetectable toxicity. *Chemical Reviews* 1322–1335.

Busch A, Richter AS, Backofen R. 2008. IntaRNA: Efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. *Bioinformatics* **24**: 2849–2856.

Cao S, Chen S-J. 2006a. Free energy landscapes of RNA/RNA complexes: With applications to snRNA complexes in spliceosomes. *Journal of molecular biology* **357**: 292–312.

Cao S, Chen S-J. 2012. Predicting kissing interactions in microRNA–target complex and assessment of microRNA activity. *Nucleic acids research* **40**: 4681–4690.

Cao S, Chen S-J. 2005. Predicting RNA folding thermodynamics with a reduced chain representation model. *Rna* **11**: 1884–1897.

Cao S, Chen S-J. 2006b. Predicting RNA pseudoknot folding thermodynamics. *Nucleic acids research* **34**: 2634–2652.

Cao S, Chen S-J. 2009. Predicting structures and stabilities for h-type pseudoknots with interhelix loops. *RNA* **15**: 696–706.

Cao S, Chen S-J. 2011. Structure and stability of RNA/RNA kissing complex: With application to HIV dimerization initiation signal. *RNA* **17**: 2130–2143.

- Cao S, Giedroc DP, Chen S-J. 2010. Predicting loop–helix tertiary structural contacts in RNA pseudoknots. *Rna* **16**: 538–552.
- Cao S, Xu X, Chen S-J. 2014. Predicting structure and stability for RNA complexes with intermolecular loop–loop base-pairing. *RNA* **20**: 835–845.
- Chen S-J. 2008. RNA folding: Conformational statistics, folding kinetics, and ion electrostatics. *Annu Rev Biophys* **37**: 197–214.
- Chworos A, Severcan I, Koyfman AY, Weinkam P, Oroudjev E, Hansma HG, Jaeger L. 2004. Building programmable jigsaw puzzles with RNA. *Science* **306**: 2068–2072.
- Darty K, Denise A, Ponty Y. 2009. VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics* **25**: 1974.
- Dibrov SM, McLean J, Parsons J, Hermann T. 2011. Self-assembling RNA square. *Proceedings of the National Academy of Sciences* **108**: 6405–6408.
- Dimitrov RA, Zuker M. 2004. Prediction of hybridization and melting for double-stranded nucleic acids. *Biophysical Journal* **87**: 215–226.
- Dirks RM, Bois JS, Schaeffer JM, Winfree E, Pierce NA. 2007. Thermodynamic analysis of interacting nucleic acid strands. *SIAM review* **49**: 65–88.
- Dutta T, Srivastava S. 2018. Small RNA-mediated regulation in bacteria: A growing palette of diverse mechanisms. *Gene* **656**: 60–72.
- Griffith LG, Swartz MA. 2006. Capturing complex 3D tissue physiology in vitro. *Nature reviews Molecular cell biology* **7**: 211–224.
- Guil S, Esteller M. 2015. RNA–RNA interactions in gene regulation: The coding and noncoding players. *Trends in biochemical sciences* **40**: 248–256.
- Guo P. 2010. The emerging field of RNA nanotechnology. *Nature nanotechnology* **5**: 833–842.
- Hofacker IL, Fontana W, Stadler PF, Bonhoeffer LS, Tacker M, Schuster P. 1994. Fast folding and comparison of RNA secondary structures. *Monatshefte für Chemie/Chemical Monthly* **125**: 167–188.
- Jasinski D, Haque F, Binzel DW, Guo P. 2017. Advancement of the emerging field of RNA nanotechnology. *ACS nano* **11**: 1142–1164.
- Kato Y, Mori T, Sato K, Maegawa S, Hosokawa H, Akutsu T. 2017. An accessibility-incorporated method for accurate prediction of RNA–RNA interactions from sequence data. *Bioinformatics* **33**: 202–209.
- Legendre A, Angel E, Tahi F. 2019. RCPred: RNA complex prediction as a constrained maximum weight clique problem. *BMC bioinformatics* **20**: 53–62.

- Li H, Lee T, Dziubla T, Pi F, Guo S, Xu J, Li C, Haque F, Liang X-J, Guo P. 2015. RNA as a stable polymer to build controllable and defined nanostructures for material and biomedical applications. *Nano today* **10**: 631–655.
- Liu B, Diamond JM, Mathews DH, Turner DH. 2011. Fluorescence competition and optical melting measurements of RNA three-way multibranch loops provide a revised model for thermodynamic parameters. *Biochemistry* **50**: 640–653.
- Lorenz R, Bernhart SH, Zu Siederdisen CH, Tafer H, Flamm C, Stadler PF, Hofacker IL. 2011. ViennaRNA package 2.0. *Algorithms for molecular biology* **6**: 1–14.
- 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. *Proceedings of the National Academy of Sciences* **101**: 7287–7292.
- Montaseri S, Zare-Mirakabad F, Moghadam-Charkari N. 2014. RNA-RNA interaction prediction using genetic algorithm. *Algorithms for Molecular Biology* **9**: 1–7.
- Nguyen THD, Galej WP, Bai X, Oubridge C, Newman AJ, Scheres SH, Nagai K. 2016. Cryo-EM structure of the yeast U4/U6. U5 tri-snRNP at 3.7 Å resolution. *Nature* **530**: 298–302.
- Paillart J-C, Shehu-Xhilaga M, Marquet R, Mak J. 2004. Dimerization of retroviral RNA genomes: An inseparable pair. *Nature Reviews Microbiology* **2**: 461–472.
- Rehmsmeier M, Steffen P, Höchsmann M, Giegerich R. 2004. Fast and effective prediction of microRNA/target duplexes. *Rna* **10**: 1507–1517.
- Reuter JS, Mathews DH. 2010. RNAstructure: Software for RNA secondary structure prediction and analysis. *BMC bioinformatics* **11**: 1–9.
- Saeed AF, Chan C, Guan H, Gong B, Guo P, Cheng X, Ouyang S. 2021. Structural insights into gp16 ATPase in the bacteriophage  $\phi$ 29 DNA packaging motor. *Biochemistry* **60**: 886–897.
- Serra MJ, Turner DH. 1995. [11] predicting thermodynamic properties of RNA. *Methods in enzymology* **259**: 242–261.
- Sharma A, Haque F, Pi F, Shlyakhtenko LS, Evers BM, Guo P. 2016. Controllable self-assembly of RNA dendrimers. *Nanomedicine: Nanotechnology, Biology and Medicine* **12**: 835–844.
- Shu D, Shu Y, Haque F, Abdelmawla S, Guo P. 2011. Thermodynamically stable RNA three-way junction for constructing multifunctional nanoparticles for delivery of therapeutics. *Nature nanotechnology* **6**: 658–667.
- Shukla GC, Haque F, Tor Y, Wilhelmsson LM, Toulme J-J, Isambert H, Guo P, Rossi JJ, Tenenbaum SA, Shapiro BA. 2011. A boost for the emerging field of RNA nanotechnology: Report on the first international conference on RNA nanotechnology.



- Sun L-Z, Zhang D, Chen S-J. 2017. Theory and modeling of RNA structure and interactions with metal ions and small molecules. *Annual review of biophysics* **46**: 227–246.
- Turner DH, Mathews DH. 2010. NNDB: The nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. *Nucleic acids research* **38**: D280–D282.
- Will CL, Lührmann R. 2011. Spliceosome structure and function. *Cold Spring Harbor perspectives in biology* **3**: a003707.
- Xia T, SantaLucia Jr J, Burkard ME, Kierzek R, Schroeder SJ, Jiao X, Cox C, Turner DH. 1998. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with watson- crick base pairs. *Biochemistry* **37**: 14719–14735.
- Xu X, Chen S-J. 2016a. A method to predict the structure and stability of RNA/RNA complexes. In *RNA structure determination*, pp. 63–72, Springer.
- Xu X, Chen S-J. 2016b. VfoldCPX server: Predicting RNA-RNA complex structure and stability. *PLoS One* **11**: e0163454.
- Xu X, Zhao P, Chen S-J. 2014. Vfold: A web server for RNA structure and folding thermodynamics prediction. *PloS one* **9**: e107504.
- Yingling YG, Shapiro BA. 2007. Computational design of an RNA hexagonal nanoring and an RNA nanotube. *Nano letters* **7**: 2328–2334.
- Zadeh JN, Steenberg CD, Bois JS, Wolfe BR, Pierce MB, Khan AR, Dirks RM, Pierce NA. 2011. NUPACK: Analysis and design of nucleic acid systems. *Journal of computational chemistry* **32**: 170–173.
- Zhang H, Endrizzi JA, Shu Y, Haque F, Sauter C, Shlyakhtenko LS, Lyubchenko Y, Guo P, Chi Y-I. 2013. Crystal structure of 3WJ core revealing divalent ion-promoted thermostability and assembly of the Phi29 hexameric motor pRNA. *Rna* **19**: 1226–1237.
- Zhao C, Xu X, Chen S-J. 2017. Predicting RNA structure with vfold. In *Functional genomics*, pp. 3–15, Springer.
- Zok T, Antczak M, Zurkowski M, Popenda M, Blazewicz J, Adamiak RW, Szachniuk M. 2018. RNApdbee 2.0: Multifunctional tool for RNA structure annotation. *Nucleic acids research* **46**: W30–W35.
- Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic acids research* **31**: 3406–3415.

## Figure Captions

**Figure 1:** Secondary structure (left) and arc (right) diagrams for (a) a four-strand structure without crossing base pairs, (b) a three-strand structure with crossing base pairs (a H-type pseudoknot motif). VfoldMCPX uses the phantom linkers (shown as three red dots) to connect the RNA strands. In the arc diagrams, RNA sequences are shown as black horizontal lines while base pairs are shown as blue arcs. The crossing arcs indicate the crossing base pairs. Specifically labeled motifs are in black boldface.

**Figure 2:** (a) A schematic diagram for the H-type pseudoknot motif, which is divided into three single-stranded loops ( $L_1$ ,  $L_2$  and  $L_3$ , green and orange) and two helix stems ( $S_1$  and  $S_2$ , blue) that may contain an internal/bulge loop. (b-i) Schematic diagrams for the H-type pseudoknot motif with one or more phantom linkers in different positions. Structures that contain the linkers are labeled. The phantom linkers are depicted as red dots.

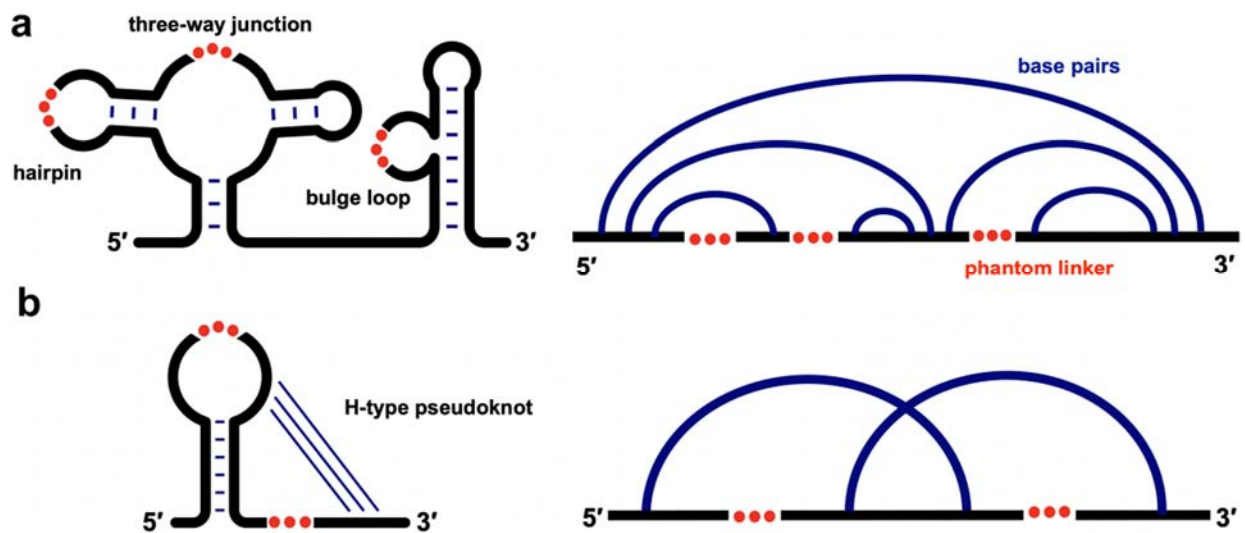
**Figure 3:** Comparison of the Positive Predictive Value (PPV), Sensitivity (STY), and  $F_1$  score between the predicted structures from VfoldMCPX, RNAcofold (Bernhart et al. 2006; Lorenz et al. 2011; Hofacker et al. 1994), bifold (Reuter and Mathews 2010), MultiRNAFold (Andronescu et al. 2005, 2007) and NUPACK (Dirks et al. 2007; Zadeh et al. 2011). The data set consist of 29 RNA-only, pseudoknot-free structures with two unique strands, obtained from the PDB database (Berman et al. 2000). Mean values are labeled and depicted as dots. Error bars are depicted using confidence intervals at 95% confidence level. Models are ranked from top to bottom according to the mean values. All the models show nearly the same performance with VfoldMCPX slightly outperforming other models in terms of mean PPV, STY, and  $F_1$  score.

**Figure 4:** Predicted melting curve (a) and secondary structures (b) for the 3WJ-forming two-strand system (Liu et al. 2011). (a) The heat capacity is computed from the total partition function of all the possible 2D structures and the total concentration of two strands ( $23.5\mu M$ ). The first peak at  $T = 59.5^\circ C$  matches the experiment-determined melting temperature  $T = 57.6^\circ C$  (Liu et al. 2011). (b) The base pairing probability-inferred structures at  $T = 37^\circ C$ ,  $T = 55^\circ C$ ,  $T = 70^\circ C$ , and  $T = 85^\circ C$ , respectively. Two strands are colored by blue and red.

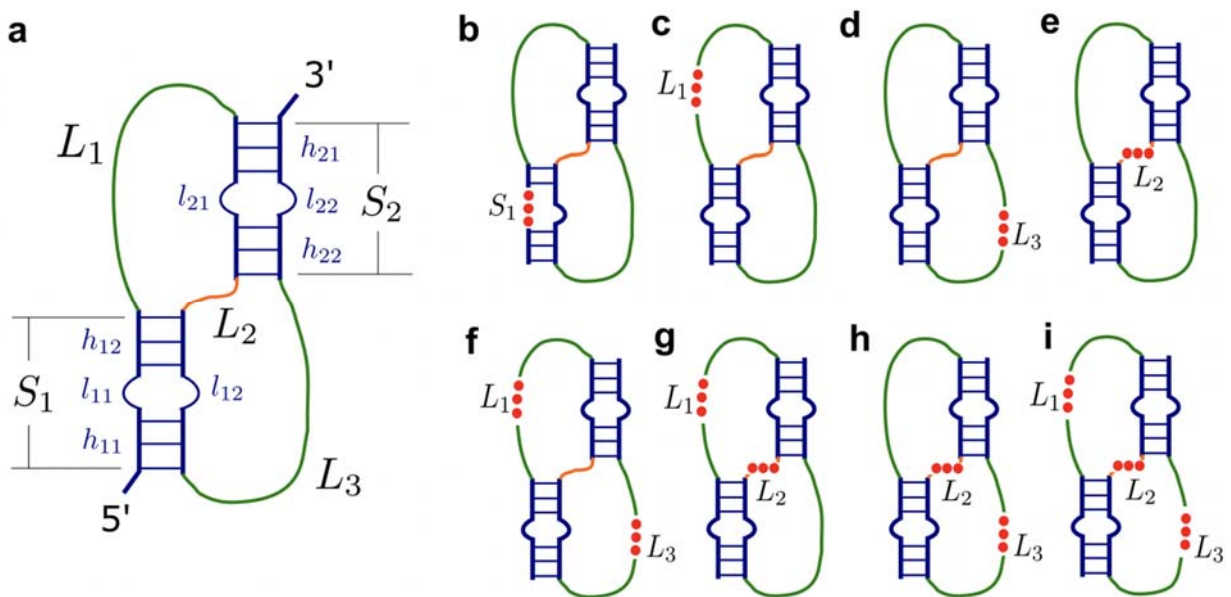
**Figure 5:** Predicted 2D structures for (a) pRNA-3WJ, (b) pRNA-3WJ\_reverse and (c) "Module-2" agree with the experimentally determined structures (Sharma et al. 2016). (a) The predicted structure for pRNA-3WJ with three RNA strands: 3WJ-a (blue), 3WJ-b (red) and 3WJ-c (green); (b) The predicted structure for pRNA-3WJ\_reverse with three RNA strands: 3WJ-a\_reverse (blue), 3WJ-b\_reverse (red) and 3WJ-c\_reverse (green); (c) The predicted structure for "Module-2" with nine RNA strands shown in blue and red.

**Figure 6:** Experimentally determined (a) and theoretically predicted (b) structures of the body region of the yeast U4/U6.U5 tri-snRNP, which contains three RNA strands: U4 snRNA (green), U5 snRNA (blue) and U6 snRNA (red). Structures in (a) are extracted from the cryo-electron microscopy structure (PDB 5GAP) (Nguyen et al. 2016). Experimentally determined 2D structure in (a) is similar to the predicted 2D structure (b).

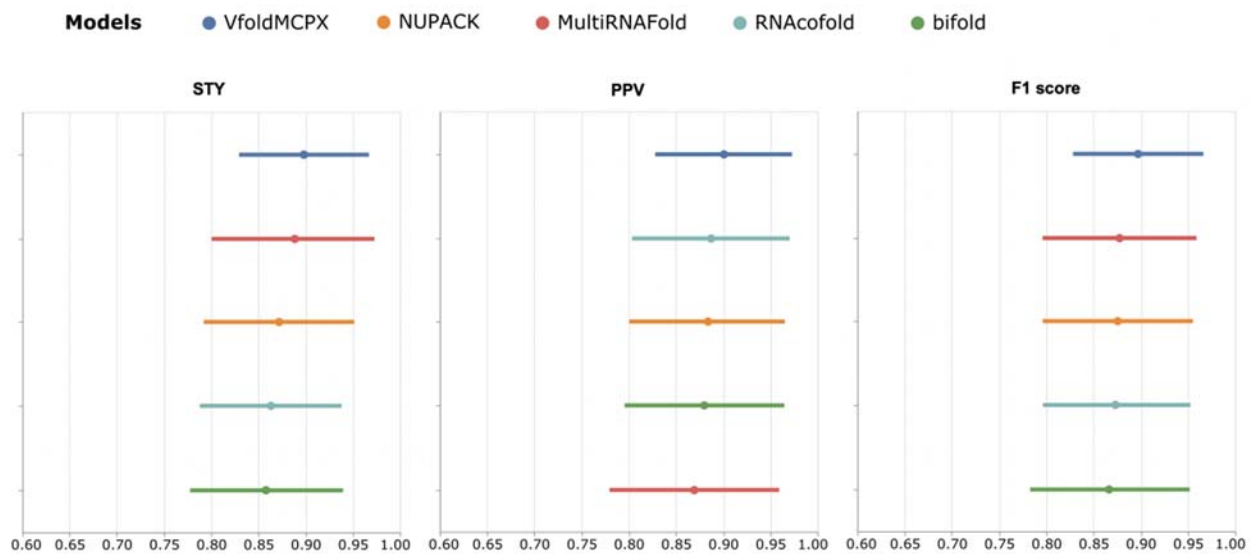
**Figure 7:** A screenshot of the 2D structure prediction result for phi29 pRNA 3WJ core (PDB 4KZ2) (Zhang et al. 2013), generated by VfoldMCPX web server. The web server displays the job information, the predicted structures, and the links to the downloadable files for the predicted structures and the corresponding base pairing probabilities. The links and the corresponding files are indicated with arrows and boxes. The results for structures with and without pseudoknots are listed separately.



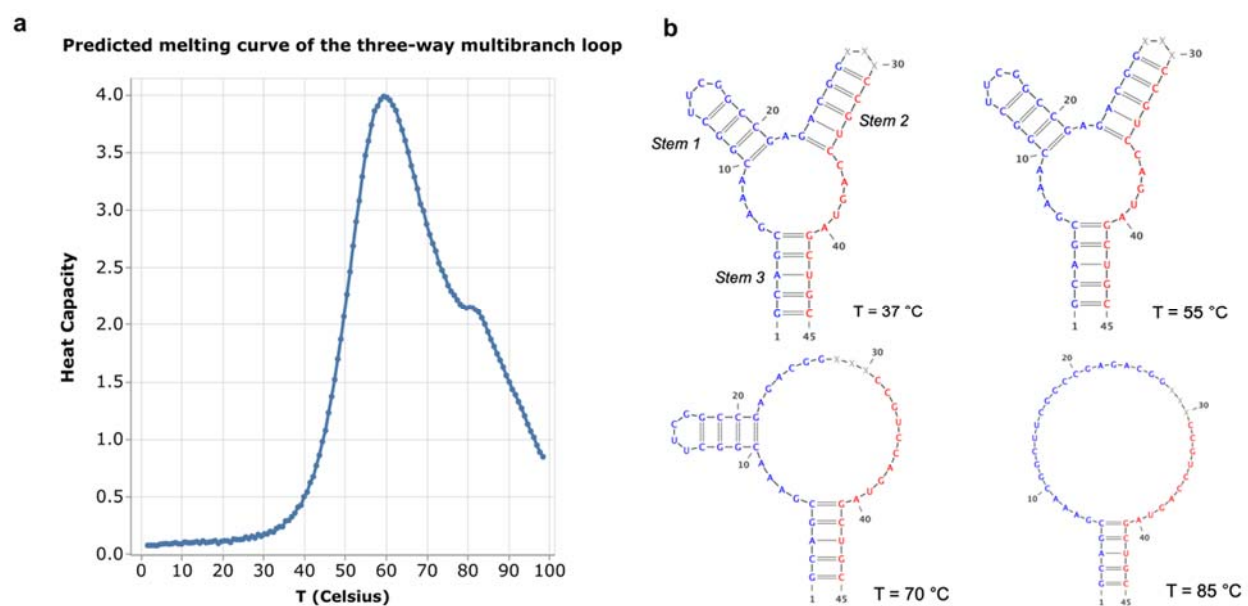
**Figure 1:** Secondary structure (left) and arc (right) diagrams for (a) a four-strand structure without crossing base pairs, (b) a three-strand structure with crossing base pairs (a H-type pseudoknot motif). VfoldMCPX uses the phantom linkers (shown as three red dots) to connect the RNA strands. In the arc diagrams, RNA sequences are shown as black horizontal lines while base pairs are shown as blue arcs. The crossing arcs indicate the crossing base pairs. Specifically labeled motifs are in black boldface.



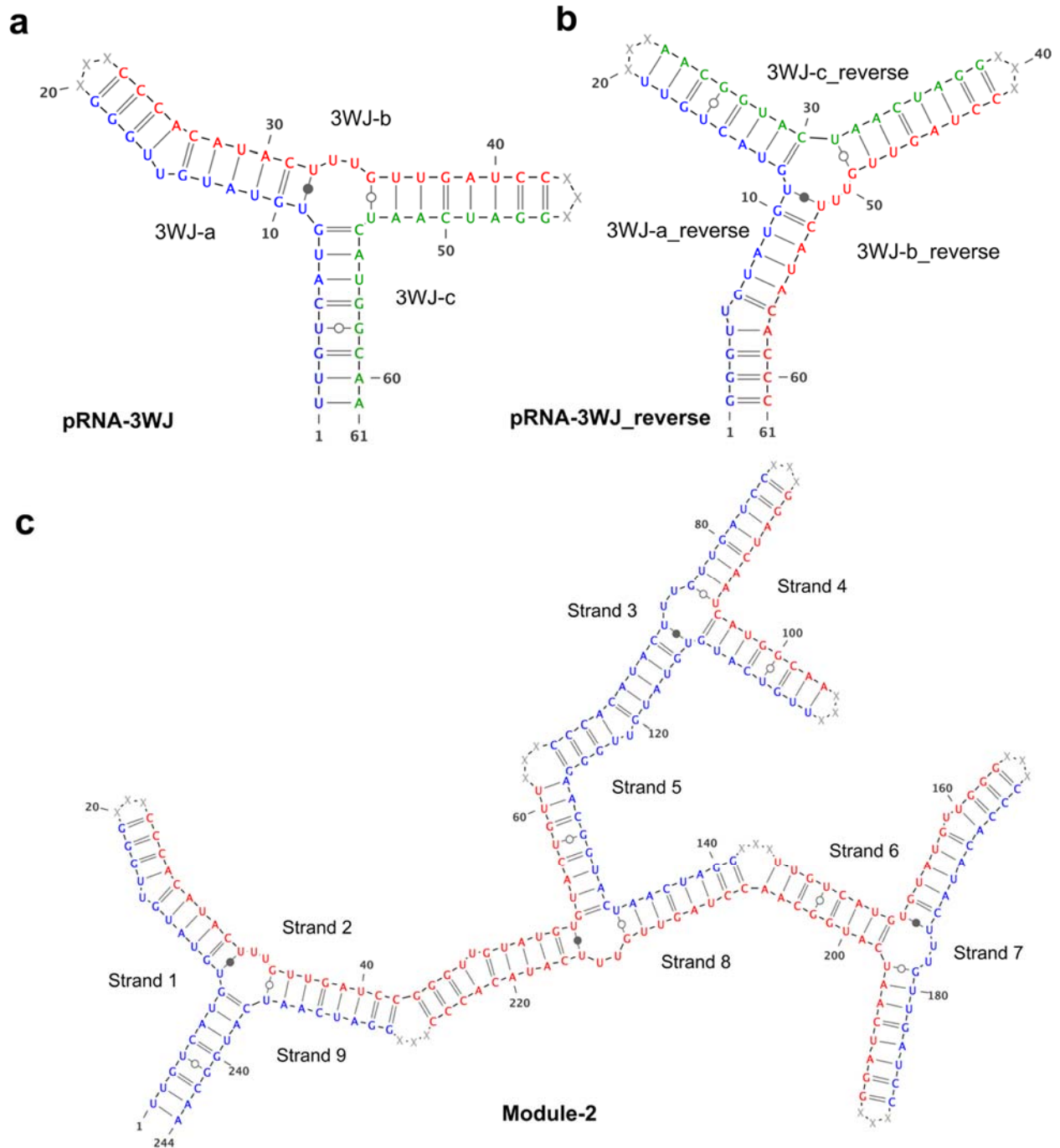
**Figure 2:** (a) A schematic diagram for the H-type pseudoknot motif, which is divided into three single-stranded loops ( $L_1$ ,  $L_2$  and  $L_3$ , green and orange) and two helix stems ( $S_1$  and  $S_2$ , blue) that may contain an internal/bulge loop. (b-i) Schematic diagrams for the H-type pseudoknot motif with one or more phantom linkers in different positions. Structures that contain the linkers are labeled. The phantom linkers are depicted as red dots.



**Figure 3:** Comparison of the Positive Predictive Value (PPV), Sensitivity (STY), and  $F_1$  score between the predicted structures from VfoldMCPX, RNAcofold (Bernhart et al. 2006; Lorenz et al. 2011; Hofacker et al. 1994), bifold (Reuter and Mathews 2010), MultiRNAFold (Andronescu et al. 2005, 2007) and NUPACK (Dirks et al. 2007; Zadeh et al. 2011). The data set consist of 29 RNA-only, pseudoknot-free structures with two unique strands, obtained from the PDB database (Berman et al. 2000). Mean values are labeled and depicted as dots. Error bars are depicted using confidence intervals at 95% confidence level. Models are ranked from top to bottom according to the mean values. All the models show nearly the same performance with VfoldMCPX slightly outperforming other models in terms of mean PPV, STY, and  $F_1$  score.

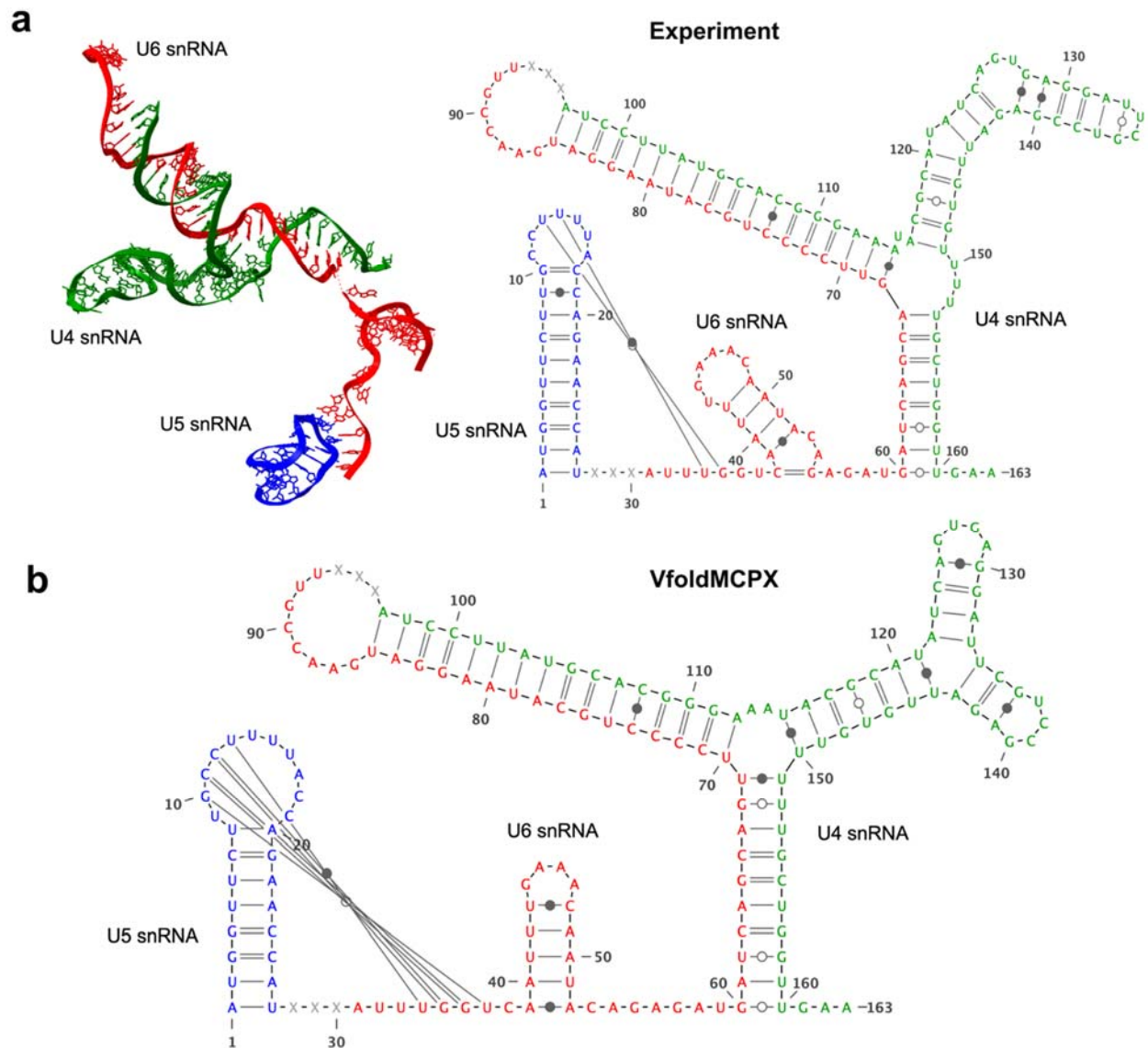


**Figure 4:** Predicted melting curve (a) and secondary structures (b) for the 3WJ-forming two-strand system (Liu et al. 2011). (a) The heat capacity is computed from the total partition function of all the possible 2D structures and the total concentration of two strands ( $23.5\mu M$ ). The first peak at  $T = 59.5^\circ C$  matches the experiment-determined melting temperature  $T = 57.6^\circ C$  (Liu et al. 2011). (b) The base pairing probability-inferred structures at  $T = 37^\circ C$ ,  $T = 55^\circ C$ ,  $T = 70^\circ C$ , and  $T = 85^\circ C$ , respectively. Two strands are colored by blue and red.



**Figure 5:** Predicted 2D structures for (a) pRNA-3WJ, (b) pRNA-3WJ\_reverse and (c) "Module-2" agree with the experimentally determined structures (Sharma et al. 2016). (a) The predicted structure for pRNA-3WJ with three RNA strands: 3WJ-a (blue), 3WJ-b (red) and 3WJ-c (green); (b) The predicted structure for pRNA-3WJ\_reverse with three RNA strands: 3WJ-a\_reverse (blue), 3WJ-b\_reverse (red) and 3WJ-c\_reverse (green); (c) The predicted structure for "Module-2" with nine RNA strands shown in blue and red.





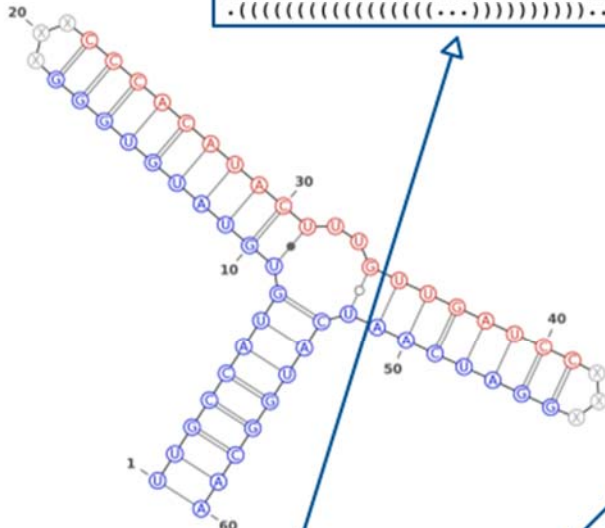
**Figure 6:** Experimentally determined (a) and theoretically predicted (b) structures of the body region of the yeast U4/U6.U5 tri-snRNP, which contains three RNA strands: U4 snRNA (green), U5 snRNA (blue) and U6 snRNA (red). Structures in (a) are extracted from the cryo-electron microscopy structure (PDB 5GAP) (Nguyen et al. 2016). Experimentally determined 2D structure in (a) is similar to the predicted 2D structure (b).

**Your input:**

- Number of RNA: 3
- Temperature: 37 °C
- Energy parameters: from the Turner parameters (04 version)
- Full RNA sequence: UUGCCAUGUGUAUGUGGGXXXCCACAUACUUUGUUGAUCCXXXGGAUCAAUCAUGGCAA
- Job name: 4KZ2
- Prediction without H-type pseudoknotted structures
- Prediction for the specific sequence order

**VfoldMCPX result:**

```
UUGCCAUGUGUAUGUGGGXXXCCACAUACUUUGUUGAUCCXXXGGAUCAAUCAUGGCAA free-energy(kcal/mol)
((((((((((((((((((((.....)))))))))))).((((((((((((((((((((.....)))))))))))). -43.99
UUGCCAUGUGUAUGUGGGXXXCCACAUACUUUGUUGAUCCXXXGGAUCAAUCAUGGCAA free-energy(kcal/mol)
.((((((((((((((((((((.....)))))))))))).((((((((((((((((((((.....)))))))))))). -43.09
```



TTC= 37 G= -44.2175		
1	U - A	60 0.80506
2	U - A	59 0.99360
3	G - C	58 0.99997
4	C - G	57 0.99997
5	C - G	56 0.99997
6	A - U	55 0.99982
7	U - A	54 0.99607
8	G - U	32 0.00311
8	G - C	53 0.92074
9	U - U	31 0.95782
9	U - U	52 0.00940
10	G - C	30 0.99725
11	U - A	29 0.99727
12	A - U	28 0.99723
13	U - A	27 0.99727
14	G - C	26 0.99728
15	U - A	25 0.99728
16	G - C	24 0.99725
17	G - C	23 0.99723
18	G - C	22 0.99238
33	U - C	53 0.07290
33	U - A	54 0.00105
34	G - U	52 0.97574
34	G - C	53 0.00105
35	U - A	51 0.99859
36	U - A	50 0.99965
37	G - C	49 0.99996
38	A - U	48 0.99996
39	U - A	47 0.99994
40	C - G	46 0.99985
41	C - G	45 0.99499

UUGCCAUGUGUAUGUGGGXXXCCACAUACUUUGUUGAUCCXXXGGAUCAAUCAUGGCAA  
((((((((((((((((((((.....)))))))))))).((((((((((((((((((((.....)))))))))))).

**(1) Top 2 prediction among secondary structure ensemble.**

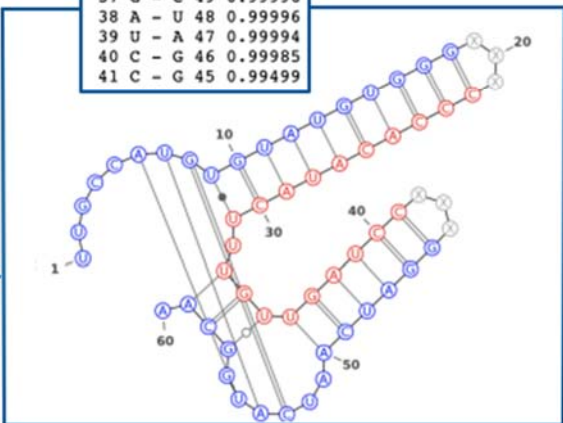
- Top 1
- Top 2

[List of predicted structures](#)  
[Predicted base pairing probabilities](#)

**(2) Top 5 prediction among H-type pseudoknotted ensemble.**

- Top 1
- Top 2
- Top 3
- Top 4
- Top 5

[List of predicted structures](#)  
[Predicted base pairing probabilities](#)



**Figure 7:** A screenshot of the 2D structure prediction result for phi29 pRNA 3WJ core (PDB 4KZ2) (Zhang et al. 2013), generated by VfoldMCPX web server. The web server displays the job information, the predicted structures, and the links to the downloadable files for the predicted structures and the corresponding base pairing probabilities. The links and the corresponding files are indicated with arrows and boxes. The results for structures with and without pseudoknots are listed separately.