

# Predicting loop–helix tertiary structural contacts in RNA pseudoknots

SONG CAO,<sup>1,2</sup> DAVID P. GIEDROC,<sup>3</sup> and SHI-JIE CHEN<sup>1,2</sup>

<sup>1</sup>Department of Physics, University of Missouri, Columbia, Missouri 65211, USA

<sup>2</sup>Department of Biochemistry, University of Missouri, Columbia, Missouri 65211, USA

<sup>3</sup>Department of Chemistry, Indiana University, Bloomington, Indiana 47405-7102, USA

## ABSTRACT

Tertiary interactions between loops and helical stems play critical roles in the biological function of many RNA pseudoknots. However, quantitative predictions for RNA tertiary interactions remain elusive. Here we report a statistical mechanical model for the prediction of noncanonical loop–stem base-pairing interactions in RNA pseudoknots. Central to the model is the evaluation of the conformational entropy for the pseudoknotted folds with defined loop–stem tertiary structural contacts. We develop an RNA virtual bond-based conformational model (*Vfold* model), which permits a rigorous computation of the conformational entropy for a given fold that contains loop–stem tertiary contacts. With the entropy parameters predicted from the *Vfold* model and the energy parameters for the tertiary contacts as inserted parameters, we can then predict the RNA folding thermodynamics, from which we can extract the tertiary contact thermodynamic parameters from theory–experimental comparisons. These comparisons reveal a contact enthalpy ( $\Delta H$ ) of  $-14$  kcal/mol and a contact entropy ( $\Delta S$ ) of  $-38$  cal/mol/K for a protonated  $C^+(G-C)$  base triple at pH 7.0, and ( $\Delta H = -7$  kcal/mol,  $\Delta S = -19$  cal/mol/K) for an unprotonated base triple. Tests of the model for a series of pseudoknots show good theory–experiment agreement. Based on the extracted energy parameters for the tertiary structural contacts, the model enables predictions for the structure, stability, and folding pathways for RNA pseudoknots with known or postulated loop–stem tertiary contacts from the nucleotide sequence alone.

**Keywords:** tertiary interaction; RNA pseudoknot; folding thermodynamics; RNA structure prediction

## INTRODUCTION

Pseudoknots are tied to many biological functions from ribosomal frameshifting (Draper 1990; Gesteland and Atkins 1996; Giedroc et al. 2000; Staple and Butcher 2005; Giedroc and Cornish 2009) to human/yeast telomerase RNA (hTR) activity (Comolli et al. 2002; Theimer et al. 2003, 2005; Chen and Greider 2005; Cao and Chen 2007; Shefer et al. 2007; Qiao and Cech 2008) to metabolite-sensing riboswitches (Kang et al. 2009; Klein et al. 2009; Spitale et al. 2009). In a pseudoknot, the close physical proximity between the loop and stem often induces loop–stem tertiary structural interactions or contacts (Kim et al. 1999; Su et al. 1999; Cornish et al. 2005; Theimer et al. 2005; Cornish and Giedroc 2006; Chen et al. 2007; Ulyanov et al. 2007). Mutations that disrupt loop–stem tertiary contacts have been found to dramatically reduce biological

activity, often due to a measurable reduction in the thermodynamic stability of the native pseudoknot relative to other partially folded conformations (Nixon et al. 2002a; Theimer et al. 2005).

In hairpin-type (H-type) pseudoknots, loops  $L_1$  and  $L_2$  ( $L_2$  is alternatively denoted  $L_3$  with  $L_2$  the short loop between the two stems at the helical junction) lie in major and minor grooves of stems  $S_2$  and  $S_1$ , respectively (Giedroc et al. 2000). The most frequently occurring loop–stem tertiary interaction is a base triple interaction, wherein a nucleotide in one of the loops forms specific hydrogen bonds with a Watson–Crick base pair in the adjacent helical stem. In general, loop nucleotides can approach the base pair either from the major groove side or the minor groove side of the base pair. Although a wide range of interactions is possible, most often, loop pyrimidines (C or U) form  $L_1$ – $S_2$  major groove Hoogsteen base pairs with purine (G or A, respectively)  $N_7$ – $N_6/O_6$  group to create  $U \cdot (A-U)$  and  $C^+ \cdot (G-C)$  base triples; in contrast, nucleotides from the minor groove side are in close proximity to the 2'-OH group and  $N_3/O_2$ – $N_2$  side of one of the bases in the base pair and form  $L_2$ – $S_1$  base triples.

**Reprint requests to:** Shi-Jie Chen, Department of Physics, University of Missouri, Columbia, MO 65211, USA; e-mail: chenshi@missouri.edu; fax: (573) 882-4195.

Article published online ahead of print. Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.1800210>.

Not surprisingly, experimental studies reveal that specific tertiary structural hydrogen-bonding interactions depend strongly on nucleotide sequence conservation and solution conditions, notably pH, monovalent salt, and/or divalent salt concentration (Soto et al. 2007). For example, protonation of the cytidine N<sub>3</sub> strongly stabilizes a major groove C<sup>+</sup>•G base pair in the context of a C<sup>+</sup>•(G–C) base triple by 2–3 kcal/mol to the overall pseudoknot stability. This value can be dependent on the context of the interaction and may or may not be linked to Mg<sup>2+</sup> binding to the folded RNA (Nixon and Giedroc 2000; Nixon et al. 2002a; Nakano and Bevilacqua 2007). Minor groove-derived adenosine or cytidine-based hydrogen-bonding interactions to 2'-OH donors are also stabilizing and may function cooperatively with other base-pairing interactions (Cornish et al. 2005; Cornish and Giedroc, 2006). Finally, these loop–stem interactions may be augmented by pseudoknot L<sub>1</sub> loop–L<sub>2</sub> loop (typically U–A) Hoogsteen base pairs (Nixon et al. 2002b; Theimer et al. 2005; Giedroc and Cornish 2009; Kang et al. 2009; Klein et al. 2009; Spitale et al. 2009).

Despite the extensive experimental studies on the loop–stem tertiary interaction energetics, quantitative predictions for the structures and folding stabilities for loop–stem tertiary contacts from the nucleotide sequence remain an unsolved problem (Batey et al. 1999; Hermann and Patel 1999; Leontis et al. 2006; Mathews and Turner 2006; Das and Baker 2007; Shapiro et al. 2007; Chauhan and Woodson 2008; Ding et al. 2008; Parisien and Major 2008; Xin et al. 2008; Zhang et al. 2008). A major complication is the strong coupling of tertiary structural folding to divalent ion concentration; however, for the simple H-type pseudoknots discussed here, the folded state is achieved in the absence of Mg<sup>2+</sup> and can be stabilized fully by monovalent salt (Soto et al. 2007). Even in these systems, however, another difficulty is a lack of energetics parameters for the tertiary contacts and the inability to accurately compute the conformational entropy for a pseudoknotted tertiary structural fold. Previous quantum mechanical computations for the tertiary contact energies provide useful estimations for the base triple energies (Oliva et al. 2006). However, the explicit nucleotide-solvent interactions, which can be important both enthalpically and entropically, were not considered in these calculations. An alternative approach is to extract the parameters from the experimental thermodynamic unfolding data. To do so requires a statistical mechanical model that bridges the macroscopic thermodynamics and the site-specific energetics parameters for individual microscopic interactions. In this paper we report the development of such a model for loop–stem tertiary interactions in pseudoknotted RNAs.

Previous molecular dynamics (MD) simulations on the BWYV (Saszar et al. 2001) and hTR pseudoknots (Yingling and Shapiro 2006) have uncovered local structural perturbations associated with the folded state. The predictions for the structure and thermal stability, however, involve large

conformational changes and a large conformational ensemble, from fully unfolded, to partially folded intermediates to fully folded structures. In addition, unlike protein folding, the RNA folding energy landscape may well be rugged and may involve multiple structural rearrangements in the folding process (Chen and Dill 2000; Chen 2008; Hyeon and Thirumalai 2008). Tertiary structural contacts can sometimes drive a rearrangement of the secondary structure for RNAs such as the P5abc domain of the group I intron (Wu and Tinoco 1994; Thirumalai and Woodson 1996). It is therefore necessary to consider the complete conformational ensemble, including structures with the formation of various tertiary contacts for different secondary structures. Herein, we develop such a statistical mechanical approach that allows us to consider the complete conformational ensemble.

Existing theories for the stability of RNA secondary structure are not applicable to the tertiary interactions in pseudoknotted RNA structures. A notable difference in the energetics for the secondary versus tertiary structure derives from the additivity of the free energy. For secondary structures, the nearest-neighbor model (Serra and Turner 1995) posits that the total free energy, enthalpy, and entropy of the structure is equal to the sum of free energies of individual substructural units (base stacks, loops, etc.). For tertiary folds, however, the additivity principle fails because different parts of the structure are brought together by long-range tertiary contacts that essentially cross-link the structure; as a result, the total free energy and entropy of the folded structure is changed in a way that will not be equal to the sum of the free energy and entropy for each individual substructure. Therefore, a method is required that can explicitly treat how folding different substructures influences the energetics of the tertiary structural fold.

In an RNA pseudoknot, a minimum of two loops (L<sub>1</sub> and L<sub>2</sub>) span the corresponding helical stem (S<sub>2</sub> and S<sub>1</sub>, respectively). A loop and a helix can become correlated through, for example, the excluded volume interactions, that is, a loop and helix cannot occupy the same space. The structure of the helix, including the major and minor grooves, can significantly influence the loop conformational variability simply through the loop–stem excluded volume interactions, imposing a strong restriction on the conformational space and thus the entropy of the loop. As a result, the total free energy of a pseudoknot cannot be computed from the sum of the free energy of helices and individual loops; instead, the helix and loop together must be considered as an integral unit when the loop entropy is calculated.

Loop–stem tertiary structural interactions would add additional significant restrictions on the conformational entropy of a loop, with the resultant entropic decrease expected to destabilize the pseudoknotted structure. However, the hydrogen bonds associated with loop–stem contacts will tend to “stick” the loop into the stem, which results in net stabilization of the tertiary structural contacts.

The hydrogen bond-induced stabilization competes with the entropy-induced destabilization, leading to a complex stability of the structures and the free energy landscape for the structures.

The theory developed here aims to predict the native structure, conformational ensemble, and folding stability from the nucleotide sequence alone. To do this, we employ a virtual bond-based conformational model (*Vfold* model) to describe an RNA tertiary structural fold. In order to provide a precise description for tertiary interactions, the model uses the sugar-base orientations in addition to the traditional backbone conformations (Leontis and Westhof 2001; Walberer et al. 2003) to describe the structure. This study focuses on the loop–stem base triple interactions, the most frequently occurring tertiary contacts in an RNA pseudoknot. In order to predict the tertiary interactions from the nucleotide sequence, we first use the *Vfold* model (Cao and Chen 2005, 2006) to compute the entropy for the different structures that contain various loop–stem contacts. Then, for a given sequence, we enumerate all the possible conformations, including component  $S_1$  and  $S_2$  hairpin structures and the pseudoknots, with or without loop–stem tertiary contacts. For each structure, we compute the free energy based on the pre-calculated entropy from the *Vfold* model. Summing over all the possible structures gives the partition function, from which we predict the folding stability and the structural distributions, with the energy parameters for the tertiary structural contacts as input parameters. Through theory–experiment comparison of the folding thermodynamics, we can fit the enthalpy and entropy parameters for the tertiary contacts. As an application of this tertiary folding model, we investigate the stable structures, folding stabilities, and equilibrium unfolding pathways for a series of experimentally studied pseudoknots. Theory–experiment tests have shown good agreements. As a further application of the model, we predict the stability for several telomerase RNA pseudoknots and find a wealth of tertiary interactions that can significantly alter the pseudoknot stability.

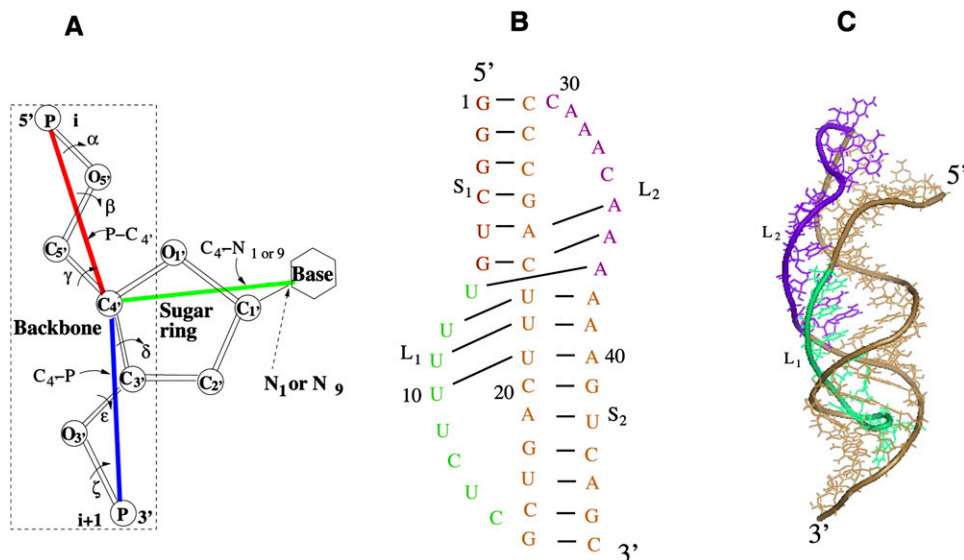
## THEORY AND MODEL

### Structural model for tertiary interaction

Because the C–O bonds in the polynucleotide backbone tend to adopt a fixed conformational state, the conformational degree of freedom can be approximately represented by two virtual bonds,  $5'P-C_4'3'$  and  $5'C_4'-P3'$ , respectively (Olson 1980). Statistics of the known RNA three-dimensional (3D) structure show that the torsion angles  $\beta$  and  $\epsilon$  (Fig. 1A) are close to the *trans* rotameric state (Murthy et al. 1999; Murray et al. 2003; Richardson et al. 2008). Note that the distribution of the  $\epsilon$  torsion is centered at  $-150^\circ$  with a broad shoulder and the  $\beta$  torsion is not exactly centered at  $180^\circ$  (Murthy et al. 1999). Therefore, the two-bond

representation of the backbone conformation is a coarse approximation to the 46 discrete conformers of the RNA backbone (Richardson et al. 2008). Our previous studies indicate that the virtual-bond model can give reliable estimations for the conformational entropy. For instance, the calculated entropies for the hairpin and internal and bulge loops are quite close to the experimental measurements (Cao and Chen 2005) and the predicted loop entropy parameters for pseudoknots lead to accurate predictions for pseudoknot stabilities. Because the exhaustive enumeration of the conformations using the 46 conformers per nucleotide backbone or the all-atom sampling for the full conformational ensemble is computationally infeasible, the simplified virtual bond approach may provide a viable approach to the computation of the conformational entropy. The underpinning idea is that the all-atom free energy landscape can have a discrete number of local minima and the different structures at the different local minima may be approximately characterized by the ensemble of virtual bond conformations. Therefore, we can estimate the all-atom conformational entropy from the conformational ensemble of the virtual bond structures. Based on the virtual bond representation of RNA conformations, we recently developed an RNA folding model (*Vfold* model). The development of the *Vfold* model is motivated by the need to have a simplified conformational model for RNAs that can capture the realism of the key features of the structures, such as the major and minor grooves of the helices and the atomistic conformations of the tertiary contacts. Surveys of the experimentally determined structures suggest that the conformations of the virtual bonds are rotameric (Olson and Flory 1972; Olson 1980; Duarte and Pyle 1998; Wadley et al. 2007). Such discrete nature of the conformations leads to significant simplifications for modeling RNA structures (Wadley et al. 2007; Cao and Chen 2009).

A traditional virtual bond model for nucleic structure employs two virtual bonds and accounts for the conformation of the backbone only (Olson and Flory 1972; Olson 1980). Recently, motivated by the importance of the sugar-base orientation in tertiary interactions, we developed a new virtual bond-based RNA folding model (*Vfold* model) by considering the orientation of the base. Specifically, besides the original two backbone virtual bonds, we include a third bond  $C_4'-N_1$  (pyrimidine) or  $C_4'-N_9$  (purine) to describe the sugar-base configuration (see Fig. 1A). From the known PDB database (Michiels et al. 2001; Theimer et al. 2005) for RNA pseudoknots, we find that the distance ( $D_{CN}$ ) between  $N_1$  ( $N_9$ ) and  $C_4'$  atoms is close to a fixed value  $3.9 \text{ \AA}$  (the average value of  $D_{CN}$  is  $3.4 \text{ \AA}$ ). In Supplemental Figure S1, we test the sensitivity of the entropy parameter on  $D_{CN}$  and find that the difference in the conformational entropies based on the two different distances is not significant. In addition, we also find that the torsion angle between plane  $P^i-C_4'-P^{i+1}$  and plane  $P^i-C_4'-N_1$  ( $N_9$ ) is close to the fixed  $g^{-1}$  isomeric state (Cao and Chen 2009). Therefore, for each



**FIGURE 1.** (A) Virtual bonds for a nucleotide: 5'P–C<sub>4'</sub>3', 5'C<sub>4'</sub>–P3', and C<sub>4'</sub>–N<sub>1</sub> (primidine) or C<sub>4'</sub>–N<sub>9</sub> (purine). (B) The secondary structure for hTR pseudoknot (Theimer et al. 2005). (C) The three-dimensional structure of the hTR pseudoknot (PDB code: 1YMO) (Theimer et al. 2005). Loops L<sub>1</sub> and L<sub>2</sub> span across the major and minor grooves of helices S<sub>2</sub> and S<sub>1</sub>, respectively.

$P^i$ –C<sub>4'</sub>–P<sup>*i*+1</sup> backbone conformation, we can fix the N<sub>1</sub> (N<sub>9</sub>) position. A previous statistical study on the RNA molecule also supports a rigid base orientation for a given backbone conformation (Olson and Flory 1972).

A simple H-type RNA pseudoknot consists of two loops (L<sub>1</sub> and L<sub>2</sub>) that span the major and minor grooves, respectively, of S<sub>2</sub> and S<sub>1</sub> helical stems (Fig. 1B,C). Based on the atomic coordinates for an A-form RNA helix, we can reduce an all-atom helix structure into a reduced virtual bond structure. For loops, which are less rigid than helices, we use the usual *gauche*<sup>+</sup> (*g*<sup>+</sup>), *trans* (*t*), and *gauche*<sup>–</sup> (*g*<sup>–</sup>) rotational isomeric states for a polymer (Flory 1969) to represent the torsions of the backbone virtual bonds. A great simplification of the virtual bond-based modeling comes from the fact that the above rotameric virtual bond conformations can be conveniently configured in a diamond lattice by treating a lattice bond as virtual bond (Cao and Chen 2005, 2006). As a result, we can generate flexible loop conformations through self-avoiding random walks in the diamond lattice.

#### Helix stems

For a generic H-type RNA pseudoknot, the intervening junction region between the two helical stems (see Fig. 1) is often bent and/or over-rotated and may contain zero, one, or many intervening nucleotides (Giedroc et al. 2000; Pflingsten et al. 2007; Klein et al. 2009). The deviation of the two-stem structure away from the coaxially stacked continuous A-form helix is due to the conformational propensity of the intervening junction/loop, the loop–stem tertiary contacts, and the conformational viability (chain

connectivity) of the loops that span across the respective helices. As a result, it will not, in most cases, be appropriate to model the two-stem system simply as a rigid continuous A-form helix. For example, as shown in Figure 1B, the hTR pseudoknot contains two helices (S<sub>1</sub> and S<sub>2</sub>) of lengths equal to 6 base pairs (bp) (S<sub>1</sub>) and 9 bp (S<sub>2</sub>), respectively, and an array of loop–stem tertiary interactions as defined by the solution structure (Theimer et al. 2005). For the hTR, the two stems are bent by ~30° (Theimer et al. 2005). In addition, the loop–helix interaction can also cause minor distortion of the helices away from perfect A-form RNA helical geometry. As a result, replacing the helices by a continuous A-form helix would cause a large distortion of the structure with a root-mean-square deviation (RMSD) of 6.2 Å, predominantly due to the noncoaxial orientation of the two helices. In the calculation, we use the best molecular fit program (Ferro and Hermans 1971) to calculate the RMSD.

We use the two helical stems in the hTR pseudoknot as a template for modeling the two-helix structure for a general pseudoknot that contains loop–stem tertiary interactions. Specifically, we use the P, C<sub>4'</sub>, and N<sub>1</sub> (N<sub>9</sub>) coordinates in the solution structure of the hTR pseudoknot (Theimer et al. 2005) to generate the virtual bond representation for the two helices as a unit. The major motivation here is to capture any noncoaxial orientations of the two helices. We model each of the two stems by an A-form RNA helix, which causes small RMSDs equal to 1.6 and 1.4 Å for the 6-bp stem 1 and the 9-bp stem 2, respectively. To develop a versatile template system for modeling pseudoknots of various helix lengths, we elongate each helix to 10 bp to generate a two-stem structure with

each helix length equal to 10 bp. Using this two-stem structure as a template, we can construct the 3D structure for pseudoknot helices of any stem lengths  $\leq 10$  bp. To test the validity of the approach, we computed the RMSD between the template and two other pseudoknots, including the sugarcane yellow leaf virus (ScYLV) (Cornish et al. 2005) and beet western yellows virus (BWYV) P1-P2 plant viral frameshifting pseudoknots (Su et al. 1999), each of which contains different loop–stem base triple interactions. The resultant RMSDs are 3.8 Å and 3.6 Å for ScYLV and BWYV pseudoknots, respectively. This suggests that the template structure may provide an approximate description for the two-helix configurations for a general pseudoknot that contains loop–stem tertiary interactions. Here the RMSDs are calculated over all the P, C<sub>4'</sub>, and N<sub>1</sub> (N<sub>9</sub>) atoms that define the virtual bonds.

### Loops

We model the loop conformations by self-avoiding walks in diamond lattice (Cao and Chen 2005, 2006), where each virtual bond is represented by a lattice bond. For a given two-stem structure, we count the number of loop conformations  $\Omega_{\text{loop}}$  through exhaustive computer enumeration and calculate the loop entropy as  $\Delta S_{\text{loop}} = k_B \ln \Omega_{\text{loop}}$ . Ignoring correlation between the two loops, we then compute the total loop entropy as the sum of each individual loop entropy. For each loop, the conformational entropy is determined by three factors: the volume exclusion between the different nucleotides within the loop, between the loop and helix nucleotides, and the formation of the base triple contacts between loop and helix nucleotides. In the enumeration of loop conformations, we take into account all three of these factors, which, in addition to the chain connectivity of whole molecule, impose rather strong restrictions on the loop conformation.

Because our template for the two-helix structure is derived from a specific pseudoknot (the hTR pseudoknot), we next determined the degree to which the template leads to reliable loop entropies for other pseudoknots. As a first test, we replaced the structure for the two helices in the hTR pseudoknot with the corresponding two-helix structure of the T2 gene 32 pseudoknot (Holland et al. 1999) and then recalculated the loop entropy of the chimeric structure. Our results for the loop entropy show that altering the helix structures leads to a minor error of 10% in total loop entropy. Therefore, the hTR pseudoknot-derived helix template would appear to return reliable estimations of the loop entropy that could be used for structure predictions.

As discussed above, different base triples can be formed between the loop and the stem. A given S<sub>1</sub>–L<sub>2</sub> or S<sub>2</sub>–L<sub>1</sub> helix–loop pair often contains multiple base triples. We denote a set of (multiple) loop–stem base triples as a “tertiary contact mode.” For a given helix and loop, we enumerate all possible modes and compute the loop

entropy for each given mode using the *Vfold* model. We then tabulate the loop entropy parameters as a function of the helix length, loop length, and the mode. Because the L<sub>1</sub> and L<sub>2</sub> loops are nonsymmetric and span the major and minor helix grooves of the S<sub>2</sub> and S<sub>1</sub> stems, respectively, we make two separate tables for the two loops. With such loop entropy tables for individual tertiary structural folds, we can readily compute the entropy for any pseudoknot with a given loop–stem base triple tertiary contact architecture.

### Base triples

A base triple is formed when a loop residue forms hydrogen bonds, for example, Hoogsteen hydrogen bond into the major groove, with a base pair in the helix (see Figs. 1B, 2). Solution NMR and crystallographic structures for a variety of RNAs, including the sugarcane yellow leaf virus (ScYLV) (Cornish et al. 2005) and BWYV (Su et al. 1999) plant frameshifting pseudoknots and the hTR (Theimer et al. 2005) pseudoknot, reveal eight types of frequently occurring loop–stem base triples, which can be designated A•(G–C), C•(G–C), A•(C–G), C•(C–G), A•(A–U), U•(A–U), A•(U–A), and U•(U–A) (Fig. 2A–G for seven base triples). In the current coarse-grained model, no distinction is made as to whether an interaction occurs in the minor or major groove of the helical stem, i.e., we allow these base triples to be formed in both minor and major grooves. Furthermore, the model makes no effort to distinguish base-pairing patterns in two similar base triples such as C•(G–C) or C•(C–G) (Nixon et al. 2002b).

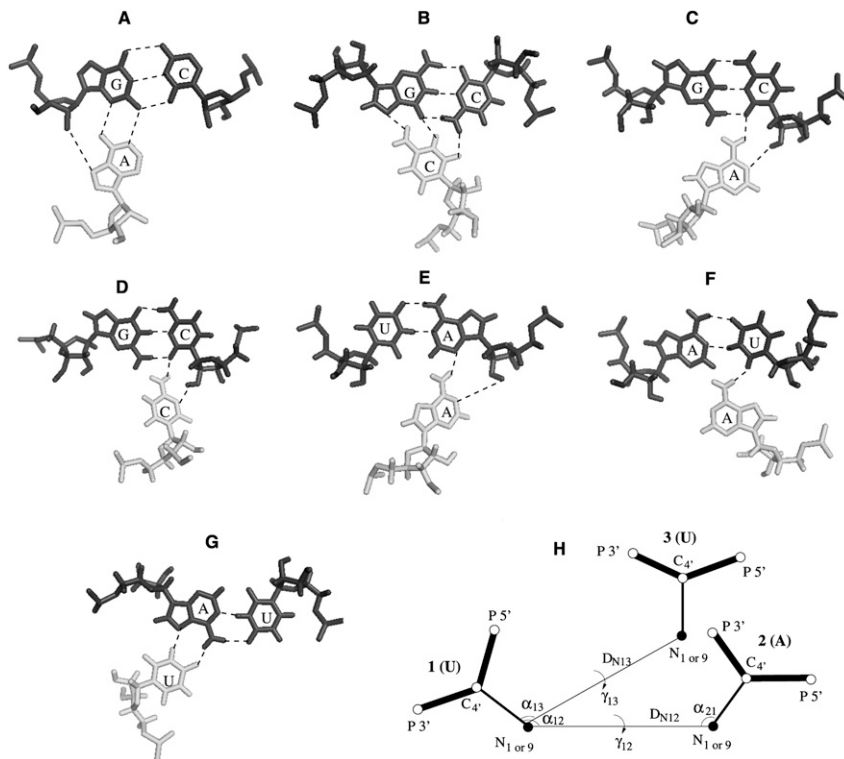
Since RNA conformations are represented by virtual bonds in the *Vfold* model, a method is required to identify base triples at the level of virtual bonds. For illustration, we use the U•(A–U) major groove Hoogsteen base triple of the hTR pseudoknot (Fig. 2G). The hydrogen bonds restrict the flexibility of the nucleotides in the loops. Our survey of the known pseudoknot structures leads to the following operationally defined rules for the virtual bond configurations for the formation of a base triple (see Fig. 2H; Table 1):

1. The close distance between the nucleotide in the loop and the nucleotides in the helix:  $5.0 \text{ \AA} \leq D_{N_{12}} \text{ or } D_{N_{13}} \leq 14.0 \text{ \AA}$ ;
2. The relative orientations of the sugar–base virtual bonds:  $(\alpha_{12}, \alpha_{13}, \alpha_{21}, \text{ and } \alpha_{31}) \leq 120^\circ$ ;
3. The torsion angle  $|\gamma_{12} \text{ or } \gamma_{13}| \leq 120^\circ$ .

Using these rules, we can identify the base triples for a given virtual bond structure of RNA pseudoknot.

### Partition function

Central to the prediction of the structure and folding stability from the nucleotide sequence is the partition function  $Q$ , which is the weighted sum over all the possible structures:



**FIGURE 2.** (A–G) The configurations for seven base triples selected from the BWYV (Su et al. 1999), ScYLV (Cornish et al. 2005), and hTR pseudoknots (Theimer et al. 2005). (H) A simplified configuration for the base triple U•(A–U). Nucleotide 3 (U) forms a base triple interaction with base pair 2 (A)–1 (U), and we use three atoms (P, C<sub>4'</sub>, and N<sub>1</sub> or N<sub>9</sub>) to represent a nucleotide. D<sub>N12</sub> or D<sub>N13</sub> denotes the N<sub>1</sub> or <sub>9</sub>–N<sub>1</sub> or <sub>9</sub> distance between two nucleotides.  $\gamma$  is the torsion angle C<sub>4'</sub>–N<sub>1</sub> or <sub>9</sub>–C<sub>4'</sub>–N<sub>1</sub> or <sub>9</sub>.  $\alpha$  is the bond angle for C<sub>4'</sub>–N<sub>1</sub> or <sub>9</sub>–C<sub>4'</sub>.

$$Q = \sum_{\text{struct}} e^{-\Delta G_s/k_B T}, \quad (1)$$

where  $\Delta G_s$  is the free energy of a given structure  $s$ . To compute the partition function from the nucleotide sequence, we enumerate all the possible secondary structures and pseudoknotted structures with and without loop-stem base triple contacts. For each pseudoknot, we exhaustively enumerate all the possible tertiary contact modes (combinations of different base triples) between stem  $S_1$  and loop  $L_2$  and between stem  $S_2$  and loop  $L_1$ , respectively. The number of tertiary contact modes  $N_{\text{mode}}$  for the base triples between an  $m$ -nucleotide (nt) loop and an  $n$ -nt helix can be estimated from the following formula:

$$N_{\text{mode}} = \sum_{l=1}^{\min(m,n)} \binom{m}{l} \binom{n}{l},$$

where  $l$  is the number of the base triples in the mode. The (combinatorial) number of modes and hence the computational time grows exponentially with the loop and stem lengths. Our benchmark test for the computational time (Fig. 3) for the different loop lengths shows that it requires  $\sim 20$  h to enumerate the loop conformations for all the

tertiary modes for a pseudoknot with  $S_1 = 9$  bp and  $L_2 = 9$  nt. Thus, in order to efficiently enumerate all the tertiary modes for pseudoknots, we restrict the stem length to  $\leq 9$  bp and loop length  $\leq 9$  nt in the conformational ensemble for each partition function calculation.

For a given pseudoknot, we use the nearest-neighbor model and the empirical thermodynamic parameters (Turner rules) (Serra and Turner 1995) to calculate the free energy  $\Delta G_{\text{stacks}}$  for the stems as the sum of the free energies for each base stack. Using the pre-calculated loop entropy tables, we obtain the loop entropy from the loop conformational count  $\Omega_m$  for a tertiary contact mode  $m$ . We compute the partition function  $Q_{\text{PK}}$  from the sum over all the different tertiary contact modes (denoted by  $m_1$  for contacts between  $L_2$  and  $S_1$  and  $m_2$  for contacts between  $L_1$  and  $S_2$ ):

$$Q_{\text{PK}}(S_1, S_2, L_1, L_2) = e^{-\Delta G_{\text{stack}}/k_B T} \sum_{m_1} \Omega_{m_1} e^{-E_{m_1}/k_B T} \sum_{m_2} \Omega_{m_2} e^{-E_{m_2}/k_B T}. \quad (2)$$

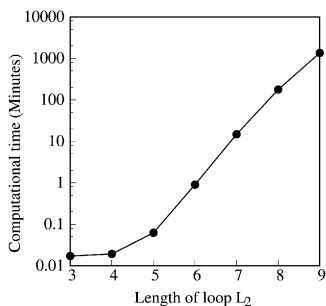
Here the tertiary contact free energy  $E_m$  for a given tertiary contact mode  $m$  accounts for the free energy change, excluding the loop entropy change, upon the formation of the base triples. As the lowest-order approximation, we assume that the total free energy of the tertiary contacts for a given mode  $m$  is the sum of the free energy for each individual contact:

$$E_m = \sum_i (\Delta h_i - T \Delta s_i), \quad (3)$$

**TABLE 1.** Distances, torsion angles, and bond angles for the known triple bases

PDB ID	1	2	3	D <sub>N12</sub>	D <sub>N13</sub>	$\alpha_{12}$	$\alpha_{21}$	$\alpha_{13}$	$\alpha_{31}$	$\gamma_{12}$	$\gamma_{13}$
1YG4	G4	C17	A21	10.8	9.8	82	33	97	41	16	–15
	U5	A16	A22	8.7	9.4	84	57	70	22	–19	29
	G6	C15	A24	9.8	9.3	81	53	73	21	–10	–6
	G7	C14	A27	10.9	7.6	68	50	87	33	–36	43
	G12	C28	C8	9.2	8.6	28	14	47	49	24	–12
1YMO	G6	C24	A36	8.2	11.2	99	42	51	26	60	13
	U21	A40	U10	11.8	7.2	20	25	68	41	–54	13
	U22	A39	U9	12.0	6.8	4	23	64	37	–78	25
	U23	A38	U8	12.5	7.1	8	21	58	32	–37	19

We obtain these triple bases from two NMR structures, 1YG4 (Cornish et al. 2005) and 1YMO (Theimer et al. 2005).



**FIGURE 3.** Plot of the CPU time for the enumeration of the complete ensemble of tertiary contacts for different loop length ( $L_2$ ) and fixed stem length ( $S_1 = 9$  bp). Note that the  $y$ -axis is in log scale.

where  $\Delta h_i$  and  $\Delta s_i$  are the enthalpy and entropy, respectively, for the formation of the  $i$ -th base triple. The enthalpy is mainly derived from the hydrogen-bonding interactions, while the entropy comes from the reduced conformational freedom for the bases and the change of the solvent entropy associated with the formation of the base triple. As shown below, Equation 2 allows us to optimize (fit) the values of the enthalpy and entropy for different types of base triples from the experimental thermodynamic data for a training set of selected pseudoknots. We do not take into account the coaxial stacking in the prediction due to the bent configuration of the two stems, nor do we explicitly consider any potential loop  $L_1$ –loop  $L_2$  base-pairing interactions.

Using a recursive algorithm, we can efficiently calculate the total partition function ( $Q_{\text{tot}}$ ) for a given sequence (Cao and Chen 2006). From the conditional partition function ( $Q_{ij}$ ) with the condition that nucleotides  $i$  and  $j$  form a base pair, we calculate the probability for nucleotides  $i$  and  $j$  to form a base pair:

$$P_{ij} = Q_{ij}/Q_{\text{tot}}. \quad (4)$$

From  $P_{ij}$  for all the possible ( $i, j$ ) pairs, we predict the structure and the equilibrium folding pathway as a function of temperature as well as a heat capacity melting curve  $C(T)$ , or melting profile, directly from the partition function  $Q_{\text{tot}}$ :

$$C(T) = \frac{\partial}{\partial T} \left[ k_B T^2 \left( \frac{\partial}{\partial T} \right) \ln Q_{\text{tot}} \right]. \quad (5)$$

We can evaluate our theory through theory-experiment comparisons of the calculated and experimentally determined heat capacity thermal unfolding curves for different RNA sequences.

## RESULTS AND DISCUSSION

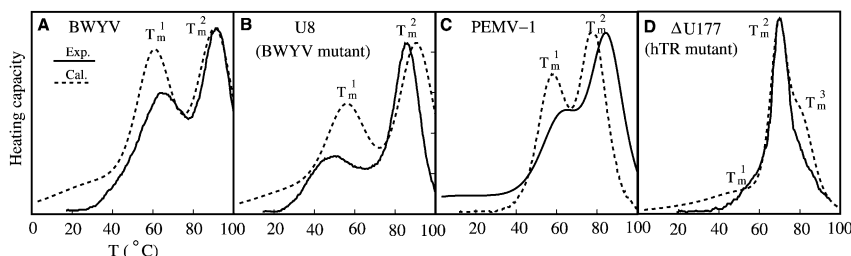
### Energetic parameters for base triples

The *Vfold* model developed in the present study allows us to extract the tertiary contact energetic parameters from

experimental data reported for a training set of RNA pseudoknots. We note that the approach is different from a simple curve fitting exercise because the theory is based on the rigorous statistical mechanical properties of the system. As a first-order approximation, we classify the base triples into two types: protonated [ $C^+(C-G)$  or  $C^+(G-C)$ ] and unprotonated base triples [ $A(A-U)$ ,  $A(U-A)$ ,  $U(A-U)$ ,  $U(U-A)$ ,  $A(C-G)$ ,  $A(G-C)$ ]. The classification is based on the protonation state of the cytosine  $N_3$  at pH 7. A protonated base triple can have a higher stability than an unprotonated base triple (Nixon et al. 2002b; Giedroc and Cornish 2009). We note that the base-pairing structure of the protonated  $C^+G$  Hoogsteen base pair is identical in both the  $C^+(C-G)$  and  $C^+(G-C)$  base triples, with the  $C^+(C-G)$  base triple formed by an S-turn in the polynucleotide chain (Nixon et al. 2002b; Giedroc and Cornish 2009).

We next optimized the ( $\Delta h$ ,  $\Delta s$ ) parameters (Eq. 3) from direct theory–experiment comparisons for the melting thermodynamics for four different RNAs in the training set. These include the PEMV-1 (Nixon et al. 2002a), BWYV wild-type, and BWYV U8 frameshifting plant viral pseudoknots (Nixon and Giedroc 2000; Soto et al. 2007), and the  $\Delta U177$  hTR pseudoknot (Theimer et al. 2005). These systems were selected because they contain a variety of loop–stem base triples and their folding thermodynamics have been studied experimentally. Thermodynamic measurements were performed at different solution conditions. For example, 0.5 M KCl was used for plant viral frameshifting RNAs, and 0.2 M NaCl for the  $\Delta U177$  hTR mutant, with the empirical base-stacking energy parameters used to evaluate the free energies for the formation of helical stems compiled at 1 M NaCl “standard” conditions (Serra and Turner 1995). To account for the effect of the different salt concentrations, we employed the following empirical correction (SantaLucia 1998; Tan and Chen 2006) for the enthalpy and entropy of individual base-pair stacks:  $\Delta H(\text{Na}^+) = \Delta H(1\text{M})$ ;  $\Delta S(\text{Na}^+) = \Delta S(1\text{M}) + 0.368 \ln [\text{Na}^+]$ , where  $\Delta H(1\text{ M})$  and  $\Delta S(1\text{ M})$  are the empirical enthalpy and entropy for a base-pair stack at 1 M NaCl (Serra and Turner 1995). Previous tests on RNA hairpins suggest that the above correction is quite reliable for salt concentrations that are not too low ( $\geq$  millimolar) (Tan and Chen 2006; Zhang and Chen 2006).

The experimental melting profiles acquired at pH 7.0 show multiple melting transitions (Fig. 4). For the plant viral RNAs, previous studies reveal three partially overlapping, sequential unfolding transitions that were assigned to the disruption of the tertiary loop–stem interactions, followed by successive unfolding of each of the helical stems to the unfolded state (Nixon and Giedroc 2000; Nixon et al. 2002a; Soto et al. 2007). However, for purposes of extracting best-fit thermodynamic parameters, these melting profiles were simply “fit” to two sequential unfolding steps defined by  $T_m^1$  and  $T_m^2$  (Fig. 4A–C). For the hTR  $\Delta U177$  pseudoknot (see Fig. 1) three discrete



**FIGURE 4.** Comparisons between the calculated (dashed lines) and the experimental (solid lines) thermal unfolding curves for (A) the BWYV wild-type pseudoknot (Nixon and Giedroc 2000), (B) the C8U BWYV pseudoknot (Nixon and Giedroc 2000), (C) the PEMV frameshifting pseudoknot (Nixon et al. 2002a), and (D) the  $\Delta$ U177 hTR RNA pseudoknot (Theimer et al. 2005). The calculated and experimental melting curves for PEMV, BWYV wild-type, and BWYV U8 mutant are measured at 0.5 KCl (pH 7.0), while for the  $\Delta$ U177 hTR RNA, the salt concentration is 0.2 NaCl (pH 7.0).

unfolding steps are clearly observed in the unfolding profiles, and thus it was modeled as such (Fig. 4D). The optimal fitting for the melting temperatures for the two transitions  $T_m^1$  and  $T_m^2$  of BWYV, U8 and PEMV, shown in Figure 4, leads to the following results for the tertiary contact parameters:

$$(\Delta h, \Delta s) = (-7 \text{ kcal/mol}, -19 \text{ cal/mol/K})$$

for (A or U)•(A–U or U–A) (6)

and (A)•(C–G or G–C),

and

$$(\Delta h, \Delta s) = (-14 \text{ kcal/mol}, -38 \text{ cal/mol/K})$$

(pH 7.0) for (C)<sup>+</sup>•(C–G or G–C). (7)

The uncertainties for  $(\Delta h, \Delta s)$  are about  $(\pm 0.5 \text{ kcal/mol}, \pm 1 \text{ cal/mol/K})$ . The above parameters are derived from a simultaneous fit to BWYV, U8, and PEMV-1 data sets. As shown in Figure 5, the melting temperature is sensitive to the  $(\Delta h, \Delta s)$ . Due to the nature of the *Vfold* model, the parameters are nucleotide-based instead of atom-based; as a result, the parameters are not specific to the chemical structure or atomistic details. In addition, to estimate the use of the coarse-grained conformational model may also contribute to the uncertainty of the fitted  $(\Delta h, \Delta s)$ . How sensitive is the fitted  $(\Delta h, \Delta s)$  to the conformational entropy model? Our test results show that a change of the conformational entropy by  $1 k_B$  would lead to the new best-fit parameters  $(\Delta h, \Delta s) = (-6 \text{ kcal/mol/K}, -15 \text{ cal/mol/K})$  for the unprotonated base triples and  $(-13 \text{ kcal/mol/K}, -35 \text{ cal/mol/K})$  for the protonated base triples. The small change in the conformational entropy does not cause significant changes in the melting curve profile (see Supplemental Fig. S2).

The above parameter set is consistent with the experimental finding that turning off the protonation would lead

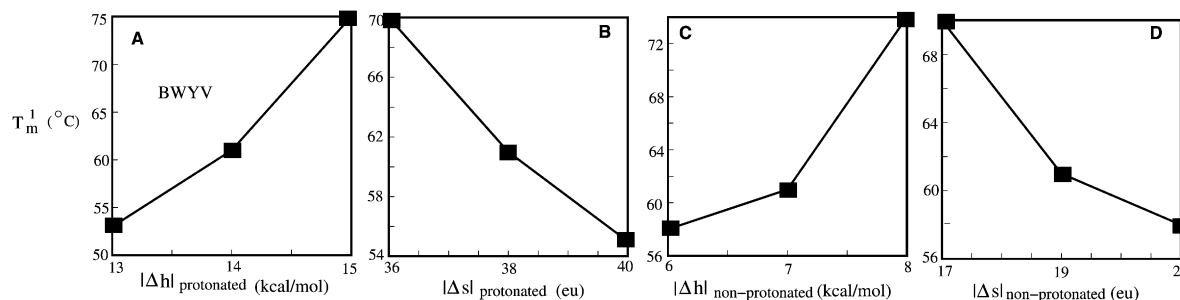
to an enthalpic increase of  $\sim 7 \text{ kcal/mol}$  and a free energy increase of  $1.5 \text{ kcal/mol}$  (Cornish et al. 2005). Our above fitted parameter  $\Delta h = 7 \text{ kcal/mol}$  is also close to the previous experimental result  $\Delta h = 5 \text{ kcal/mol}$  for a U•(A–U) base triple (Krakauer and Sturtevant 1968). The  $2.0 \text{ kcal/mol}^{-1}$  theory–experiment discrepancy ( $\Delta h = 7 \text{ kcal/mol}$  versus  $5 \text{ kcal/mol}$ ) is probably caused by the simplifying assumption that distinct non-protonated base triples irrespective of structure are forced to be characterized by the same thermodynamic parameters.

In general, the predicted melting curves agree quite well with the experimental melting profiles. For example, both theoretical predictions and experimental measurements give two peaks in the melting curves for PEMV-1, BWYV, and U8 pseudoknots, which correspond, in this simplified model, to the disruption of the tertiary interactions following helical unfolding structures, respectively. For the BWYV U8 mutant RNA, the C<sup>+</sup>•(G–C) base triple in the wild type is disrupted by the C8-to-U8 mutant. The melting temperature for the first unfolding transition, which is derived from the nearly coincident disruption of tertiary structural interaction as well as stem  $S_2$  (Nixon and Giedroc 2000), is  $\sim 10^\circ\text{C}$  lower for the U8 mutant than for the BWYV wild-type RNA.

To test the robustness of the extracted parameters, we examined the sensitivity of the melting temperature ( $T_m^1$ ) to the tertiary energy parameters  $(\Delta h, \Delta s)$ . As shown in Figure 5, the melting temperatures are quite sensitive to the magnitude of  $(\Delta h, \Delta s)$ , with  $T_m^1$  rapidly decreasing with a decreased enthalpy  $|\Delta h|$  or increased entropy  $|\Delta s|$ . Therefore, the fact that the single set of the fitted  $(\Delta h, \Delta s)$  values is sufficient to predict the unfolding behavior of four different RNA systems (Fig. 4) suggests that the extracted parameters in Equations 6 and 7 are robust and thus could be used for predictive purposes.

For the  $\Delta$ U177 hTR RNA, a loop–loop U7•A37 Hoogsteen base pair (see Figs. 1B; 2A) is found in the solution structure (Theimer et al. 2005); an exactly analogous U9•A27 base pair (Nixon et al. 2002a,b; Giedroc and Cornish 2009) is found in the PEMV-1 RNA structure as well. In the hTR pseudoknot, the loop–loop base pair plays a significant role in the stabilization of the pseudoknot, causing the elevation of the melting temperature for the tertiary interactions by nearly  $20^\circ\text{C}$  (Theimer et al. 2005). In the present form of the model, loop–loop contacts are not explicitly considered; as a result, our model cannot predict the U7•A37 Hoogsteen base pair. In order to account for such loop–loop contact, we simply assign an interaction energy for a conventional Watson–Crick U7–A37 base pair. Using a  $\Delta h$  of  $-8 \text{ kcal/mol}$  (Krakauer and Sturtevant 1968), the





**FIGURE 5.** The predicted melting temperature ( $T_m^1$ ) for the disruption of the tertiary contacts in the BWYV pseudoknot with different enthalpy and entropy parameters. (A,B) We fix the energy parameter for the nonprotonated base triples at (enthalpy, entropy) = (−7 kcal/mol, −19 cal/mol/K) and vary the entropy and enthalpy parameters for the protonated base triples. (C,D) We fix the energy parameter for the protonated base triples at (enthalpy, entropy) = (−14 kcal/mol, −38 cal/mol/K) and vary the entropy and enthalpy parameters for nonprotonated base triples. We find that the melting temperature for the disruption of the tertiary interactions is sensitive to the enthalpy and entropy parameters. The experimentally determined low-temperature maximum in the melting profile under these conditions is  $T_m^1 = 63^\circ\text{C}$  (Nixon and Giedroc 2000).

fitted  $\Delta s$  from the experimental melting curve (Fig. 4D) is  $-16$  cal/mol/K.

### Equilibrium unfolding pathways

#### *BWYV pseudoknot*

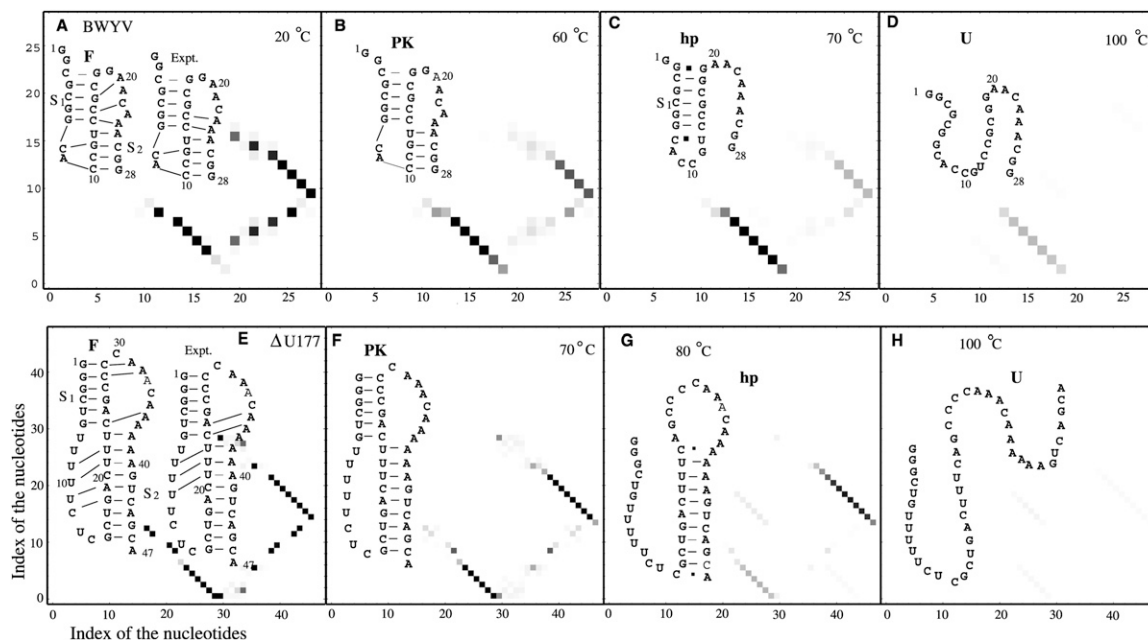
Previous experimental studies (Nixon and Giedroc 2000; Soto et al. 2007) suggested a three-step equilibrium folding pathway for the wild-type BWYV RNA: the native folded state (denoted F), to a pseudoknotted state with the tertiary interactions disrupted (PK), to the most stable component hairpin (hp or S) intermediate, to the fully unfolded state (U). To use our model to predict the folding pathways from the sequence, we first compute the partition functions (Eqs. 1 and 2) at different temperatures, from which we predict the melting curve (Eq. 5). From the melting curve, we identify the melting temperatures for each structural transition. In each different temperature regime, we predict the dominant (most probable) structure ( $s$ ) from the base-pairing probability (Eq. 4). The change of the dominant structures as temperature changes gives the predicted equilibrium unfolding pathway.

Figure 6A shows the theoretically predicted folding pathway for the wild-type BWYV pseudoknot. We find that at low temperature ( $T \sim 20^\circ\text{C}$ ), both loops are involved in base triple interactions, with the model predicting that the protonated  $\text{C}8^+ \cdot \text{G}12$  base pair contributes about  $-3.0$  kcal/mol to the overall folding stability at  $T \sim 20^\circ\text{C}$  as dictated by the fitted entropy and enthalpy parameters for a protonated base triple (see Eq. 7). We note that the fact that this value is very close to the experimentally determined  $\Delta\Delta G$  for protonation is not unexpected since the BWYV RNA was one of the RNAs in the training set. Moreover, we find that loop residues A24 and C22 form base triples with base pairs G7–C14 and G6–C15, respectively. These predicted base triples are consistent with the crystallographic structure of this RNA (Su et al. 1999). This structure reveals that C22 is flipped toward the base pairs

G6–C15 and C5–G16 and A20 also forms a base triple with G4–C17 and C5–G16. Our calculation also predicts that A20 forms a base triple only with the C5–G16 base pair. This minor difference does not alter the global stability of the pseudoknot to any measurable degree. As the temperature increases, the tertiary interactions are disrupted first. This is because breaking the tertiary interactions causes a dramatic increase in the entropy due to the change in the conformational entropy for the backbone, bases, and solvent molecules. In addition, we find that stem  $S_1$  is the most stable structural component, which is the last structural component to be disrupted. Unfolding of stem 1 results in a sharp peak at  $T \sim 90^\circ\text{C}$  in the melting curve (Fig. 4A). The three-step sequential unfolding pathway for wild-type BWYV pseudoknot is consistent with that proposed previously (Nixon and Giedroc 2000; Soto et al. 2007), despite the fact that  $\Delta h$  and  $\Delta s$  for tertiary structural contacts were optimized assuming a two-step unfolding curve for this RNA.

#### *$\Delta U177$ hTR pseudoknot*

Thermodynamic unfolding experiments show that the tertiary interactions are disrupted at  $T \sim 60^\circ\text{C}$ . NMR experiments (Theimer et al. 2005) indicate three Hoogsteen base pairs are found in the  $\text{U} \cdot (\text{U}-\text{A})$  base triple between loop  $L_1$  and the major groove of stem  $S_2$ . Our model correctly predicts that U9 and U10 form tertiary contacts with U22–A39 and U21–A40, respectively, at low temperature (see Fig. 6E–H). The previous molecular dynamics simulation failed to predict the correct tertiary interactions for the hTR pseudoknot (Yingling and Shapiro 2006). In contrast, our *Vfold* model appears to give much improved predictions for these tertiary interactions. However, the *Vfold* model cannot account for the all-atom details, relying instead on discrete rotameric states for the backbone conformation. As a result, the model cannot predict base triples with atomistic details not considered in the *Vfold* model. For instance, the model does not predict the



**FIGURE 6.** The predicted equilibrium unfolding pathways for (A–D) BWAYV and (E–H)  $\Delta$ U177 hTR RNAs. We also show the experimentally observed loop–stem base triples between the loops and stems. The results support the F to PK to hp to U folding pathway (Nixon and Giedroc 2000), where F is the folded state with the tertiary interactions, PK is the pseudoknot structure without the tertiary interactions, hp is a folding intermediate (usually a stem–loop structure), and U is the unfolded state. In the density plots, the  $x$ - and  $y$ -axes denote the indices of the nucleotides  $i$  and  $j$ , and the darkness of the points denote the base-pairing probability  $P_{ij}$  (from Eq. 4).

U8•(U21–A40) and A35•(U5–A25) base triples in the  $\Delta$ U177 hTR pseudoknot (Theimer et al. 2005).

From the temperature dependence of the ensemble of structures, we are able to predict the unfolding pathway for the  $\Delta$ U177 hTR RNA. We find that the tertiary interactions in  $\Delta$ U177 are completely disrupted as the temperature approaches 70°C (Fig. 4D). In accordance with the experimental findings, our model predicts that a shoulder at  $\sim$ 60°C in the melting curve in Figure 4D corresponds to the disruption of the tertiary interactions. Further unfolding steps involve the unfolding of stem  $S_1$ , followed by unfolding of stem  $S_2$ . The above two transitions result in two peaks at temperatures 70°C and 80°C, respectively. The resultant  $\Delta$ U177 supports the general pathway.

### Predicting loop–helix tertiary interactions

Using the *Vfold* model, we can also predict the tertiary structural interactions for RNA pseudoknots that are not in the training set. We first tested our *Vfold* model for the prediction of simple nonpseudoknot RNA conformations such as hairpins. Our test results (Supplemental Table S1) indicate that the model can accurately predict the base pairs for the tested sequences. The accuracy of the theoretical predictions is measured by two parameters: sensitivity SE and specificity SP:  $SE = N_p^{(c)}/N_e$ ,  $SP = N_p^{(c)}/N_p$ , where  $N_p^{(c)}$  is the number of correctly predicted canonical base pairs, and  $N_e$  and  $N_p$  are the total numbers of canonical

base pairs in the experimentally determined and theoretically predicted structures, respectively.

We then tested the theory for the predictions of pseudoknotted structures. Specifically, we investigated the structures, stabilities, and unfolding pathways for three additional plant viral frameshifting RNA pseudoknots from sugarcane yellow leaf virus (ScYLV) and potato leaf roll virus (PLRV), which are of known structure (Cornish et al. 2005; Pallan et al. 2005), and cucurbit aphid-borne yellows virus (CABYV) (of unknown structure) and compared those predictions for the refined structure of the PEMV-1 RNA, which is one of the RNAs in the training set.

For the ScYLV RNA, the model correctly predicts that the C8<sup>+</sup> and C27 form major- and minor-groove-derived tertiary structural contacts with G12–C28 and G7–C14 base pairs, respectively. In addition, the model predicts that nucleotides A20, A21, and A22 from  $L_2$  are capable of forming minor groove tertiary contacts with the G4–C17 base pair, while A22, A23, and A24 can form tertiary contacts with the U5–A16 base pair. Inspection of the solution structure reveals that A21 and A22 form base triple interactions with G4–C17 and U5–A16, respectively (Cornish et al. 2005). Thus, although the prediction is unable to pinpoint the precise hydrogen bonding interactions with certainty, the model predicts two  $L_2$ – $S_1$  interactions at the top of the molecule, in agreement with the solution structure (Cornish et al. 2005).

For the CABYV RNA, Figure 7B shows that we can correctly predict the pseudoknot structure only if we consider the pseudoknot stem-loop interactions since a model (Cao and Chen 2006) that ignores the tertiary structural interactions predicts the  $S_1$  hairpin rather than the pseudoknot. This finding reveals that tertiary structural interactions play a vital role for the structure and stability of the CABYV pseudoknot, as previously documented experimentally for the BWYV pseudoknot (Nixon et al. 2002a). For the refined solution structural model of the PEMV-1 pseudoknot (Giedroc and Cornish 2009), we predict that the  $C10^+$ ,  $C24$ , and  $A26$  will form tertiary structural contacts with the  $C13$ – $G28$ ,  $G7$ – $C16$ , and  $G8$ – $C15$  base pairs, respectively, each of which is found in the refined solution structure of this RNA (Giedroc and Cornish 2009).

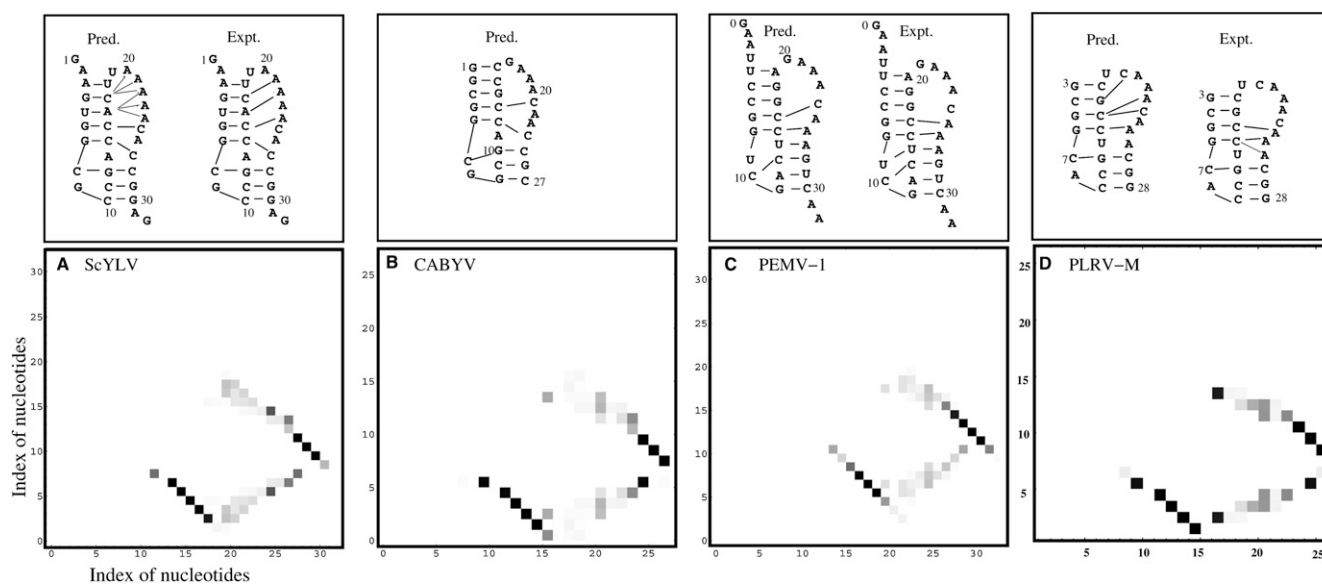
We next used our methods to predict tertiary structural loop–stem interactions in a mutant luteoviral PLRV pseudoknot, designated PLRV-M, whose crystallographic structure is known to high resolution (1.35 Å) (Pallan et al. 2005). The *Vfold* model (Fig. 7D) correctly predicts that nucleotide  $C7$  and  $A25/A24$  ( $A25$  is involved in two H bonds;  $A24$  is involved in just one) form base triple interactions with  $G11$ – $C26$  and  $G6$ – $C13$  base pairs, respectively. However, the model does not predict the hydrogen bonding interactions between  $A23$  and the  $G5$ – $C14$  base pair. In addition, our model cannot predict the detailed tertiary interaction between  $A20$  and the two base pairs  $G3$ – $C16$  and  $C4$ – $G15$  (Pallan et al. 2005) because such tertiary interactions are not considered in the theory.

Finally, we have applied our model to predict the probabilities for the formation of loop–stem tertiary in-

teractions in telomerase pseudoknots in species other than human, each of which are of unknown high-resolution structure (Ulyanov et al. 2007). Figure 8 shows our calculated total partition functions  $Q_{\text{tot}}$  for all the possible structures as well as the conditional partition function  $Q_2$  for all the structures without loop–stem tertiary interactions. A large  $Q_{\text{tot}}$  over  $Q_2$  is indicative of significantly stabilizing tertiary structural interactions between loop and stem in the RNA. From Figure 8, we conclude that there exist significant tertiary structural interactions in four of the six telomerase RNAs, these from *Colpidium colpoda*, *Tetrahymena thermophila*, *Glaucoma chattoni*, and *Euplotes aediculatus*. Similar to the human telomerase RNA pseudoknot (Theimer et al. 2005), loop–stem tertiary interactions may play important functional roles in the telomerase activity of these species (Ulyanov et al. 2007). The *Vfold* model can predict the loop–stem tertiary contacts (base triples) for the four telomerase sequences (see Fig. 9). The model predicts a conserved major-groove tertiary interaction between nucleotide A or U and a U–A base pair for the four telomerase sequences. The result is consistent with the predictions from molecular dynamics simulation (Ulyanov et al. 2007). In addition, we also predicted the minor-groove tertiary interaction between loop  $L_2$  and stem  $S_1$  in sequences *G. chattoni*, *T. thermophila*, and *E. aediculatus* (Fig. 9).

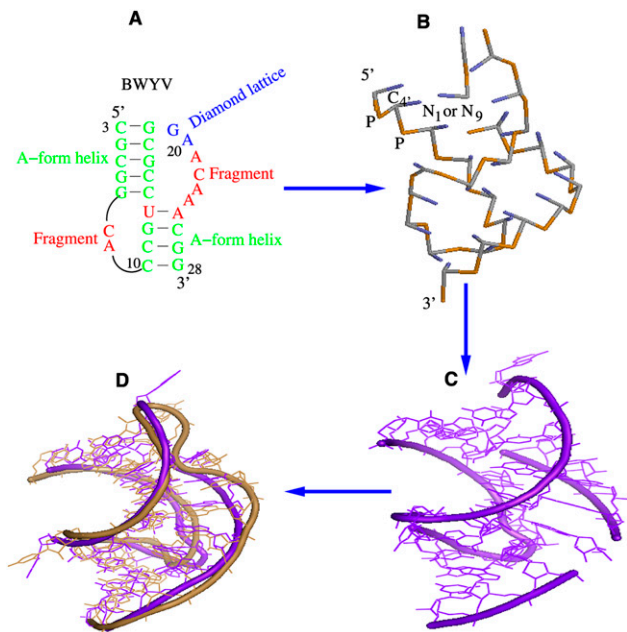
### Predicting all-atom three-dimensional structures for pseudoknots

Predicting the high-resolution 3D RNA structure from sequence is not a solved problem (Major et al. 1991; Macke



**FIGURE 7.** The predicted tertiary interactions and the density plots for the base-pairing probability for three pseudoknot-forming sequences: (A) ScYLV RNA (Cornish et al. 2005), (B) CABYV RNA, (C) PEMV-1 RNA, and (D) PLRV-M (Pallan et al. 2005) at 20°C. In C we number the PEMV-1 sequence from  $G_0$  to  $A_{32}$  in order to be consistent with the experiment (Nixon et al. 2002b; Giedroc and Cornish 2009).





**FIGURE 10.** (A) The predicted 2D structure for the BWYV pseudoknot; (B) the predicted virtual-bond level 3D structure; (C) the all-atom structure constructed from the virtual bond structure; and (D) the refined all-atom structure using Amber 9.

a two-state cooperative process or a noncooperative multistep process.

- Loop–stem tertiary contacts are known to be crucial for the stabilization of pseudoknots (Nixon et al. 2002a). Our calculations reveal that destabilizing loop–stem base triples, particularly in the context of pseudoknots with short  $S_2$  stems, may lead to complete change of the structure of the RNA, to a partially folded hairpin loop.
- We predict that there exist significant loop–stem tertiary interactions in the telomerase RNA pseudoknots of *C. colpoda*, *T. thermophila*, *G. chattoni*, and *E. aediculatus*. This implies that the tertiary structural interactions are significantly conserved and may therefore play a functional role in the activity of these telomerase RNAs (Theimer et al. 2005).

Development of the theory described here provides a starting point with which to further refine the thermodynamic parameters for the chemically different noncanonical base pairs within loop–stem base triples. Moreover, the current model assumes the additivity of the base triple interaction energies, for which there is no a priori justification. Improvement of the theory should account for the possible interferences (such as stacking) between neighboring base triples like that which has been documented to occur in many RNA pseudoknots (Cornish et al. 2005; Klein et al. 2009). Furthermore, an accurate prediction for the base triple interactions obviously requires an atomistic descrip-

tion of the structure. One plausible approach to enhance the utility of the current algorithm might be to use the *Vfold* model-predicted structure as a starting scaffold onto which one could add atomic details to further refine the structure. Finally, although we only treat the tertiary contacts in RNA pseudoknots in this study, the same methodology can be systematically extended to other more complex tertiary structures that incorporate other loop–stem base pairs, such as in the HDV pseudoknot (Ferré-D'Amaré et al. 1998), and other A-minor motifs (Nissen et al. 2001), as well as the influence of base pairing to exogenous nucleobase metabolites, like that present in the adenine, guanine, and preQ1 sensing riboswitches (Gilbert et al. 2009; Kang et al. 2009; Klein et al. 2009; Spitale et al. 2009).

## SUPPLEMENTAL MATERIAL

Supplemental material can be found at <http://www.rnajournal.org>.

## ACKNOWLEDGMENTS

We acknowledge two anonymous reviewers for careful readings and constructive suggestions. This research was supported by NIH grants GM063732 (to S.-J.C.) and AI067416 (to D.P.G.). Most of the numerical calculations were performed on the HPC resources at the University of Missouri Bioinformatics Consortium (UMBC).

Received June 26, 2009; accepted November 24, 2009.

## REFERENCES

- Batey RT, Rambo RP, Doudna JA. 1999. Tertiary motifs in RNA structure and folding. *Angew Chem Int Ed* **38**: 2326–2343.
- Burks J, Zwieb C, Mller F, Wower I, Wower J. 2005. Comparative 3-D modeling of tmRNA. *BMC Mol Biol* **6**: 14. doi: 10.1186/1471-2199-6-14.
- 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. 2006. Predicting RNA pseudoknot folding thermodynamics. *Nucleic Acids Res* **34**: 2634–2652.
- Cao S, Chen S-J. 2007. Biphasic folding kinetics of RNA pseudoknots and telomerase RNA activity. *J Mol Biol* **367**: 909–924.
- Cao S, Chen S-J. 2009. Predicting the folding thermodynamics for the pseudoknot with interstem loops. *RNA* **15**: 696–706.
- Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B, Woods RJ. 2005. The Amber biomolecular simulation programs. *J Comput Chem* **26**: 1668–1688.
- Case DA, Darden TA, Cheatham TE III, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM, Pearlman DA, Crowley M, et al. 2006. *AMBER 9*. University of California, San Francisco.
- Chauhan S, Woodson SA. 2008. Tertiary interactions determine the accuracy of RNA folding. *J Am Chem Soc* **130**: 1296–1303.
- Chen S-J. 2008. RNA folding: Conformational statistics, folding kinetics, and ion electrostatics. *Annu Rev Biophys* **37**: 197–214.
- Chen S-J, Dill KA. 2000. RNA folding energy landscapes. *Proc Natl Acad Sci* **97**: 646–651.
- Chen JL, Greider CW. 2005. Functional analysis of the pseudoknot structure in human telomerase RNA. *Proc Natl Acad Sci* **102**: 8080–8085.

- Chen G, Wen JD, Tinoco I Jr. 2007. Single-molecule mechanical unfolding and folding of a pseudoknot in human telomerase RNA. *RNA* **13**: 2175–2188.
- Comolli LR, Smirnov I, Xu L, Blackburn EH, James TL. 2002. A molecular switch underlies a human telomerase disease. *Proc Natl Acad Sci* **99**: 16998–17003.
- Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. 1995. A 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic molecules. *J Am Chem Soc* **117**: 5179–5197.
- Cornish PV, Giedroc DP. 2006. Pairwise coupling analysis of helical junction hydrogen bonding interactions in luteoviral RNA pseudoknots. *Biochemistry* **45**: 11162–11171.
- Cornish PV, Hennig M, Giedroc DP. 2005. A loop 2 cytidine-stem 1 minor groove interaction as a positive determinant for pseudoknot-stimulated –1 ribosomal frameshifting. *Proc Natl Acad Sci* **102**: 12694–12699.
- Das R, Baker D. 2007. Automated de novo prediction of native-like RNA tertiary structures. *Proc Natl Acad Sci* **104**: 14664–14669.
- Ding F, Sharma S, Chalasani P, Demidov VV, Broude NE, Dokholyan NV. 2008. Ab initio RNA folding by discrete molecular dynamics: From structure prediction to folding mechanisms. *RNA* **14**: 1164–1173.
- Draper DE. 1990. Pseudoknots and the control of protein synthesis. *Curr Opin Cell Biol* **2**: 1099–1103.
- Duarte CM, Pyle AM. 1998. Stepping through an RNA structure: A novel approach to conformational analysis. *J Mol Biol* **284**: 1465–1478.
- Ferré-D'Amaré AR, Zhou KH, Doudna JA. 1998. Crystal structure of a hepatitis delta virus ribozyme. *Nature* **395**: 567–574.
- Ferro DR, Hermans J. 1971. A different best rigid-body molecular fit routine. *Acta Crystallogr A* **33**: 345–347.
- Flory PJ. 1969. *Statistical mechanics of chain molecules*. Wiley, New York.
- Gesteland RF, Atkins JF. 1996. Recoding: Dynamic reprogramming of translation. *Annu Rev Biochem* **65**: 741–768.
- Giedroc DP, Cornish PV. 2009. Frameshifting RNA pseudoknots: Structure and mechanism. *Virus Res* **139**: 193–208.
- Giedroc DP, Theimer CA, Nixon PL. 2000. Structure, stability and function of RNA pseudoknots involved in stimulating ribosomal frameshifting. *J Mol Biol* **298**: 167–185.
- Gilbert SD, Reyes FE, Edwards AL, Batey RT. 2009. Adaptive ligand binding by the purine riboswitch in the recognition of guanine and adenine analogs. *Structure* **17**: 857–868.
- Hermann T, Patel DJ. Stitching together RNA tertiary architectures. 1999. *J. Mol. Biol.* **294**: 829–849.
- Holland JA, Hansen MR, Du ZH, Hoffman DW. 1999. An examination of coaxial stacking of helical stems in a pseudoknot motif: The gene 32 messenger RNA pseudoknot of bacteriophage T2. *RNA* **5**: 257–271.
- Hyeon C, Thirumalai D. 2008. Multiple probes are required to explore and control the rugged energy landscapes of RNA hairpins. *J Am Chem Soc* **130**: 1538–1539.
- Jonikas MA, Radmer RJ, Laederach A, Das R, Pearlman S, Herschlag D, Altman RB. 2009. Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters. *RNA* **15**: 189–199.
- Jossinet F, Westhof E. 2005. Sequence to structure (S2S): Display, manipulate, and interconnect RNA data from sequence to structure. *Bioinformatics* **21**: 3320–3321.
- Kang M, Peterson R, Feigon J. 2009. Structural insights into riboswitch control of the biosynthesis of queuosine, a modified nucleotide found in the anticodon of tRNA. *Mol Cell* **33**: 784–790.
- Kim YG, Su L, Maas S, O'Neill A, Rich A. 1999. Specific mutations in a viral RNA pseudoknot drastically change ribosomal frameshifting efficiency. *Proc Natl Acad Sci* **96**: 14234–14239.
- Klein DJ, Edwards TE, Ferré-D'Amaré AR. 2009. Cocrystal structure of a class I preQ1 riboswitch reveals a pseudoknot recognizing an essential hypermodified nucleobase. *Nat Struct Mol Biol* **16**: 343–344.
- Krakauer H, Sturtevant JM. 1968. Heats of the helix-coil transitions of the poly A–poly U complexes. *Biopolymers* **6**: 491–512.
- Leontis NB, Westhof E. 2001. Geometric nomenclature and classification of RNA base pairs. *RNA* **7**: 499–512.
- Leontis NB, Lescoute A, Westhof E. 2006. The building blocks and motifs of RNA architecture. *Curr Opin Struct Biol* **16**: 279–287.
- Macke T, Case DA. 1998. Modeling unusual nucleic acid structures. In *Molecular modeling of nucleic acids* (ed. NB Leontes, J SantaLucia Jr), pp. 379–393. American Chemical Society, Washington, DC.
- Major F, Turcotte M, Gautheret D, Lapalme G, Fillion E, Cedergren R. 1991. The combination of symbolic and numerical computation for three-dimensional modeling of RNA. *Science* **253**: 1255–1260.
- Martinez HM, Maizel JV Jr, Shapiro BA. 2008. RNA2D3D: A program for generating, viewing and comparing three-dimensional models of RNA. *J Biomol Struct Dyn* **25**: 669–683.
- Mathews DH, Turner DH. 2006. Prediction of RNA secondary structure by free energy minimization. *Curr Opin Struct Biol* **16**: 270–278.
- Michiels PJA, Versleijen AAM, Verlaan PW, Pleij CWA, Hilbers CW, Heus HA. 2001. Solution structure of the pseudoknot of SRV-1 RNA, involved in ribosomal frameshifting. *J Mol Biol* **310**: 1109–1123.
- Murray LJ, Arendall WB III, Richardson DC, Richardson JS. 2003. RNA backbone is rotameric. *Proc Natl Acad Sci* **100**: 13904–13909.
- Murthy VL, Srinivasan R, Draper DE, Rose GD. 1999. A complete conformational map for RNA. *J Mol Biol* **291**: 313–327.
- Nakano S, Bevilacqua PC. 2007. Mechanistic characterization of the HDV genomic ribozyme: A mutant of the C41 motif provides insight into the positioning and thermodynamic linkage of metal ions and protons. *Biochemistry* **46**: 3001–3012.
- Nissen P, Ippolito JA, Ban N, Moore PB, Steitz TA. 2001. RNA tertiary interactions in the large ribosomal subunit: The A-minor motif. *Proc Natl Acad Sci* **98**: 4899–4903.
- Nixon PL, Giedroc DP. 2000. Energetics of a strongly pH dependent RNA tertiary structure in a frameshifting pseudoknot. *J Mol Biol* **296**: 659–671.
- Nixon PL, Cornish PV, Suram SV, Giedroc DP. 2002a. Thermodynamic analysis of conserved loop–stem interactions in P1-P2 frameshifting RNA pseudoknots from plant Luteoviridae. *Biochemistry* **41**: 10665–10674.
- Nixon PL, Rangan A, Kim YG, Rich A, Hoffman DW, Hennig M, Giedroc DP. 2002b. Solution structure of a luteoviral P1-P2 frameshifting mRNA pseudoknot. *J Mol Biol* **322**: 621–633.
- Oliva R, Cavallo L, Tramontano A. 2006. Accurate energies of hydrogen bonded nucleic acid base pairs and triplets in tRNA tertiary interactions. *Nucleic Acids Res* **34**: 865–879.
- Olson WK. 1980. Configurational statistics of polynucleotide chains: An updated virtual bond model to treat effects of base stacking. *Macromolecules* **13**: 721–728.
- Olson WK, Flory PJ. 1972. Spatial configurations of polynucleotide chains. I, Steric interactions in polyribonucleotides: A virtual bond model. *Biopolymers* **11**: 1–23.
- Pallan PS, Marshall WS, Harp J, Jewett FC III, Wawrzak Z, Brown BA II, Rich A, Egli M. 2005. Crystal structure of a luteoviral RNA pseudoknot and model for a minimal ribosomal frameshifting motif. *Biochemistry* **44**: 11315–11322.
- Parisien M, Major F. 2008. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* **452**: 51–55.
- Pfingsten JS, Costantino DA, Kieft JS. 2007. Conservation and diversity among the three-dimensional folds of the Dicistroviridae Intergenic Region IRESes. *J Mol Biol* **370**: 856–869.
- Qiao F, Cech TR. 2008. Triple-helix structure in telomerase RNA contributes to catalysis. *Nat Struct Mol Biol* **15**: 634–640.
- Richardson JS, Schneider B, Murray LW, Kapral GJ, Immormino RM, Headd JJ, Richardson DC, Ham D, Hershkovits E, Williams LD, et al. 2008. RNA backbone: Consensus all-angle conformers and

- modular string nomenclature (an RNA Ontology Consortium contribution). *RNA* **14**: 465–481.
- SantaLucia J Jr. 1998. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc Natl Acad Sci* **95**: 1460–1465.
- Saszar K, Spackova N, Stefl R, Sponer J, Leontis NB. 2001. Molecular dynamics of the frame-shifting pseudoknot from beet western yellow virus: The role of non-Watson–Crick base-pairing, ordered hydration, cation binding and base mutations on stability and unfolding. *J Mol Biol* **313**: 1073–1091.
- Serra MJ, Turner DH. 1995. Predicting thermodynamic properties of RNA. *Methods Enzymol* **259**: 242–261.
- Shapiro BA, Yingling YG, Kasprzak W, Bindewald E. 2007. Bridging the gap in RNA structure prediction. *Curr Opin Struct Biol* **17**: 157–165.
- Shefer K, Brown Y, Gorkovoy V, Nussbaum T, Ulyanov NB, Tzfati Y. 2007. A triple helix within a pseudoknot is a conserved and essential element of telomerase RNA. *Mol Cell Biol* **27**: 2130–2143.
- Soto AM, Misra V, Draper DE. 2007. Tertiary structure of an RNA pseudoknot is stabilized by ‘diffuse’ Mg<sup>2+</sup> ions. *Biochemistry* **46**: 2973–2983.
- Spitale RC, Torelli AT, Krucinska J, Bandarian V, Wedekind JE. 2009. The structural basis for recognition of the PreQ0 metabolite by an unusually small riboswitch aptamer domain. *J Biol Chem* **284**: 11012–11016.
- Staple DW, Butcher SE. 2005. Pseudoknots: RNA structures with diverse functions. *PLoS Biol* **3**: 956–959.
- Su L, Chen LQ, Egli M, Berger JM, Rich A. 1999. Minor groove RNA triplex in the crystal structure of a ribosomal frameshifting viral pseudoknot. *Nat Struct Biol* **6**: 285–292.
- Tan Z-J, Chen S-J. 2006. Nucleic acid helix stability: Effects of salt concentration, cation valence and size, and chain length. *Biophys J* **90**: 1175–1190.
- Tan RKZ, Petrov AS, Harvey SC. 2006. YUP: A molecular simulation program for coarse-grained and multiscaled models. *J Chem Theory Comput* **2**: 529–540.
- Theimer CA, Finger LD, Trantirek L, Feigon J. 2003. Mutations linked to dyskeratosis congenita cause changes in the structural equilibrium in telomerase RNA. *Proc Natl Acad Sci* **100**: 449–454.
- Theimer CA, Blois CA, Feigon J. 2005. Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function. *Mol Cell* **17**: 671–682.
- Thirumalai D, Woodson SA. 1996. Kinetics of folding of proteins and RNA. *Acc Chem Res* **29**: 433–439.
- Ulyanov NB, Shefer K, James TL, Tzfati Y. 2007. Pseudoknot structures with conserved base triples in telomerase RNAs of ciliates. *Nucleic Acids Res* **18**: 6150–6160.
- Wadley LM, Keating KS, Duarte CM, Plye AM. 2007. Evaluating and learning from RNA pseudotorsional space: Quantitative validation of a reduced representation for RNA structure. *J Mol Biol* **372**: 942–957.
- Walberer BJ, Cheng AC, Frankel AD. 2003. Structural diversity and isomorphism of hydrogen-bonded base interactions in nucleic acids. *J Mol Biol* **327**: 767–780.
- Wang JM, Cieplak P, Kollman PA. 2000. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J Comput Chem* **21**: 1049–1074.
- Wu M, Tinoco I Jr. 1994. RNA folding energy causes secondary structure rearrangement. *Proc Natl Acad Sci* **95**: 11555–11560.
- Xin Y, Laing C, Leontis NB, Schlick T. 2008. Annotation of tertiary interactions in RNA structures reveals variations and correlations. *RNA* **14**: 2465–2477.
- Yingling YG, Shapiro BA. 2006. The prediction of the wild-type telomerase RNA pseudoknot structure and the pivotal role of the bulge in its formation. *J Mol Graph Model* **25**: 261–274.
- Zhang WB, Chen S-J. 2006. Exploring the complex folding kinetics of RNA hairpins. 1. General folding kinetics analysis. *Biophys J* **90**: 765–777.
- Zhang J, Lin M, Chen R, Wang W, Liang J. 2008. Discrete state model and accurate estimation of loop entropy of RNA secondary structures. *J Chem Phys* **128**: 125107. doi: 10.1063/1.2895050.