Prediction of RNA multiloop and pseudoknot conformations from a lattice-based, coarse-grain tertiary structure model

Daniel Jost, and Ralf Everaers

Citation: The Journal of Chemical Physics **132**, 095101 (2010); doi: 10.1063/1.3330906
View online: http://dx.doi.org/10.1063/1.3330906
View Table of Contents: http://aip.scitation.org/toc/jcp/132/9
Published by the American Institute of Physics

Prediction of RNA multiloop and pseudoknot conformations from a lattice-based, coarse-grain tertiary structure model

Daniel Jost[a] and Ralf Everaers
Laboratoire de Physique and Centre Blaise Pascal of the École Normale Supérieure de Lyon, Université de Lyon, CNRS UMR 5672, 46 allée d'Italie, 69364 Lyon Cedex 07, France

(Received 15 December 2009; accepted 3 February 2010; published online 3 March 2010)

We present a semiquantitative lattice model of RNA folding, which is able to reproduce complex folded structures such as multiloops and pseudoknots without relying on the frequently employed *ad hoc* generalization of the Jacobson–Stockmayer loop entropy. We derive the model parameters from the Turner description of simple secondary structural elements and pay particular attention to the unification of mismatch and coaxial stacking parameters as well as of border and nonlocal loop parameters, resulting in a reduced, unified parameter set for simple loops of arbitrary type and size. For elementary structures, the predictive power of the model is comparable to the standard secondary structure approaches, from which its parameters are derived. For complex structures, our approach offers a systematic treatment of generic effects of chain connectivity as well as of excluded volume or attractive interactions between and within all elements of the secondary structure. We reproduce the native structures of tRNA multiloops and of viral frameshift signal pseudoknots. © *2010 American Institute of Physics*. [doi:10.1063/1.3330906]

I. INTRODUCTION

A quantitative understanding of the folding and opening processes of the DNA or RNA strands is relevant for many biological functions such as transcription, replication,[1] proteins production (mRNA and tRNA), *in vivo* reaction catalysis (ribozymes)[2] or gene silencing,[3] as well as recent biotechnological applications such as DNA microarrays,[4] DNA self-assembly,[5] or DNA origami.[6] Considerable efforts have been made to predict the thermodynamic properties, secondary, and tertiary structure of RNA molecules using chemical physics and bioinformatics based approaches (for a review, see Refs. 7 and 8). Most secondary structure approaches are based on the nearest-neighbor (NN)-like Turner model,[9] which is solved using different computational techniques: dynamic algorithms exploiting recurrence relations in simplified partition functions with[10] or without[11–16] experimentally determined structural constraints, genetic algorithms,[17,18] or kinetic Monte Carlo (MC) schemes.[19–21] Several different strategies have been employed for tertiary structure models: secondary structure predictions coupled to three-dimensional (3D) modeling,[22,23] or *de novo* parameterization of coarse-grained models based on thermodynamic[24–28] or structural[29,30] data, which can be complemented by chemically determined contact maps.[31]

The evaluation of the free-energy difference $\Delta G = G_f - G_u$ between a folded structure (free energy G_f) and the corresponding denaturated state (G_u), is essential to predict thermodynamics and native structures of nucleic acids. In this paper, we follow the tradition of Poland–Sheraga (Ref. 32) to consider polynucleotides as polymers with free-energy contributions associated to specific local interactions and conformations (formation of double helical segments, forking, and capping) and generic polymer effects (long range excluded volume interactions, looping entropies).[9] In the case of RNA, secondary structure prediction[11–13,15,33,34] is based on the Turner model,[9] which provides of the order of 650 experimentally determined parameters to describe the sequence-dependent specific interactions.

However, the generic contributions may not be neglected and play a major role in the thermodynamics of RNA folding.[16,35] There exists a large variety of loop structures ranging from simple hairpins and internal loops or bulges [Figs. 1(c), 1(d), and 1(g)] to multibranch loops connected to at least three stems [Fig. 8(c)] and pseudoknots (Fig. 10) containing at least two base pairs (i, j and k, l, $i<j$ and $k<l$) which do not follow the nesting convention $i<k<l<j$ or $i<j<k<l$.[40] For small loops, standard secondary structure models[11–13] essentially use tabulated experimental data. For larger loops, they require a model to account for the loop length dependence of the nonlocal loop formation free energy Δg_{loop}. For simple structures (such as hairpins, internal loops, or bulges), Δg_{loop} is well modelized by a Jacobson–Stockmayer equation[41]

$$\Delta g^{\text{JS}}_{\text{loop}}(n) = -T(\Delta s_{\text{loop}} - k_B c \log n), \tag{1}$$

where n is the loop size, $-T\Delta s_{\text{loop}}$ is the entropic nucleation energy depending on the substructure, and c is an exponent characteristic of the polymer nature of loops. For simple loops (hairpin, internal or bulge loops), the Turner model[9] takes $c=1.75$ which only accounts for intra-loop excluded volume interactions.[42] Note, however, in the case of small loops connected with long stems, interactions with the rest of the chain are not negligible and $c \approx 2.1$.[43,44] For complex structures (multibranch loops, pseudoknots), whose substructures can interact by excluded volume effects, no such universal and well-parameterized equation exists. Mainly, stan-

[a]Electronic mail: daniel.jost@ens-lyon.fr.

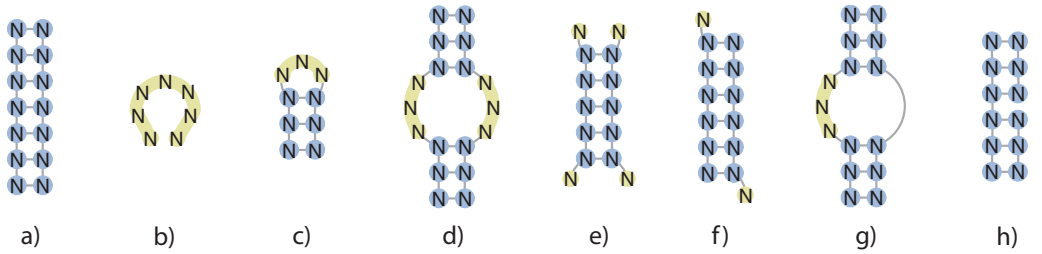

FIG. 1. Typical secondary structures: (from left to right) double-strand, single strand, hairpin, internal loop, double-strand with forks, double-strand with dangles, bulge loop, and nicked double-strand. We use PseudoViewer Web Application (Ref. 36) to draw these structures. Nucleotides colored in blue are paired, those in yellow are unpaired.

dard programs[11,13,33,34] approximate this free energy by a generalized Jacobson–Stockmayer equation

$$\Delta g_{\mathrm{loop}}^{\mathrm{genJS}}(n,h) = k_B T(a + b \times n + c \log n + d \times h), \qquad (2)$$

where n is the number of unpaired nucleotides in the loop, h is the number of branching stems, and a, b, and d are fitted parameters depending on the kind of substructures. For multibranch loops, the dependence of Eq. (2) in the central loop length is often discarded for algorithmic reason ($b=0$ and $c=0$), although a logarithmic dependence in the loop size (under the assumption $b=0$ and $c=1.76$) could be included for free-energy minimization.[9,11,13] Heuristically changing the exponent c in Eq. (1) or (2) has drastic consequences for the predicted melting behavior[16] even though care should be taken to readjust the nucleation penalty.[45,46] Note that specific generalized Jacobson–Stockmayer relations may be required for each class of pseudoknots (for example, one for H-pseudoknots without interhelix loops,[39] another for H-pseudoknots with interhelix loops,[47] and so on). Alternatively, there is a recent proposal[48,49] to model more complex topologies by further generalizing Eq. (2) to account for the topological genus of the graph representing a RNA secondary structure.

Being generic polymer effects, looping entropies and excluded volume interactions between different parts of a chain molecule can be studied using simple model systems. For example, exact enumerations of typical core units (loops, pseudoknots) on cubic or squared lattices have been employed to estimate the polymeric loop length dependence of Δg_{loop}.[50–52] Extensive simulations of simple lattice models[53–56] have been used to study melting properties of nucleic acids, such as the predicted[43] change in the order of the DNA melting transition.

In this work, we follow the logic of Ref. 57 to introduce sequence-specific local interactions into a generic lattice model of RNA folding. In particular, we relate the parameters of the lattice model to the experimentally determined parameters of standard secondary structure descriptions, allowing us to reproduce the dominant contributions to the folding free energy with state-of-the-art accuracy. In addition, the model accounts for subdominant, yet important, free-energy contributions due to chain connectivity and generic polymer interactions by treating an ensemble of coarse-grain, 3D structures. In spirit, our ansatz is similar to the Kinefold model[19–21] but goes beyond the inclusion of connectivity effects. The model strongly relies on the hierarchical nature of RNA folding and should be viewed as an attempt to extend and systematically improve secondary structure based descriptions. In particular, we do not try to resolve the internal structure of secondary structure elements or to derive or explain the sequence-dependent local parameters of the NN model. While these are extremely interesting questions in themselves, we feel that the predictive power of more microscopic approaches[24,26–28,58] for large RNA molecules is quickly lost, if the RNA oligomer thermodynamics is not reproduced with a precision exceeding the performance of standard NN-model.

The paper is structured as follows: In a first part, we define the lattice model and compare it to the extensively used Turner model.[9] After a detailed analysis of gauge freedoms in the Turner model and a unification of loop and forking parameters in the secondary structure description, we derive the parameters of the lattice model and outline the computational techniques used to simulate the model. In a second part, we first validate the model for simple hairpins and test its predictive power for more complex conformations such as multibranched loops and pseudoknots. We find that it faithfully reproduces the native structures of tRNAs and viral frameshift signals. As applications, we study the equilibrium folding pathways of tRNA-phe of yeast, determine loop destabilizing free energies required as input for standard secondary structures programs, and study the tetraloop/tetraloop-receptor,[59] as an example for the influence of a specific tertiary contact on RNA folding.

II. MODEL AND METHODS

To the best of our knowledge, no convincing lattice model already studied succeeds in accurately predicting the thermodynamic properties of the folding of RNA heteropolymers by comparison with experiments. That is why we develop our lattice model with the constraint to not only fully account for polymeric effects but also to quantitatively describe the experimental data.

Hence, in Sec. II A, we present the semiquantitative parameterization of a simple lattice model.[53,57] We compare it to the Turner model,[9,60] the most widely employed model to describe nucleic acids chains at the secondary structure level. We illustrate the definition of the lattice model with several typical examples. In Sec. II B and II C, we analyze in detail the loop parameters and explain how we derive the parameters of the lattice model from those of the Turner model. Finally, in Sec. II D, we define the advanced MC techniques used to simulate the model.

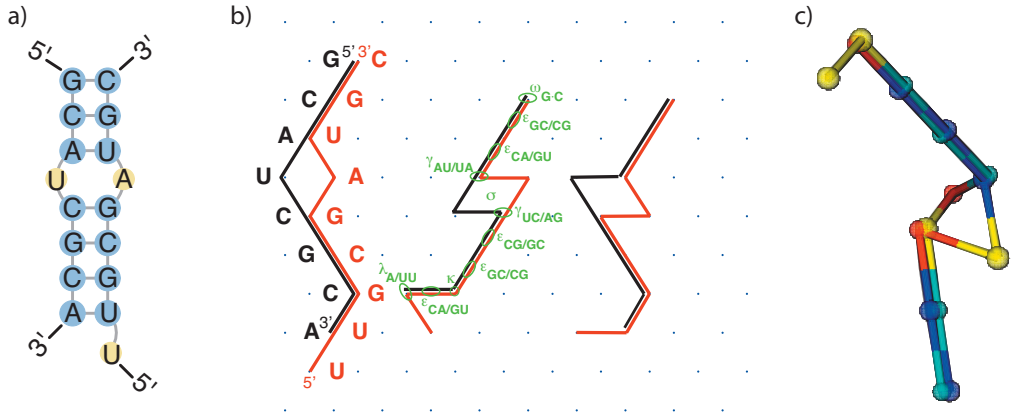

FIG. 2. (a) Example of a secondary structure for the RNA complex $(GCAUCGCA)\cdot(UUGCGAUGC)$. (b) 2D-projection of possible conformations corresponding to the same secondary structure than the molecule in (a) and definition of the various lattice free-energy contributions.(c) 3D conformation on the fcc lattice. Red dots represent adenosine nucleotides, cyan dots are guanines, yellow dots are uracyles and blue dots are cytosine. Two paired nucleotides occupy the same lattice site.

A. Definition of the model

1. Interactions in the lattice and the Turner models

In the Turner model, a secondary structure $\mathcal{S}$ is viewed as a succession of stems and loops [see Figs. 1 and 2(a)]. In contrast, the lattice model provides a coarse-grain description of a 3D conformation $\mathcal{C}$. A nucleic acid strand is modeled as a self-avoiding walk (SAW) on a regular lattice [Figs. 2(b) and 2(c)]. The possible positions of bases are the lattice sites separated by a distance b. Two bases are allowed to overlap, if and only if, they can form a Watson–Crick base pair (antiparallel and complementary) A/T (DNA), A/U (RNA) or G/C (DNA/RNA), or the antiparallel wobble pair G/U (RNA). The secondary structure $\mathcal{S}(\mathcal{C})$ of an RNA conformation in the lattice model is easily defined via these contacts. In turn, a given secondary structure is represented by a set $\{\mathcal{C}\}_{\mathcal{S}}$ of lattice conformations. In general, there is a large number $\Omega(\mathcal{S})$ of conformations per secondary structure.

One can think of each conformation of the lattice model as representing a large number of microstates of the RNA/ionic solvent system. As in the Turner model, the Hamiltonian of the lattice model is hence a temperature-dependent *free* energy and not an energy. Both models are defined on the same length scale and have, on first sight, nearly identical parameters. The free energy of a secondary structure $\mathcal{S}$ in the Turner model and of a conformation $\mathcal{C}$ in the lattice model can be decomposed into the sum of NN and nonlocal terms:

- The stacking (or pairing) free energy [Turner model: $\Delta g^{st}_{\mathrm{NN}}=\Delta h^{st}_{\mathrm{NN}}-T\Delta s^{st}_{\mathrm{NN}}$; lattice model: $\epsilon(T)=\epsilon_H-T\epsilon_S$] accounts for the stacked neighboring base pairs and depends on the ten (DNA) or 21 (RNA) different possible steps.[61]
- The capping (or terminal) free energy [Turner model: $\Delta g_{\mathrm{term}}=\Delta h_{\mathrm{term}}-T\Delta s_{\mathrm{term}}$; lattice model: $\omega(T)=\omega_H-T\omega_S$] accounts for paired ends and depends on the A/T or G/C nature (DNA) or on the A/U, G/C, or G/U nature (RNA) of the ending base pair.
- The forking (or interfacial, or terminal mismatch) free energy [Turner model: $\Delta g^{tm}_{\mathrm{NN}}=\Delta h^{tm}_{\mathrm{NN}}-T\Delta s^{tm}_{\mathrm{NN}}$; and lattice model: $\gamma(T)=\gamma_S-T\gamma_H$] accounts for interfaces between paired and unpaired sections in the complex and depends on the four bases forming the fork.
- The dangling free energy [Turner model: $\Delta g^{dg}_{\mathrm{NN}}=\Delta h^{dg}_{\mathrm{NN}}-T\Delta s^{dg}_{\mathrm{NN}}$; lattice model: $\lambda(T)=\lambda_H-T\lambda_S$] accounts for an end paired with a no-end base (by comparison, the capping free-energy concerns two paired ends) and depends on the three surrounding bases.
- An entropic, nonlocal, loop nucleation free-energy penalty [Turner model: $-T\Delta s^{T}_{\mathrm{loop}}$; lattice model: $\sigma(T)=\sigma_H-T\sigma_S$], which is assumed to be independent of the loop composition. For steric reasons, hairpin loops with less than two nucleotides are excluded.[40]
- A coaxial stacking free energy [Turner model: Δg_{coax}; lattice model: $\chi(T)=\chi_H-T\chi_S$] accounts for the favorable interaction of two double-helix stems stacked end to end and depends on the corresponding stacked step.
- An intermolecular mixing (or initiation) free energy (Turner model: Δg_{init}; lattice model: G_{mix}) for multistrand complexes, which included translational and rotational entropy loss upon association.
- Finally, only for the lattice model, to account for the rigidity of the double-helix, we introduce a bending free-energy $\kappa(T,\Psi)=\kappa_H(\Psi)-T\kappa_S(\Psi)$. Ψ is the angle where the double-helix is bended. On the face-centered-cubic (fcc) lattice, there are only four possible angles: 0° (1 possibility), 60° (4), 90° (2), and 120° (4). The backward possibility $\Psi=180°$ is excluded. In the example configuration of Fig. 2, $\Psi=60°$.

Unpaired nucleotides adjacent to many stems cannot participate to more than one NN-stacking (dangling or forking). Noncanonical interactions such as triplet interactions or *trans* Watson–Crick base pairs[62] could, in principle, be included in an analogous fashion in the lattice model. In the Turner model, the forking and nucleation penalty parameters depend on the nature of the adjacent loop (hairpin, internal, bulge, and multibranched). The above list illustrates that there is a fairly close correspondence between the parameters of the

lattice model and those of Turner-like descriptions of secondary structure free energies. A notable exception are universal parameters (exponents) accounting for the polymeric nature of nucleic acid chains.[16,42,43,45,63] In particular, the Jacobson–Stockmayer relations [Eqs. (1) and (2)] have no equivalent in the lattice model. Rather, the underlying excluded volume interactions are treated explicitly via the condition that unpaired bases cannot occupy the same lattice site. Their consequences for the free energy of a secondary structure are thus calculated without further approximation. The model does *not* limit excluded volume interactions to individual loops, but fully accounts for their long-range nature (in the sense of chemical not spatial distance). The only approximation entering the calculation of the free energy of a secondary structure is the neglect of the fine structure of the corresponding 3D RNA conformations including the double-helical nature of RNA stems.

2. Secondary structure entropies in the lattice model

Within the lattice model, the free energy of a secondary structure $\mathcal{S}$ is given by the sum of the interaction free energies described in Sec. II A 1 and of a conformational entropic free energy $-k_BT \log \Omega(\mathcal{S})$, where $\Omega(\mathcal{S})$ is the number of conformations corresponding to $\mathcal{S}$, the position of the first nucleotide being fixed. To illustrate this point, we present some typical secondary structures and their corresponding free energies. In the Turner model, the free energy of a structure is given relatively to its denaturated state (i.e., it is in fact a free-energy difference). In the lattice model, we treat the single strands (i.e., the denatured states) as SAWs and the free energy of a conformation is defined within this convention. Note that the Turner initiation free energies Δg_{init} contain a contribution from the entropy of mixing. We will deal with the corresponding term for the lattice model in Sec. II C.

- Short double-strand [Fig. 1(a)] composed by N base-pair steps: Turner model: $\Delta G_{ds}=N\Delta g^{st}_{\text{NN}}+2\Delta g_{\text{term}}+\Delta g_{\text{init}}$; lattice model: $G_{ds}=N\epsilon+2\omega-k_BT \log(z)$, where z is the number of possible nearest-neighbors for a lattice site.
- Single strand [Fig. 1(b)] composed by N steps: Turner model: $\Delta G_{ss}=0$; lattice model: $G_{ss}=-k_BT \log(f_s\mu^N N^{c'})$, where $f_s\mu^N N^{c'}$ accounts for the total number of SAW for a N-polymer chain. μ is the effective number of possible NNs and $c' \approx 0.16$ is a universal exponent.[64,65]
- Hairpin [Fig. 1(c)] composed by a stem with N base-pair steps and a loop with M steps: Turner model: $\Delta G_{hp}=N\Delta g^{st}_{\text{NN}}+\Delta g_{\text{term}}+\Delta g^{tm}_{\text{NN}}+\Delta g^{JS}_{\text{loop}}(M)$; lattice model: $G_{hp}=N\epsilon+\omega+\gamma+\sigma-k_BT \log((z-2)f_l\mu^M M^{-c})$. The number of hairpin conformations is given by the product of the number of self-avoiding polygons for a M-polymer chain, $f_l\mu^M M^{-c}$, and the number of possibilities to attach a stem to a loop, which is approximately $(z-2)$. $c \approx 1.76$ is a universal exponent.[42,64]
- Double-strand with an internal bubble [Fig. 1(d)] composed by a loop of $2M$ steps between two stems of N base-pair steps each: Turner model: $\Delta G_{\text{int}}=2N\Delta g^{st}_{\text{NN}}+2\Delta g_{\text{term}}+2\Delta g^{tm}_{\text{NN}}+\Delta g^{JS}_{\text{loop}}(2M)+\Delta g_{\text{init}}$; lattice model: $G_{\text{int}}=2N\epsilon+2\omega+2\gamma+\sigma-k_BT \log((z-2)^2 f_l\mu^{2M}(2M)^{-c})$, where we have followed the same logic as in the hairpin case.
- Stem with two terminal forks [Fig. 1(e)] composed of free ends with one nucleotide: Turner model: $\Delta G_{\text{ext}}=N\Delta g^{st}_{\text{NN}}+2\Delta g^{tm}_{\text{NN}}+\Delta g_{\text{init}}$; lattice model: $G_{\text{ext}}=N\epsilon+2\gamma-k_BT \log(z(z-1)^2(z-2)^2)$, where a factor $(z-1)(z-2)$ accounts for the number of possibilities to place the two mutually avoiding free ends next to a stem.
- Stem with two dangling ends [Fig. 1(f)] composed by free ends with one nucleotide: Turner model: $\Delta G_{dg}=N\Delta g^{st}_{\text{NN}}+2\Delta g^{dg}_{\text{NN}}+\Delta g_{\text{init}}$; lattice model: $G_{dg}=N\epsilon+2\lambda-k_BT \log(z(z-1)^2)$, where we have followed the same logic as for terminal forks.
- Bulge structure [Fig. 1(g)] composed by two stems (length N) and a loop (length $M \gg 1$): Turner model: $\Delta G_{\text{bulge}}=2N\Delta g^{st}_{\text{NN}}+2\Delta g_{\text{term}}+\Delta g^{\text{JS}}_{\text{loop}}(M)+\Delta g_{\text{init}}$ (coaxial stacking is neglected for large loops); lattice model: $G_{\text{bulge}}=2N\epsilon+2\omega+\sigma+\chi-k_BT \log[f_l\mu^{M+1}(M+1)^{-c}(z-2)^2]$, where we have followed the same logic as in the internal case.
- Nicked structure [Fig. 1(h)] composed by two stems (length N) connected by a single step: Turner model: $\Delta G_{\text{nick}}=2N\Delta g^{st}_{\text{NN}}+2\Delta g_{\text{term}}+\Delta g_{\text{coax}}+2\Delta g_{\text{init}}$; lattice model: $G_{\text{nick}}=2N\epsilon+2\omega+\chi-k_BT \log[z(z-1)^2]$, where a factor $z(z-1)^2$ approximately accounts for the number of possibilities to attach three rigid rods consecutively.

Whereas the geometric parameters z, μ, f_s, and f_l depend on the nature of the lattice (simple cubic, fcc, etc.), c and c' are universal exponents characteristic of the polymer nature of nucleic acids chains.

In the following, we choose to work on a fcc lattice [$z=12$, $\mu=10.035$, $f_s=1.14$, and $f_l=0.25$ (Ref. 66)]. The choice of the fcc lattice instead of the original simple cubic lattice[57] is motivated by the higher symmetry of the lattice and by the possibility to describe loops of any lengths (whereas for the simple cubic lattice, only the $(2\times n+2)$-loops are allowed).

B. Parameters in the Turner model

Before attacking the lattice model parameterization, we want to discuss some subtle features of the Turner model. In Secs. II B 1–II B 3, we propose a unification of forking, coaxial and mismatch parameters at the secondary structure level. Overall, this leads to a drastic decrease in the number of independent Turner-like models parameters ($\sim 650 \rightarrow \sim 350$). As a last step, in Sec. II B 4, we show that the Turner model is in fact defined modulo a constant which does not modify the model predictions. This will turn out to be useful for the suppression of nonlocal terms in the parameterization of the lattice model.

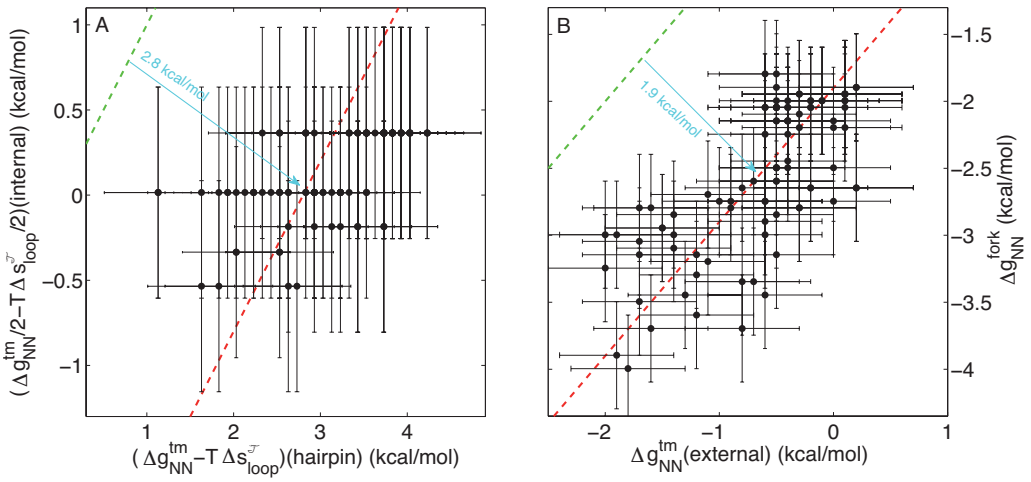

FIG. 3. (a) Correlation between the internal and hairpin Turner loop parameters. The observed shift equals half the unified loop nucleation penalty $-T_{37}\Delta s_{\text{loop}}$ (see text). (b) Correlation between the Turner free energies for external 1 bp forks and the corresponding unified forking parameters for large loops. We attribute the average difference of 1.9 kcal/mol to next-nearest neighbor interactions.

1. Unification of forking and loop nucleation entropies

In contrast to forking enthalpies, the available (Turner) entropy penalties for loop nucleation and forking are loop-type depended. This causes no harm in their usual context of application, but is not convenient for our lattice model as the calculation of free-energy changes associated with a *local* structural rearrangement would require a *global* analysis of the secondary structure. Moreover, this dependence is surprising from a physical point of view: why should the local forking free energies be different for large hairpins, internal loops and fraying ends?

In the following, we show that the Turner secondary structure free energies can be written in terms of *nonspecific* entropic penalties for loop nucleation and forking. An arbitrary gauge ϕ_1 does not affect the total entropy penalty for the hairpin loop, $(\Delta s^{\mathcal{T}}_{\text{loop}}(\text{hairpin}) - \phi_1) + (\Delta s^{tm}_{\text{NN}}(\text{hairpin}) + \phi_1)$. Similarly, we can introduce an arbitrary shift between the nucleation and the total forking penalties for internal loops without changing the measurable total entropy penalty: $(\Delta s^{\mathcal{T}}_{\text{loop}}(\text{internal}) - \phi_2) + 2(\Delta s^{tm}_{\text{NN}}(\text{internal}) + \phi_2/2)$. It is straightforward to solve ϕ_1 and ϕ_2 for

$$\Delta s^{\mathcal{T}}_{\text{loop}}(\text{hairpin}) - \phi_1 = \Delta s^{\mathcal{T}}_{\text{loop}}(\text{internal}) - \phi_2 \equiv \Delta s_{\text{loop}}, \quad (3)$$

$$\Delta s^{tm}_{\text{NN}}(\text{hairpin}) + \phi_1 = \Delta s^{tm}_{\text{NN}}(\text{internal}) + \phi_2/2 \equiv \Delta s^{\text{fork}}_{\text{NN}}, \quad (4)$$

where Δs_{loop} and $\Delta s^{\text{fork}}_{\text{NN}}$ are the corresponding nonspecific parameters. Our working hypothesis implies a correlation between $\Delta s^{\mathcal{T}}_{\text{loop}}(\text{hairpin}) + \Delta s^{tm}_{\text{NN}}(\text{hairpin}) = \Delta s_{\text{loop}} + \Delta s^{\text{fork}}_{\text{NN}}$ and $\Delta s^{\mathcal{T}}_{\text{loop}}(\text{internal})/2 + \Delta s^{tm}_{\text{NN}}(\text{internal}) = \Delta s_{\text{loop}}/2 + \Delta s^{\text{fork}}_{\text{NN}}$. Within the experimental error the correlation is clearly borne out by the available parameters [see Fig. 3(a)]. Using least-squared methods[67] and the version 3.0 of Turner parameters,[9] we have evaluated $\Delta s^{\text{fork}}_{\text{NN}}$, Δs_{loop} values by fitting Eqs. (3) and (4). The corresponding value of the incomplete gamma function Q is nearly 1, signature of a good fit.[67] We find

$$\Delta s_{\text{loop}} = -9.2 \pm 2k_B = -18.3 \pm 4 \text{ cal mol K}. \quad (5)$$

As indicated above, the enthalpies for the hairpin and the internal forking energies are equal in the Turner model, therefore, $\Delta h^{\text{fork}}_{\text{NN}} = \Delta h^{tm}_{\text{NN}}$. Data for $\Delta g^{\text{fork}}_{\text{NN}}$ at 37 °C are shown in Table I. Turner has proposed a further reduction in the number of parameters by noting that, for internal loops, the forks closing by an *AU*, *UA*, *GU*, or *UG* base pair and having *XY* as a first mismatch (with $X, Y \in \{A, C, G, U\}$ and $XY \neq AG$, *GA*, and *UU*) get the same free energy[9] at 37 °C (*idem* for forks closing by a *GC* or *CG* base pair). This observation applied to the independently measured hairpin parameters gives very small Q-values ($Q \sim 5\ 10^{-2}$ for *AU*/*UA*/*GU*/*UG* closing forks and $Q \sim 5\ 10^{-7}$ for *GC*/*CG* closing forks), signature of a wrong hypothesis. We have therefore kept the 96 independent parameters in Table I.

2. Small versus large forks

Following the arguments in Sec. II B 2, the Turner entropy penalty for the external fork of a fraying end should be given by $\Delta s^{\text{fork}}_{\text{NN}}$. Figure 3(b) shows a strong correlation, but an important and unexpected offset ($\sim 3.1 k_B T = 1.9$ kcal/mol). To understand the origin of this apparent failure of our unification scheme, it is necessary to go back to the actual experiments underlying the Turner forking parameter and to critically review the assumption made in deriving them.

In contrast to the forking parameters for hairpins or internal loops, the Turner external loop parameters have been parameterized using experimental data for conformations with only two single free ends [for an example, see Fig. 1(e)]. There is thus a hidden assumption that in fraying ends, all nucleotides except the pair adjacent to the double-helical stem experience identical environments as in a denatured single stranded chains [a corresponding assumption is not made for small hairpins or loops, whose entropic cost are not calculated from Eq. (1), but tabulated]. It seems more likely that the effect should extend a small distance along the chain with free energy correction $\Delta g^{i}_{\text{fork}}$ relative to the single strand

TABLE I. Unified forking free energies $\Delta g_{\mathrm{NN}}^{\mathrm{fork}}$ in kcal/mol (1 kcal/mol$\approx 1.6k_B$) at 37 °C (error: 0.4 kcal/mol).

	5′−AX−3′ / 3′−UY−5′				5′−CX−3′ / 3′−GY−5′				5′−GX−3′ / 3′−CY−5′			
X \ Y	A	C	G	U	A	C	G	U	A	C	G	U
A	−2.1	−2.2	−2.6	−2.1	−3.0	−3.0	−3.5	−3.2	−2.8	−3.0	−3.5	−3.3
C	−2.0	−2.0	−2.7	−2.0	−2.8	−2.7	−3.7	−2.7	−2.8	−2.6	−3.5	−2.5
G	−3.0	−2.5	−2.0	−1.8	−3.9	−3.3	−3.1	−2.8	−4.0	−3.7	−3.0	−2.9
U	−2.1	−2.1	−2.2	−2.8	−3.1	−3.0	−3.2	−3.6	−3.2	−2.8	−3.4	−3.4

	5′−GX−3′ / 3′−UY−5′				5′−UX−3′ / 3′−AY−5′				5′−UX−3′ / 3′−GY−5′			
X \ Y	A	C	G	U	A	C	G	U	A	C	G	U
A	−1.8	−2.2	−2.6	−2.1	−2.2	−2.1	−2.8	−2.2	−2.2	−2.1	−2.8	−2.2
C	−2.0	−2.0	−2.7	−2.0	−2.0	−2.0	−2.5	−1.9	−2.0	−2.0	−2.8	−1.9
G	−2.9	−2.5	−2.1	−1.9	−3.2	−2.5	−2.3	−2.0	−2.9	−2.5	−2.1	−2.3
U	−2.1	−2.1	−2.1	−2.8	−2.1	−2.0	−2.2	−2.7	−2.2	−2.0	−2.2	−2.7

rapidly decreasing with distance i from the fork. In this case, the forking energy for long strands would be given by $\Delta g_{\mathrm{fork}}=\Sigma_{i=1}^{\infty}\Delta g_{\mathrm{fork}}^{i}$. In particular, we argue that $\Delta g_{\mathrm{NN}}^{tm}(\mathrm{external})=\Delta g_{\mathrm{fork}}^{1}$, while Δg_{fork} should be given by the nonloop-specific parameter determined in Sec II B 1. Since we ignore the distance dependence, we propose to set $\Delta g_{\mathrm{fork}}^{2}=\Delta g_{\mathrm{fork}}-\Delta g_{\mathrm{fork}}^{1}$ and $\Delta g_{\mathrm{fork}}^{i}\equiv 0$ for all $i>2$. Figure 3(b) suggests $\Delta g_{\mathrm{fork}}^{2}=-3.1\pm 0.8k_BT$ ($Q=0.98$). For the enthalpies, the corresponding analysis yields $\Delta h_{\mathrm{fork}}^{2}=-6.8\pm 6.5k_BT$ ($Q=0.77$). With these corrections, forking energies in the Turner model become independent of the type of the adjacent loop.

Similar arguments apply to dangling ends. Hence, we can define $\Delta g_{\mathrm{dangle}}^{2}=\Delta g_{\mathrm{dangle}}-\Delta g_{\mathrm{dangle}}^{1}$ and $\Delta g_{\mathrm{dangle}}^{i}\equiv 0$ for all $i>2$. Moreover, we could reasonably assume that the correction $\Delta g_{\mathrm{fork}}^{2}$ for the forks is the double than the correction $\Delta g_{\mathrm{dangle}}^{2}$ for one dangle. We note that the numerical value of the sequence-independent correction term $\Delta g_{\mathrm{dangle}}^{2}=-1.6k_BT$ is consistent with previous observations.[68]

3. Coaxial stacking and mismatches parameters

Experiments involving coaxial stacking[69–71] observed a stabilization of the complexes due to the end-to-end stacking of two stems. Data are available for structures with or without an intervening mismatch. From these free energies, one can easily extract the contribution due to coaxial stacking by subtracting from the total free energy all the other contributions (stacking, capping, initiation, etc.).[9] Using our unified loop and forking parameters, for coaxial stacking without an intervening mismatch and no strand extensions beyond the interface [Fig. 4(a)], we find that Δg_{coax} is highly correlated with the corresponding NN-stacking $\Delta g_{\mathrm{NN}}^{st}$ and the excess stability $\Delta g_{\mathrm{coax}}-\Delta g_{\mathrm{NN}}^{st}=-2.9\pm 0.6k_BT\equiv -T\Delta s_c^{0}$ ($Q=0.99$). Interestingly, a nick stabilizes the helix. For interfaces followed by one strand extension [Fig. 4(b)], we find $\Delta g_{\mathrm{coax}}-\Delta g_{\mathrm{NN}}^{st}=-2.3\pm 0.6k_BT\equiv -T\Delta s_c^{1}$ ($Q=0.99$), and for interfaces followed by two strand extensions [Fig. 4(c)], $\Delta g_{\mathrm{coax}}-\Delta g_{\mathrm{NN}}^{st}=-0.8\pm 0.6k_BT\equiv -T\Delta s_c^{2}$ ($Q=0.96$). Note that in the Turner model, contributions for situations with one or two strand extensions are supposed to be equal. This hypothesis leads to a smaller Q-value of 0.4. For coaxial stacking cases with an intervening mismatch [Fig. 4(d)], we observe a correlation between Δg_{coax} and the forking free energies between the base pairs at the end of the stems and the mismatch. The excess penalty $\Delta g_{\mathrm{coax}}-2\Delta g_{\mathrm{NN}}^{\mathrm{fork}}=2.9\pm 0.4k_BT\equiv -T\Delta s_c^{im}$ ($Q=0.97$) and does not depend on possible strand extensions beyond the interface.

For bulge, internal loop mismatches or hairpin loops with a small number of nucleotides (1 nt bulge, or 1 nt × 1 nt, 1 nt × 2 nt and 2 nt × 2 nt internal loops, or 3 nt hairpin loops), free energy parameters are different from larger loops.[9] The main reason is the nonperturbation of the double-helix by the mismatched nucleotides, i.e., the two adjacent stems and the small loop belong to the same double-helical structure.[72,73] For the 1 nt bulge [Fig. 4(e)], one finds[9] $\Delta g-\Delta g_{\mathrm{NN}}^{st}=6.2k_BT\equiv -T\Delta s_b^{1}$. Using Turner data, for the inter-

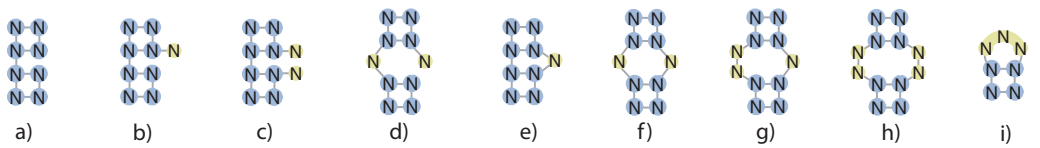

FIG. 4. Coaxial stacking and small loops situations: (from left to right) coaxial stacking without an intervening mismatch and without strand extension (a), with one strand extension (b) and with two strand extensions (c), coaxial stacking with an intervening mismatch (d), 1nt bulge (e), 1 nt × 1 nt loop (f), 1 nt × 2 nt loop (g), 2 nt × 2 nt loop (h), and 3 nt hairpin loop (i).

nal mismatches, we observe a correlation between the total loop free energy Δg and the forking free energies between the base pairs at the end of the stems and the mismatch ($Q \sim 1$ in all cases), and we find that the nucleation penalty is $\Delta g - 2\Delta g_{\mathrm{NN}}^{\mathrm{fork}} = 10.2 \pm 1.4 k_B T \equiv -T\Delta s_{\mathrm{loop}}^{1,1}$ for 1 nt×1 nt [Fig. 4(f)], $15 \pm 2 k_B T \equiv -T\Delta s_{\mathrm{loop}}^{1,2}$ for 1 nt×2 nt [Fig. 4(g)] and $11.1 \pm 2 k_B T \equiv -T\Delta s_{\mathrm{loop}}^{2,2}$ for 2 nt×2 nt [Fig. 4(h)]. For hairpin loops containing three unpaired nucleotides [Fig. 4(i)], the loop free energy only contains a nucleation penalty term[9] (forking is not considered) $-T\Delta s_{\mathrm{loop}}^{3} = 6.8 k_B T$.

The total number of possible mismatch parameters is of the order of 10^4. In the Turner model, they are derived from a large set of experimental measured parameters (~280) using a small set of assumed parameters (~10). The preceding relations between mismatch parameters and forking parameters lead to a drastic reduction in the number of independent parameters (280→4 parameters).

4. Gauge freedom for boundary and initiation terms in Turner-like models

In the Turner model, the parameterization of stacking ($\Delta g_{\mathrm{NN}}^{st}$), terminal ($\Delta g_{\mathrm{term}}$) and initiation ($\Delta g_{\mathrm{init}}$) free energies is derived from experimental results on short oligomers.[60] However, from melting experiments, it is not possible to uniquely (i.e., independently) determine the three possible Δg_{term} $(A/U, G/C, G/U)$ and Δg_{init}.[61,74,75] In the Turner gauge, $\Delta g_{\mathrm{term}}(G/C)=0$. It turns out that a different choice is more convenient in simulations of lattice models, as it allows to eliminate nonlocal loop nucleation entropies. In the following, we take a closer look at the underlying gauge freedom, i.e., we ask if there exists a set of interrelated constants Φ_{init}, Φ_{nuc}, Φ_{bound}, and Φ_{term} by which initiation, nucleation, boundary and terminal energies may be shifted without affecting predictions of the Turner model for measurable observables. In particular, we demand that for an arbitrary secondary structure composed of n_s strands forming n_{loop} loops with n_{term} external boundary terms (terminal, dangling, or external forking) and n_{bound} inner boundary terms (internal, hairpin or bulge forking)

$$\Delta G_\Phi = (n_s - 1)\Phi_{\mathrm{init}} + n_{\mathrm{bound}}\Phi_{\mathrm{bound}} + n_{\mathrm{term}}\Phi_{\mathrm{term}} + n_{\mathrm{loop}}\Phi_{\mathrm{nuc}}, \tag{6}$$

be equal to zero.

To find the relation between the various constants, it is useful to consider a number of simple cases. From the invariance of the association equilibrium of complementary oligomers, it follows immediately that $\Phi_{\mathrm{term}} = -\Phi_{\mathrm{init}}/2$. Similarly, the choice of a gauge may neither affect the free energy cost of the creation of internal bubble in a double-helical domain (so that $\Phi_{\mathrm{nuc}} = -2\Phi_{\mathrm{bound}}$) nor the formation of a hairpin in a single strand (implying $\Phi_{\mathrm{nuc}} + \Phi_{\mathrm{term}} + \Phi_{\mathrm{bound}} = 0$). These conditions are automatically fulfilled by the choice

$$\Phi_{\mathrm{init}} = \Phi_{\mathrm{nuc}} = \Phi, \tag{7}$$

$$\Phi_{\mathrm{bound}} = \Phi_{\mathrm{term}} = -\Phi/2. \tag{8}$$

Note, that Eq. (8) is consistent with our suggestion to assign identical free-energy penalties to terminal and internal forking. Using this choice and the topological property that the number of stems in a secondary structure is equal to $n_{\mathrm{loop}} + n_s - 1$ and to $(n_{\mathrm{bound}} + n_{\mathrm{term}})/2$, we find as expected

$$\Delta G_\Phi = (n_s + n_{\mathrm{loop}} - 1 - (n_{\mathrm{bound}} + n_{\mathrm{term}})/2)\Phi = 0. \tag{9}$$

To prove the topological property, one can first show by induction on the number of strands and on the number of stems that $n_{\mathrm{loop}} = n_{\mathrm{stem}} - (n_s - 1)$. In a second step, one can use the trivial property that the number of boundary terms is equal to two times the number of stems.

To summarize, using Eqs. (7) and (8), one may shift *all* initiation, nucleation and boundary terms in the Turner model by an arbitray offset of Φ and $-\Phi/2$, respectively, without affecting predictions for measurable quantities. In the Turner gauge, this is used to set $\Delta g_{\mathrm{term}}(G/C)=0$, but alternatively $\Phi = T\Delta s_{\mathrm{loop}}$ may be used to eliminate the unified nonlocal loop nucleation entropy introduced above.

C. Parameterization of the lattice model

A first naive parameterization can be obtained by simply equating corresponding lattice and Turner parameters. However, since $\Omega(\mathcal{S})$ depends on the lattice nature, thermodynamic results then also become lattice-nature-dependent. This is not physically satisfying. Moreover, the resulting predictions are far from experimental data. For example, for short hairpins, the two-state character of the melting transition is not reproduced by the naive parameterization and the computed melting temperatures are about 45 K smaller than the experimental results (see Fig. 7). Better results can be obtained by demanding that the system of interest (RNA and DNA) and our lattice model show identical behavior on the secondary structure level. This implies that the parameters in the lattice model have to be inferred from standard secondary-structure descriptions[9] by equating Turner free energies to free energies obtained from partition functions for appropriate groupings of microstates of the lattice model.[57] In the following, before relating the lattice model to DNA/RNA Turner parameters (Sec. II C 2), it is instructive to consider the adjustment of lattice parameters required by a change in the underlying lattice (Sec. II C 1).

1. Correspondences between two different lattices

Consider two versions of the lattice model (L_1 and L_2) on different lattices (simple cubic, fcc, etc.). Walks on the lattices are characterized by two different sets of constants $\{z_i, \mu_i, f_{s_i}, f_{l_i}\}$ $(i=1,2)$ defined in Sec. II A 2. For a given lattice i, the conformational entropy difference between a secondary structure $\mathcal{S}$ and the denatured state $\mathcal{S}_0$ is given by $\Delta s_i = k_B \log[\Omega_i(\mathcal{S})/\Omega_i(\mathcal{S}_0)]$. The introduction of lattice derived free-energy differences into the Turner model[15] is delicate for two reasons. First, because $\Delta s_{\mathrm{loop}}^{i} = k_B \log(f_{l_i}/f_{s_i})$ depends on the chosen lattice. Second, because the effect of supplanting $\Delta s_{\mathrm{loop}}^{\mathcal{T}}$ by $\Delta s_{\mathrm{loop}}^{i}$ is *different* in equivalent formulations of the Turner model, which make use of the various gauge freedoms discussed in Secs. II B 1 and II B 4.

To obtain a lattice independent behavior, we require that the free-energy differences between any two secondary structures be invariant under a change in the employed lattice. For

example, for the two-state equilibrium between a double-stranded stem and the two corresponding denatured single strands, we obtain the condition

$$G_{ds}(1) - 2G_{ss}(1) + G_{\mathrm{mix}_1} = G_{ds}(2) - 2G_{ss}(2) + G_{\mathrm{mix}_2}. \tag{10}$$

From Eq. (10), we obtain relations between the two sets of parameters,

$$\epsilon_2 = \epsilon_1 + 2k_BT \log(\mu_1/\mu_2), \tag{11}$$

$$\omega_2 = \omega_1 + k_BT \log\left[\frac{f_{s_1}}{f_{s_2}}\left(\frac{z_2}{z_1}\right)^{1/2}\right] + \frac{\Phi_{12}}{2}, \tag{12}$$

where $\Phi_{12}=G_{\mathrm{mix}_1}-G_{\mathrm{mix}_2}$. Applied to other typical examples described before, the same procedure yields

$$\gamma_2 = \gamma_1 + k_BT \log\left[\frac{z_2-2}{z_1-2}\left(\frac{z_1}{z_2}\right)^{1/2}\right] + \frac{\Phi_{12}}{2}, \tag{13}$$

$$\sigma_2 = \sigma_1 + k_BT \log(f_{l_2}/f_{l_1}) - \Phi_{12}, \tag{14}$$

$$\lambda_2 = \lambda_1 + k_BT \log\left[\frac{z_2-1}{z_1-1}\left(\frac{z_1}{z_2}\right)^{1/2}\right] + \frac{\Phi_{12}}{2}, \tag{15}$$

$$\chi_2 = \chi_1 + k_BT \log\left[\frac{z_1}{z_2}\left(\frac{z_2-1}{z_1-1}\right)^{2}\right] + \Phi_{12}. \tag{16}$$

The two sets of parameters are related only via the lattice constants $\{z_i, \mu_i, f_{s_i}, f_{l_i}\}$ and $\Phi_{12}=G_{\mathrm{mix}_1}-G_{\mathrm{mix}_2}$. Equation (14) shows that the loop nucleation parameter σ cannot be neglected, since f_l is a nonuniversal, lattice-independent quantity. This is a subtle point, but it highlights the importance of properly accounting for a nonlocal loop nucleation free energy in tertiary structure description of nucleic acids. In practice, it is cumbersome to work with $\sigma \neq 0$, since local changes in the association of base pairs then need to be analyzed for a change in the global topological state. Fortunately, the above analysis implies that the lattice model exhibits the same gauge freedom as the Turner model, which can be exploited to set $\sigma \equiv 0$ in a suitable "lattice loop" or $\sigma=0$ gauge.

2. Parameterization

To parameterize the lattice model, we need to equate the free energy contributions to the well-known NN-parameters.[9,60] Given the similar nature of the interactions in the lattice and the NN model, corresponding parameters are readily identified. The necessary corrections due to the conformational entropy of secondary structures in the lattice model are obtained by grouping and counting the corresponding microstates for a number of simple cases.[57] For example, for the two-state equilibrium between a double-stranded complex and the two single strands

$$G_{ds} - 2G_{ss} + G_{\mathrm{mix}} = N\Delta g^{st}_{\mathrm{NN}} + 2\Delta g_{\mathrm{term}} + \Delta g_{\mathrm{init}}. \tag{17}$$

Grouping N-dependent and boundary terms, it follows:

$$\epsilon = \Delta g^{st}_{\mathrm{NN}} - 2k_BT \log \mu, \tag{18}$$

$$\omega = \Delta g_{\mathrm{term}} + \frac{\Delta g_{\mathrm{init}} - G_{\mathrm{mix}}}{2} - k_BT \log(f_s \bar{N}^{c'}/z^{1/2}), \tag{19}$$

where $\bar{N} \approx 10$ is the typical strand size used in the NN-parameterization process.[57]

As mentioned in Sec. II C 1, a gauge freedom exists for boundary, nucleation and initiation terms in the lattice model. From now on, we work in the "lattice-loop gauge" defined in Sec. II C 1. This has the advantage to make the model strictly local and to avoid an analysis of the secondary structure for the evaluation of the lattice energy. We iterate, that this choice does not influence the computation of the thermodynamic properties. Using the same procedure as for the double-strand equilibrium in other examples described in Figs. 1 and 4, we obtain the lattice parameterization

$$\sigma \equiv 0, \tag{20}$$

$$G_{\mathrm{mix}}(T) = \Delta g_{\mathrm{init}}(T) + T\Delta s_{\mathrm{loop}} - k_BT \log f_l, \tag{21}$$

$$\epsilon(T) = \Delta g^{st}_{\mathrm{NN}}(T) - 2k_BT \log \mu, \tag{22}$$

$$\omega(T) = \Delta g_{\mathrm{term}}(T) - \frac{T\Delta s_{\mathrm{loop}}}{2} + k_BT \log\left[\frac{(f_l z)^{1/2}}{f_s \bar{N}^{c'}}\right], \tag{23}$$

$$\gamma(T) = \Delta g_{\mathrm{fork}}(T) - \frac{T\Delta s_{\mathrm{loop}}}{2} + k_BT \log\left[\left(\frac{f_l}{z}\right)^{1/2}(z-2)\right], \tag{24}$$

$$\lambda(T) = \Delta g_{\mathrm{dangle}}(T) - \frac{T\Delta s_{\mathrm{loop}}}{2} + k_BT \log\left[\left(\frac{f_l}{z}\right)^{1/2}\left(\frac{z-1}{\bar{N}^{c'}}\right)\right], \tag{25}$$

$$\chi(T) = \Delta g_{\mathrm{coax}}(T) - T\Delta s_{\mathrm{loop}} + k_BT \log\left(\frac{f_l(z-1)^2}{z\bar{N}^{c'}}\right), \tag{26}$$

where Δg_{fork} ($\Delta g_{\mathrm{dangle}}$) depends on the size of the fork (dangle) (see Sec. II B 2) and Δg_{coax} depends on the kind of coaxial stacking (with or without intervening mismatch or strand extension) (see Sec. II B 3). Generally, the lattice parameters are given by the corresponding Turner parameters modified by the entropy of the equivalent microstates in the lattice model.

For the small bulge, internal or hairpin loops studied in Sec. II B 3, we use

$$\sigma^1_b(T) = -T(\Delta s^1_b - \Delta s^2_c) + T\Delta s_{\mathrm{loop}} + k_BT \log\left[\left(\frac{z-2}{z-1}\right)^2 \bar{N}^{c'} 3^{-c}\right], \tag{27}$$

$$\sigma_{1,1}(T) = -T\Delta s^{1,1}_{\mathrm{loop}} + T\Delta s_{\mathrm{loop}}, \tag{28}$$

$$\sigma_{1,2}(T) = -T\Delta s^{1,2}_{\mathrm{loop}} + T\Delta s_{\mathrm{loop}}, \tag{29}$$

$$\sigma_{2,2}(T) = -T\Delta s^{2,2}_{\mathrm{loop}} + T\Delta s_{\mathrm{loop}}, \tag{30}$$

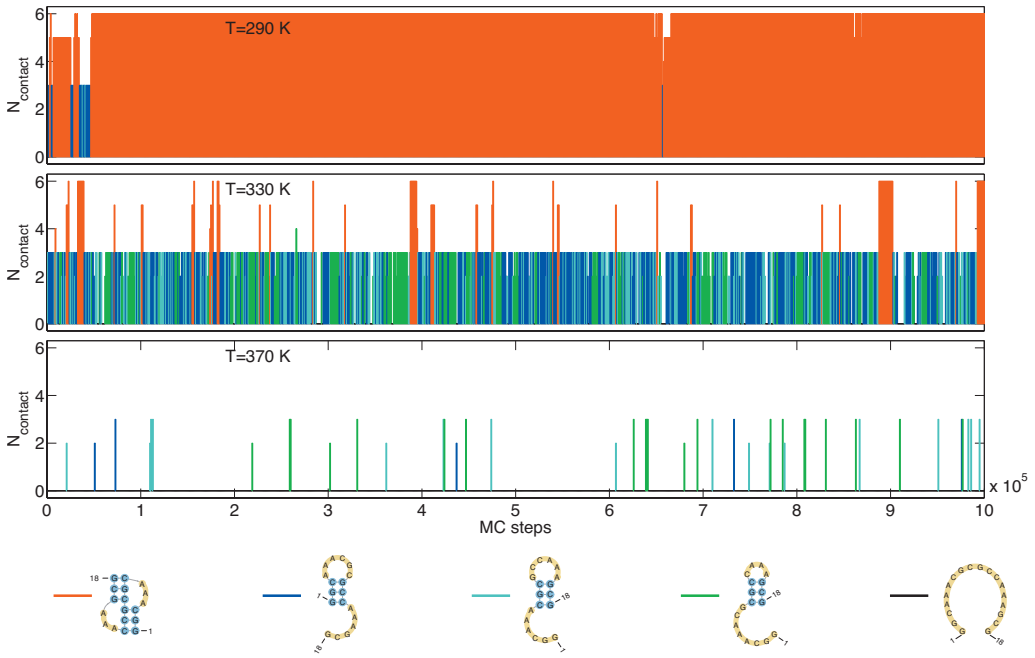

FIG. 5. Evolution of the current structure during MC runs at different temperatures for the pseudoknot structure *GGCAAACGCGCCAAAGCG* (see also Fig. 10). For every conformation, we plot its current number of base pairs N_{contact} and its secondary structure group (see color legend). At $T=290$ K, the pseudoknot structure (red) is predominant. At $T=330$ K, the current structure fluctuates a lot between the pseudoknot (red), the native hairpins (blue and cyan), and the misfolded hairpin (green). At $T=370$ K, the pseudoknot structure is no more stable and the current structure is mostly the denaturated state.

$$\sigma_3(T) = -T\Delta s^3_{\text{loop}} + \frac{T\Delta s_{\text{loop}}}{2} + k_B T \log\left(\frac{z-2}{z^{1/2}}\right), \tag{31}$$

where σ_b^1 is the nucleation penalty for the 1 nt bulge, $\sigma_{1,1}$, $\sigma_{1,2}$, and $\sigma_{2,2}$ for the internal mismatch loops, and σ_3 for the 3 nt hairpin loop.

$\Delta g^{st}_{\text{NN}}$, Δg_{init}, Δg_{term}, and $\Delta g^1_{\text{dangle}}$ are computed with the version 3.0 of the Turner RNA parameters.[9] Data for $\Delta g^{\text{fork}}_{\text{NN}}$ at 37 °C are given in Table I and data for $\Delta h^{\text{fork}}_{\text{NN}}$ are extracted from version 2.3 of the Turner parameters. Δs_{loop} is given in Eq. (5). Δg^2_{fork}, $\Delta g^2_{\text{dangle}}$, coaxial and small loops contributions are produced in Secs. II B 2 and II B 3. All these parameters are given for a salt concentration of 1*M* and to the best of our knowledge, no efficient salt correction rule exists for the RNA parameters. Like in many standard parameterizations,[9,60,79,80] enthalpies and entropies are considered temperature-independent.

Concerning the bending free energy, we modelize both the entropic and the enthalpic parts by the base-independent standard form

$$\kappa(\Psi) = \bar{\kappa}(1-\cos\Psi), \tag{32}$$

with $\bar{\kappa}(T)=\bar{\kappa}_h - T\bar{\kappa}_s$.

Moreover, using a current polymer calculation,[81] one can show that

$$c_\infty \equiv \frac{l_K}{b} = \frac{1+\langle\cos\Psi\rangle}{1-\langle\cos\Psi\rangle}, \tag{33}$$

where l_K is the Kuhn lenght[81] (two times the persistence length) and b the lattice distance between two nucleotides. For nucleic acids, at $T=300$ K and $c_\infty \approx 300$.[82]

For the fcc lattice

$$\langle\cos\Psi\rangle = \frac{1+2e^{-\beta\bar{\kappa}(T)/2}-2e^{-3\beta\bar{\kappa}(T)/2}}{1+4e^{-\beta\bar{\kappa}(T)/2}+2e^{-\beta\bar{\kappa}(T)}+4e^{-3\beta\bar{\kappa}(T)/2}}. \tag{34}$$

Then, with Eq. (33), $x \equiv e^{-\beta\bar{\kappa}(T)/2}$ has to verified

$$(3c_\infty - 1)x^3 + (c_\infty - 1)x^2 + (c_\infty - 3)x - 1 = 0. \tag{35}$$

If we assume that c_∞ follows the law

$$c_\infty(T) = \frac{300\ \text{K}}{T} c_\infty(300\ \text{K}), \tag{36}$$

we can numerically solve Eq. (35) and we find that

$$\bar{\kappa}_h = 646 \pm 3\ k_B\ \text{K} \quad \text{and} \quad \bar{\kappa}_s = -9.23 \pm 0.01\ k_B. \tag{37}$$

D. Numerical simulation techniques

To sample the canonical ensemble corresponding to the described lattice model at a given temperature T, we use importance sampling MC techniques.[83] From a random initial conformation, successive trial configurations are generated using the pivot algorithm,[84] short segment transformations[85] or reptation moves.[86] In the pivot algorithm, a trial move consists of randomly picking a nucleotide along the chain and applying a symmetry operation of the lattice (reflection and rotation) to the segment between the chosen nucleotide and the end of the chain. The short segment transformations modify chain segments between the two randomly chosen points. There are different classes of transformation: inversion of the sequence steps, reflection or interchange of step coordinates. Reptation moves slide local conformations along the chain. A trial move is accepted according to the Metropolis rules.[87] An MC step is composed by one pivot move, one short segment transformation, and

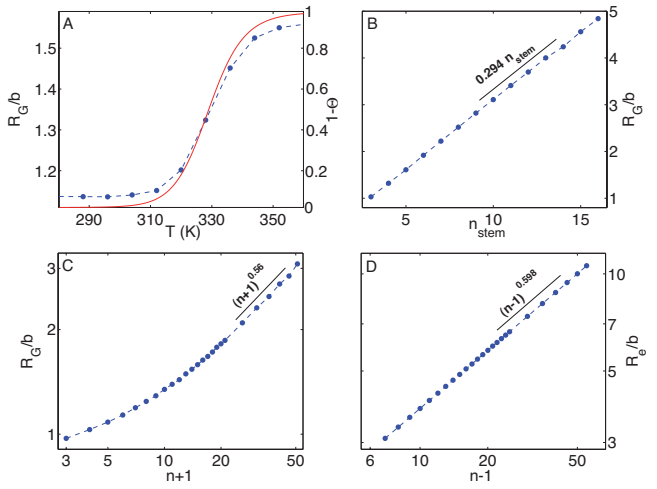

FIG. 6. (a) Evolution of the radius of gyration R_G (in lattice distance unit b, blue dots) and of the opening probability $1-\Theta$ (red line) as a function of T for the native hairpin sequence $GGCA_5GCC$. During melting, conformations change from a native compact structure to a random coil. (b) R_G/b as a function of the stem length n_{stem} for the hairpin sequences $GC_nGCAAAGUG_nC$, at low temperatures where only the native structure is present. R_G depends linearly on n_{stem} with a slope of $0.294b$, characteristic of rigid rods. Indeed, considered stems are smaller than the persistence length of nucleic acids ($\sim$150 bp) and can be considered as a rigid rod. (c) Log-log plot of R_G/b as a function of the loop size $n+1$ for the native structures of the hairpin sequences $GGCA_nGCC$, at low temperatures where only the native structure is present. We find that approximately $R_G \sim b n_{\mathrm{loop}}^{0.56}$. The exponent (0.56) is smaller than the purely self-avoiding loop exponent (0.593) (Refs. 66 and 88) due to the relative compactification imposed by the rigid stem. (d) Log-log plot of the mean end-to-end distance R_e as a function of $n-1$ for denatured state containing n nt at high temperatures, where only the denatured structure is present. We retrieve the SAW scaling law $R_e \sim (n-1)^{c/3}$.

one reptation move. Using this algorithm, we are able to measure global and local, thermodynamic as well as structural properties of the chain at a given temperature T. Figure 5 illustrates the evolution of the secondary structure of a pseudoknot sequence during typical MC runs at different temperatures. The corresponding correlation times are $\sim 10^5$ MC steps (with a average acceptance rate for a trial move of 0.27) at 290 K, $\sim 10^4$ MC steps (0.33) at 330 K, and $\sim 10^3$ MC steps (0.53) at 370 K. Figure 6 shows that we reproduce the expected polymeric behavior of structural properties such as the radius of gyration or the end-to-end distance.

By repeating the procedure for several temperatures distributed over the whole temperature region of interest, we extract the number of state for all relevant secondary structures with a multiple histograms reweighting method.[83,89] At each temperature T_i, we store the histogram $\mathcal{N}_i(\mathcal{S}_j)$ of the visited secondary structures $\mathcal{S}_j$

$$\frac{\mathcal{N}_i(\mathcal{S}_j)}{\mathcal{N}_i^{\mathrm{tot}}} = \frac{\Omega(\mathcal{S}_j)\exp\{-\beta_i[h_{\mathrm{latt}}(\mathcal{S}_j) - T_i s_{\mathrm{latt}}(\mathcal{S}_j)]\}}{Z_i}, \tag{38}$$

where $\mathcal{N}_i^{\mathrm{tot}}$ is the total number of entries in the histogram, and $h_{\mathrm{latt}}(\mathcal{S}_j)$ and $s_{\mathrm{latt}}(\mathcal{S}_j)$ are the enthalpy and the entropy of $\mathcal{S}_j$ evaluated with the lattice parameters (Sec. II C 2). For example, for the hairpin in Fig. 1(c), $h_{\mathrm{latt}} = \omega_h + 2\epsilon_h + \gamma_h + \sigma_h$ (*idem* for s_{latt}). We combine all the histograms to compute the number of lattice conformations, $\Omega(\mathcal{S}_j)$, of each encountered secondary structure $\mathcal{S}_j$ of a given oligonucleotide

$$\Omega(\mathcal{S}_j) = \frac{\Sigma_{T_i}\mathcal{N}_i(\mathcal{S}_j)}{\Sigma_{T_i}\mathcal{N}_i^{\mathrm{tot}}\exp\{-\beta_i[h_{\mathrm{latt}}(\mathcal{S}_j) - T_i s_{\mathrm{latt}}(\mathcal{S}_j)]\}/Z_i}, \tag{39}$$

where the Z_i are defined by the set of implicit equations

$$Z_i = \sum_{\mathcal{S}_j}\left(\frac{\Sigma_{T_k}\mathcal{N}_k(\mathcal{S}_j)}{\Sigma_{T_k}\mathcal{N}_k^{\mathrm{tot}}\exp\{-\beta_k[h_{\mathrm{latt}}(\mathcal{S}_j) - T_k s_{\mathrm{latt}}(\mathcal{S}_j)]\}/Z_k}\right) \times \exp\{-\beta_i[h_{\mathrm{latt}}(\mathcal{S}_j) - T_i s_{\mathrm{latt}}(\mathcal{S}_j)]\}. \tag{40}$$

Note, that we can eliminate the, *a priori* unknown prefactor in $\Omega(\mathcal{S})$, by the constraint that $\Omega(\mathcal{S}_0)$ for the denatured, unpaired single strands equals the known number $\Omega_0(N)$ of SAWs on the lattice in question. Using this information, properties such as temperature dependent contact maps can be computed for an arbitrary temperature without requiring any additional simulations. In particular, for an arbitrary temperature T, the free-energy difference ΔG between a secondary structure $\mathcal{S}$ and the denatured state $\mathcal{S}_0$ is given by the sum of the corresponding interaction free energies (stacking, forking, nucleation,…) and of the conformational free-energy difference

$$\Delta G(\mathcal{S},T) = h_{\mathrm{latt}}(\mathcal{S}) - T[s_{\mathrm{latt}}(\mathcal{S}) + k_B \log(\Omega(\mathcal{S})/\Omega_0)]. \tag{41}$$

While, the multiple histograms reweighting method is classically applied to system characterized by a potential energy described by one or several order parameters, the originality of our approach is to construct the histograms relatively to the secondary structures and not the energy, and to work with a Hamiltonian depending on the temperature.

Typically in our simulations, the studied temperatures range is from 273 to 380 K every 5–8 K. For each temperature, we accumulate the histograms of 5–10 different runs containing 10^7 MC steps each. This corresponds to at least 10^3 independent measurements (by comparison to the correlation times). For a sequence with 20 nucleotides, the multiple histograms method takes $\sim$2 hours to run on a 2.4 GHz computer for 13 scanned temperatures and 5×10^7 MC steps at each temperature. Such conditions of simulation lead to small statistical errors. For example, typical statistical error for the free energy of a given secondary structure is $0.05k_BT$, reaching to a typical statistical error of 0.2 K for melting temperatures.

The present algorithm becomes inefficient for sequence length longer than $\sim$50–70 nts, where chains become trapped in secondary minima of the free energy landscape.[40] Using our trial moves, it is not possible to leave these states without crossing high free energy barriers due to transition states containing unstable short stems with high nucleation free-energy penalties.

III. RESULTS AND DISCUSSIONS

A. Validation of the model on simple hairpins

Hairpins are among the simplest secondary structures for nucleic acids. They are composed of a double-stranded stem closed by a single-stranded loop. Their biological roles are numerous and important: regulation in transcription and replication, *in vivo* nucleic acids protection and stabilization, mutagenesis facilitation as well as tertiary contact initiation

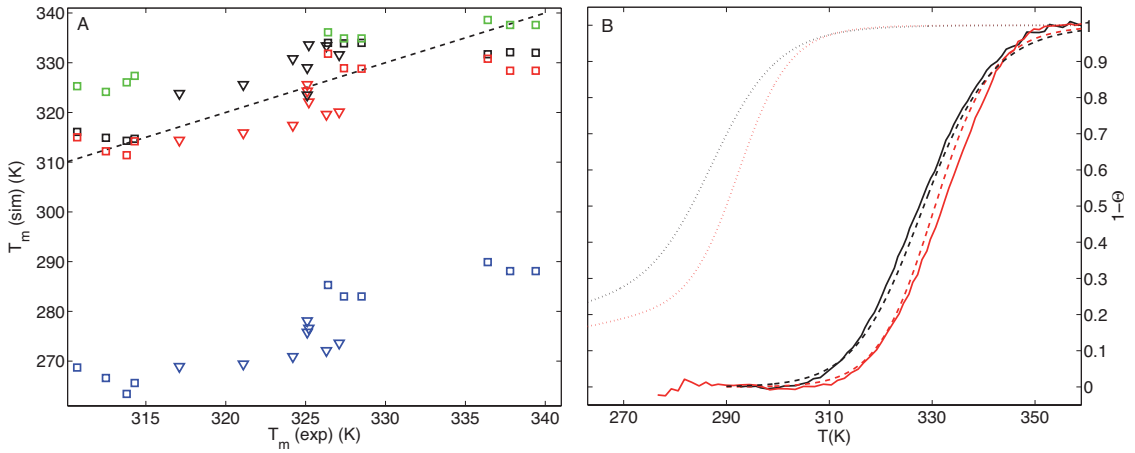

FIG. 7. (a) Computed melting temperatures T_m(sim), using the lattice model with the naive parameterization (blue) or the correct parameterization (red) or standard programs: RNAfold (Ref. 13) or DINAmelt (Ref. 12) (black) and Kinefold (Ref. 21) (green), as a function of the experimental melting temperatures T_m(exp) for ten RNA hairpins experimented by Serra *et al.* in Ref. 76 (squares) and eight DNA hairpins experimented by Hilbers *et al.* in Ref. 77 (triangles). $\sigma_{T_m} \equiv [\langle (T_m(\text{sim}) - T_m(\text{exp}))^2 \rangle]^{1/2} = 46.7$ K for the lattice model data using the naive parameterization, 5.4 K using the correct parameterization, 5.5 K for RNAfold, and 9.3 K for Kinefold. (b) Experimental (full lines) or computed (correct parameterization: dashed lines, naive parameterization: dotted lines) probability $1-\Theta$ that the hairpin is open for the sequences *GGCAUUAGCC* (red) and *GGGAUAUACCC* (black). Experimental data are extracted from Ref. 76 and normalized using the proceeding described in Ref. 78.

in ribozymes.[90–93] Hairpins have been extensively studied thermodynamically[76,77] and mechanically[94,95] to parameterize the NN-model. Testing our model on these simple molecules is therefore an important step in its validation.

Short hairpins exhibit a two-state transition from the paired complex to the denatured single strand.[60] Figure 7(a) compares experimental melting temperatures for DNA and RNA hairpins with a two-state melting character to computed values using the lattice model, typical programs using the full Turner model [DINAMELT (Ref. 12) or the VIENNARNA package (Ref. 13)] and the Kinefold application.[21] The typical deviations from the experiments are approximately equal to 5 K for the lattice and the full Turner model and to more than 8 K for the Kinefold model. The excellent agreement between experimental results and the predictions of the lattice model is underlined by Fig. 7(b), which shows temperature-dependent melting curves for two RNA hairpins. To justify our lengthy derivation of the parameters of the lattice model, we have also included result for a "naive" parameterization, which simply identifies interaction free energies on the lattice with the corresponding NN terms without correcting for the conformational entropy of secondary structures in the lattice model. Using this naive parameterization, the model predicts a non-two-state transition at temperatures which deviate by more than 40 K from the experimental value.

B. Multibranch loops

Multibranch loops are present in many natural nucleic acids molecules such as tRNA,[2] and particularly in native structures of long sequences where the complexity of the secondary structure involves many interconnected stem-loop substructures. Figure 8 shows the evolution of the free-energy difference ΔG between the native structure and the denatured state at $T=310$ K for three kinds of loops with a increasing number of connected stems (hairpin, internal, and multibranch) as a function of the size of the central loop, computed with the lattice model, standard programs using the full Turner model (RNAfold and Mfold), Kinefold, and Vfold.

For hairpin and internal loops [Figs. 8(a) and 8(b)], the lattice and the Turner model (via RNAfold) give similar results whatever the employed version of Turner parameters. This confirms our derivation of the lattice forking and loop nucleation parameters from the corresponding Turner parameters. Kinefold and Vfold, which continue to use the older version 2.3 (Ref. 37) of the Turner parameters, provide reasonable estimates for hairpin and internal loops. We note however, that some care must be taken when upgrading Kinefold and Vfold to new versions of the Turner parameters. In both models the free energy is calculated as $\Delta G = \Delta G_{\text{stack}} - T\Delta s_{\text{loop}}$, where the employed estimates of the loop conformational entropy Δs_{loop} (including the nucleation penalty) are computed independently of the Turner parameters and the employed gauge, and ΔG_{stack} contains gauge-dependent stacking parameters (such as association or forking free energies). Obviously, $\Delta s_{\text{loop}} \approx \Delta s_{\text{loop}}^{T} - k_B c \log N_{\text{loop}}$ in version 2.3 of the Turner parameters where $\Delta s_{\text{loop}}^{T} = 6k_B$ for internal loops and $4.6k_B$ for hairpin loops. However, in the present version 3.0,[9] $\Delta s_{\text{loop}}^{T} = 0.1k_B$ for internal loops and $6.5k_B$ for hairpin loops, i.e., the corresponding substitution by an invariant estimate Δs_{loop} is bound to lead to much stronger deviations from the RNAfold estimate.

For multibranch loops [Fig. 8(c)], for lack of theoretical input, Turner models typically *assume* a ΔG independent of loop size. This is *not* confirmed by the lattice model, which predicts that ΔG can vary by $3-4k_BT$ between $N_{\text{loop}}=9$ and $N_{\text{loop}}=30$ with an even stronger variation for shorter loops. Interestingly, Kinefold reproduces the absolute magnitudes of the looping free energies for intermediate loop sizes fairly well. Not surprisingly, in all three cases we observe $c \approx 1.5$, i.e., the looping exponent for random walks.

Closer inspection of Figs. 8(a) and 8(b) shows that the lattice model clearly supports the description of large hair-

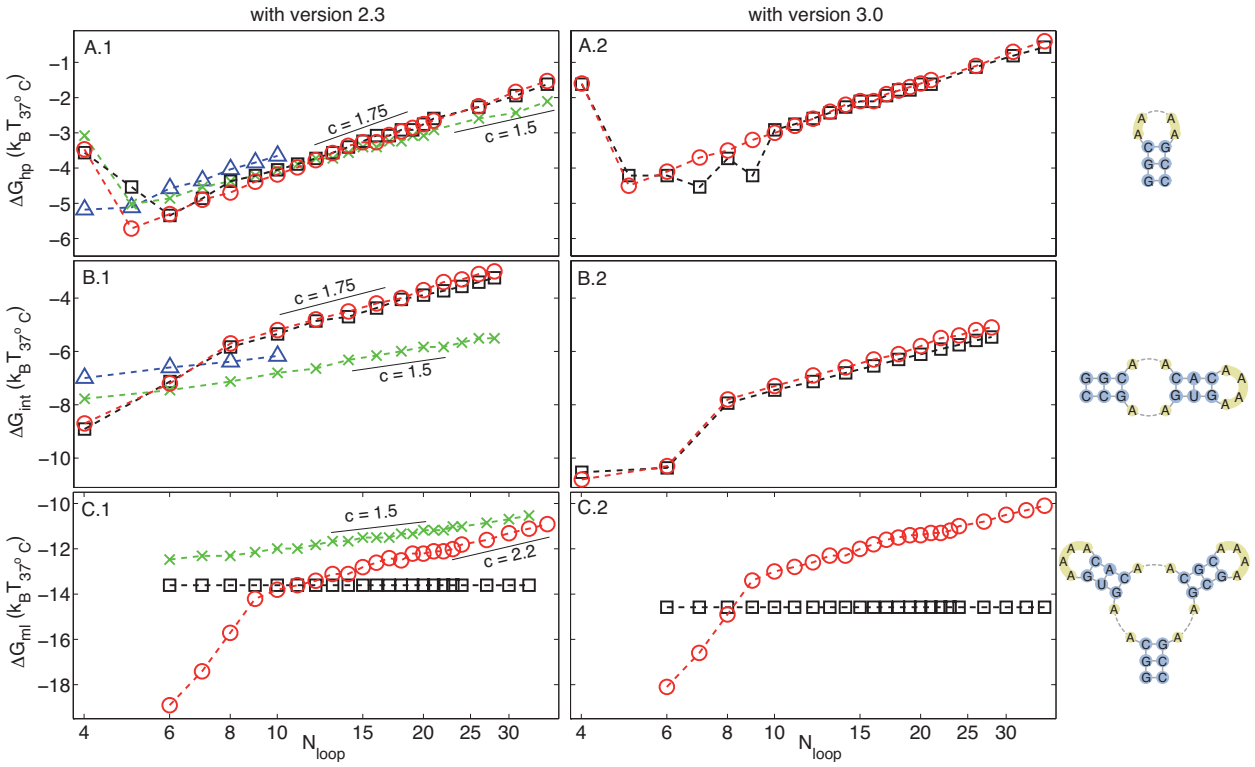

FIG. 8. Free energy ΔG as a function of the loop size N_{loop} for three kinds of branch loop and for two versions of the Turner parameters: hairpin (a), internal (b), and multibranch (c), computed with the lattice model (red circles), the RNAfold model (Ref. 13) (black squares), the Kinefold model (Ref. 21) (green crosses) and the Vfold model (Ref. 15) (blue triangles) and using version 2.3 (Ref. 37) (first column) or version 3.0 (Ref. 9) (second column) of the Turner parameters. Data for RNAfold and Kinefold were obtained from their respective web servers. Data for Vfold were computed by adding the contribution of the canonical and mismatch stacking, and of the conformational entropic free energy extracted from Fig. 3b in Ref. 15.

pins and large internal bubbles via a Jacobson–Stockmayer equation with $c=1.76$ (Ref. 42) (the deviations for hairpins with $N_{\mathrm{loop}}=7$ and 9 steps are due to short loop corrections in the Turner model not accounted for in the lattice model). In the present example of a multibranch loop [Fig. 8(c)], the situation is less clear, but the observed effective exponent of $c\sim 2.2$ is in good agreement with theoretical predictions for chains with sterically interacting loops.[43,63] From simulations for different kinds of multiloops, we find that Δg_{loop} for multiloop approximately verifies the generalized Jacobson–Stockmayer equation

$$\Delta g_{\mathrm{loop}}^{\mathrm{genJS}} = (7.2 + 0.65h + 2.2 \log N_{\mathrm{loop}}) k_B T, \tag{42}$$

$$= (4.4 + 0.4h + 1.36 \log N_{\mathrm{loop}}) \ \text{kcal/mol at } 37\ ^\circ\mathrm{C}, \tag{43}$$

where N_{loop} is the loop size and h is the number of connected stems. Equation (42) is valid with the unified set of forking energies defined in Sec. II B 1. However, it is not obvious that a single asymptotic[43] or heuristic[16] value of c faithfully describes arbitrary multiloop geometries. For example, the effective value of c should depend on the relative size of the central loop and the surrounding substructures with c again tending to the self-avoiding loop exponent 1.76 for large central loops.

For multibranch loops, we test the predictive power of the lattice model on two tRNA sequences by computing the sensitivity *SE* and the specificity *SP* of the base pairs in the most stable secondary structure predicted by the lattice model compared with the experimentally determined native structure. *SE* is defined as the ratio between the number of correctly predicted base pairs and the number of measured base pairs in the most stable structure; and *SP* is defined as the ratio between the number of correctly predicted base pairs and the number of predicted base pairs. The native secondary structures of such sequences are very well predicted, even better than other models (see Table II).

Using the lattice model, we are not limited to predictions

TABLE II. Predictive power (sensitivity *SE*/specificity *SP*) of the lattice model and other standard methods tested on two transfer RNAs at $T=37\ ^\circ$C.

Sequence	Lattice model	RNAfold[a]	Kinefold[b]	PknotsRG[c]	Nupack[d]
tRNA-phe1 of yeast	1/1	0.95/1	0.9/0.83	0.24/0.24	0.95/0.87
tRNA-ala1 of human	1/0.95	1/1	1/0.84	0.95/0.95	0.95/0.8

[a]Reference 13.
[b]Reference 21.
[c]Reference 33.
[d]Reference 34.

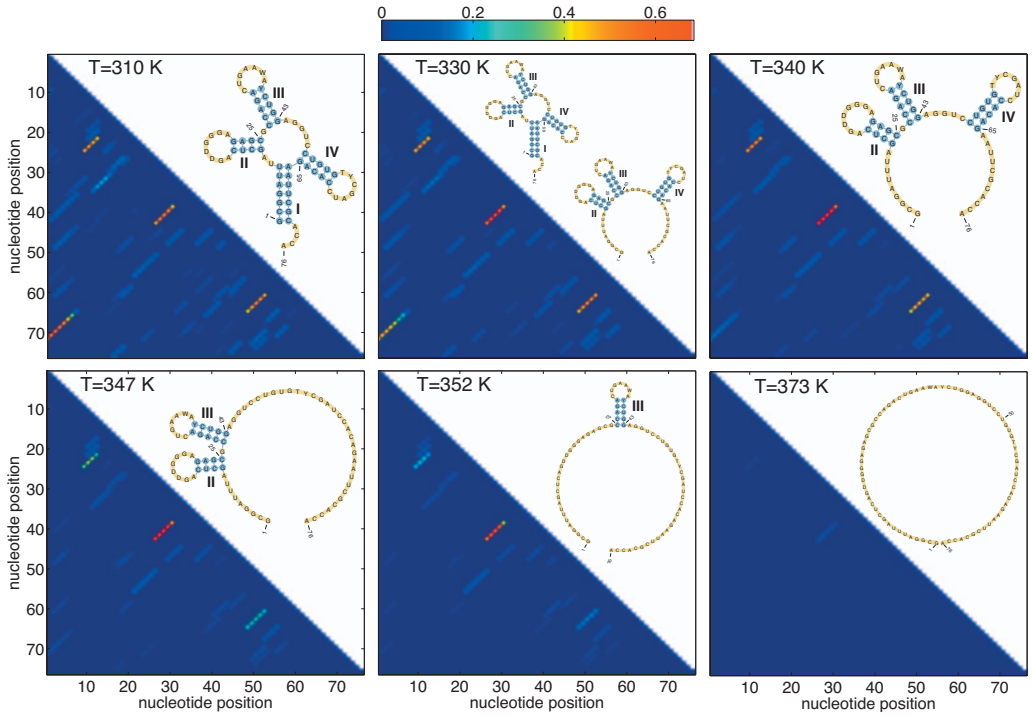

FIG. 9. Illustration of the equilibrium folding pathway of tRNA-phe1 of yeast. At different temperatures, we plot the contact map and the more stable structures. The contact map represents the probability that two nucleotides are paired in a given configuration. The color legend is given at the top of the figure.

of native structures, but we can study the equilibrium folding pathway of tRNA-phe1 of yeast. Figure 9 shows the contact map and the most stable structure at different temperatures. Around 37 °C, the four native stems present in the experimental native structure are well predicted by the lattice model. As we increase the temperature, we observe the unfolding of the molecule. This process is shown to occur in 4 steps characterized by the successive opening of the native stems. (1) Around 57 °C, stem I opens. At a first sight, it is surprising that the longest stem (7 bps), which is moreover coaxially stacked with stem IV, should denature first. The explanation is found in the analysis of the entropic contributions. Breaking stem I opens the central multibranched loop which carries a large entropic penalty. In contrast, the denaturation of one of the other stems would only increase the size of the multibranch loop and increase the entropic penalty [see Eq. (42)]. (2) Around 72 °C, stem IV opens. Stems II and III are still stabilized by coaxial stacking. (3) Then, at 76 °C, stem II denatures, while the nucleotides in stem III retain a high contact probability (~0.5). (4) Finally, around 82 °C, the random coil state dominates. Therefore, along the equilibrium folding pathway, the anticodon loop (stem III) is the first formed. Then, the TϕC loop (stem II), the D loop (stem IV) and finally, the acceptor stem (stem I).

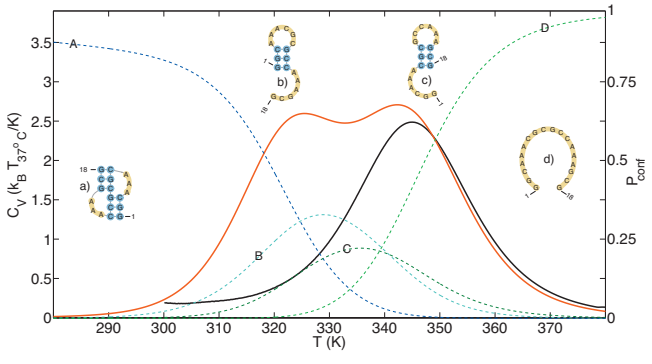

FIG. 10. Evolution of the heat capacity C_V (full lines) for the H-pseudoknot sequence *GGCAAACGCGCCAAAGCG* computed with the lattice model (red) or with the RNAheat program (Ref. 13) (black). The dashed lines represent the probability for the occurrence of the most important secondary structures: pseudoknot A (blue), hairpin B (cyan), hairpin C (dark-green), and random coil D (light-green). The two peaks in the C_V curve correspond to the transition between the native pseudoknot (a) and the hairpin (b) and to the transition between hairpins [(b) and (c)] and the random coil (d).

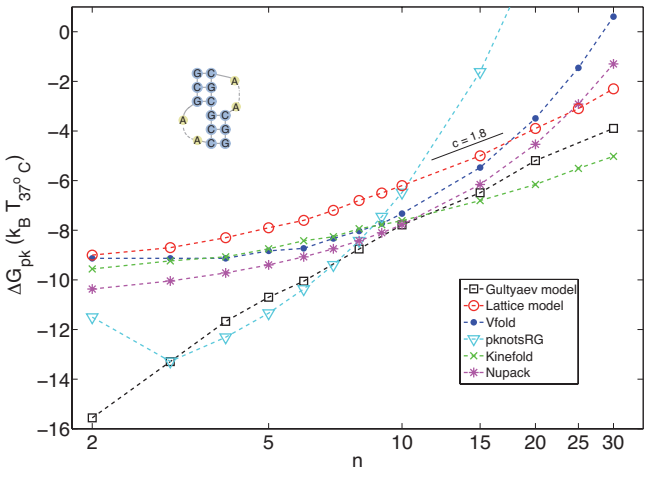

FIG. 11. Free energy ΔG_{pk} as a function of the loop size n of the H-pseudoknot $GGCA_nUCGGCCA_nCGA$ (n_1=n_2=3 bp and l_1=l_2=n nt) computed with different models: the lattice model, Gultyaev *et al.* model (Ref. 38), Vfold model (Ref. 39), PknotsRG model (Ref. 33), Kinefold model (Ref. 21), and Nupack model (Ref. 34). The typical error bars for ΔG_{pk} is approximately $2k_BT_{37\,°C}$ and is mainly due to parameters uncertainty. Data for pknotsRG, Kinefold and Nupack were obtained from their respective web servers. Data for Gultyaev and Vfold models were computed following the rules described in Refs. 38 and 39, respectively.

TABLE III. Predictive power (sensitivity *SE*/specificity *SP*) of the lattice model (LM) and other standard methods tested on truncated sequences of viral frameshift signals (Ref. 103). We give the computed free-energy difference ΔG between the predicted stablest structure and the H-pseudoknot experimental native structure described by $n_1/l_1/n_2/l_2$ (* presence of an unpaired nucleotide between the two stems).

Abbreviation	T (°C)	Truncated sequences	$n_1/l_1/n_2/l_2$	SE/SP (LM)	ΔG ($k_BT_{37°}$)	RNAfold[a]	Kinefold[b]	Vfold[c]	PknotsRG[d]	Nupack[e]
BChV	25	*G*1595–*C*1620	4/1/4/8*	1/1	0	0/0	0.5/0.57	1/1	1/1	1/1
BLV	37	*G*1604–*U*1630	6/5/3/4	0.67/1	5.5	0.67/1	0.67/1	0.67/0.86	1/1	0.67/1
BWYV	25	*C*1566–*G*1591	5/2/4/6	1/1	0	0.56/1	1/1	1/1	1/1	1/1
BYDV-NY-RPV	25	*G*1706–*C*1732	5/2/4/7	0.33/0.33	1.9	0/0	0/0	1/1	0/0	1/1
CABYV	25	*G*1494–*C*1520	5/2/3/8*	0.38/0.38	5.5	0/0	0/0	0/0	1/1	1/1
EIAV	37	*G*1797–*C*1831	6/3/4/12	1/1	0	0.5/0.71	1/1	1/1	1/1	0.9/1
FIV	37	*G*1893–*C*1927	5/2/6/11	0.82/1	3.6	0.45/1	1/1	1/1	1/1	1/1
MMTVgag/pro	37	*G*2090–*U*2123	5/1/8/8*	0/0	3.8	0/0	0.92/1	1/1	1/1	0.42/0.5
PEMV	25	*U*2042–*C*2069	6/2/4/6	0.9/1	0.3	0.6/1	0.9/1	0.9/1	1/1	0.9/1
PLRV-S	25	*G*1781–*G*1806	4/2/4/8	1/1	0	0.5/1	0.5/0.57	1/1	1/1	1/1
PLRV-W	25	*G*1676–*G*1701	4/2/3/9*	0/0	2.2	0.5/1	0/0	1/0.88	1/1	1/1
SRV1gag/pro	37	*G*2337–*C*2373	6/1/6/12	0.83/1	6.8	0/0	1/1	1/1	1/1	1/1

[a]Reference 13.
[b]Reference 21.
[c]Reference 39.
[d]Reference 33.
[e]Reference 34.

C. Pseudoknots

Pseudoknots are more complex molecules. They are present in catalytic cores in ribozymes, self-splicing introns and telomerases, or inducement of ribosomal frameshifts.[96–100] The presence of base pairs which violate of the nesting convention complicates the secondary structure prediction by highly increasing the considered configuration space.

Standard examples are H-type pseudoknots (Fig. 10). They are composed of two stems *S*1 and *S*2 (containing, respectively, n_1 and n_2 base pairs), and of two loops *L*1 and *L*2 (containing, respectively, l_1 and l_2 nucleotides). RNA molecules forming such structures often exhibit a non-two-state thermal transition[101] between the native H-pseudoknot, the two corresponding hairpins, the denaturated state or other intermediate states. Figure 10 illustrates this behavior for an example studied via the lattice model. Figure 11 shows the loop-size dependence of ΔG_{pk}, the free-energy difference be-

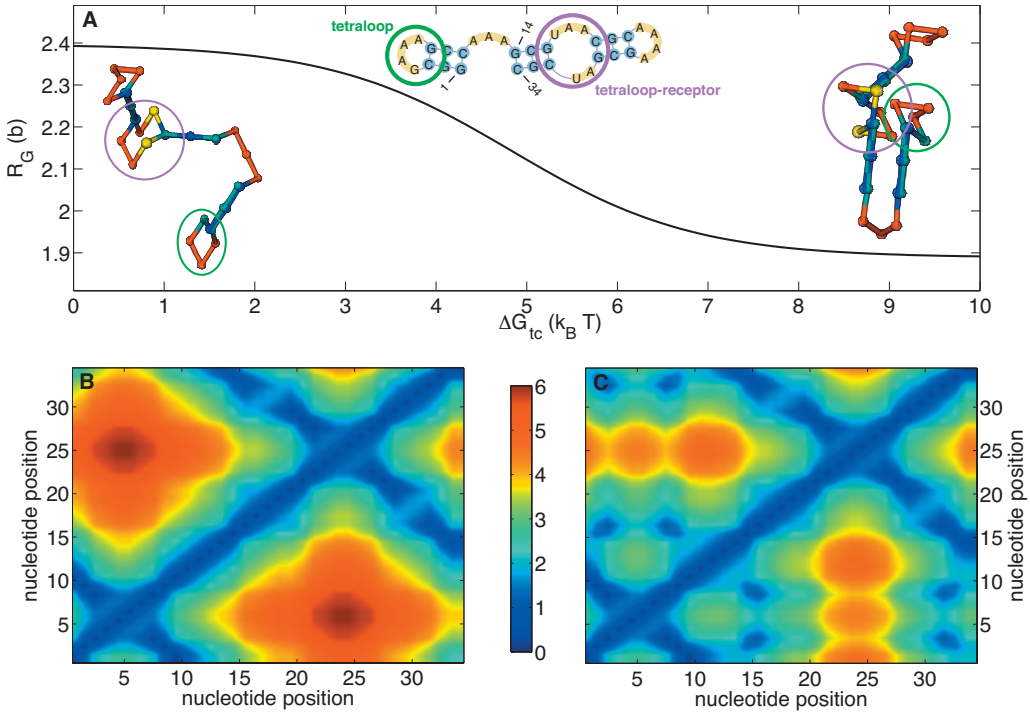

FIG. 12. (a) Two-state transition for the radius of gyration R_G (in b unit) as a function of the tertiary contact interaction free energy ΔG_{tc}. The secondary structure of the studied T/Tr_3 sequence is drawn in the middle. Examples of tertiary structure representation with $\Delta G_{tc}=0k_BT$ (left) and $-10k_BT$ (right) are also illustrated. [(b) and (c)] Average distance (in b unit) between all pairs of nucleotides for $\Delta G_{tc}=0k_BT$ (b) and $-10k_BT$ (c).

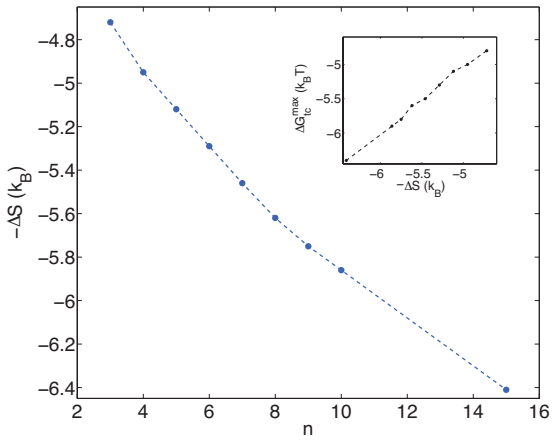

FIG. 13. Entropy difference $-\Delta S$ between the packed and the unpacked molecules as a function of the linker size n. (Inset) The tetraloop/tetraloop-receptor interaction $\Delta G_{tc}^{\max}$ at the collapse-transition as a function of $-\Delta S$.

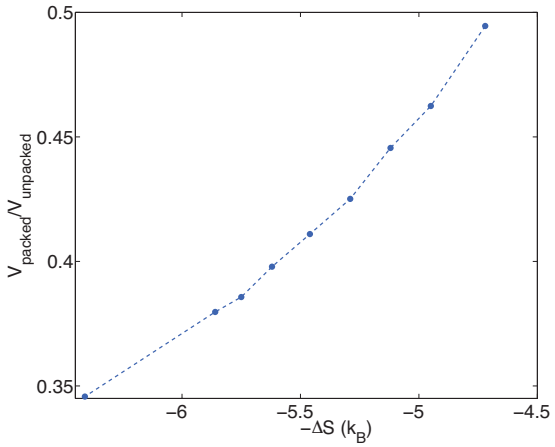

FIG. 14. Packing volume ratio between the packed and the unpacked molecules as a function of the entropy difference $-\Delta S$.

tween the pseudoknotted structure and the coil state, for $n_1 = n_2 = 3$. In the same figure, we have also included results from other models. To appreciate the excellent overall agreement, it is worthwhile to recall the different methods used to parameterize the models: The parameters of Gultyaev *et al.*,[38] of the pknotsRG model,[33] and of the Nupack model[34] are purely phenomenological and were adjusted to correctly predict experimentally or phylogenetically known pseudoknotted secondary structures and to avoid the prediction of spurious pseudoknots in unpseudoknotted test secondary structures. Vfold[39,47] parameters are derived from a more detailed microscopic lattice model,[39,47] while Kinefold[19] makes predictions on conformational entropies by modeling stems as rigid rods and unpaired loops as Gaussian chains, neglecting excluded volume interactions. For loop sizes $n \approx 10$, all methods yield nearly identical predictions. For smaller loops, the predictions of the lattice model agree with those from Refs. 21, 34, and 39 and shows a similar large difference to the predictions from Refs. 33 and 38. For large pseudoknots, the lattice model predicts in agreement with Ref. 38 that the asymptotic behavior follows a Jocobson–Stockmayer equation with an exponent $c \approx 1.8$, very close to the single self-avoiding loop exponent. This means that there is no significant steric interaction between $L1$ and $L2$. As expected, Kinefold[21] predicts a weaker loop size dependence with $c = 1.5$. PknotsRG,[33] Vfold,[39,47] and Nupack[34] predict larger free-energy penalties for the pseudoknot formation.

Frameshifting consists in inducing ribosomes to slide into alternative reading frames to product alternative proteins. This phenomenon is often found in retroviruses and could be caused by the presence of a pseudoknot downstream of the ribosome.[100,102] To test the predictive power of the lattice model on pseudoknots, we study the native structure of short, experimentally known, pseudoknots inducing frameshifting. We choose the same set of short truncated sequences of viral ribosomal frameshift signals taken from the Pseudobase database[103] than Vfold. We search for their native structure at 37 °C for animal virus and at 25 °C for plant virus.

In all cases, the native structure is generated with good statistics in our ensemble. Table III compares predicted and measured structures by computing the sensitivity *SE* and the specificity *SP*. We remark that pknotsRG and Nupack perform better in predicting the experimentally observed native structure than the lattice model, Kinefold or Vfold. However, the comparison is biased because, the present sequences were most likely among those used to fit the parameters of both models. We also note the weaker performance of pknotsRG and Nupack in the prediction of multiloop structures (Table II). In particular, Nupack predicts a lowest-energy state containing a pseudoknot, which is not observed experimentally.

Among the methods based on physical models for conformational entropies, Vfold performs best, followed by our lattice model and Kinefold.

Taking a closer look at the lattice model, we remark that it fully predicts four native structures (BChV, BWYV, EIAV, and PLRV-S), and almost reproduces the correct secondary structure for three other frameshift signals (FIV, PEMV, and SRV1gag/pro). For these seven sequences, no spurious base pairs are predicted, while a small number of base pairs are missing, because they are highly geometrically constrained in the lattice model. For other sequences, the lattice model fails partially (BLV, BYDV-NY-RPV, and CABYV) or completely (MMTVgag/pro and PLRV-W) to predict the native structure.

One possible reason for these misleading results could be the omission of stabilizing effects. For CABYV, only the models including fitted pseudoknot parameters (Nupack and pknotsRG) well predict its native structure. Unsatisfactory predictions of the lattice model, Kinefold and Vfold could be due to neglected weak tertiary interactions (such as base triples) whose effects could be effectively incorporated in the fitted parameters of Nupak and pknotsRG. For BLV, almost all the models fail to predict the native structure meaning the possible omission of stronger tertiary interactions. For BYDV-NY-RPV and PLRV-W, while Vfold succeeds, the lattice model and Kinefold fail. This could mean that ingredients present in Vfold (such as the excluded volume at the helix-loop junction or the description of the 3D structure of

RNA helix) and neglected in the lattice model and in Kinefold could explain their misleading results for these sequences.

Another reason could be the underestimation of the free energy for very short pseudoknot loops induced by the geometrical constraints imposed by the lattice. This apply for MMTVgag/pro whose known secondary structure is a H-pseudoknot with $n_1=5$, $n_2=8$, $l_1=1$, and $l_2=8$. Due to the geometry of the fcc lattice, constructing conformations on the lattice describing such a secondary structure would need the insertion of kinks in $S1$ or $S2$. These kinks are lattice artifacts and are energetically highly unfavorable, preventing the acceptance of such conformations.

D. Effect of tertiary contacts on RNA folding

To illustrate the advantage of using a 3D structure modeling, we briefly study the effect of tertiary contacts to stabilize a compact native structure. The RNA folding is partly hierarchical: under typical conditions (temperature and salt concentration) the single strands first fold into intermediate states characterized by elements of the double-helical secondary structure. Then, if the salt concentration is higher enough, tertiary contacts are formed between secondary substructures and lead to the fully folded tertiary structure.[104,105] Possible tertiary contacts are numerous[62] and mostly are very sensitive to the cation salt concentration [especially Mg^{2+} (Refs. 104 and 106)]. An example of a relevant tertiary contact is the tetraloop/tetraloop-receptor contact present in the P4-P6 domain of the extensively studied *Tetrahymena* group I ribozyme.[59,107,108] It connects the L5b tetraloop (a *GAAA* hairpin loop) to the J6a/b receptor (a *UAA*/*AU* internal loop).[59]

To be as general as possible, we study the model tetraloop/tetraloop-receptor (T/Tr) sequence $GGCGA_3GCCA_nGCGUAACGC_4GCGAUCGC$ (T/Tr_n), where n allows us to vary the size of the linker between the tetraloop and the receptor (see Fig. 12). We assume that a tertiary contact is possible when the two structures are close enough (we arbitrarily choose a distance of b between the barycenters of the tetraloop and of the general observations are not affected by this choice). Figure 12(a) shows the compaction of the complex as the T/Tr interaction free energy ΔG_{tc} increases. The evolution of the radius of gyration is characteristic of a two-state collapse transition. Figures 12(b) and 12(c) represent the effect of the compaction on the average distances between nucleotides. We observe that the inter-nucleotide distances decrease as the strength of the T/Tr interaction increases, signature of a general packing of the molecule. In particular, in the packed structure [Fig. 12(c)], stems are preferentially parallel, in agreement with experimental observations of the P4-P6 domain of *Tetrahymena* group I ribozyme.[59,107]

For each linker size n, we can estimate the ratio r of conformations on the lattice, where a T/Tr link can be established, i.e., the conformations where the distance between the tetraloop and the receptor is smaller than b, and which are considered as packed conformations. Then, we define the collapse entropy difference between the packed and the unpacked structures as $-\Delta S=-\log r$. Figure 13 shows the decreasing evolution of $-\Delta S$ as a function of n. This means that the probability to form a T/Tr contact is small when the number of nucleotides between the tetraloop and the receptor is high. We also remark that the value of ΔG_{tc} at the collapse-transition (ΔG_{tc}^{m}) is equal to $-T\Delta S$ (see inset of Fig. 13). In other words, the minimum contact energy to get at least half of the conformations with a T/Tr contact, is exactly the entropic cost of passing from a unpacked to a packed structure.

We evaluate the packing of the molecule by computing the ratio between the volume occupied when $\Delta G_{tc}=0$ (V_{unpacked}) and when ΔG_{tc} is sufficiently strong to get all the conformations packed (V_{packed}). Figure 14 shows that an important compaction of the molecule is observed (more than 50% of the original volume). This is compatible with the experimental observation of compaction of *Tetrahymena* group I ribozyme.[59]

IV. CONCLUSION

To summarize, we have introduced a semiquantitative lattice model of RNA folding whose parameters are systematically derived from experimental data on short molecules via the Turner secondary structure model.[9] Like the latter, we include the formation free energies of double-helical sections, forks, dangling ends, etc. via RNA-specific parameters, without resolving the internal (helical) structure of the strands on the level of bases or the sugar-phosphate backbone[24] or atoms.[25,26] Our results can nevertheless provide relevant input for methods, which generate atomistic models for given coarse-grained or secondary structures.[22]

With the lattice and the Turner models defined on the same length scale, corresponding parameters are readily identified. We have shown in detail, how the necessary corrections due to the conformational entropy of secondary structures in the lattice model are obtained by grouping and counting the corresponding microstates for simple cases.[57] From a practical point of view, the parameterization and simulation of the lattice model is greatly facilitated by a number of simplifications and unifications of the standard Turner parameters, which we have proposed on general physical grounds (Sec. II B). In particular, it is possible to avoid the nonlocal secondary structure analysis of a conformation by (i) unifying the loop-type dependent forking energies and (ii) suppressing the loop nucleation free-energy penalties for *individual* loops via a suitable gauge transformation (Sec. II C 1). Our considerations might already be of interest on the level of the secondary structure description. They should equally apply to other attempts[15,21] to integrate independently calculated conformational entropy estimates with the Turner model. We emphasize that parameters and the predictive power of such models should evolve with the standard secondary structure description. Simple tests of the internal coherence of a parameterization scheme are similar behavior for different versions of the Turner parameter set and the invariance of model predictions under the various gauge freedoms of the Turner model.

Contrary to secondary structure approaches, we have *no*

free parameters accounting for the conformational entropy of the folded molecule, for example, in the form of a generalized Jacobson–Stockmayer relation Eq. (2). Rather the coarse-grain representation of the molecule's 3D structure allows us to *predict* the generic (polymer) contributions to nonlocal loop formation free energy Δg_{loop} of arbitrary secondary structures. In particular, we include the effects of connectivity (important for pseudoknots) and of excluded volume interactions within and between all elements of the secondary structure, a feature which is essential for the systematic treatment of multiloops. For comparison, Kinefold[21] only includes connectivity effects. Vfold works at higher spatial resolution, but is limited to a slowly growing number pseudoknot architectures.[39,47]

In this paper, we have tested the predictive power of the lattice model for simple hairpins as well as for complex structures such as tRNAs or pseudoknots. While specialized applications achieve a slightly better performance for the prediction of pseudoknot groundstates, the lattice model provides a complete statistical mechanical description and shows a consistent reliability independently of the type of structure studied. Currently, our main limitation is the computational cost to obtain the density of state for long heterogenenous sequences (>80 nts). Using more sophisticated MC methods, the lattice model should provide an efficient framework to systematically study nonlocal effects on RNA folding including spatial confinement[109] or tertiary interactions.[62] At the same time, it remains an interesting challenge to extend our parameterization scheme to more detailed representations of the 3D structure of RNA.[15,24]

ACKNOWLEDGMENTS

We thank H. Isambert for some explanations on the Kinefold model. R.E. acknowledges support from the chair of excellence program of the ANR (France) and discussion with Ch. Simm of some technical points.

[1] C. Calladine, H. Drew, B. Luisi, and A. Travers, *Understanding DNA: The Molecule and How It Works* (Elsevier, San Diego, 2004).
[2] R. Gesteland, T. R. Cech, and J. F. Atkins, *RNA World* (Cold Spring Harbor Laboratory, New York, 2005).
[3] A. Fire, S. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver, and C. C. Mello, Nature (London) **391**, 806 (1998).
[4] D. D. Shoemaker, E. E. Schadt, C. D. Armour, Y. D. He, P. Garrett-Engele, P. D. McDonagh, P. M. Loerch, A. Leonardson, P. Y. Lum, G. Cavet, L. F. Wu, S. J. Altschuler, S. Edwards, J. King, J. S. Tsang, G. Schimmack, J. M. Schelter, J. Koch, M. Ziman, M. J. Marton, B. Li, P. Cundiff, T. Ward, J. Castle, M. Krolewski, M. R. Meyer, M. Mao, J. Burchard, M. J. Kidd, H. Dai, J. W. Phillips, P. S. Linsley, R. Stoughton, S. Scherer, and M. S. Boguski, Nature (London) **409**, 922 (2001).
[5] N. C. Seeman, Biochemistry **42**, 7259 (2003).
[6] P. W. K. Rothemund, Nature (London) **440**, 297 (2006).
[7] B. A. Shapiro, Y. G. Yingling, W. Kasprzak, and E. Bindewald, Curr. Opin. Struct. Biol. **17**, 157 (2007).
[8] E. Capriotti and M. A. Marti-Renom, Curr. Bioinformatics **3**, 32 (2008).
[9] D. H. Mathews, J. Sabina, M. Zuker, and D. H. Turner, J. Mol. Biol. **288**, 911 (1999).
[10] D. H. Mathews, M. D. Disney, J. L. Childs, S. J. Schroeder, M. Zuker, and D. H. Turner, Proc. Natl. Acad. Sci. U.S.A. **101**, 7287 (2004).
[11] M. Zuker, Nucleic Acids Res. **31**, 3406 (2003).
[12] N. R. Markham and M. Zuker, Nucleic Acids Res. **33**, W577 (2005).
[13] I. L. Hofacker, Nucleic Acids Res. **31**, 3429 (2003).
[14] E. Rivas and S. R. Eddy, J. Mol. Biol. **285**, 2053 (1999).
[15] S. Cao and S. J. Chen, RNA **11**, 1884 (2005).
[16] T. R. Einert, P. Nager, H. Orland, and R. R. Netz, Phys. Rev. Lett. **101**, 048103 (2008).
[17] A. P. Gultyaev, F. H. van Batenburg, and C. W. Pleij, J. Mol. Biol. **250**, 37 (1995).
[18] B. A. Shapiro, W. Kasprzak, C. Grunewald, and J. Aman, J. Mol. Graphics Modell. **25**, 514 (2006).
[19] H. Isambert and E. D. Siggia, Proc. Natl. Acad. Sci. U.S.A. **97**, 6515 (2000).
[20] A. Xayaphoummine, T. Bucher, F. Thalmann, and H. Isambert, Proc. Natl. Acad. Sci. U.S.A. **100**, 15310 (2003).
[21] A. Xayaphoummine, T. Bucher, F. Thalmann, and H. Isambert, Nucleic Acids Res. **33**, W605 (2005).
[22] Y. G. Yingling and B. A. Shapiro, J. Mol. Graphics Modell. **25**, 261 (2006).
[23] J. Burks, C. Zwieb, F. Muler, I. Wower, and J. Wower, BMC Mol. Biol. **6**, 14 (2005).
[24] T. A. Knotts, N. Rathore, C. Schwartz, and J. J. de Pablo, J. Chem. Phys. **126**, 084901 (2007).
[25] M. Feig and B. M. Pettitt, Biophys. J. **75**, 134 (1998).
[26] F. Merzel, F. Fontaine-Vive, M. R. Johnson, and G. J. Kearley, Phys. Rev. E **76**, 031917 (2007).
[27] C. Hyeon and D. Thirumalai, Biophys. J. **92**, 731 (2007).
[28] F. Ding, S. Sharma, P. Chalasani, V. V. Demidov, N. E. Broude, and N. V. Dokholyan, RNA **14**, 1164 (2008).
[29] R. Das and D. Baker, Proc. Natl. Acad. Sci. U.S.A. **104**, 14664 (2007).
[30] M. Parisien and F. Major, Nature (London) **452**, 51 (2008).
[31] R. Das, M. Kudaravalli, M. Jonikas, A. Laederach, R. Fong, J. P. Schwans, D. Baker, J. A. Piccirilli, R. B. Altman, and D. Herschlag, Proc. Natl. Acad. Sci. U.S.A. **105**, 4144 (2008).
[32] D. Poland and H. A. Scheraga, J. Chem. Phys. **45**, 1456 (1966).
[33] J. Reeder and R. Giegerich, BMC Bioinf. **5**, 104 (2004).
[34] R. M. Dirks and N. A. Pierce, J. Comput. Chem. **24**, 1664 (2003).
[35] I. Tinoco, Jr. and C. Bustamante, J. Mol. Biol. **293**, 271 (1999).
[36] Y. Byun and K. Han, Nucleic Acids Res. **34**, W416 (2006).
[37] M. J. Serra and D. H. Turner, Methods Enzymol. **259**, 242 (1995).
[38] A. P. Gultyaev, F. H. van Batenburg, and C. W. Pleij, RNA **5**, 609 (1999).
[39] S. Cao and S. J. Chen, Nucleic Acids Res. **34**, 2634 (2006).
[40] P. Schuster, Rep. Prog. Phys. **69**, 1419 (2006).
[41] H. Jacobson and W. H. Stockmayer, J. Chem. Phys. **18**, 1600 (1950).
[42] M. E. Fisher, J. Chem. Phys. **45**, 1469 (1966).
[43] Y. Kafri, D. Mukamel, and L. Peliti, Eur. Phys. J. B **27**, 135 (2002).
[44] E. Carlon, E. Orlandini, and A. L. Stella, Phys. Rev. Lett. **88**, 198101 (2002).
[45] R. Blossey and E. Carlon, Phys. Rev. E **68**, 061911 (2003).
[46] D. Jost and R. Everaers, J. Phys.: Condens. Matter **21**, 034108 (2009).
[47] S. Cao and S. J. Chen, RNA **15**, 696 (2009).
[48] G. Vernizzi, H. Orland, and A. Zee, e-print arXiv:q-bio/0405014.
[49] M. Bon, G. Vernizzi, H. Orland, and A. Zee, J. Mol. Biol. **379**, 900 (2008).
[50] A. Lucas and K. A. Dill, J. Chem. Phys. **119**, 2414 (2003).
[51] Z. Kopeikin and S. J. Chen, J. Chem. Phys. **122**, 094909 (2005).
[52] Z. Kopeikin and S. J. Chen, J. Chem. Phys. **124**, 154903 (2006).
[53] M. S. Causo, B. Coluzzi, and P. Grassberger, Phys. Rev. E **62**, 3958 (2000).
[54] P. Leoni and C. Vanderzande, Phys. Rev. E **68**, 051904 (2003).
[55] A. Kabakçioğlu and A. L. Stella, Phys. Rev. E **70**, 011802 (2004).
[56] M. Sales-Pardo, R. Guimera, A. A. Moreira, J. Widom, and L. A. N. Amaral, Phys. Rev. E **71**, 051902 (2005).
[57] R. Everaers, S. Kumar, and C. Simm, Phys. Rev. E **75**, 041918 (2007).
[58] T. E. Ouldridge, I. G. Johnston, A. A. Louis, and J. P. K. Doye, J. Chem. Phys. **130**, 065101 (2009).
[59] K. Takamoto, R. Das, Q. He, S. Doniach, M. Brenowitz, D. Herschlag, and M. R. Chance, J. Mol. Biol. **343**, 1195 (2004).
[60] N. Xia, J. SantaLucia, Jr., M. E. Burkard, R. Kierzek, S. J. Schroeder, X. Jiao, C. Cox, and D. H. Turner, Biochemistry **37**, 14719 (1998).
[61] D. M. Gray and I. Tinoco, Jr., Biopolymers **9**, 223 (1970).
[62] N. B. Leontis and E. Westhof, RNA **7**, 499 (2001).
[63] A. Dkhissi, G. Renvez, and R. Blossey, J. Phys.: Condens. Matter **21**, 034115 (2009).
[64] P.-G. de Gennes, *Scaling Concepts in Polymer Physics* (Cornell University Press, Ithaca, 1979).
[65] C. Vanderzande, *Lattice Models of Polymers* (Cambridge University Press, Cambridge, UK, 1998).

[66] J. L. Martin, M. F. Sykes, and F. T. Hioe, J. Chem. Phys. **46**, 3478 (1967).
[67] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in Fortran 77: The Art of Scientific Computing* (Cambridge University Press, Cambridge, England, 1996).
[68] T. Ohmichi, S.-I. Nakano, D. Miyoshi, and N. Sugimoto, J. Am. Chem. Soc. **124**, 10367 (2002).
[69] A. E. Walter, D. H. Turner, J. Kim, M. H. Lyttle, P. Muller, D. H. Mathews, and M. Zuker, Proc. Natl. Acad. Sci. U.S.A. **91**, 9218 (1994).
[70] A. E. Walter and D. H. Turner, Biochemistry **33**, 12715 (1994).
[71] J. Kim, A. E. Walter, and D. H. Turner, Biochemistry **35**, 13753 (1996).
[72] B. M. Znosko, M. E. Burkard, S. J. Schroeder, T. R. Krugh, and D. H. Turner, Biochemistry **41**, 14969 (2002).
[73] S. J. Johnson and L. S. Beese, Cell **116**, 803 (2004).
[74] D. M. Gray, Biopolymers **42**, 783 (1997).
[75] D. M. Gray, Biopolymers **42**, 795 (1997).
[76] M. J. Serra, T. W. Barnes, K. Betschart, M. J. Gutierrez, K. J. Sprouse, C. K. Riley, L. Stewart, and R. E. Temel, Biochemistry **36**, 4844 (1997).
[77] C. W. Hilbers, C. A. G. Haasnoot, S. H. de Bruin, J. J. M. Joordens, G. A. Van Der Marel, and J. H. Van Boom, Biochimie **67**, 685 (1985).
[78] R. M. Wartell and A. S. Benight, Phys. Rep. **126**, 67 (1985).
[79] R. D. Blake and S. G. Delcourt, Nucleic Acids Res. **26**, 3323 (1998).
[80] J. SantaLucia, Jr., Proc. Natl. Acad. Sci. U.S.A. **95**, 1460 (1998).
[81] M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics* (Oxford University Press, New York, 1986).
[82] C. Bustamante, J. F. Marko, E. D. Siggia, and S. Smith, Science **265**, 1599 (1994).
[83] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications* (Academic, San Diego, 2002).
[84] N. Madras and A. D. Sokal, J. Stat. Phys. **50**, 109 (1988).
[85] N. Madras, A. Orlitsky, and L. A. Shepp, J. Stat. Phys. **58**, 159 (1990).
[86] A. D. Sokal, e-print arXiv:hep-lat/9509032.
[87] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
[88] E. J. Janse van Rensburg, S. G. Whittington, and N. Madras, J. Phys. A **23**, 1589 (1990).
[89] A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **63**, 1195 (1989).
[90] G. Varani, Annu. Rev. Biophys. Biomol. Struct. **24**, 379 (1995).
[91] M. A. Glucksmann-Kuis, X. Dai, P. Markiewicz, and L. B. Rothman-Denes, Cell **84**, 147 (1996).
[92] H. Lodish, A. Berk, S. L. Zipursky, P. Matsudaira, D. Baltimore, and J. E. Darnell, *Molecular Cell Biology* (W. H. Freeman and Company, New York, 2000).
[93] O. C. Uhlenbeck, Nature (London) **346**, 613 (1990).
[94] J. Liphardt, B. Onoa, S. B. Smith, I. Tinoco, Jr., and C. Bustamante, Science **292**, 733 (2001).
[95] D. Collin, F. Ritort, C. Jarzynski, S. B. Smith, I. Tinoco, Jr., and C. Bustamante, Nature (London) **437**, 231 (2005).
[96] A. Ke, K. Zhou, F. Ding, J. H. Cate, and J. A. Doudna, Nature (London) **429**, 201 (2004).
[97] P. L. Adams, M. R. Stahley, A. B. Kosek, J. Wang, and S. A. Strobel, Nature (London) **430**, 45 (2004).
[98] C. A. Theimer, C. A. Blois, and J. Feigon, Mol. Cell **17**, 671 (2005).
[99] L. X. Shen and I. Tinoco, Jr., J. Mol. Biol. **247**, 963 (1995).
[100] D. W. Staple and S. E. Butcher, PLoS Biol. **3**, e213 (2005).
[101] C. A. Theimer and D. P. Giedroc, J. Mol. Biol. **289**, 1283 (1999).
[102] S. Cao and S.-J. Chen, Phys. Biol. **5**, 016002 (2008).
[103] F. H. van Batenburg, A. P. Gultyaev, and C. W. Pleij, Nucleic Acids Res. **28**, 201 (2000).
[104] V. K. Misra, R. Shiman, and D. E. Draper, Biopolymers **69**, 118 (2003).
[105] D. Thirumalai and C. Hyeon, Biochemistry **44**, 4957 (2005).
[106] D. E. Draper, RNA **10**, 335 (2004).
[107] J. H. Cate, A. R. Gooding, E. Podell, K. Zhou, B. L. Golden, C. E. Kundrot, T. R. Cech, and J. A. Doudna, Science **273**, 1678 (1996).
[108] F. L. Murphy and T. R. Cech, Biochemistry **32**, 5291 (1993).
[109] A. M. Yoffe, P. Prinsen, A. Gopal, C. M. Knobler, W. M. Gelbart, and A. Ben-Shaul, Proc. Natl. Acad. Sci. U.S.A. **105**, 16153 (2008).