Computing the conformational entropy for RNA folds

Liang Liu and Shi-Jie Chen[a)]
Department of Physics and Astronomy and Department of Biochemistry, University of Missouri, Columbia, Missouri 65211, USA

(Received 18 March 2010; accepted 14 May 2010; published online 16 June 2010)

We develop a polymer physics-based method to compute the conformational entropy for RNA tertiary folds, namely, conformations consisting of multiple helices connected through (cross-linked) loops. The theory is based on a virtual bond conformational model for the nucleotide chain. A key issue in the calculation of the entropy is how to treat the excluded volume interactions. The weak excluded volume interference between the different loops leads to the decomposition of the whole structure into a number of three-body building blocks, each consisting of a loop and two helices connected to the two ends of the loop. The simple construct of the three-body system allows an accurate computation for the conformational entropy for each building block. The assembly of the building blocks gives the entropy of the whole structure. This approach enables treatment of molten globule-like folds (partially unfolded tertiary structures) for RNAs. Extensive tests against experiments and exact computer enumerations indicate that the method can give accurate results for the entropy. The method developed here provides a solid first step toward a systematic development of a theory for the entropy and free energy landscape for complex tertiary folds for RNAs and proteins. © *2010 American Institute of Physics*. [doi:10.1063/1.3447385]

I. INTRODUCTION

RNAs often fold through a hierarchical pathway, namely, RNA secondary structures (helices and loops) are formed first followed by the final consolidation of the tertiary contacts between the different secondary structural subunits. For a thermal equilibrium state, what determines the structural formation is the free energy landscape—the free energy as a function of the different structures. The free energy of the system is a net result of several competing or cooperative factors, such as base stacking interactions, conformational entropy effect, electrostatic effects, and tertiary interactions. Although these different effects are often coupled, as a first and essential step, it is useful to have a method that can evaluate the conformational entropy for a given fold.

RNA native state is a complex, highly compact tertiary structure. However, what determines the stability of the native state is also the unfolded state. Moreover, in RNA folding and RNA functions, RNAs often undergo multiple transitions between the partially unfolded intermediate states. The formation of such intermediate states can be critical for the folding of RNA. These partially unfolded intermediate states contain helix stems and loops with negligible helix-helix and loop-helix tertiary contacts. We call the structures as molten globule (MG)-like folds. Such partially disordered states have been observed for RNAs in small-angle x-ray scattering experiments.[1] The primary focus of the present study is the conformational entropy for the MG-like states.

A MG-like fold can be classified into two types according to the ways in which the different helices are linked. These two types can be conveniently defined in terms of the graphical representation of the structure [also called the "two-dimensional (2D) structure"; see Figs. 1(a) and 1(b)]. The first type contains no-crossing links and the second type contains crossing links (cross-linked loops). We call the first type of structures as the secondary structures and the second type as the tertiary structures.

For a secondary structure, due to the absence of cross-linked loops, the helix assembly can be treated as a train-track-like or tree-like structure, where the total structure can be decomposed as independent subunits connected in series. For a RNA structure, a structural subunit can be a helix stem, a hairpin loop, an internal loop, or a bulge loop. The absence of the cross-links between the different subunits causes the additivity for a secondary structure: The total conformational entropy can be estimated as an additive sum of the entropy for each subunit.

For tertiary structures (with cross-linked loops), however, the above additivity principle for the conformational entropy fails. This is because the cross-linked loops can fold back to make contact with the preceding helices, causing coupling between the loop and the preceding helices. The coupling would result in interdependence of the different structural subunits. As a result, the entropy for a subunit such as a loop would be sensitive to its structural context (structural environment). This would lead to the nonadditivity of the entropy: The total entropy of the structure is not equal to the sum of the entropy for each individual loop or helix. Because of the effects from the local structural context on the loop entropy, for the purpose of the prediction for the tertiary structural folding, it is neither possible nor useful to determine the entropy parameter for an individual loop using thermodynamic experiments. The entropy (and free energy) can only come from theoretical calculations instead of empirical parameters for any specific structures. This approach to the

[a)]Author to whom correspondence should be addressed. Electronic mail: chenshi@missouri.edu.

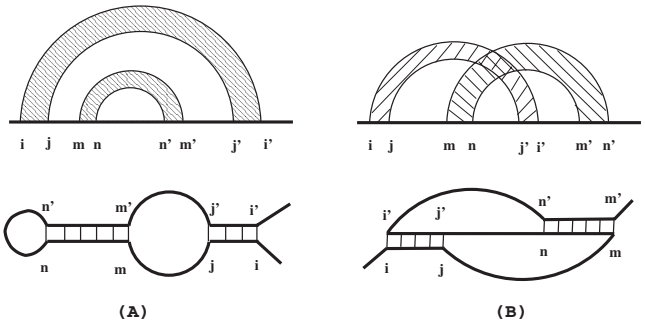

FIG. 1. The two types of MG-like fold: (a) A RNA secondary structure with no-crossing links. The figure shows a structure that contains a hairpin and an internal loop. (b) A RNA tertiary structure with crossing links. The figure shows a three-loop H-type pseudoknot.

entropy and free energy for a tertiary fold is in sharp contrast to the approach for a secondary structure, where one can rely on empirical entropy and enthalpy parameters for each individual loop or base stack to predict the free energy of the whole (secondary) structure. Physics-based RNA folding theories can be classified into two types: statistical mechanical modeling[2–9] and computer simulation modeling.[10–12] Both types of approaches have provided useful insights into the folding mechanisms from the nucleotide sequence. Previous attempts at computing the conformational entropy by building spatial models of helix stems and loops using the Monte Carlo sampling have led to promising results.[13,14] In addition, analytical methods for the entropy have been developed based on the conformational statistics for the complete conformational ensemble, including conformations with cross-linked intrachain contacts.[15,16] These analytical methods are based on perturbation expansions according to increasing structural complexity. However, application of the theories to realistic three-dimensional (3D) RNA tertiary folds has not been possible due to the complexity of the structures. In the present study, based on the complete conformation ensemble, we develop an entropy theory that can treat realistic 3D RNA tertiary folds.

For a MG-like structure with cross-linked loops, the available methods, which are based on the additivity[17] or weak nonadditivity,[2,3,16] are not applicable. The main concern of the present paper is to develop a method to treat MG-like structures that contain cross-linked loops. Because the conformational entropy is about the statistics of the 3D conformations of the chain, the evaluation of the conformational entropy is intrinsically a 3D conformational statistics problem. The basic approach here is to develop a new analytical method for counting the 3D conformations that involve strong coupling/nonadditivity.

Our approach is based on a "dividing and conquering" strategy. A MG-like fold can contain multiple helices connected by multiple loops. For a tertiary fold, the loop can be cross-linked. The conformational entropy is determined by the number of 3D conformations. The key issue in the conformational count is how to treat the volume exclusion between the different segments of the nucleotide chain. For a MG-like fold, the excluded volume interaction can be highly complex due to the complex chain connectivity by the loops. Our method for the entropy calculation is based on two key approximations. First, in the 3D space, loop-loop volume exclusion is relatively weak (and thus negligible) as compared to loop-helix and helix-helix volume exclusion. This approximation leads to the decoupling between the different loops. Second, for a given loop, the primary excluded volume and chain connectivity effect on the loop conformation come from the (two) helices that are attached to the (two ends) of the loop. The above two approximations allow for the factorization of the conformational count of the whole structure into the product of the conformational count of structural building blocks,[15] each consisting of a loop and the two helices that are connected to the loop (a *three-body system*). An attractive feature of the method is that because the system for each building block is simple and manageable, one can focus the computational resources on each building block to calculate their conformational entropies in great detail, including the volume exclusion between the loop and its nearby helices, and evaluate the (loop) entropy for the whole structure efficiently as a sum of the entropies of the building blocks (loops).

II. FREE ENERGY MODEL FOR MG-LIKE STRUCTURES

A MG-like state (structure) is defined by the base pairs [=graph; see Figs. 1(a) and 1(b)]. Equivalently speaking, a MG-like state is defined by the helices (formed by the base pairs) and the loops (=unpaired regions). In a MG-like state, we assume that helix-helix, helix-loop, or loop-loop contacts have not yet formed. The full free energy landscape of the chain is determined by the ensemble average over all the possible structures,

$$\Delta G_{\mathrm{MG}} = -k_{\mathrm{B}}T \ln \sum_{\mathrm{structure}} e^{-\Delta G/k_{\mathrm{B}}T}, \quad \Delta G = \Delta G_{\mathrm{helix}} + \Delta G_{\mathrm{loop}},$$

where ΔG is the free energy for a given structure. The sum over all the possible structures ($\Sigma_{\mathrm{structure}}$) includes all the MG-like structures for a given chain. Without loop-helix, loop-loop, and helix-helix tertiary contacts, the free energy of a MG-like fold can be parsed into the free energies for the helices $\Delta G_{\mathrm{helix}}$ and for the loop ΔG_{loop}. In the absence of helix-helix interactions, the free energy of the helices $\Delta G_{\mathrm{helix}}$ can be calculated as the sum of the free energy for each helix, which is given by the nearest neighbor model combined with the empirical thermodynamic parameters (Turner rules) of base stacking and pairing.[18] $\Delta G_{\mathrm{helix}}$ is sequence dependent. The free energy for the loop regions is mainly entropic: $\Delta G_{\mathrm{loop}} = -T\Delta S_{\mathrm{loops}}$, where $\Delta S_{\mathrm{loops}}$ is the change in the conformational entropy of the loop segments as the chain folds from the fully unfolded state to the MG-like state. The computation for the entropy $\Delta S_{\mathrm{loops}}$ for the loop regions requires a model.

A MG-like state is generally disordered and corresponds to a large ensemble of 3D conformations. For a given MG-like fold, the configurations for the multiple helices are not fixed. Because helices are tethered by the intervening loops of given lengths and a helix cannot bump into another helix or a loop, helix configurations can be quite restricted. Different configurations of the helices lead to different loop con-

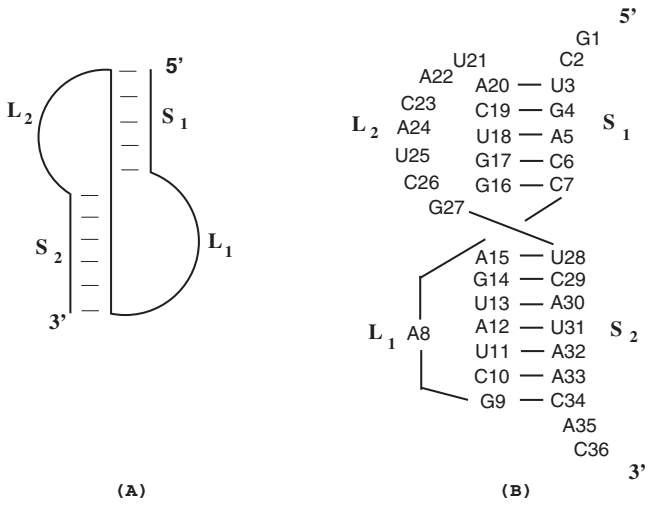

FIG. 2. (a) A schematic for a simplest (two-loop) H-type RNA pseudoknot, consisting of two helix stems (S1 and S2) and two crossing links (L1 and L2). (b) The secondary structure for the 32 mRNA pseudoknot of bacteriophage T2 (PDB ID: 2TPK), which is used as a template for the helix structures for a simple H-type pseudoknot.

formational entropies. How do we evaluate the number of all the viable 3D conformations (conformational entropy) for a given MG-like state?

The entropy for a given MG structure comes from two parts: (a) the ensemble of the 3D configurations (positions/orientations) of the helices ($\Sigma_{\text{helix config}}$) and (b) the ensemble of loop conformations for each given 3D configuration of the helices. We first focus on the computation of the loop entropy for a given configuration of the helices. The combination of the above two conformational ensembles gives the following loop entropy:

$$\Delta S_{\text{loop}} = k_{\text{B}} \ln \sum_{\text{helix config}} \Omega_{\text{loop}}, \tag{1}$$

where Ω_{loop} is the number of loop conformations for a given 3D configuration of the helices.

Our focus is on the MG-like tertiary structures, namely, structures with cross-linked loops. Loops in RNA secondary structures,[19] which contain no cross-linked loops, are connected to the helices in series. Loops in a tertiary MG-like fold, which contains cross-linked loops, however, often span across helices. This causes the strong interdependence between loop and helix conformations. Such conformational interdependence is unique to tertiary fold. The loop-helix interdependence comes from two effects. First, the end-to-end distance for a loop conformation is set by the configuration of the two helices that are connected to the loop. Second, the loop and helix conformations cannot bump into each other in the 3D space.

The key issue surrounding the entropy calculation for cross-linked loops is the structural dependence between a loop and the surrounding helices. To develop an intuitive grasp of our strategy, it is instructive to consider the simplest cross-linked loops, which occur in a simplest two-loop H-type pseudoknot (Fig. 2). A pseudoknot consists of two helix stems (S_1 and S_2) and two cross-linked loops (L_1 and L_2). The two helix stems in a pseudoknot tend to coaxially stack on each other to form a rigid continuous helix rod. Due to the helix-loop volume exclusion and the loop end-to-end distance set by the helices, the flexibility of a loop can be sensitive to the 3D configurations of the helices. Therefore, given the presence of the helices, the loop entropy is effectively a conditional entropy.

Polymer physics-based methods have been proposed for the calculation of the pseudoknot loop entropy.[4,20–27] These methods, especially the method based on the explicit conformation enumeration,[23] have led to several useful predictions for pseudoknot structures and folding stabilities. For a simple canonical H-type pseudoknot (Fig. 2), there is only one fixed 3D configuration (=a quasicontinuous long helix) for the helices. The fixed helix configuration makes it computationally feasible to produce an entropy table for different helix and loop lengths using exhaustive computer enumeration. However, for a MG-like fold that contains multiple helices whose orientations can be flexible, it would be practically impossible to enumerate all the possible 3D configurations of the helices and tabulate the loop entropy for each helix configuration. What we need here is a new computational method that can give efficient evaluation of the (loop) entropy for any given 3D configuration of the helices.

A. Virtual bond-based RNA conformational model

Motivated by the need to sample the complete set of the chain conformations for nucleic acids while keeping the computational efficiency, Cao and Chen[4] developed a virtual bond-based RNA folding model (called the "Vfold" model). The development of the Vfold model is based on two observations about the known RNA structures. First, the C–O torsions in the nucleotide backbone of RNA tend to be in the *trans* (t) rotational isometric state [Fig. 3(a)]. Consequently, both the $P-O_5-C_5-C_4$ bonds and the $C_4-C_3-O_3-P$ bonds in the nucleotide backbone are approximately planar.[28] This makes it possible to reduce the nucleotide backbone conformations into two effective virtual bonds $P-C_4$ and C_4-P.[29–31] Our calculation shows that the length of each backbone virtual bond is about 3.9 Å. Second, RNA backbones and virtual bonds are rotameric. The survey of the known structures suggests that the virtual bonds [Figs. 3(a) and 3(b)] adopt rotamerlike configurations.[32–37] Therefore, RNA backbone conformations can be approximately described by discrete rotational isometric states of the virtual bonds. For the backbone virtual bonds connecting the C_4 and P atoms, the pseudotorsional angles [θ and η in Figs. 3(a) and 3(b)] can adopt the conventional isomeric states [*gauche*$^+$ (g^+), *trans* (t), and *gauche*$^-$ (g^-)].[38] Because these three rotational isomeric states can be represented by the diamond lattice bonds, we can generate loop conformations through self-avoiding random walks of the virtual bonds in a diamond lattice[4] with bond length of 3.9 Å. In addition, the bond angle in a diamond lattice is fixed as 109.5° in diamond lattice,[4,14] which lies well within the bond angle range of 90°–120° of virtual bonds (β_{C} and β_{P} in Fig. 3) in the experimentally measured RNA structures. The virtual bond RNA model (Vfold model) provides an effective scaffold for the all-atom RNA conformations as described below.

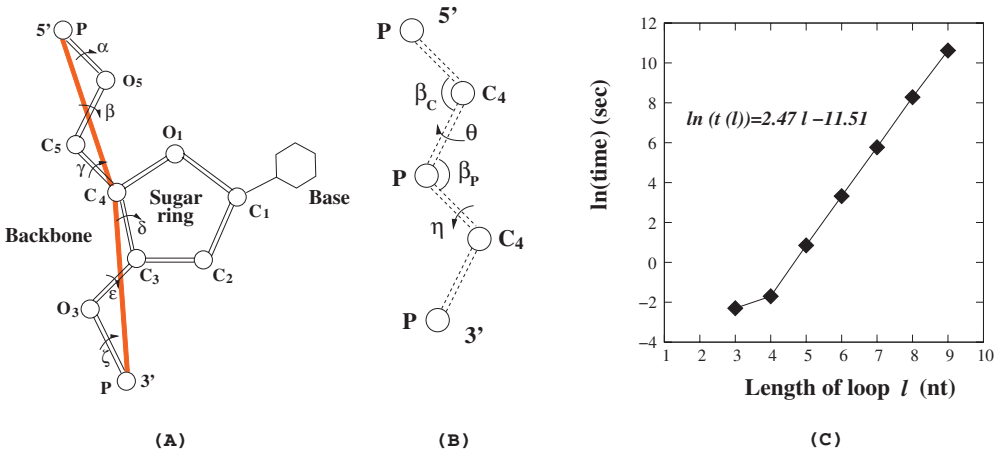

FIG. 3. (a) The virtual bond scheme for RNA nucleotides. (b) The bond angles (β_C,β_P) and the (pseudo)torsional angles (θ, η). (c) The computer time for the exhaustive conformational enumeration for a single-stranded chain. The *y*-axis represents the logarithm of the time ($t(l)$) and the *x*-axis is the chain length of loop. The computer time increases with the length of loop exponentially: $\ln t(l)=2.47l-11.51$ s. The computations are performed on a P4 3.06 GHz Intel processor PC with 1.5 Gbytes main memory.

1. Helix

We use the atomic coordinates for the A-form helix to configure the virtual bond structure of a helix. To use the diamond lattice to configure the all-atom coordinates, we move the P and C_4 atoms in an A-form RNA helix to the nearest diamond lattice sites.[4] The positions of these atoms determine the configurations of the virtual bonds in the helix. Such off-lattice to on-lattice transformation causes a small distortion of the helix. We found that for an A-form RNA helix, the use of the diamond lattice would cause a root-mean-square deviation (RMSD) of about 0.4 Å for a base pair.

2. Loop

The loop part is often flexible and can form an ensemble of conformations in the 3D space. Several previous models have been developed to describe the loop conformations using the Cartesian coordinate system,[39] torsion-angle,[40] pseudotorsion angle,[32,33] or distance representations.[41] These different methods employ various simplified chain representations in order to reduce the conformational degrees of freedom. We use our virtual bond scheme because it can conveniently generate the conformations through the pseudotorsion angles.[4,29–31] With the virtual bond RNA conformational model, we construct the loop conformational ensemble through self-avoiding random walks of the virtual bonds in the diamond lattice. To generate loop conformations using the virtual bond model is computationally more efficient than other available models. For instance, the virtual bond model gives $2\times3=6$ torsional angles for a trinucleotide system. This is in contrast to the 54, 18, and 153 variables used in the Cartesian coordinate-based model, the torsion-angle representation, and the distance representation, respectively.[41]

3. Loop-helix connection

The helices in a MG-like fold are connected by intervening loops. In order to ensure the loop-helix connectivity, we allow a loop and a helix connected to the loop to have a overlap nucleotide (see Fig. 4). The overlap nucleotide is a terminal nucleotide of the helix. For a helix, the configuration of the terminal (overlap) nucleotide uniquely determines the configuration of the pairing nucleotide and hence the configuration of the helix. Therefore, the ensemble of helix configurations can be effectively generated from the configurations of the corresponding overlap nucleotide (=loop terminal).

In practice, for a given configuration of the loop terminal nucleotide (=the overlap nucleotide), we perform 3D rotation of the (all-atom) helix to identify the optimal helix configuration with the minimum RMSD. Here the RMSD is the deviation between the diamond lattice coordinates and the off-lattice atomic coordinates of the overlap nucleotide. Our test shows that the minimum RMSD is around 0.4 Å (see Fig. 4).

The Vfold model allows us to sample the conformational space with reduced chain representation. The model has two notable advantages. First, unlike the simulation approaches, the Vfold model enables statistical mechanical calculations based on the complete ensemble of the reduced conformations.[4,9,23,27] Second, it provides a reliable low-resolution structural model, which can serve as a scaffold for further all-atom refinement through molecular dynamics computations.[42]

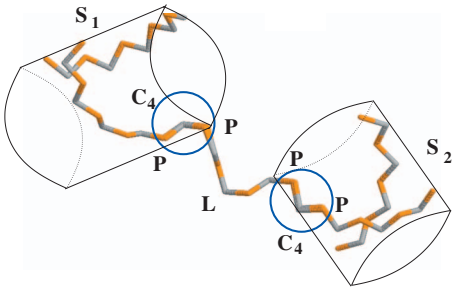

FIG. 4. The connection between a loop ($L=3$ nt) and two helices ($S_1=5$ bp and $S_2=4$ bp). The two helices are connected to the two ends of an intervening loop through the overlap terminal nucleotides (shown in the circles). The minimal RMSD for the overlap atoms ($P-C_4-P$) is $\simeq0.4$ Å (each nucleotide). The small RMSD ensures the uniqueness of the helix configuration for a given loop conformation.

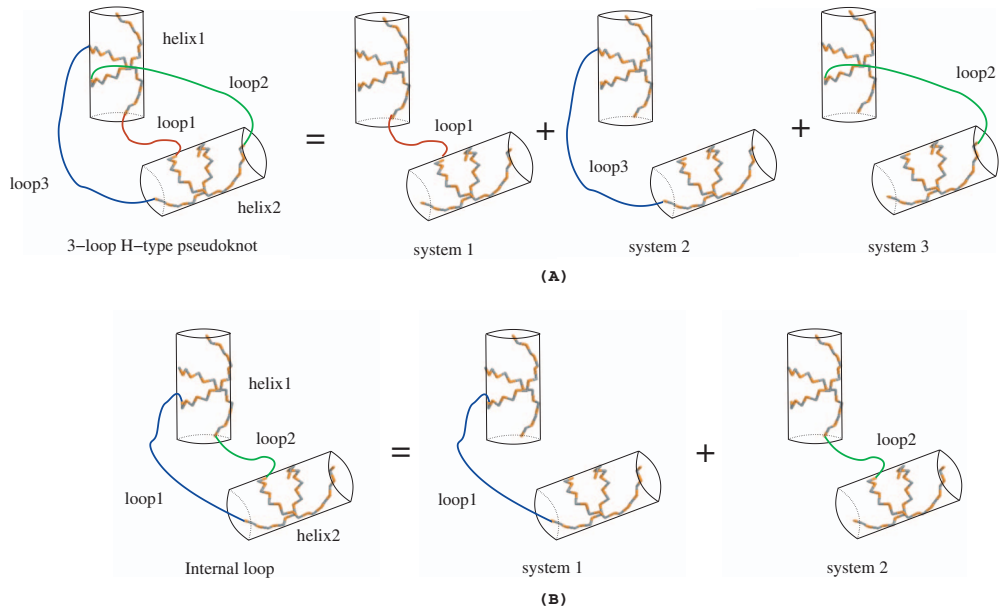

FIG. 5. The calculation of the loop entropy for a given helix configuration can be reduced into the calculation of a set of three-body systems. (a) shows a three-loop H-type pseudoknot and (b) shows an internal loop. The strategy [Eq. (2)] is based on the observation that the loop-loop excluded volume clash is weak and the impact of helices on the conformational viability of a loop predominantly comes from the helices that are directly tethered to the two ends of the loop.

B. Three-body model for loop entropy

One of the key issues in the computation of the free energy for a MG-like state is the loop entropy. For a MG-like fold with a given set of helix configuration, the conformational entropy corresponds to the size of the loop conformational ensemble [Ω_{loop} in Eq. (1)]. Brute force enumeration of the loop conformations is not viable, not only because of the exceedingly long computational time to enumerate the conformational for even a simple loop [Fig. 3(c)] but also because of the constraints from the environment (helices, loops). The problem is more pronounced for a MG-like fold with multiple loops and multiple helices, which makes exact computer enumeration not viable.

The conformational viability of a loop is restricted by two factors: (a) the loop-helix and loop-loop excluded volume interactions and (b) the fixed coordinates of the loop terminal nucleotides as determined by the helix configuration. The difficulty for the evaluation of Ω_{loop} in Eq. (1) comes from the fact that the conformational viability of a loop is the spatial constraints, which mainly come from the nearby helices and loops. Therefore, the problem in essence is a many-body problem.

However, for a MG-like structure, the excluded volume clash between the different loops is weak. For a MG-like fold which is not tightly compact, the different loops can have many different ways to be configured in the 3D space, resulting in a weak volume exclusion between each other. This is in contrast to 2D loops which have much limited ways to be configured in the (2D) space and hence loops can easily bump into each. The weak volume exclusion correlation between the loops causes the decorrelation between the different loops. This is remarkable because it leads to the factorization of Ω_{loop}, namely, the loop conformational count Ω_{loop} can be evaluated as the product of the conformational count ω_i of each loop i,

$$\Omega_{\text{loop}} = \Pi_i \omega_i. \tag{2}$$

This approach reduces the conformation count for multiple loops into the conformational enumeration for one loop at a time, resulting in a dramatic reduction in the computer time from $t(\Sigma_n L_n)$ to $\Sigma_n t(L_n)$, where L_n is the length of the nth loop and $t(L)$ is the computer time for counting conformations for a loop of length L [Fig. 3(c)].

In addition, for a MG-like structure, which is not tightly compact, the impact of helices on the conformational viability of a loop is predominantly dictated by the helices that are directly tethered to the two ends of the loop. The observation is important because it suggests that for each loop, the complex problem of the conformational enumeration [ω_i in Eq. (2)] for a loop i in a multiple-helix and multiple-loop environment can be reduced to a three-body system: the loop and the two helices directly connected to the two ends of the loop (Fig. 5).

Based on the three-body model, our basic strategy here is to evaluate the conformational count ω_i for each loop (junction) in the context of the three-body system and to compute the entropy Ω_{loop} from Eq. (2). The major steps to compute the conformational count for the three-body system can be summarized as the following [Fig. 6(a)].

(1) We first model the conformations of the loop by self-avoiding random walks of the virtual bonds (the P and C_4 atoms) in a diamond lattice. In this step, the loop is treated as a free chain.
(2) For each enumerated conformation, we construct the three-body system by attaching the two A-form helix stems to the two ends of the loop, respectively, using the loop-helix connection method described above.
(3) We map the two A-form helix stems onto the diamond lattice. We examine the excluded volume effect be-

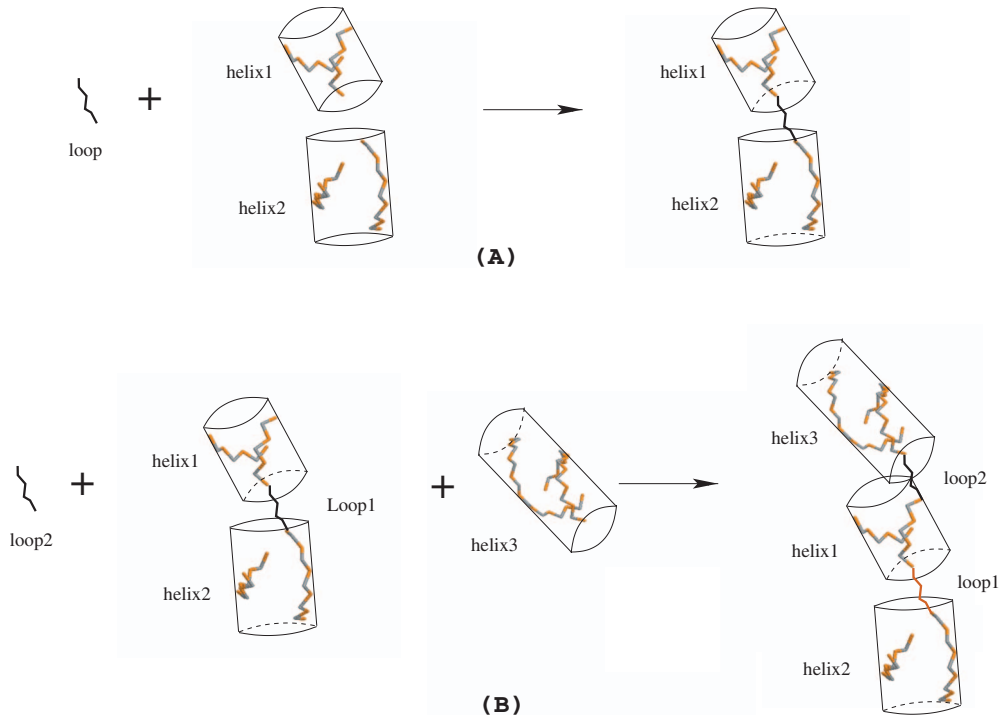

FIG. 6. (a) The strategy to compute the conformational count for a three-body system: (a) enumerating the loop conformations, (b) attaching the two helices to the two ends of the loop with the $5' \rightarrow 3'$ directionality preserved, (c) examining the excluded volume effect. By classifying the conformations according to the conformational types (see Fig. 7), we compute and tabulate the conformational count of the loop for the different types of helix configurations. (b) By repeating the above procedures [(a)–(c)] for the next loop, we configure the next helix (helix 3).

tween the loop and helix by disallowing conformations with two atoms occupying the same lattice site.

To exhaustively compute and tabulate the entropy of three-body system would be prohibited by the rapid increase of the required computational time [Fig. 3(c)]. To enhance the computational efficiency, we choose a set of reduced parameters to describe the 3D configurations of the two helices and the loop. The purpose here is to represent the loop conformational count ω_i as a function of the reduced parameters.

The configurations of the two helices can be conveniently described by (a) the end-to-end distance (x) and (b) the (relative) orientation of the two end nucleotides of the loop. How many different (relative) orientations can the two end nucleotides have? In the Vfold model, each terminal nucleotide of the loop is represented by two virtual bonds. In a diamond lattice, each bond can have four possible orientations. Therefore, for a nucleotide (=2 virtual bonds) there are $\binom{4}{2}=6$ possible configurations. Considering the $5 \rightarrow 3$ polarity, there are a total of $2 \times 6 = 12$ relative orientations between the two terminal nucleotides of a loop. Because the orientation of a terminal nucleotide (of a helix) uniquely determines the orientation of the helix, these 12 orientations correspond to 12 relative orientations of the two helices. Therefore, according to the relative orientations of the helices, we can classify the conformational entropy for a loop (in the three-body context) into 12 different types (Fig. 7). For each type t, we denote the number of loop conformations as $\omega(l,t,x)$, where l is the loop length and x is the end-to-end (P-P) distance [Fig. 8(a)].

By using the above classification scheme, we assume that other factors, such as the lengths of the helix stems, may

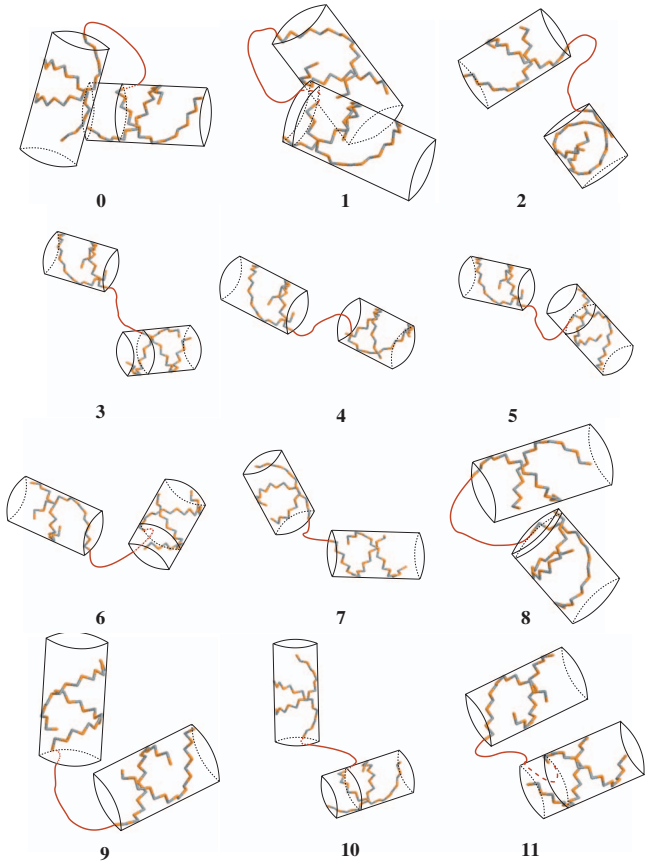

FIG. 7. The 12 types of the configurations of the three-body system classified according to the relative orientations of any two helices. In the figures, the cylinders represent the orientations of the helices.

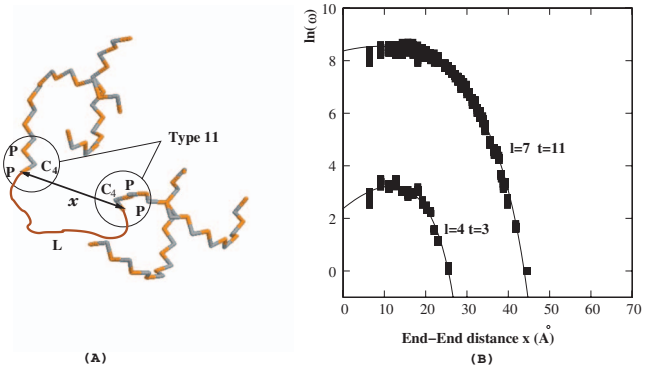

FIG. 8. (a) The conformation entropy of a three-body system is sensitive to the chain length (l), end-end (P-P) distance (x), and connection types (t). (b) Based on the loop conformations from the exact computer enumerations (■), the parameters in Eq. (3) (—) can be fitted. The helix lengths (S_1 and S_2) in the two cases are in the range of 2–9 bp and the loop lengths (l) are 4 and 7 nt. The connection types of the helices are $t=3$ and $t=7$, respectively. This figure also shows that the conformation entropy of a three-body system is sensitive to (l,t,x) and is only weakly dependent on the lengths of the helix stems.

not play a dominant role in determining the loop entropy. Indeed, the assumption is supported by our test results. As shown in Fig. 8(b) for test systems with helix stems of 2–9 bp (base pairs) and a loop of 4 or 7 nt (nucleotides), we find that the conformation entropy of a three-body system is sensitive to (l,t,x) and is only weakly dependent on the lengths of the helix stems.

Furthermore, for longer loops ($l \geq 9$ nt), it becomes possible and necessary to develop an analytical formula for $\omega(l,t,x)$ in order to avoid exact computer enumeration which is computationally demanding. Through extrapolation from the results for the shorter loop lengths [Fig. 8(b)], we find the following fitted formula for $\omega(l,t,x)$:

$$\ln \omega_{\text{system}}(l,x,t) = (0.1 + 2.14l) - \left[A(l,t)\ln\left(l - \frac{x}{d} + 1\right) + B(l,t)\left(l - \frac{x}{d} + 1\right) + C(l,t) \right], \quad (3)$$

where $d=6.4$ Å is the length of a nucleotide (=the distance between two consecutive P atoms). In the diamond lattice, $A(l,t)$, $B(l,t)$, and $C(l,t)$ are functions of the loop length (l) and type (t),

$$A(l,t) = a_{1t} l^{1/4} + a_{2t}, \quad B(l,t) = \frac{b_{1t}}{l} + b_{2t}, \quad C(l,t) = c_{1t} l + c_{2t}. \quad (4)$$

In Table I we show the data for a_{1t}, a_{2t}, b_{1t}, b_{2t}, c_{1t}, and c_{2t}. Note that a_{1t}, a_{2t}, b_{1t}, b_{2t}, c_{1t}, and c_{2t} are functions of the orientation type t only.

For short loops with $l=1$ and 2 nt, we compute the loop entropy from the exhaustive enumeration for the conformations of the three-body system. For loops with $l=3$ and 4 nt, we compute $\omega_{\text{system}}(l,x,t)$ from the analytical formula Eq. (3) with the parameters given by Table II. For longer loops with $l \geq 5$ nt, we compute the loop entropy from the analytical formula Eqs. (3) and (4) with the parameters from Table I.

III. ILLUSTRATIVE CALCULATIONS

To test the validity and to illustrate the application of the above approach, we compute the conformational entropy for several paradigmatic systems with increasing complexity. The theoretical predictions are compared to the results from the exact computer enumerations and from the thermodynamic experiments.

The entropy ΔS_{loop} in Eq. (1) involves the enumeration of the 3D configurations of the three helices. We generate the helix configurations using the terminal nucleotides of the loops. We first compute and tabulate the possible end-to-end coordinates/configurations for different loop lengths. For any given MG-like fold, we start from the shortest loop and enumerate all the possible end-end configurations of the loop from the pretabulated results. Each given end-end configuration of the loop fixes the configuration of the helices attached to the loop. Repeating the same procedure for the next loop determines the configuration of the next helix [Fig. 6(b)]. Here, the $5' \rightarrow 3'$ polarity of the chain for both the loop and the helices must be preserved.

TABLE I. Parameters in Eq. (4) (for loop length ≥5 nt).

Type	a_{1t}	a_{2t}	b_{1t}	b_{2t}	c_{1t}	c_{2t}
0	−16.6169	18.3553	2.583 48	0.815 093	2.603 00	−3.273 45
1	−16.3055	17.8392	2.766 86	0.785 653	2.603 55	−3.267 95
2	−14.9979	15.4767	4.208 68	0.642 991	2.639 34	−3.552 73
3	−15.8370	16.4620	4.409 97	0.780 402	2.676 46	−4.033 36
4	−12.9740	11.6886	6.891 27	0.429 394	2.696 58	−4.057 99
5	−14.3778	14.0218	5.797 66	0.582 219	2.695 38	−4.090 29
6	−17.0547	19.0807	1.603 46	0.887 777	2.595 57	−2.990 63
7	−17.5967	20.0118	1.239 63	0.934 679	2.600 26	−3.070 64
8	−17.1778	19.4606	1.247 04	0.894 766	2.572 40	−2.830 05
9	−15.2644	15.8595	4.300 44	0.661 839	2.652 71	−3.700 54
10	−14.3481	14.1437	5.286 88	0.592 249	2.667 38	−3.841 77
11	−15.8476	16.5521	4.239 53	0.761 426	2.669 99	−3.886 06

TABLE II. Parameters in Eq. (3) (loop lengths are 3 and 4 nt).

Type	$A(3,t)$	$B(3,t)$	$C(3,t)$	$A(4,t)$	$B(4,t)$	$C(4,t)$
0	−4.205 74	1.630 16	4.785 98	−6.156 70	1.788 47	7.021 03
1	−5.070 35	2.076 93	4.370 69	−6.112 98	1.722 77	7.094 49
2	−6.432 25	2.836 52	3.648 89	−6.735 80	1.978 77	6.896 76
3	−5.676 48	2.534 53	3.742 06	−6.801 54	2.114 36	6.579 37
4	−6.834 93	3.122 32	3.328 32	−7.536 42	2.393 85	6.487 51
5	−6.650 30	3.003 56	3.442 79	−7.389 51	2.341 60	6.512 66
6	−5.248 81	2.077 86	4.445 22	−5.772 94	1.459 01	7.430 10
7	−5.328 70	2.136 72	4.393 34	−5.563 47	1.371 52	7.480 03
8	−5.779 96	2.485 09	4.045 68	−5.494 94	1.340 87	7.541 47
9	−6.532 14	2.933 80	3.525 82	−6.627 58	1.950 42	6.885 91
10	−6.773 07	3.048 56	3.422 80	−6.953 25	2.095 50	6.751 72
11	−6.267 85	2.796 60	3.661 63	−6.751 58	2.028 76	6.789 05

A. Hairpin loop

A RNA hairpin consists of a helix and a loop connected to the two strands of the helix [Fig. 1(a)]. From a structural point of view, hairpin loop should be excluded from our three-body model, as it cannot be factored into two helices tethered to a loop. Nevertheless, the configuration of the loop (helix) terminal nucleotides (=the overlap nucleotides in the loop-helix junctions) falls into the eighth type ($t=8$) of the 12 types of the relative orientations between two nucleotides (Fig. 7). By fitting the helix stem onto the diamond lattice, we find that the end-to-end (P-P) distance of the loop (helix strands) in the diamond lattice is equal to $x=16.9$ Å.[43] From Eq. (3), we can compute the loop conformational entropy as a function of the loop length l. As shown in Fig. 9(a), the theoretical predictions show good agreements with the experimental data[18] and with the results from the exact computer enumerations. The agreement suggests the validity of our three-body model for the loop entropy. In our calculation, the conformational entropy of the coil state (Ω_{coil}) is computed from the exact enumeration of the virtual bond conformations in a diamond lattice.

B. Two-loop H-type pseudoknot

A two-loop H-type pseudoknot contains two helix stems that are coaxially stacked to form a quasicontinuous helix (S_1 and S_2 in Fig. 2). For the fixed helix configuration, we can compute the loop entropy based on the three-body system. We first generate the P and C_4 coordinates for the coaxially stacked helices using a template structure. Specifically, we use gene 32 mRNA of bacteriophage T2 pseudoknot (PDB ID: 2TPK, Fig. 2) as a template. Our tests for choosing different templates to model the coaxially stacked helices show that the loop entropy results are not sensitive to the template (data not shown).

By reducing the all-atom structure of the coaxially stacked helices into a virtual bond structure and mapping the virtual bond structure onto a diamond lattice, we can determine the connection types (t) and the end-to-end distance (x) of the loops. The connection type (t) and end-end distance (x) are dependent on the helix lengths (Table III) and are uniquely determined for a given pseudoknot. For the given connection type (t) and the end-end distance (x), Eqs. (3) and (4) give the loop entropy for loops L_1 and L_2.

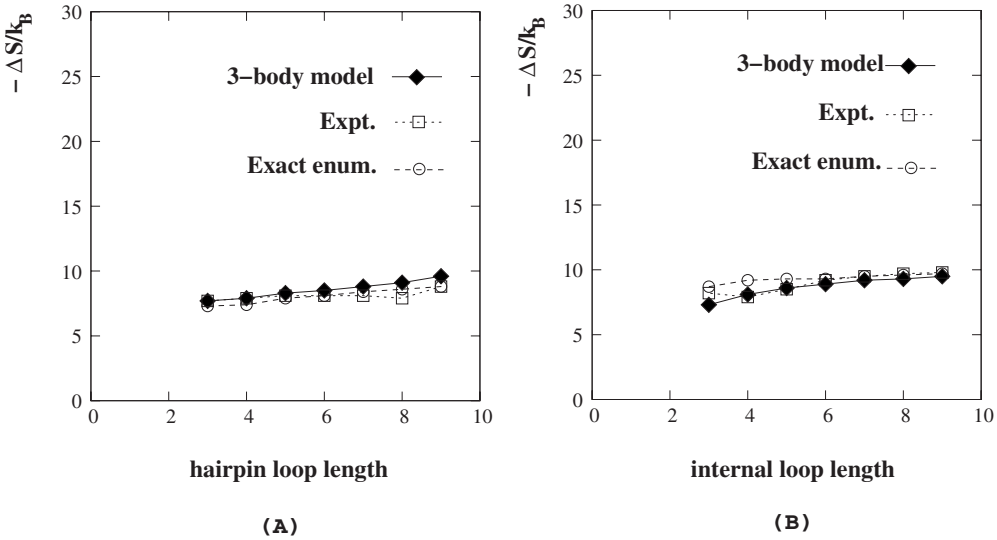

FIG. 9. Comparison between the loop entropies from the experiment (□), the exact enumerations (○), and our analytical theory (♦) for (a) hairpin loops and (b) internal loops. The experimental results are from Ref. 18. $-\Delta S$ is computed as $k_B \ln(\Omega_{\text{coil}}/\Omega_{\text{loop}})$, where Ω_{coil} and Ω_{loop} are the numbers of the conformations for the coil and the loop, respectively.

TABLE III. The connection type (t) and end-to-end distance (d) for a two-loop H-type pseudoknot with the different helix lengths (S_1 and S_2).

S_1	t	d	S_2	t	d
2	10	16.9	2	11	19.1
3	10	21.1	3	9	12.7
4	10	24.7	4	9	12.7
5	9	26.3	5	9	6.4
6	9	33.1	6	10	6.4

Figure 10 shows the calculated conformational entropy of the two-loop H-type pseudoknot and the comparisons with the results from exact computer enumerations. Our results based on the three-body system yield better results for loop L_1, which spans across the narrow (major) groove S_2 than for loop L_2 which spans across the wide (minor) groove S_1. The difference may be due to the more severe/complex excluded volume interactions between L_2 and S_1 than L_1 and S_2. Nevertheless, our analytical formula leads to reliable overall predictions for the entropies.

C. Three-helix model system

An important assumption for our three-body approach is that the loop-helix excluded volume mainly comes from the helices that are connected to the loop. To test this assumption, we choose a model system with three helices ($S_1=S_2=S_3=3$ bp; see Fig. 11) to calculate the loop entropy. We compare the loop entropy predicted from the three-body analytical theory [Eq. (1)] with the result from the exact computer enumeration, which considers the volume exclusion between the loop and all the three helices.

We randomly select two configurations for the three helices, one with connection type $t=3$ and the other with type $t=5$ for the loop. As shown in Fig. 11, the analytical results given by Eqs. (3) and (4) agree with the results from the exact computer enumerations. This suggests that the excluded volume effect on the loop conformations may predominantly come from the helices (S_1 and S_3) that connect to the loop (l) instead of from the other helices (S_2).

D. Internal loop

An internal loop consists of two helix stems and two (unpaired) strands [Fig. 1(a)]. A recent attempt[4] based on the Vfold model enables evaluation of the entropy for internal loops through exhaustive computer enumeration of the conformations. However, as the two strands in the loop can be flexible, the exact enumeration of the conformations can be computationally time demanding [Fig. 3(c)]. For example,

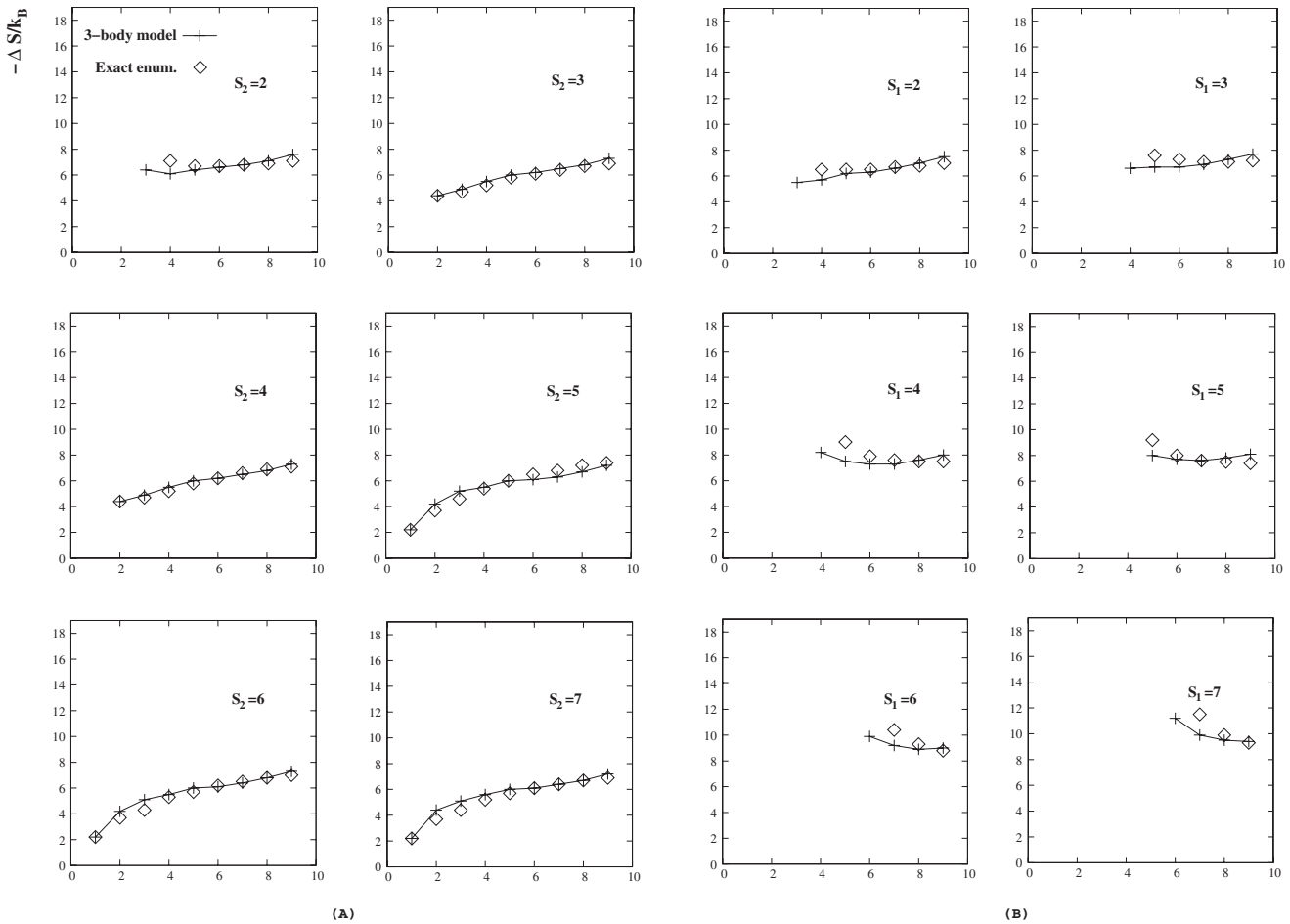

FIG. 10. Comparison between the loop conformational entropies from exact enumeration (◇) and from our three-body model (+) for a two-loop H-type pseudoknot: (a) loop L_1 across the narrow (major) groove of S_2 and (b) loop L_2 across shallow (minor) groove of S_1. The P and C_4 atomic coordinates for the coaxially stacking helices are from a template H-type pseudoknot (ID: 2TPK) in the PDB database (Fig. 2) (Ref. 39).

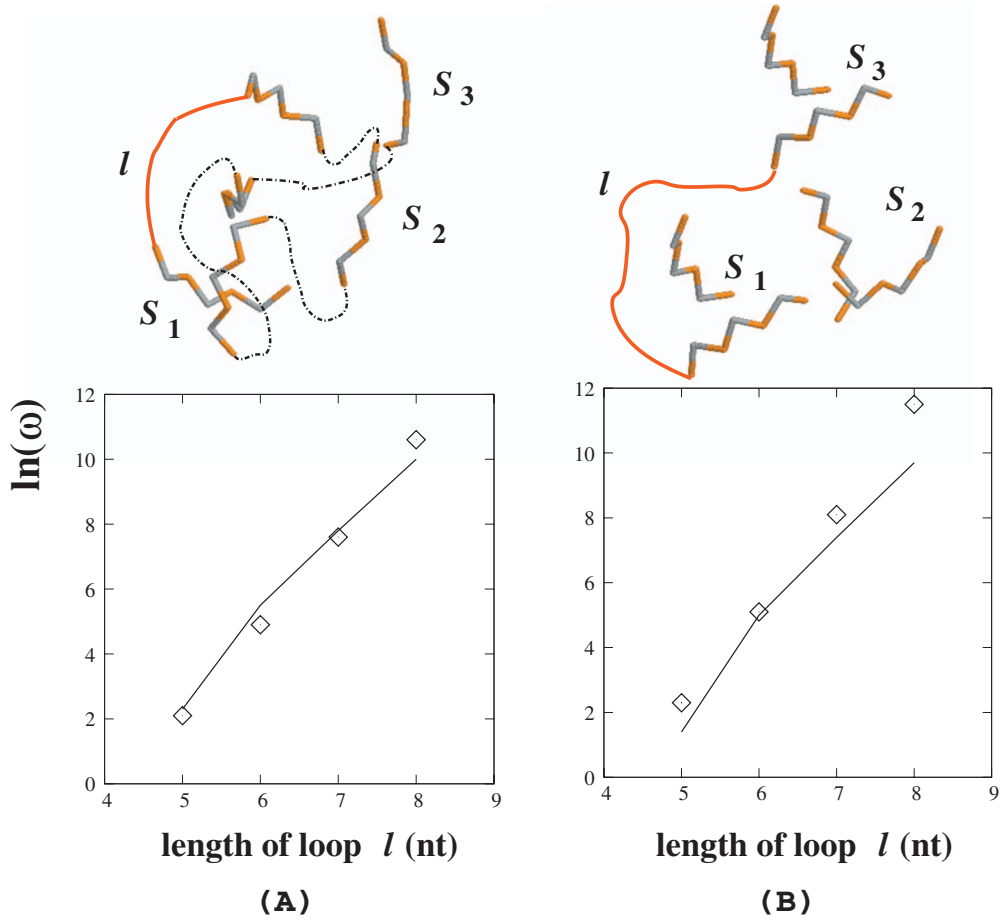

FIG. 11. Comparison between the conformational entropies for a three-helix RNA evaluated from the exact computer enumerations (◇) and from the three-body model (—). The helix stems are $S_1=3$ bp, $S_2=3$ bp, and $S_3=3$ bp, while the connection types (t) are (a) $t=3$ and (b) $t=5$, respectively.

the exact enumeration of a $l_1 \times l_2$ internal loop will cost the computer time as $t(l_1+l_2)$, which can be exceedingly large even for an internal loop of moderate size. Here $t(l)$ is the computational time for enumerating an l nt chain and is an exponential function of l [see Fig. 3(c)].

Our three-body approach avoids the exhaustive enumeration and is thus computationally more efficient than the previous method. The internal loop can be factored into two three-body systems, each containing two helices and one (loop) strand [Fig. 5(b)]. The three-body model [Eq. (2)] gives the conformational count for the internal loop as $\Omega_{\text{internal loop}}=\omega_{\text{system 1}} \times \omega_{\text{system 2}}$.

For an illustrative calculation and a test of our theory, we choose an internal loop with $l_1=4$ nt, $l_2=5$ nt, $s_1=5$ bp, and $s_2=5$ bp (namely, "internal 4-5-5-5"). This internal loop is factored into two three-body systems, one consisting of $l=4$ nt, $s_1=5$ bp, and $s_2=5$ bp (named as "4-5-5 system") and another "5-5-5 system." For illustrative purpose, we randomly select a helix configuration for the two helices. For the helix configuration that we choose, we have $(l,t,x)=(4\text{ nt},6,25.5\text{ Å})$ and (5 nt, 4,9.0 Å) as the (loop length, connect type, end-to-end distance) for the 4-5-5 system and the 5-5-5 system, respectively. Equation (4) gives $[A(4,6)=-5.77, B(4,6)=1.46, C(4,6)=7.43]$ and $[A(5,4)=-7.71, B(5,4)=1.81, C(5,4)=9.42]$. With these parameters, Eq. (3) gives the conformational count for the two three-body systems as $\ln \omega_{\text{4-5-5 system}}=0.16$ and $\ln \omega_{\text{5-5-5 system}}=4.82$. Thus the internal loop conformation for the given configuration is estimated as [Eq. (2)]

$$\ln \Omega^{\text{pred}}_{\text{internal 4-5-5-5}} = \ln(\omega_{\text{4-5-5 system}} \times \omega_{\text{5-5-5 system}}) = 4.98.$$

Exact enumeration of the internal loop conformations in the diamond lattice gives $\Omega^{\text{exat}}_{\text{internal 4-5-5-5}}=104$ and $\ln \Omega^{\text{exat}}_{\text{internal 4-5-5-5}}=4.64$. Therefore, the result from our analytical theory is close to the one from the exact enumeration. This suggests that the theory, which drastically reduces the conformational enumeration time from $t(9)\simeq 4.1\times 10^4$ s to $t(4)+t(5)\simeq 2.5$ s [Fig. 3(c)], may be quite reliable. Furthermore, in the practical calculation, the use of the analytical formula is even more computationally efficient because it does not involve any conformational enumeration.

By summing up the conformations of all helices configurations [Eq. (1)], the entropy for the internal loop can be obtained. Figure 9(b) shows a systematic theory-experiment comparison[18] for the entropy of internal loops with different sizes. The theory-experiment agreement further suggests the validity of our three-body model.

E. Three-loop H-type pseudoknot

Three-loop H-type pseudoknot consists of two helix stems and three loops. Unlike a two-loop pseudoknot where the two helix stems form a quasicontinuous helix, the helix stems here are separated by an intervening loop. This loop

disrupts the coaxial stacking interaction between the two helix stems and allows for more flexible orientations for the helices [Figs. 1(b) and 5(a)]. The large number of the possible configurations for the helices causes extremely long computer time $t(l_1+l_2+l_3)$ for the exact enumeration of the entropy. This highlights the necessity to evaluate the loop entropy using a computationally efficient method such as the three-body method developed above.

Indeed, the three-body model can readily give the loop entropy for a three-loop pseudoknot [Fig. 5(a)]. For example, a pseudoknot consisting of loops $l_1=3$ nt, $l_2=4$ nt, and $l_3=5$ nt and helices $s_1=s_2=5$ bp ("pseudoknot 3-4-5-5-5") can be factored into three three-body systems: 3-5-5, 4-5-5, and 5-5-5 systems (named in the way described above). We randomly select a 3D configuration of the helices for the illustrative purpose. We choose $(l,t,x)=(3\ \text{nt},0,16.9\ \text{Å})$, (4 nt, 2, 24.7 Å), and (5 nt, 0,19.1 Å) for the 3-5-5, 4-5-5, and 5-5-5 systems, respectively. Equation (4) gives the following parameters: $[A(3,0)=-4.21, B(3,0)=1.63, C(3,0)=4.79]$, $[A(4,2)=-6.74, B(4,2)=1.98, C(4,2)=6.90]$, and $[A(5,0)=-6.49, B(5,0)=1.33, C(5,0)=9.69]$. With the above parameters, Eq. (3) gives the conformational count for the three three-body systems as $\ln \omega_{3\text{-}5\text{-}5\ \text{system}}=0.81$, $\ln \omega_{4\text{-}5\text{-}5\ \text{system}}=0.41$, $\ln \omega_{5\text{-}5\text{-}5\ \text{system}}=4.25$. The loop conformational count for the pseudoknot of the given three-helix configuration is [Eq. (2)]

$$\ln \Omega^{\text{pred}}_{\text{pseudoknot 3-4-5-5-5}} = \ln(\omega_{3\text{-}5\text{-}5\ \text{system}} \times \omega_{4\text{-}5\text{-}5\ \text{system}} \times \omega_{5\text{-}5\text{-}5\ \text{system}}) = 5.47.$$

We aim to compare the above analytical result with the result from the exact conformational enumeration. However, because of the large total size of loops ($l_{\text{eff}}=3+4+5$), to exhaustively enumerate all the conformations of this three-loop H-type pseudoknot is practically impossible [Fig. 3(c)], even for a fixed helices configuration. Therefore, we compute the conformational counts for the (4 nt) loop 2, and the (5 nt) loop 3 separately with the presence of the (3 nt) loop 1. For each given helix configuration, we enumerate the conformations for loop 1. For each conformation of loop 1, we enumerate the conformations for loops 2 and 3 independently. We found $\Omega^{\text{exat}}_{\text{pseudoknot 3-4-5-5-5}}=149$ and $\ln \Omega^{\text{exat}}_{\text{pseudoknot 3-4-5-5-5}}=5.00$, which is close to the result from the above analytical calculation. Our three-body approach can reduce the computational time for loop conformational enumeration from $t(3+4+5)\simeq 5.9\times 10^7$ s to $t(3)+t(4)+t(5)\simeq 2.6$ s, and the use of the analytical formulas, which avoids the conformational enumeration, would lead to further drastic enhancement in the computational efficiency.

To compute the total entropy for the pseudoknot, we need to generate the 3D configurations for the helices. Summing over the loop conformational count for the different helix configurations [Eq. (1)] gives the total loop conformational count for the pseudoknot as $\ln \Omega^{\text{pred}}_{\text{tot}}=12.0$ from the three-body analytical model and $\ln \Omega^{\text{exact}}_{\text{tot}}=10.6$ from the exact computer enumeration described above, respectively. The result indicates that the three-body model is likely reliable for an efficient estimation of the conformational entropy.

IV. SUMMARY

We develop a three-body approach to compute the conformational entropy for RNA MG-like folds. The key idea for the approach is to factor the whole structure into three-body systems, which consist of a loop and the two helices that are connected to the two ends of the loop. Such a factorization converts a complex many-body problem into a set of three-body problems and hence leads to significant enhancement of computational efficiency. The approach involves two approximations.

(1) The excluded volume interaction between the different loops is weak. Therefore, we can neglect correlation between the loops, which allows us to compute the total loop entropy as a sum of the entropy of each loop.
(2) The entropy for a loop is a function of its local structural environment. For a MG-like structure, which is not tightly compact, the strong effect of the local environment on the loop conformation mainly comes from the helices that are connected to the loop. This suggests the use of a simple three-body system (one loop and the two helices) for the entropy calculation for a loop.

Our extensive tests against exact computer enumerations and experiments indicate that the model is valid for the systems tested. Furthermore, the method leads to a set of analytical formulas for the efficient computation of the loop entropy. With the nearest neighbor model and the empirical energy parameters (Turner's rule)[18] for the helix free energies, the method can be used to predict RNA structures. As a caveat, it is important to note that the nearest neighbor model assumes additivity and the empirical energy parameters are based on the two-state approximation and are valid only for $1M$ NaCl solution at 37 °C.

The above two approximations are likely valid for MG-like folding intermediates, where the structure is not highly compact and no tertiary contact has been established/stabilized. However, for the evaluation of the free energy of a tightly compact structure such as the native structure, the above approximations should be modified. For instance, helices other than the ones directly connected to the loop may affect the loop entropy and should be considered. The conformational entropy for a tertiary fold may be used as an auxiliary measure for the compactness of a RNA structure. In contrast to the radius of gyration, which describes the global average size of the structure, the conformational entropy, which gives the conformational freedom of the structure, may contain more "structural" information about the cross-linked entanglement of helices and loops.

Another improvement for the model should address the efficient enumeration of the helix configurations. Our major focus of the present study is on the loops (for a given configuration of the helices). For complex large RNA tertiary folds, the enumeration of the 3D configurations for the helices is a great challenge. The Monte Carlo sampling[13] might provide a viable method for the generation of the 3D configurations for a many-helix system. Nevertheless, applications of the model to simple tertiary folds suggest that the model may be useful for the entropy and free energy com-

putations for more complex RNA tertiary folds. The current form of the theory may be a solid first step for further systematic development.

ACKNOWLEDGMENTS

This research was supported by the NIH through Grant No. GM063732 (to S.-J.C.). Part of the numeric calculations involved in this research were performed on the HPC resources at the University of Missouri Bioinformatics Consortium (UMBC). We thank Dr. Song Cao and Dr. Zhi-Jie Tan for useful discussions.

[1] R. Russell, I. S. Millett, S. Doniach, and D. Herschlag, Nature (London) **7**, 367 (2000).
[2] S.-J. Chen and K. A. Dill, J. Chem. Phys. **103**, 5802 (1995).
[3] S.-J. Chen and K. A. Dill, J. Chem. Phys. **109**, 4602 (1998).
[4] S. Cao and S.-J. Chen, RNA **11**, 1884 (2005).
[5] S.-J. Chen and K. A. Dill, Proc. Natl. Acad. Sci. U.S.A. **97**, 646 (2000).
[6] W. Zhang and S.-J. Chen, J. Chem. Phys. **114**, 7669 (2001).
[7] Z. Kopeikin and S.-J. Chen, J. Chem. Phys. **122**, 094909 (2005).
[8] Z. Kopeikin and S.-J. Chen, J. Chem. Phys. **124**, 154903 (2006).
[9] S.-J. Chen, Annu. Rev. Biophys. **37**, 197 (2008).
[10] C. Hyeon and D. Thirumalai, Biophys. J. **92**, 731 (2007).
[11] A. Villa and G. Stock, J. Chem. Theory Comput. **2**, 1228 (2006).
[12] E. J. Sorin, Y. M. Rhee, B. J. Nakatani, and V. S. Pande, Biophys. J. **85**, 790 (2003).
[13] J. Zhang, J. Dundas, M. Lin, R. Chen, W. Wang, and J. Liang, RNA **15**, 2248 (2009).
[14] J. Zhang, M. Lin, R. Chen, W. Wang, and J. Liang, J. Chem. Phys. **128**, 125107 (2008).
[15] O. K. Vorov, D. R. Livesay, and D. J. Jacobs, Entropy **10**, 285 (2008).
[16] M. Muthukumar and B. G. Nickel, J. Chem. Phys. **80**, 5839 (1984).
[17] K. A. Dill, J. Biol. Chem. **272**, 701 (1997).
[18] M. J. Serra and D. H. Turner, Methods Enzymol. **259**, 242 (1995).
[19] M. Chastain and I. Tinoco, Jr., Prog. Nucleic Acid Res. Mol. Biol. **41**, 131 (1991).
[20] A. P. Gultyaev, F. H. D. Van Batenburg, and C. W. A. Pleij, RNA **5**, 609 (1999).
[21] D. P. Aalberts and N. O. Hodas, Nucleic Acids Res. **33**, 2210 (2005).
[22] A. Xayaphoummine, T. Bucher, F. Thalmann, and H. Isambert, Proc. Natl. Acad. Sci. U.S.A. **100**, 15310 (2003).
[23] S. Cao and S.-J. Chen, Nucleic Acids Res. **34**, 2634 (2006).
[24] S. Cao and S.-J. Chen, J. Mol. Biol. **367**, 909 (2007).
[25] S. Cao and S.-J. Chen, J. Mol. Biol. **357**, 292 (2006).
[26] S. Cao and S.-J. Chen, Phys. Biol. **5**, 016002 (2008).
[27] S. Cao and S.-J. Chen, RNA **15**, 696 (2009).
[28] V. A. Blomfield, D. M. Crothers, and I. Tinoco, Jr., *Nucleic Acids: Structures, Properties, and Functions* (University Science Books, Sausalito, CA, 2000).
[29] W. K. Olson and P. J. Flory, Biopolymers **11**, 1 (1972).
[30] W. K. Olson, Macromolecules **8**, 272 (1975).
[31] W. K. Olson, Macromolecules **13**, 721 (1980).
[32] C. M. Duarte and A. M. Pyle, J. Mol. Biol. **351**, 26 (1998).
[33] C. M. Duarte, L. M. Wadley, and A. M. Pyle, Nucleic Acids Res. **31**, 4755 (2003).
[34] L. J. W. Murray, W. B. Arendall III, D. C. Richardson, and J. S. Richardson, Proc. Natl. Acad. Sci. U.S.A. **100**, 13904 (2003).
[35] V. L. Murthy, R. Srinivasan, D. E. Draper, and G. D. Rose, J. Mol. Biol. **291**, 313 (1999).
[36] L. M. Wadley, K. S. Keating, D. M. Duarte, and A. M. Plye, J. Mol. Biol. **372**, 942 (2007).
[37] J. S. Richardson, B. Schneider, L. W. Murray, G. J. Kapral, R. M. Immormino, J. J. Headd, D. C. Richardson, D. Ham, E. Hershkovits, L. D. Williams, K. S. Keating, A. M. Pyle, D. Micallef, J. Westbrook, and H. M. Berman, RNA **14**, 465 (2008).
[38] P. J. Flory, *Statistical Mechanics of Chain Molecules* (Wiley, New York, 1969).
[39] F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, E. F. Meyer, Jr., M. D. Brice, J. R. Rogers, O. Kennard, T. Shimanouchi, and M. Tasumi, J. Mol. Biol. **112**, 535 (1977).
[40] E. Hershkovitz, E. Tannenbaum, S. B. Howerton, A. Sheth, A. Tannenbaum, and L. D. Williams, Nucleic Acids Res. **31**, 6249 (2003).
[41] T. H. Reijmers, R. Wehrens, and L. M. Buydens, J. Chem. Inf. Comput. Sci. **41**, 1388 (2001).
[42] S. Cao and S.-J. Chen, RNA **16**, 538 (2010).
[43] Z.-J. Tan and S.-J. Chen, Biophys. J. **95**, 738 (2008).