

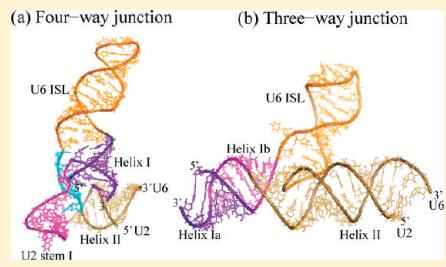
# Physics-Based De Novo Prediction of RNA 3D Structures

Song Cao and Shi-Jie Chen\*

Department of Physics and Department of Biochemistry, University of Missouri, Columbia, Missouri 65211, United States

S Supporting Information

**ABSTRACT:** Current experiments on structural determination cannot keep up the pace with the steadily emerging RNA sequences and new functions. This underscores the request for an accurate model for RNA three-dimensional (3D) structural prediction. Although considerable progress has been made in mechanistic studies, accurate prediction for RNA tertiary folding from sequence remains an unsolved problem. The first and most important requirement for the prediction of RNA structure from physical principles is an accurate free energy model. A recently developed three-vector virtual bond-based RNA folding model (“Vfold”) has allowed us to compute the chain entropy and predict folding free energies and structures for RNA secondary structures and simple pseudoknots. Here we develop a free energy-based method to predict larger more complex RNA tertiary folds. The approach is based on a multiscaling strategy: from the nucleotide sequence, we predict the two-dimensional (2D) structures (defined by the base pairs and tertiary contacts); based on the 2D structure, we construct a 3D scaffold; with the 3D scaffold as the initial state, we combine AMBER energy minimization and PDB-based fragment search to predict the all-atom structure. A key advantage of the approach is the statistical mechanical calculation for the conformational entropy of RNA structures, including those with cross-linked loops. Benchmark tests show that the model leads to significant improvements in RNA 3D structure prediction.



## INTRODUCTION

RNA three-dimensional (3D) structure is critical for RNA cellular functions. For example, the 3D structure of a microRNA-target complex is crucial for the binding affinity and the efficacy of the microRNA in gene regulation by silencing the mRNA in the 3' untranslated region (UTR).<sup>1,2</sup> The widespread biological significance of RNA 3D structures draws a strong demand of structure determination from the sequence. However, the laborious, time-consuming structural measurements alone cannot catch up the pace with the increasing number of biologically significant RNAs such as regulatory RNAs. Therefore, RNA structural genomics cannot just rely on experimental determination of the structures. We also need a reliable theoretical/computational method for structure determination.

Recent developments in de novo prediction of RNA 3D structures have led to promising results.<sup>3–17</sup> Several of these methods are based on knowledge-based scoring functions in combination with atomistic computations or input from auxiliary experimental results about the structure. For example, the FARNA model can predict the 3D structures for hairpins, duplexes, and pseudoknots for short sequences of length  $\leq 30$  nts.<sup>9</sup> Recently, the model<sup>13</sup> was extended to predict the high-resolution structures for the different types of base pairs as defined by Leontis and Westhof<sup>6</sup> and for large RNAs with the experimentally determined structural constraints as input information.<sup>7</sup> In another model (MC-Fold/MC-SYM pipeline), the 3D structures for a variety of RNA 3D folds from hairpins to structures with junctions and pseudoknots<sup>10</sup> can be predicted. These methods are highly useful for structures with known

homologous folds or large structures with available auxiliary structural data. In addition to knowledge-based free energy approach, computer simulations<sup>3,18–23</sup> and phylogenetic analysis<sup>24</sup> have also proven to be useful for structure prediction. For example, molecular dynamics simulation can predict a 75-nt tRNA structure with 4.0 Å rmsd.<sup>4</sup> In general, purely bioinformatics-based methods can be quite efficient and effective in predicting the native structure from a near-native structural ensemble.

Here we develop a physics-based method to predict RNA 3D tertiary folds from the sequence without using the experimental constraints as input information. At the center of theory is the calculation of the free energy landscape from RNA sequence. The free energy landscape gives not only the native state as the global minimum but also the metastable states as the local minima. Such information is essential for understanding RNA functions. Many important RNA structures have been solved,<sup>25,26</sup> but the ability to predict RNA functions<sup>27</sup> from these native structures remains limited. This is partly because RNA functions are often determined not only by the native state but also by the conformational switches between the different states.<sup>28,29</sup> For example, gene expression in RNA viruses is often known or proposed to be linked to structural changes between alternative or competing RNA conformations.<sup>30</sup> In another example, a U2 spliceosomal small nuclear RNA (snRNA)

**Received:** December 20, 2010

**Revised:** February 28, 2011

**Published:** March 17, 2011

molecule can undergo seven different structural rearrangements, several of which are catalytically important.<sup>31</sup> Therefore, one of the key issues in modeling RNA functions is the prediction of the free energy landscape, from which we can predict the stable structures, folding stabilities, and structural changes.

Following the different levels of structural complexity, we develop a multiscaling approach to predict RNA free energy landscapes and 3D structures from the sequence. We start with 2D structures, which are defined by the assignments of bases pairs formed in the structures. From the free energy landscape for the ensemble of (2D) structures, we identify the low free energy (2D) structures. For each low free energy state, we construct a 3D coarse-grained structure as a scaffold based on the fragments selected from the structural database. We then add all atoms to the coarse-grained scaffold. Finally, we run AMBER energy minimization to compute the final atomistic 3D structure.

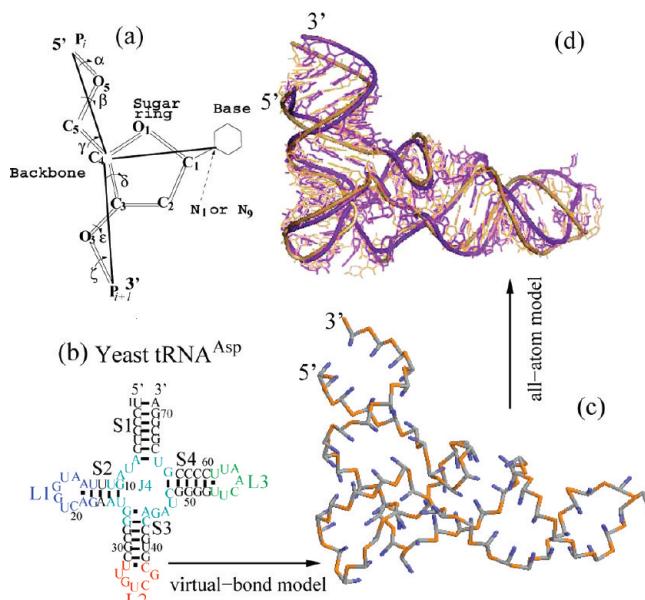
In contrast to the previous de novo methods, our method is built upon the free energy landscape, especially for conformations that contain cross-linked contacts (base pairs), also called tertiary contacts. We call structures with/without cross-linked contacts (base pairs) as tertiary/secondary structures.<sup>32</sup> Previous free energy-based RNA folding models mainly focus on the 2D secondary structures, and these models have led to many successful predictions.<sup>33–38</sup> For example, free energy minimization of the 2D structures can predict the helical coaxial stacking in multibranched structures.<sup>39</sup> In contrast, due to the difficulty to account for the effect of the nonlocal correlation between the different structural subunits (helices, loops), the ability to predict RNA tertiary structural folding, at either 2D or 3D structural level, is quite limited. In the present study, we simplify the conformational analysis by using a reduced RNA conformational model (virtual bond model) and based on the reduced conformational model, we develop a theory to evaluate the conformation entropy for RNA tertiary folds.

There are several advantages for the structure prediction method developed here. First, the method is based on statistical mechanical calculations for the conformational entropy for RNA tertiary folds. To date, no such computational method has been applied to compute the conformational entropy for RNA tertiary folds. Second, the energy landscape approach can potentially map out all the low-lying 3D structures. Third, we use the coarse-grained structure as the initial state for the all-atom energy minimization, which can significantly enhance the computational speed compared to other simulational methods.<sup>40</sup> Extensive benchmark tests against other de novo methods suggest that the new model developed here leads to much improved accuracy in 3D structural prediction. Moreover, the model enables predictions of large conformational changes which can be biologically significant.<sup>31,41,42</sup>

To illustrate the applications of the model, we predict the 3D all-atom structures for a set of complex tertiary folds and for the conformational switches for the dimerization initiation signal (DIS) of HIV-1 strain Lai kissing complexes and for the catalytic core in yeast U2–U6 small nuclear RNA (protein-free) complexes.

## THEORY AND MODEL

**A Virtual Bond-Based RNA Folding Model (Vfold Model).** Based on the rotameric properties of RNA backbone,<sup>43–45</sup> we describe each nucleotide by 3-vector virtual bonds (see Figure 1a)<sup>46</sup> instead of the original seven torsional angles while keeping the realism of the conformational complexity and



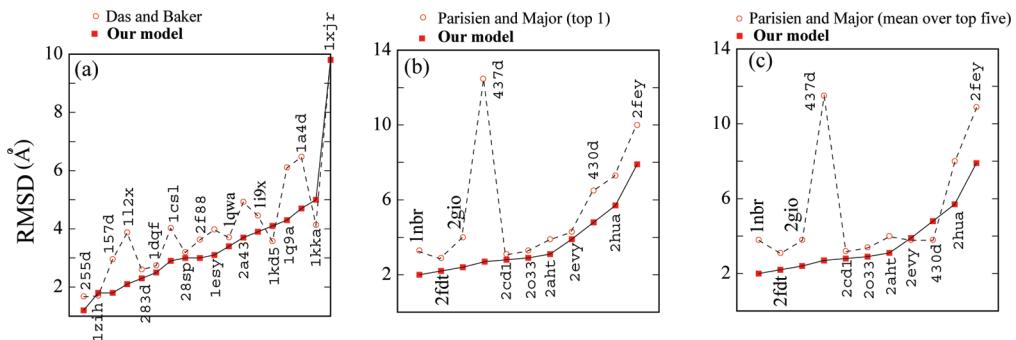
**Figure 1.** (a) A 3-vector virtual bond model for RNA nucleotides. (b) The predicted 2D structure for tRNA<sup>Asp</sup>. (c) The virtual bond scaffold. (d) The predicted all-atom structure (purple blue) and the experimentally determined structure (sand). The rmsd over all heavy atoms is 4.2 Å.

freedom. The three-vector model is an extension of the original two-vector virtual bond model for nucleic acid structures.<sup>43,46</sup> The reduced conformational complexity enables conformational sampling through exact enumeration for the conformations and analytical calculation for the conformational entropy.<sup>46,47</sup>

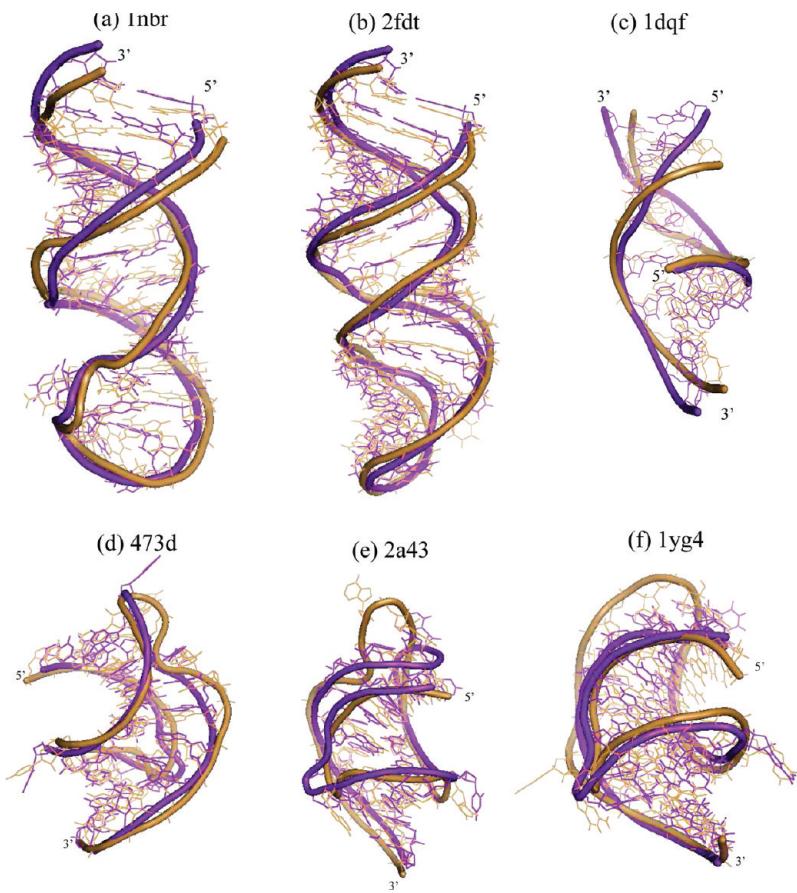
Both the conformational entropy calculation and the construction of a 3D scaffold require the construction of virtual bond structures from the 2D structures (base pairs). We model a helix stem as an A-form RNA helix using the experimentally determined atomic coordinates.<sup>48</sup> For loops, which can be flexible, we use the usual gauche<sup>+</sup> (g<sup>+</sup>), trans (t), and gauche<sup>-1</sup> (g<sup>-1</sup>) rotational isomeric states for polymers<sup>49</sup> to sample the backbone conformations. The fact that the three isomeric states can be exactly configured in a diamond lattice<sup>38</sup> suggests that we can effectively generate loop conformations through random walks of the virtual bonds on a diamond lattice; see details in the Supporting Information about the loop entropy calculation. In the current stage, we can exhaustively enumerate the conformations for a loop  $\leq 14$  nt.

**Free Energy Model and 2D Structural Prediction.** The free energy of a (2D) structure is estimated as  $\Delta G = \sum_{\text{helices}} \Delta G_{\text{helix}} - T \Delta S_{\text{loop}}$ , where the first term is the free energy of the helices and the second term is the loop free energy. The free energy for a helix  $\Delta G_{\text{helix}}$  is evaluated from the sequence-dependent empirical thermodynamic parameters.<sup>50</sup> A 2D structure can correspond to a large number of 3D structures due to the multiplicity of loop conformations. The loop entropy is estimated as  $\Delta S_{\text{loop}} = k_B \ln \Omega / \Omega_{\text{coil}}$ , where  $\Omega$  and  $\Omega_{\text{coil}}$  denote the numbers of loop and coil conformations, respectively. The above virtual bond conformational model (the Vfold model) allows computation of the loop entropy through enumeration of the random walks of the virtual bonds in a diamond lattice.<sup>38</sup>

For a tertiary fold, which involves cross-linked loop-helix connections, the conformational entropy  $\Delta S_{\text{loop}}$  is nonadditive. This is because for cross-linked loop-helix connections, loop conformations span across helices, causing loop conformations



**Figure 2.** (a) Comparison of the RMSDs between our 3D structural prediction model and the model from Das and Baker (2007).<sup>9</sup> We use the hairpin, duplex, and pseudoknot structures in Das and Baker<sup>9</sup> as the test cases. The RMSDs are calculated over the C<sub>4</sub> atom in the backbone. (b) and (c) Comparison of the RMSDs between our 3D model and the model from Parisien and Major (2008).<sup>10</sup> We use the hairpins and pseudoknots in Parisien and Major (2008) as the test cases. The RMSDs are calculated over all heavy atoms. We use the online tool “<http://www.major.iric.ca/MC-Fold/>” to test the accuracy of Parisien and Major’s model and calculate the rmsd for the top predicted structure as well as the mean rmsd over the top five predicted structures. In the calculation, the temperature is set to 25 °C.



**Figure 3.** The predicted structure (purple blue) for hairpins (a, b), duplex (c), and pseudoknots (d, e, f). The RMSDs between the predicted structures and the experimentally determined structures (sand) are 2.0 Å, 2.2 Å, 2.5 Å, 2.7 Å, 4.3 Å, and 4.5 Å, respectively. For the duplex (1dqf), the rmsd is calculated over the C<sub>4</sub> in order to directly compare with the result of Das and Baker (2007).<sup>9</sup> For others, the rmsd is calculated over all heavy atoms.

to be constrained by the volume exclusion from the nearby helices. An advantage of the Vfold model is that it allows us to enumerate loop conformations by accounting for the excluded volume effect from the nearby helices. We then estimate  $\Delta S_{\text{loop}}$  for a given structure as the sum of the entropy of each constituent loop of the structure. Such a strategy has led to reliable prediction for the folding thermodynamics and structure for a variety of

systems such as RNA secondary structures, pseudoknots, and simple molten globule-like folds.<sup>46,47</sup>

To predict the free energy landscape and the structure for a given sequence, we first generate an ensemble of 2D structures. In the current form of the theory, the ensemble includes pseudoknot structures and (nonpseudoknotted) secondary structures. We evaluate the free energy for each 2D structure

using the above Vfold-based method. In the 2D structure prediction, we consider terminal mismatches in the loops. While the energetic parameters of mismatches can be determined from the Turner rule,<sup>50</sup> the dramatic entropic decrease caused by the formation of mismatches are evaluated from our Vfold model. Moreover, we consider the energy contributions from a single-bulge loop and small internal loops ( $1 \times 1$ ) using the empirical parameters.<sup>51,52</sup> From the free energies, we identify the low free energy 2D structures (see Figure 1b for the predicted 2D structure for Yeast tRNA<sup>Asp</sup>). This step is critical because a correct 2D structure is a necessary condition for the prediction of a correct 3D structure. A notable advantage of our method here is the ability to treat RNA folds with cross-linked loops and helices.

**Fragment-Based Construction of a 3D Scaffold.** Based on the predicted 2D structure, we build a 3D scaffold. The 3D scaffold for the free energy minimum serves as an initial state for further structural refinement. To build a 3D scaffold, we model a predicted helix as an A-form RNA helix. Because helix stems are assumed to be rigid, the 3D global fold is determined by the loop/junction structure. As described below, we develop a method to select the optimal loop/junction structures from fragments of the known structures.

First, we build a structural template database. We download the complete 1476 PDB structures (<http://www.rcsb.org/>) and classify the structures according to the different motifs such as hairpin loops and internal/bulge loops, 3-way junctions, 4-way junctions, and pseudoknots. Second, we search for the optimal structural templates for the loops/junctions in the predicted 2D structure (see colored nucleotides in Figure 1b). We use the hairpin loop L1 (UAAUGGUCAG) in Figure 1b to illustrate our search strategy. We screen the database for hairpin loops of the same length. The optimal fragment template is the one with the minimum values of  $h_1$ , the number of the different nucleotides between L1 and the template candidate. Often this criteria leads to two or more templates. In order to distinguish these templates, we further apply the second criteria below. We define parameter  $h_2$ :  $h_2 = \sum_i h_{2i}$ , where  $h_{2i}$  is the hamming distance between nucleotide  $i$  in the selected fragment and the corresponding nucleotide in L1 through the following substitution cycles: A → G → C → U, C → U → A → G, G → A → U → C, and U → C → G → A. We find that the best template for L1 has sequence CAAUGGUCAC with  $h_1 = 2$  and  $h_2 = 4$  from the PDB structure 1f7u.<sup>53</sup> In a similar way, we find the optimal templates for hairpin loops L2 and L3 from the PDB structures 1ato<sup>54</sup> and 3cul<sup>55</sup>, respectively. The templates have sequences UCCUCGC and UUCGAAU, respectively, with  $(h_1, h_2)$  equal to (2, 4) for both. In rare cases, the templates can result in steric clash in the 3D construct. If this occurs, we will include one or more terminal base pairs of the helix (such as the U13-G22 pair for L1) into the loop sequence until a viable structure is found.

For multibranched loops (MBLs), the availability of templates in the PDB database is limited. To effectively enlarge the search space for the fragments, we allow unzipping of the terminal base pairs of the helix stems in order to relax the restriction on the lengths of the loop branches. We then screen all the known MBL structures with the given number of branches and the given length of each branch. We identify the optimal fragment as the one with the minimum  $(h_1, h_2)$ . Here the  $h_1$  and  $h_2$  values are evaluated as the sum over all the loop branches.<sup>10</sup> For instance, for junction J4 (see Figure 1a), we find the optimal template as the MBL in the PDB structure 2dr2<sup>56</sup> with the (minimum)  $(h_1, h_2)$  values equal to (12, 22). The sequences of branches in the

template MBL are 5'GUGGC3', 5'CGCG3', 5'AGGUUG3', and 5'UC3' for the loop branches between S1 and S2, S2 and S3, S3 and S4, and S4 and S1, respectively (see Figure 1b).

From the 3D structure of the templates, we extract the coordinates of P, C<sub>4</sub>, N<sub>1</sub>, or N<sub>9</sub> to construct the virtual bonds structures. We then assemble the backbone chain of the loop/junction fragments (L1, L2, L3, and J4 in Figure 1) and the helix stems (S1, S2, S3, and S4 in Figure 1) using the algorithm reported in ref 57. This step leads to a 3D (virtual bond) scaffold for the whole structure (Figure 1c). Because the templates usually contain mutated sequences, we need to revert the sequence to the “wild-type” form before implementing further refinement. The use of the virtual bond structure is ideal for our purpose here because it captures the backbone conformation and sugar-base orientations, which are critical determinants for the success of further structural refinement.

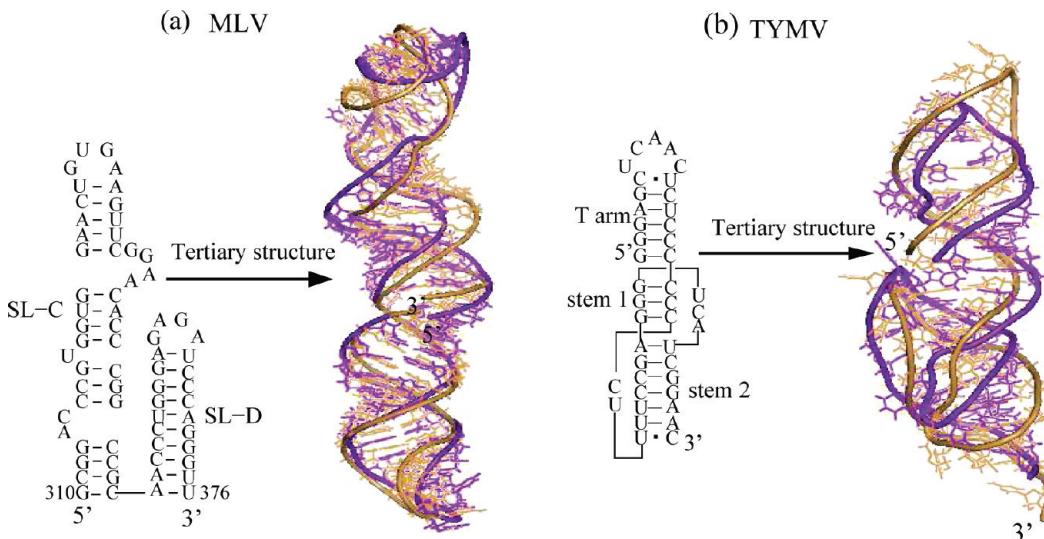
In the above method, we use two strategies to significantly augment the fragment database. First, the fragment templates are selected according to junction/loop lengths (as predicted by the Vfold) instead of the sequence identity. For instance, for the predicted tRNA<sup>Asp</sup> in Figure 1b, the sequences of the 4-way junction (5'AUAGU3', 5'AUGG3', 5'CAGAUC3', and 5'GU3') are significantly different from the sequences of the selected template (PDB ID: 2dr2). Second, the allowed variations of the junction/loop lengths would further enhance the availability of the fragment database.

The large pool of the known loop/junction structures with the different types and different lengths leads to broad applicability of the model. For instance, our model predicts not only the test cases by Das and Baker<sup>9</sup> and Parisien and Major<sup>10</sup> (see Figure 2), but also the pseudoknotted structures such as TYMV and other larger RNA structures (see Figures 4 and 5).

Unlike the homologous models based on the structural families<sup>58,59</sup> or the fragment approach based on the short segments of the loops/junctions,<sup>9,10</sup> we use fragments of the whole loop/junction. Our model has several advantages over the traditional sequence homology-based models, as described by other researchers, such as Pardi, Bax, and colleagues.<sup>58</sup> Unlike the homology-based models, the applicability of our model is not limited to homologous structural families. For only 2 (PDB ID: 1cq5 and 1wks) out of the 38 cases shown in Figure 2, the template is derived from a structure in the same family as the target structure (see also Tables S1 and S2). The template used to build tRNA<sup>Asp</sup> structure (Figure 1b) is derived from another tRNA structure since the currently known cloverleaf structure is from the tRNA family. It may require further more extensive tests to verify whether for this tRNA case, we can choose a non-tRNA template to predict the target tRNA structure. We expect the future availability of the tRNA-like 3D structure in TYMV and TMV viruses<sup>60</sup> may provide reliable templates for tRNAs.

**Atomistic 3D Structural Predictions.** Based on the virtual bond structure, we build an all-atom model for the 3D structure (Figure 1c) by adding the bases to the virtual bond backbone. We first extract the 3D coordinates of nucleotides A, U, G, and C from the known A-form helix structure. These coordinates serve as the templates for base configurations. Second, we add the bases to the virtual bond backbone according to the templates for base configurations.

We then refine the all-atom structure using AMBER energy minimization. First, we performed 2000 steps minimization with 500.0 kcal/mol restraints for all the residues. In the energy minimization, we use the mixture of the steepest descent method and conjugate gradient method. Following the 2000 steps



**Figure 4.** The predicted structure for (a) the G310-U376 domain of MLV RNA and (b) the T arm and the pseudoknot receptor of TYMV RNA. The PDB ids are 1s9s and 1a60, respectively. The RMSDs are 4.1 Å and 6.7 Å for the two structures. The sand color shows the experimentally determined structures. The RMSDs are calculated over all the heavy atoms.

minimization, we run another 2000 steps minimization without restraints. We use a 12 Å layer of TIP3PBOX water molecules to explicitly consider the solvent. In the energy minimization refinement, the backbone charge is neutralized by  $\text{Na}^+$ . We use the command ‘addions’ in AMBER 9 to add  $\text{Na}^+$  until the total charge of the whole system is zero.<sup>61</sup> The nonbonded interactions are cut at 12 Å. The energy minimization is done using the sander of AMBER 9.<sup>61</sup> In the calculation, we use the AMBER force field version ff99 for RNA.<sup>62</sup> We note that there is only a slight change for the rmsd before and after running the AMBER minimization. Figure 1d shows the refined all-atom structure (purple blue). The all-atom rmsd over all heavy atoms is 4.2 Å by comparing with the experimental structure 2tra (sand).<sup>63</sup> The main advantage of the multiscale approach is that the virtual bond tertiary structure as the initial state already lies in the free energy basin, so the all-atom simulations can avoid sampling of large structural rearrangements.

In contrast to the MC-Sym method,<sup>10</sup> our method is based on statistical mechanical calculations for the free energy landscape. In addition, the method for MBL is computationally more efficient. Furthermore, the method is deterministic in giving the optimal 3D structure, while the MC-Sym method outputs an ensemble of 3D structures and does not give the optimal structure without additional structural information (e.g., from experimental measurements).

## RESULTS AND DISCUSSION

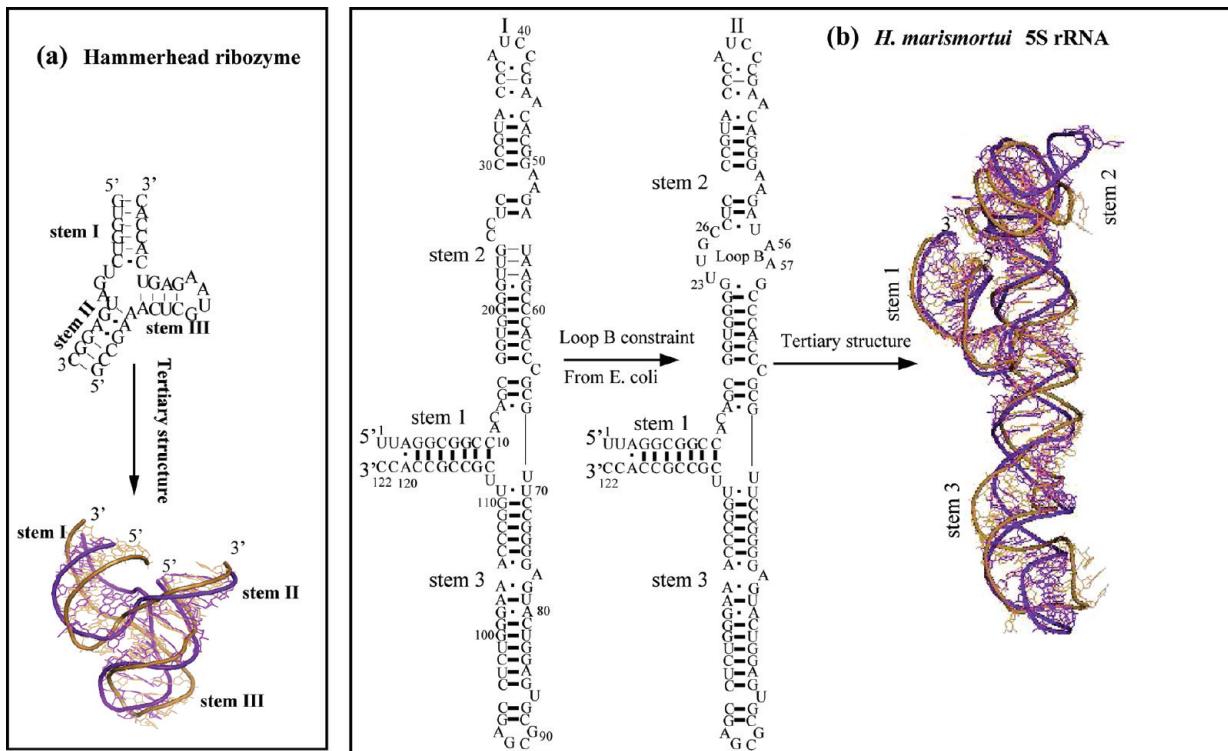
**3D Structural Prediction. Benchmark Test against Other Models.** We test our predictions against two representative *de novo* 3D structural prediction models: the MC-Fold/MC-Sym model<sup>10</sup> and the FARNA model<sup>9</sup> (see Figure 2). In MC-Fold/MC-Sym pipeline algorithm, 3D folds are predicted from 2D structures. Therefore, we first compare the accuracy of our 2D structural prediction model with MC-Fold.<sup>10</sup> We use the sensitivity (SE) and specificity (SP) parameters to measure the prediction accuracy. Here SE (SP) is the ratio between the number of the correctly predicted base pairs and the total

number of the base pairs in the experimentally determined (theoretically predicted) structure. The results in Figure S1 of the Supporting Information suggest that our model gives improved predictions than MC-Fold. We attribute the improvement to the rigorous physics-based calculations for the free energy especially the entropy in our Vfold model.<sup>46</sup>

Second, we compare our model with the MC-Fold/MC-Sym model<sup>10</sup> and the FARNA model<sup>9</sup> on 3D structural prediction. We use exactly the same sequences that were chosen as showcases in the respective publications for the two models. Figure 2a shows the comparison with the FARNA algorithm. We use the sequences for hairpins, duplexes, and pseudoknots in Das and Baker (2007)<sup>9</sup> for benchmark tests. The rmsd is calculated over all  $C_4$  atoms. In general, our model gives better predictions. The model gives a 0.6 Å improvement on the mean rmsd over the 18 sequences. In addition, our model gives more accurate predictions for 15 sequences except for 1zih, 1kka, and 1kd5.

Figure 2b and c gives the results for the comparison with the MC-Fold/MC-Sym pipeline. The test set is adopted from Parisien and Major (2008).<sup>10</sup> Unlike our model, which can predict the single native structure, the MC-Sym gives an ensemble of 3D structures. Therefore, we calculate the rmsd for both the top one structure and the mean value for the top five structures as predicted from the MC-Fold/MC-Sym algorithm (<http://www.major.iric.ca/MC-Fold/>). We find that our model gives better predictions for the tested sequences. The mean rmsd for the eleven structures is 3.6 Å for our model, which outperforms 5.5 Å for the top one structure and 5.3 Å over the top five structures from the MC-Fold/MC-Sym pipeline. Especially for the pseudoknot (437d),<sup>64</sup> MC-Fold fails to predict the native structure as shown by the large rmsd >10 Å. In contrast, our model gives a good prediction with rmsd = 2.7 Å for the sequence.

The recent benchmark tests<sup>18</sup> suggest that (a) MC-Fold/MC-Sym gives more reliable predictions than other molecular dynamics simulation-based models (such as IFoldRNA<sup>3,65</sup>) and (b) the FARNA model yields a similar accuracy as that of IFoldRNA. Our benchmark test for IFoldRNA<sup>3,65</sup> shows an average 5.0 Å rmsd for the 11 cases shown in Figure 2b. The



**Figure 5.** The predicted structure (purple blue) for (a) the hammerhead ribozyme and (b) *H. marismortui* 5S rRNA. The experimentally determined structures are shown in color sand. The RMSDs over all heavy atoms are 6.3 Å and 7.4 Å, respectively. The PDB IDs for the two structure are 1nyi and 1ffk. In (b), we show the predicted 2D structure for 5S rRNA (*H. marismortui*) without and with the loop B constraint. In the loop B constraint, nucleotides U23, U24, G25, C26, A56, and A57 are unpaired, as suggested by the structure of *E. coli* 5S rRNA. The (SE, SP) for the predicted structures I and II are (0.97, 0.90) and (0.97, 0.97), respectively.

accuracy is comparable to that of MC-Fold/MC-Sym (average 5.5 Å based on the top 1 structure and 5.3 Å based on the mean value over top 5 structures). In addition, we find that MC-Fold/MC-Sym gives better RMSDs for 7 out of 11 cases than IFoldRNA. As a comparison, our model gives an average rmsd of 3.6 Å, and the model gives better predictions for 10 out of 11 tested cases than IFoldRNA.

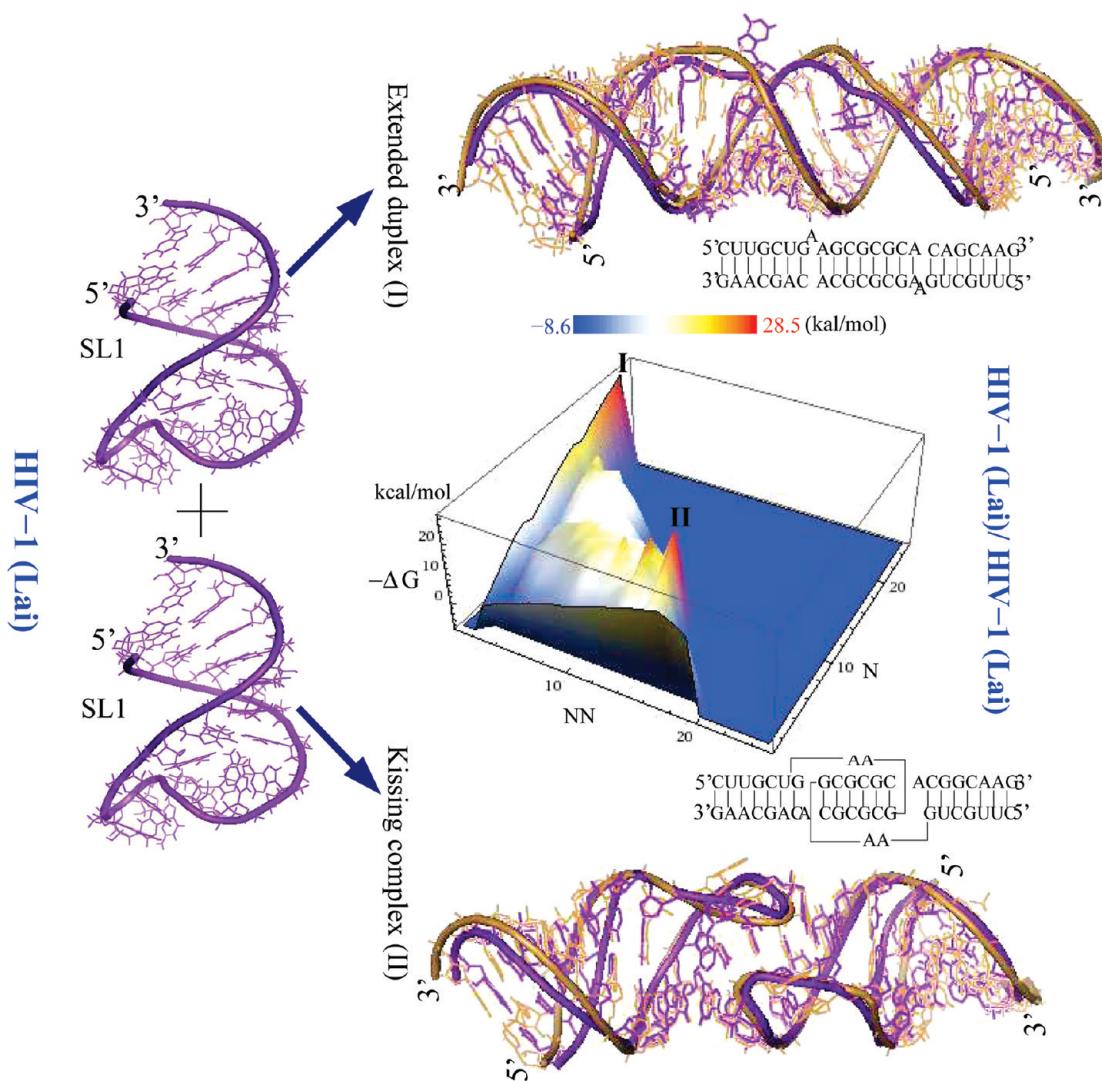
Our evaluation of the accuracy for 3D structure prediction is measured by rmsd. Recent studies<sup>66,67</sup> suggest different metrics to characterize the local configurations in the predicted structures. We here also use the new metrics (INF) proposed by Parisien et al.<sup>66</sup> to test the accuracy of our predictions. The interaction network fidelity (INF) is used to evaluate the local base pairing configurations instead of the global architecture as captured by the rmsd. Table S3 shows the comparison of INF between our model and other two models, i.e., the Parisien and Major's model<sup>10</sup> and the IFoldRNA model.<sup>3,65</sup> The benchmark test results shown in Table S3 of the Supporting Information indicate that our model gives a better INF value.

**Hairpins.** Hairpin is probably the most frequently occurring RNA structural motif. Extensive tests on Figures 2a and b show that our model is reliable for predicting the 3D structure of RNA hairpins for sequence length < 40 nts. Figures 3a and b show the predicted structures for two additional biologically important hairpins. The 29-nt hairpin (1nbr) plays a critical role in regulating iron levels through the binding with proteins.<sup>68</sup> Hairpin 2fdt is important for efficient retrotransposition.<sup>69</sup> Our calculation shows that the overall RMSDs for the two predicted structures are 2.0 Å and 2.2 Å, respectively.

**Multihairpin Structures.** Moloney Murine Leukemia Virus (MLV) RNA contains a conserved structural domain (nucleotides 278 to 374) that is essential for the genome packaging and virus replication. A structural study shows that the domain contains three stem loops (labeled SL-B, SL-C, and SL-D).<sup>70</sup> From the NMR measurement, SL-C and SL-D fold into a rigid 3D structure, while the orientation of SL-B is flexible. Here we apply our model to predict the rigid 3D structure of the domain that contains SL-C and SL-D. We find that the predicted 2D structure (Figure 4a) is consistent with the experimental data.<sup>70</sup> The predicted structure contains two adjacent stem loops (SL-C and SL-D). Based on the predicted 2D structure, we build the 3D structure using the above fragment-based method. The predicted 3D structure is in a good agreement with the experimental structure<sup>70</sup> (see Figure 4a). The rmsd over all heavy atoms is 4.1 Å.

**RNA–RNA Complexes.** The Vfold model has two unique advantages in predicting folding of RNA–RNA complexes. First, it is based on the analytical (nonsimulational) calculation for the free energy and can predict large structural rearrangements of RNAs upon RNA–RNA binding. Second, it can treat both intermolecular and intramolecular interactions (base pairs).<sup>37</sup> Previous tests for 2D structural predictions for RNA–RNA binding indicate that our theory gives better predictions than other models.<sup>37,71</sup> Here based on the 2D structures, we can predict the 3D structures for RNA–RNA complexes.

Tests on systems of simple duplexes show reliable results (see Figure 2a). In Figure 3c we show the predicted 3D structure and the experimental structure for a duplex (PDB code: 1dqf).



**Figure 6.** The predicted free energy landscape for the folding of the complex of the DIS of HIV-1 (Lai) at room temperature. In the energy landscape, N and NN are the numbers of native and non-native base pairs, respectively. The free energy landscape shows two stable structures (I, II). The switch between the structures corresponds to a large structural rearrangement for the single stranded HIV-1 hairpin upon binding to each other. The RMSDs for the predicted extended duplex (PDB ID: 2gm0) and kissing complex (PDB ID: 1xpf) are 3.2 and 3.1 Å, respectively. In the calculation, the RMSDs are evaluated over all heavy atoms.

To go beyond the simple duplexes, we predict the 3D structure for the hammerhead ribozyme.<sup>72</sup> The predicted 2D and 3D structures (Figure 5a) show good agreements with the experiment (PDB code: 1nyi). The overall rmsd between the predicted and the experimentally determined structures is 6.3 Å. Note that the predicted structure has the same global fold as the experimentally determined structure. For example, stems II and III stack coaxially to form a single quasi-continuous helical structure. Stem I splits off the quasi-continuous helical structure and branches toward stem II.

**Pseudoknots.** Pseudoknots play widespread functions in the control of viral replication<sup>30,73</sup> and regulation of telomerase activity.<sup>74</sup> A simple H-type pseudoknot consists of two stems and two loops. Figure 2 shows that our model gives much improved prediction of the pseudoknot structures than FARNA algorithm and MC-Fold/MC-Sym pipeline. Figure 3d, e, and f shows the predicted structures of three frameshifting<sup>30</sup> pseudoknots (PDB codes: 437d, 2a43, 1yg4). The overall RMSDs for the three pseudoknots are 2.7 Å, 4.3 Å, and 4.5 Å, respectively.

Turnip Yellow Mosaic Virus (TYMV) RNA contains a unique pseudoknot structure at the 3' ends. Figure 4b shows the predicted 2D structure, which is in good agreement with the experiment.<sup>75</sup> The SE and SP values are equal to 1.0 and 0.92, respectively. The predicted 3D structure and the experimental structure (PDB: 1a60) are similar in the global shape such as the coaxial stacking between T arm and stem 1 (see Figure 4b). The overall rmsd is 6.7 Å.

**Larger RNAs.** A functional RNA domain can contain a few hundreds of nucleotides. However, it is challenging to use Vfold to predict the tertiary structure of large RNAs (>100 nt) due to the computational time to sample the conformational ensemble. In this study, we attempt to predict the 3D tertiary structure of the 122-nt *H. marismortui* 5S rRNA.<sup>25</sup> The Vfold model gives a reliable prediction for the 2D structure (Figure 5b) with SE and SP values equal to 0.97 and 0.90, respectively. From the 2D structure of *E. coli* 5S rRNA,<sup>76</sup> we find that U23, U24, G25, A56, and A57 forms an internal loop, labeled as Loop B in Figure 5b. Assuming the formation of the internal loop, we can get a better

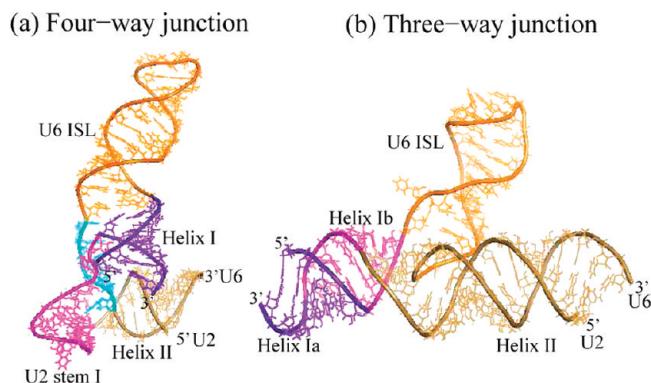
prediction with SE = 0.97 and SP = 0.97 (see Figure 5b). Based on the refined 2D structure, we predict the 3D structure with rmsd equal 7.4 Å.

We attribute the improvements from our model to several reasons. First, our method is based on a statistical mechanical calculation of the entropy and hence gives a more accurate estimation for the free energy. Second, unlike MC-Fold and FARNA models, our model can treat cross-linked loops (such as the loops in the BWYV pseudoknot “437d” in Table S3). Third, we select template structures for the whole-loop/junction rather than short piece-wise fragments of the junctions. Therefore, our approach can handle long-range effects in loop/junction conformations.

**Low-Lying Structures in the Energy Landscape.** The ultimate goal of our computational predictions is to predict not only the native structure but also the 3D structure of the local minima in the energy landscape.<sup>29</sup> The latter aim is much more difficult than the first since it requires complete sampling of the conformational ensembles. Unlike the bioinformatics-based methods, which may involve the problem of incomplete sampling, the physics-based method can treat complete conformational ensemble through the analytical theory for the conformational entropy. Thus, the theory can be used to predict the low-lying structures in the free energy landscape. Here we illustrate the computation for the free energy landscape by using two biologically significant systems (the DIS of HIV-1 (Lai) kissing complexes and the U2–U6 yeast spliceosomal snRNA complexes).

**Dimerization Initiation Signal (DIS) of HIV-1 Strain Lai Kissing Complex.** The predicted free energy landscape for the 2D structures shows two low-lying structures, corresponding to the two alternative structures (the extended duplex and the kissing complex); see Figure 6 for the free energy landscape and the predicted 3D structures. In the energy landscape, we find that the extended duplex and the kissing complex have the similar stability. The calculation shows that the free energies of the two structures are close to  $-28$  kcal/mol at room temperature. The fractional populations of structures I and II are 45% and 55%, respectively. In the calculation, we estimate the loop entropy for the kissing complex using the Vfold model. The entropy change associated with the formation of the kissing loop is about  $\Delta S \sim 17.3$  eu, corresponding to a free energy of  $T\Delta S \approx 4.3$  kcal/mol. If we neglect the contribution of the kissing loop entropy and account only for the free energy of the helical stems, we would predict the kissing complex as the only stable state with a population of nearly 100%. We note that such a result is inconsistent with the experimental observations, which shows both the kissing complex and extended duplex structures.<sup>77–79</sup> The predicted single-stranded SL1 hairpin (Figure 6) has been found in experiments to be the binding site for the DIS of HIV-1 (Lai) dimerization.<sup>77</sup> In order to form the extended duplex (I), hairpin SL1 is completely unzipped. We can predict the 3D structure of the extended duplex at 3.2 Å accuracy (PDB code: 2gm0).<sup>78</sup> Meanwhile, the formation of the kissing complex does not require unzipping the SL1 although there are local structural changes in the loop part of hairpin SL1. The rmsd for the predicted kissing complex is 3.1 Å (PDB code: 1xpf).<sup>79</sup>

**Catalytic Core Domain of U2–U6 Yeast Spliceosomal snRNA Complex.** The Vfold-predicted free energy landscape for the 2D structures shows two minima, corresponding to two alternative (2D) structures: a four-way junction structure and a three-way junction structure.<sup>37</sup> The predicted 2D structures agree with the previously proposed structures from biochemical and structural



**Figure 7.** The predicted 3D structures for (a) the four-way junction and (b) the three-way junction of the yeast spliceosomal U2–U6 complex.

studies.<sup>31,41,42</sup> The U2–U6 catalytic domain may undergo conformational switch between the two structures in the different stages of mRNA splicing. With the fragment-based model developed here, we can predict the 3D tertiary folds for the structures (Figure 7a and b). In the predicted four-way junction structure, we find that U2 stem I forms coaxial stack with helix II, while U6 ISL and helix I stacks coaxially. The predicted structure is consistent with the experimentally observed tertiary interactions between U2 and U6.<sup>80</sup> We note that the structure adopts a similar topology as the hairpin ribozyme.<sup>81</sup> As shown in Figure 7b, structure II folds into a γ-shape junction similar to the junction structure in the hammerhead ribozyme,<sup>72</sup> where helix I and helix II form a quasicontinuous helical structure and the branch of U6 ISL folds toward helix II.

## CONCLUSIONS

We develop a physics-based de novo method to predict RNA 3D tertiary folds from the RNA sequence. Systematic benchmark tests of the model show that the model gives much improved predictions for the 3D structures, as summarized below.

- 1 For a wide range of RNA motifs such as hairpin, duplex, pseudoknots, our model predicts 3D structures with a mean rmsd of about 3.5 Å.
- 2 For the complex pseudoknotted and junction structures such TYMV and hammerhead ribozyme, our model gives a medium rmsd 6.0 Å. The overall shape is in agreement with the experiments.
- 3 For a large RNA such as a 122-nt 5S rRNA domain, our prediction shows a rmsd about 7.4 Å. The predicted tertiary structure can give the correct orientation for the three helix stems.
- 4 The model can be used to predict the 3D structures for low-lying structures in the energy landscape. For example, the model can predict the structural changes during dimerization process for the dimerization initiation signal (DIS) of HIV-1 strain Lai. The prediction shows two alternative structures for the complex. In addition, for the catalytic core domain of yeast U2–U6 spliceosomal snRNA complex, the model can predict the 3D tertiary structures for the two alternative structures: a 4-way junction and a 3-way junction structure. For the 4-way junction structure, helix I and U6 ISL, U2 stem I and helix II form coaxial stacks.

The improved accuracy in the structural prediction can be attributed to the physics-based calculation for the conformational

entropy and free energy, especially for structures with cross-linked loops such as pseudoknots.

The current model can predict the structures of hairpins, two-way (see Figure 4), three-way, four-way junction structures and the structures with cross-linked contacts such as the H-type pseudoknot and RNA kissing hairpins. In addition, the model can also predict structures with a mixture of secondary and pseudo-knotted structures (see TYMV in Figure 4). In our 2D structure prediction model, tertiary interactions such loop–helix interactions in pseudoknots<sup>82</sup> and helix–helix coaxially stacking interactions<sup>38</sup> are accounted for. However, the model does not consider other more detailed tertiary contacts such as specific interactions between backbone/base atoms (e.g., A-minor motifs). For hierarchical folding of RNA structures, these detailed tertiary interactions, which are presumably consolidated after the formation of helices and loops, may not significantly alter the 2D structure. However, we note that tertiary contacts can cause large structural rearrangement for PSc stem in group I intron.<sup>83</sup> Thus, as a caveat, the current approach may not be effective for structures whose folding does not follow the hierarchical folding mechanism. Nevertheless, for nonhierarchically folded RNAs, we could use our Vfold model to find an ensemble of all the low-free energy 2D structures, which may include the native 2D structure. As a future study, it would be useful to investigate, for nonhierarchically folded RNAs, whether the model can capture the correct 3D native structure from the ensemble of the low-energy 2D structures. The specific tertiary contacts can be important for helix/loop 3D orientation and packing. As shown by our tests results, for a great variety of structures, these interactions can be accounted for by our 3D modeling method. First, the predicted cross-linked (pseudo-knotted) loops would impose dramatic restriction on the orientational freedom of the helices/loops; second, for non-cross-linked loops/junctions, such as the ones in multibranched loops, the fragment template and all-atom energy minimization may partially account for the specific tertiary contacts. For more complex folds, incorporating the detailed tertiary interactions from the known structural database<sup>18,84</sup> into the physical theory (Vfold) without losing the rigor in physics and the efficiency in computation is the next step of the theory development. The current theory may provide a first step toward an ultimate all-encompassing theory for RNA 3D structure prediction at ultimate degree of complexity and structural details.

The current theory involves several approximations that should be further examined and improved in the future studies. First, the current model cannot treat convoluted pseudoknots such as pseudoknots with pseudoknotted loops. Further developments of the 3D model should include a method to treat more complex structures such as the pseudoknotted structure in the internal ribosome entry site (IRES) of the cricket paralysis-like viruses.<sup>26</sup>

Second, the current calculation assumes standard 1 M Na<sup>+</sup> ionic solution condition. Future development of the theory should include an ion electrostatic theory to consider the ion effects, especially the Mg<sup>2+</sup> ion effects, on the folding free energy.<sup>29</sup>

Third, we have neglected the tertiary contacts in loops in the 2D structure prediction stage. For a flexible loop, we use the coarse-grained Vfold model to generate multiple viable configurations through self-avoiding walks. However, due to noncanonical tertiary interactions, short internal loops and hairpin loops (such as a stable tetraloop) are often rigid. These interactions have been neglected in the 2D structural prediction. On the other

hand, extensive tests reported by different laboratories<sup>46,85,86</sup> indicate that our loop entropy model gives better predictions for 2D structures than other more crude loop entropy models. Because these different models in comparison use the same thermodynamic parameters for the helices, the difference in the predictions comes from the different loop parameters. The test results suggest the importance of using accurate loop entropy.

The loop thermodynamic (free energy) parameter is determined by the difference between the loop state and the random coil state. For flexible loops, our Vfold-based loop model may give a reliable estimate for the loop entropy. For rigid loops, part of the noncanonical interactions (such as the single-strand base stacking) may partially cancel out when the difference between the loop and the coil is evaluated. In addition, intraloop tertiary interactions may be more pronounced for smaller loops. If the loop free energy contribution from a small loop to the global total free energy is not significant, the tertiary contacts in the small loops would not be strong enough to alter the global shape of the free energy landscape to cause rearrangements of the 2D structure. In that case, a 2D structures corresponds a basin in the free energy landscape while the tertiary contacts add more fine details to the free energy basin. Such coarse-grained free energy basins provide useful initial scaffolds for further detailed structural modeling. Further development of the theory should consider inclusion of the sequence-dependent tertiary contacts, such as more general loop–helix and loop–loop interactions,<sup>82</sup> in loop free energy modeling.

## ■ ASSOCIATED CONTENT

**S Supporting Information.** Text, Figure S1, and Tables S1-S3. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: chenshi@missouri.edu.

## ■ ACKNOWLEDGMENT

The research was supported by NIH through grant GM063732 (to S.-J.C.). Part of computations involved in this research were performed on the HPC resources at the University of Missouri Bioinformatics Consortium (UMBC).

## ■ ABBREVIATIONS:

(Vfold), virtual-bond fold; (UTR), untranslated region; (MLV), Moloney Murine Leukemia Virus; (TYMV), Turnip Yellow Mosaic Virus; (TMV), Tobacco Mosaic Virus

## ■ REFERENCES

- (1) Bartel, D. P. (2009) *Cell* **2009**, *136*, 215-233.
- (2) Kertesz, M.; Iovino, N.; Unnerstall, U.; Gaul, U.; Segal, E. *Nat. Genet.* **2007**, *39*, 1278-1284.
- (3) Ding, F.; Sharma, S.; Chalasani, P.; Demidov, V. V.; Broude, N. E.; Dokholyan, N. V. *RNA* **2008**, *14*, 1164-1173.
- (4) Gherghe, C.; Leonard, C.; Ding, F.; Dokholyan, N. V.; Weeks, K. M. (2009) *J. Am. Chem. Soc.* **2009**, *131*, 2541-2546.
- (5) Hajdin, C. E.; Ding, F.; Dokholyan, N. V.; Weeks, K. M. *RNA* **2010**, *16*, 1340-1349.
- (6) Leontis, N. B.; Westhof, E. (2001) *RNA* **2001**, *7*, 499-512.

- (7) Das, R.; Kudaravalli, M.; Jonikas, M.; Laederach, A.; Fong, R.; Schwans, J. P.; Baker, D.; Piccirilli, J. A.; Altman, R. B.; Herschlag, D. (2008) *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 4144–4149.
- (8) Shapiro, B. A.; Yingling, Y. G.; Kasprzak, W.; Bindewald E. (2007) *Curr. Opin. Struct. Biol.* **2007**, *17*, 157–165.
- (9) Das, R.; Baker, D. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 14664–14669.
- (10) Parisien, M.; Major, F. *Nature* **2008**, *452*, 51–55.
- (11) Jonikas, M. A.; Radmer, R. J.; Laederach, A.; Das, R.; Pearlman, S.; Herschlag, D.; Altman, R. B. *RNA* **2009**, *15*, 189–199.
- (12) Jonikas, M. A.; Radmer, R. J.; Altman, R. B. *Bioinformatics* **2009**, *25*, 3259–3266.
- (13) Das, R.; Karanicolas, J.; Baker, D. *Nat. Methods* **2010**, *7*, 291–294.
- (14) Yang, S.; Parisien, M.; Major, F.; Roux, B. *J. Phys. Chem. B* **2010**, *114*, 10039–10048.
- (15) Pasquali, S.; Derreumaux, P. *J. Phys. Chem. B* **2010**, *114*, 11957–11966.
- (16) Xia, Z.; Gardner, D. P.; Gutell, R. R.; Ren, P. *J. Phys. Chem. B* **2010**, *114*, 13497–13506.
- (17) Flores, S. C.; Altman, R. B. *RNA* **2010**, *16*, 1769–1778.
- (18) Laing, C.; Schlick, T. *J. Phys.: Condens. Matter* **2010**, *22*, 283101.
- (19) Cho, S. S.; Pincus, D. L.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 17349–17354.
- (20) Zhang, J.; Dundas, J.; Lin, M.; Chen, R.; Wang, W.; Liang, J. *RNA* **2009**, *15*, 2248–2263.
- (21) Macke, T.; Case, D. A. Modeling unusual nucleic acid structures. In *Molecular Modeling of Nucleic Acids*; Leontes, N. B., SantaLucia, J., Jr., Eds.; American Chemical Society: Washington, DC, 1998; pp 379–393.
- (22) Cheatham, T., III *Curr. Opin. Struct. Biol.* **2004**, *14*, 360–367.
- (23) Tan, R. K. Z.; Petrov, A. S.; Harvey, S. C. *J. Chem. Theory Comput.* **2006**, *2*, 529–540.
- (24) Michel, F.; Westhof, E. *J. Mol. Biol.* **1990**, *216*, 585–610.
- (25) Ban, N.; Nissen, P.; Hansen, J.; Moore, P. B.; Steitz, T. A. *Science* **2000**, *289*, 905–920.
- (26) Filbin, M. E.; Kieft, J. S. *Curr. Opin. Struct. Biol.* **2009**, *19*, 267–276.
- (27) *The RNA World*, 3rd ed.; Gesteland, R. F., Cech, T., Atkins, J. F., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2005.
- (28) Sosnick, T. R. *Protein Sci.* **2008**, *17*, 1308–1318.
- (29) Chen, S.-J. *Annu. Rev. Biophys.* **2008**, *37*, 197–214.
- (30) Giedroc, D. P.; Cornish, P. V. *Virus Res.* **2009**, *139*, 193–208.
- (31) Madhani, H. D.; Guthrie, C. *Cell* **1992**, *71*, 803–817.
- (32) Chastain, M.; Tinoco, I., Jr. Structural elements in RNA. *Prog. Nucleic Acid Res. Mol. Biol.* **1991**, *41*, 131–177.
- (33) Nussinov, R.; Jacobson, A. B. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 6903–6913.
- (34) Zuker, M. *Science* **1989**, *244*, 48–52.
- (35) Hofacker, I. L. *Nucl. Acids. Res.* **2003**, *31*, 3429–3431.
- (36) Mathews, D. H.; Turner, D. H. *Curr. Opin. Struct. Biol.* **2006**, *16*, 270–278.
- (37) Cao, S.; Chen, S.-J. *J. Mol. Biol.* **2006**, *357*, 292–312.
- (38) Cao, S.; Chen, S.-J. *Nucleic Acids Res.* **2006**, *34*, 2634–2652.
- (39) Tyagi, R.; Mathews, D. H. *RNA* **2007**, *13*, 939–951.
- (40) Lu, M.; Ma, J. (2008) *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 15358–15363.
- (41) Sashital, D. G.; Cornilescu, G.; Butcher, S. E. *Nat. Struct. Mol. Biol.* **2004**, *11*, 1237–1242.
- (42) Butcher, S. E.; Brow, D. A. *Biochem. Soc. Trans.* **2005**, *33*, 447–449.
- (43) Olson, W. K. *Macromolecules* **1980**, *13*, 721–728.
- (44) Duarte, C. M.; Pyle, A. M. *J. Mol. Biol.* **1998**, *284*, 1465–1478.
- (45) Richardson, J. S.; Schneider, B.; Murray, L. W.; Kapral, G. J.; Immormino, R. M.; Headd, J. J.; Richardson, D. C.; Ham, D.; Hershkovits, E.; Williams, L. D.; Keating, K. S.; Pyle, A. M.; Micallef, D.; Westbrook, J.; Berman, H. M. *RNA* **2008**, *14*, 465–481.
- (46) Cao, S.; Chen, S.-J. *RNA* **2009**, *15*, 696–706.
- (47) Liu, L.; Chen, S.-J. *J. Chem. Phys.* **2010**, *132*, 235104.
- (48) Arnott, S.; Hukins, D. W. L. *Biochem. Biophys. Res. Commun.* **1972**, *48*, 1392–1399.
- (49) Flory, P. J. *Statistical mechanics of chain molecules*; Wiley: New York, NY, 1969.
- (50) Serra, M. J.; Turner, D. H. *Methods Enzymol.* **1995**, *259*, 242–261.
- (51) Jaeger, J. A.; Turner, D. H.; Zuker, M. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 7706–7710.
- (52) Kierzek, R.; Burkard, M. E.; Turner, D. H. *Biochemistry* **1999**, *38*, 14214–14223.
- (53) Delagoutte, B.; Moras, D.; Cavarelli, J. *EMBO J.* **2000**, *19*, 5599–5610.
- (54) Kolk, M. H.; Heus, H. A.; Hilbers, C. W. *EMBO J.* **1997**, *16*, 3685–3692.
- (55) Xiao, H.; Murakami, H.; Suga, H.; Ferre-Amare, A. R. *Nature* **2008**, *454*, 358–361.
- (56) Shen, N.; Guo, L.; Yang, B.; Jin, Y.; Ding, J. *Nucleic Acids Res.* **2006**, *34*, 3246–3258.
- (57) Ferro, D. R.; Hermans, J. *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr.* **1971**, *33*, 345–347.
- (58) Grishaev, A.; Ying, J.; Canny, M. D.; Pardi, A.; Bax, A. *J. Biomol. NMR* **2008**, *42*, 99–109.
- (59) Bevilacqua, P. C.; SantaLucia, J., Jr. *ACS Chem. Biol.* **2007**, *2*, 440–444.
- (60) Hammond, J. A.; Rambo, R. P.; Filbin, M. E.; Kieft, J. S. *RNA* **2009**, *15*, 294–307.
- (61) Case, D. A.; Darden, T. A.; Cheatham, III, T. E.; Simmerling, J.; Wang, R. E.; Duke, R.; Luo, K. M. et al. AMBER 9; University of California: San Francisco, 2006.
- (62) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. (1995) *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- (63) Westhof, E.; Dumas, P.; Moras, D. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **1988**, *44*, 112–123.
- (64) Su, L.; Chen, L.; Egli, M.; Berger, J. M.; Rich, A. *Nat. Struct. Biol.* **1999**, *6*, 285–292.
- (65) Sharma, S.; Ding, F.; Dokholyan, N. V. *Bioinformatics* **2008**, *24*, 1951–1952.
- (66) Parisien, M.; Cruz, J. A.; Westhof, E.; Major, F. *RNA* **2009**, *15*, 1875–1885.
- (67) Hajdin, C. E.; Ding, F.; Dokholyan, N. V.; Weeks, K. M. *RNA* **2010**, *16*, 1340–1349.
- (68) McCallum, S. A.; Pardi, A. *J. Mol. Biol.* **2003**, *326*, 1037–1050.
- (69) Nomura, Y.; Kajikawa, M.; Baba, S.; Nakazato, S.; Imai, T.; Sakamoto, T.; Okada, N.; Kawai, G. *Nucleic Acids Res.* **2006**, *34*, 5184–5193.
- (70) D’Souza, V.; Dey, A.; Habib, D.; Summers, M. F. *J. Mol. Biol.* **2004**, *337*, 427–442.
- (71) Andronescu, M.; Zhang, Z.; Condon, A. *J. Mol. Biol.* **2005**, *345*, 987–1001.
- (72) Dunham, C. M.; Murray, J. B.; Scott, W. G. *J. Mol. Biol.* **2003**, *332*, 327–336.
- (73) Staple, D. W.; Butcher, S. E. *PLoS Biol.* **2005**, *3*, e213.
- (74) Theimer, C. A.; Feigon, J. *Curr. Opin. Struct. Biol.* **2006**, *16*, 307–318.
- (75) Kolk, M. H.; van der Graaf, M.; Wijmenga, S. S.; Pleij, C. W.; Heus, H. A.; Hilbers, C. W. *Science* **1998**, *280*, 434–438.
- (76) Speek, M.; Lind, A. *Nucl. Acids. Res.* **1982**, *10*, 947–965.
- (77) Paillart, J. C.; Shehu-Xhilaga, M.; Marquet, R.; Mak, J. *Nat. Rev. Microbiol.* **2004**, *2*, 461–472.
- (78) Ulyanov, N. B.; Mujeeb, A.; Du, Z.; Tonelli, M.; Parslow, T. G.; James, T. L. *J. Biol. Chem.* **2006**, *281*, 16168–16177.
- (79) Ennifar, E.; Dumas, P. *J. Mol. Biol.* **2006**, *356*, 771–782.
- (80) Madhani, H. D.; Guthrie, C. *Genes Dev.* **1994**, *8*, 1071–1086.
- (81) Rupert, P. R.; Ferre-D’Amare, A. R. *Nature* **2001**, *410*, 780–786.
- (82) Cao, S.; Giedroc, D. P.; Chen, S.-J. *RNA* **2010**, *16*, 538–552.

- (83) Wu, M.; Tinoco, I., Jr. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, 95, 11555–11560.
- (84) Laing, C.; Schlick, T. *J. Mol. Biol.* **2010**, 390, 547–59.
- (85) Andronescu, M. S.; Pop, C.; Condon, A. E. *RNA* **2010**, 16, 26–42.
- (86) Sperschneider, J.; Datta, A. *Nucleic Acids Res.* **2010**, 38, e103.