Cold denaturation of RNA secondary structures with loop entropy and quenched disorder

Flavio Iannelli*
Humboldt-Universität zu Berlin, Institut für Physik, Newtonstraße 15, 12481 Berlin, Germany

Yevgeni Mamasakhlisov
Department of Molecular Physics, Yerevan State University, 1 Alex Manougian Street, Yerevan 0025, Armenia

Roland R. Netz
Fachbereich Physik, Freie Universität Berlin, 14195 Berlin, Germany

(Received 5 October 2018; published 31 January 2020)

The critical behavior of ribonucleic acid (RNA) secondary structures with quenched sequence randomness is studied by means of the constrained annealing method. A thermodynamic phase transition is induced by including the conformational weight of loop structures. In addition to the expected melting at high temperature, a cold-melting transition appears when the disorder strength induces competition between favorable and unfavorable base pairs. Our results suggest that the cold denaturation of RNA found experimentally might be triggered by quenched sequence disorder. We calculate hot- and cold-melting critical temperatures for competing favorable and unfavorable base-pair energies and present a folding phase diagram as a function of the loop exponent and temperature.

DOI: 10.1103/PhysRevE.101.012502

I. INTRODUCTION

Ribonucleic acids (RNA) are biopolymers that are crucial to all living systems; they process and transmit genetic information and take part in many important cellular activities [1]. The RNA primary structure is a sequence which consists of four bases U, A, G, and C, while the secondary structure is defined by the list of base pairings that occurs.

RNA molecules fold into a native conformation when the temperature is lowered. As for proteins, the folding of RNA structures is crucial for understanding their biological functions and has been vastly studied, based mostly on the idea of hierarchical folding [2]. In this scheme, the primary structure determines the folding mechanism and the secondary structure forms independently of the tertiary structure. This drastic simplification makes it possible to exactly compute the secondary structure partition function of given RNA sequences. The molecule folding can be qualitatively and quantitatively addressed via the experimentally observable helicity degree θ defined as the average fraction of paired bases [3], which increases when lowering the temperature in the standard scenario.

A particularly interesting phenomenon that is observed for polymers and proteins is *cold denaturation* [4–7]. The denaturation of proteins and polymers with rising temperature is a consequence of the increase in configurational entropy. Denaturation when lowering the temperature is usually interpreted in terms of hydrophobic interactions. Experimentally, denaturation can be inferred by the presence of peaks in the specific heat C_V, which physically follows from an abrupt increase of the system entropy [8–11]. In this paper, we give an alternative explanation for cold denaturation in terms of quenched disorder, which itself weakens the secondary structure formation at low temperature. The associated double-peak behavior of C_V turns out to be associated with two different melting temperatures.

Here we focus on a two-letter alphabet binding energy model with symbols chosen from the subset {U, A}, in the spirit of [12–16], which corresponds to the hydrophilic-hydrophobic model for protein folding [17]. Although this approximation is an oversimplification, it has proven to reproduce in a reasonable fashion the folding thermodynamics and the glass transition [14,15,18,19].

For real RNA, the so-called Watson-Crick base pairs UA, and GC are most stable. Since each base is essentially planar and its conformations are limited, every RNA secondary structure is defined by a list of base pairings (i, j) with each base appearing, at most, once. In addition, for any two base pairings (i, j) and (k, l), we only consider nested, where $i < k < l < j$, and independent pairings, where $i < j < k < l$; see Fig. 1. A third possibility would be to include pseudoknots where one has $i < k < j < l$, but they can be excluded because they are rare in real RNA [2,20]. This defines a hierarchical structure for the RNA conformation so that a recursive equation for the partition function can be used. Besides, this choice discards all configurations that are not planar, meaning that in the case of disordered sequences, the system displays frustration [15,21]. By excluding pseudoknots, which can be accounted for by using advanced theoretical techniques [22–25], we are able to account for the combined effects of logarithmic loop entropies and RNA sequence disorder, which otherwise would not be analytically possible.

*iannelli.flavio@gmail.com

 ©2020 American Physical Society

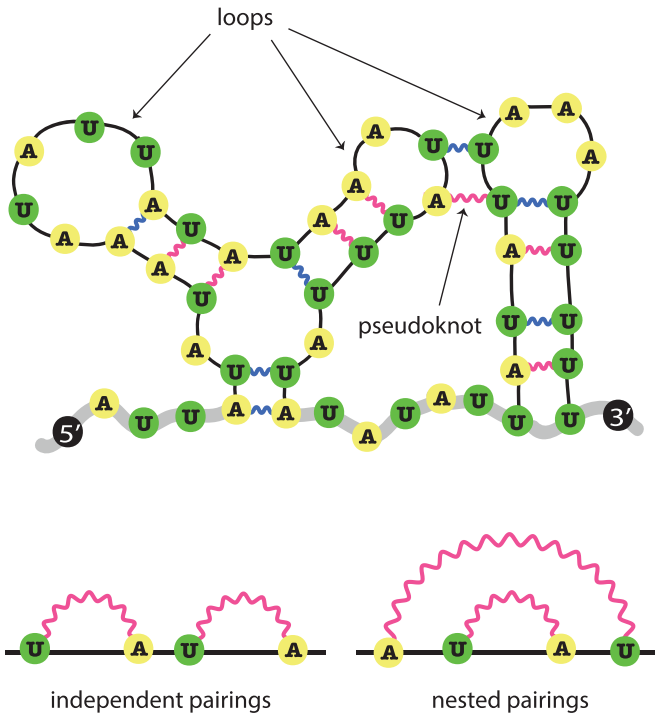


FIG. 1. Upper panel: Secondary structure representation with stacks and loops. Nucleotides are the green and yellow dots for U and A, respectively. The wavy lines identify the hydrogen bonds for favorable (red) and unfavorable (blue) base pairs, while the black solid lines are the nested backbone links. Thick gray denote the non-nested backbone links. In this secondary structure, a pseudoknot is formed between two loops. Lower panel: The two possible types of base pairings that are considered.

When the RNA molecule is far from the native state, i.e., when there is a significant amount of unpaired bases, the presence of loops in RNA secondary structures plays a crucial role. The configurational entropy contribution for loops of length n, $\Delta S_n^{\rm loop} \sim k_B \ln n^{-c}$, is characterized by the universal loop exponent c [26]. For ideal polymers, which can be modeled as simple random walks, the exponent is $c_{\rm RW} = D\nu$, where D is the spatial dimensionality and $\nu = 1/2$. Instead, for a self-avoiding walk of length n, the probability to return to the origin is $\mathcal{P}_{\rm SAW}(n) \sim n^{-c_{\rm SAW}}$, where $c_{\rm SAW} = D\nu + \gamma - 1 > c_{\rm RW}$ [27]. This value is further increased when considering real polymers since c also depends on the number of helical strands emerging from loops [28].

In the absence of disorder, i.e., for homopolymeric RNA, a phase transition from the folded to the unfolded molten state [14,29] occurs when the temperature increases up to the melting point $T_m(c)$ [30]. This happens for values of the loop exponent $2 < c \lesssim 2.479$ and the c-dependent critical exponents have been analytically obtained [31]. Folding of heteropolymer RNA in the presence of loop entropy has been studied before [32], but not usually in the additional presence of quenched disorder. The current lack of results that combine quenched disorder with loop entropy is the main motivation of this paper. By providing a variational approximation of the average free energy in terms of the homopolymer free energy with loop entropy, we are able to provide a solid framework for describing the melting transition of heteropolymer RNA.

In this paper, we study the effect of a loop exponent $c \neq 0$ on RNA secondary structures with random sequences where the disorder is quenched. This allows us to characterize a generic RNA molecule and most of the results obtained here also apply to DNA. For disordered sequences with an energetic competition between favorable and unfavorable base pairs, we find that in addition to hot melting, a cold-melting phase transition appears when including loop entropy. We describe the decrease in the number of paired bases, quantified by the helicity degree θ, in terms of the singularities of the grand-canonical partition function of a homopolymer [31]. To account for disorder, we use the constrained annealing method [33]. We find that the decrease in the helicity degree at low temperature is matched with a double-peak structure of the specific heat C_V that characterizes an abrupt increase of the system entropy close to the cold-melting point [8,9]. Our results shed light on the experimentally observed cold denaturation of proteins and provide a simple explanation of the associated phase transition in terms of the interplay between quenched disorder and loop entropy.

The paper is structured as follows. In Sec. II, we introduce the model for RNA secondary structures with base-pairing energies depending on quenched disordered variables. A recursion equation to compute the exact partition function in the presence of disorder is presented for the cases with and without loop entropy. In Sec. III, we review the known results starting from de Gennes theory for nondisordered, i.e., homopolymeric, sequences using a generating function approach. The folding of homopolymer RNA in the presence of loop entropy is discussed in the grand-canonical ensemble as a function of the loop exponent c. In Sec. IV, we present our main results. We first derive analytically the constrained annealing method for computing the disorder average of the free energy including the contribution of loop entropy. In addition to the standard high-temperature denaturation of the molecule, a cold melting phase transition appears and is explained in terms of the homopolymer loop entropy contributions. We validate numerically our theory and conclude the section by presenting our findings in a global phase diagram in the temperature-loop exponent plane. In Sec. V, we summarize our results and present our conclusions.

II. DISORDERED RNA

A. The model

For a RNA sequence h of length N, we define the base-pairing matrix $\mathcal{S}$, which completely determines the secondary structure, as the $N \times N$ symmetric matrix with components $s_{i,j}$ equal to unity if (i, j) are paired and equal to zero otherwise. The model considered here includes the energy of hydrogen bonds and the entropy of loops. The stacking free energy between neighboring base pairs can, in principle, be included by adding a constant term to the loop free energy, as was shown previously in [31,34], but this is not pursued in the present paper. If stacking energies are neglected, the Hamiltonian for a given sequence can be written as a sum over nonrepeated base pairs,

$$\mathcal{H}(\mathcal{S}, h) = \sum_{(i,j)\in\mathcal{S}} \epsilon_{i,j} = \sum_{1\leqslant i<j\leqslant N} s_{i,j}\epsilon_{i,j}. \qquad (1)$$

We assume the simplest nontrivial pairing energy function as the sum of a constant and a disorder term in the spirit of [16],

$$\epsilon_{i,j} = \epsilon_0 + \epsilon h_i h_j. \tag{2}$$

Helix stacking can, in principle, be taken into account by a suitable redefinition of base-pairing energy [2]. In this paper, we are concerned with the interplay between disorder and loop entropy at the level of secondary structures. Thus, all tertiary interactions are neglected, in line with previous works [12–15,26,31,34].

The sign of ϵ_0 defines the nature of the background interaction between nucleotides. If $\epsilon_0 > 0$, the interaction is repulsive and attractive otherwise. The second term in Eq. (2) is the product of two independent variables and similar to spin-glass models [35], and is multiplied by the additional constant $\epsilon > 0$. We assign Ising variables to each base along the chain so that $h_i = +1$ if i is the nucleotide U and $h_i = -1$ if it is A. Note that the model can be easily generalized to account for four bases.

Contrary to the base-pairing matrix elements $s_{i,j}$, which are free to evolve within the dynamics of the system, the site sequence variables h_i are frozen. Having fixed the sign of the background interaction ϵ_0, the absolute value of the energy ratio ϵ/ϵ_0 is the relevant parameter characterizing the system behavior. We take h_i as quenched random variables with a factorized probability distribution

$$\mathcal{P}(h) = \prod_{i=1}^{N} \rho(h_i). \tag{3}$$

The factorization approximation is motivated by the fact that no strong correlations are found in base sequences [15]. Defining the probability of finding the base U as $p \equiv \rho(h_i = +1)$, the probability distribution factors can be written as

$$\rho(h_i) = p\delta(h_i - 1) + (1 - p)\delta(h_i + 1). \tag{4}$$

In the two-letter model adopted here, due to symmetry we only need to explore the probability range $0.5 \leqslant p \leqslant 1$.

B. Partition function with and without loops

The partition function of a given sequence h is the sum of the statistical weight over all allowed realizations of the base paring matrix $\{\mathcal{S}\}$,

$$Z_N(h) = \sum_{\{\mathcal{S}\}} e^{-\beta \mathcal{H}(\mathcal{S},h)}. \tag{5}$$

Here, $\beta = (k_B T)^{-1}$ is the inverse temperature and $\{\mathcal{S}\}$ denotes the set of all secondary structures without pseudoknots. The free energy is obtained by performing the quenched average, denoted by $\overline{(\cdots)}$, of the disordered free energy,

$$f(h) = -\frac{1}{\beta N} \ln Z_N(h), \tag{6}$$

over the disorder distribution given by Eq. (3), yielding

$$\overline{f(h)} = \sum_{\{h\}} \mathcal{P}(h) f(h) = -\frac{1}{\beta N} \overline{\ln Z_N(h)}. \tag{7}$$

For sufficiently large chains, the physical properties of the system do not depend on the specific disorder realization h and the free energy self-averages [21],

$$\lim_{N\to\infty} f(h) = \overline{f(h)}. \tag{8}$$

The full partition function for a given sequence h can, in the absence of pseudoknots, be obtained via the recursive equation [3]

$$Z_{i,j+1} = Z_{i,j} + \sum_{k=i}^{j} w_{k,j+1} Z_{i,k-1} Z_{k+1,j}, \tag{9}$$

for the restricted partition function $Z_{i,j+1}$ of the subsequence of monomers going from i to $j+1$, as illustrated in Fig. 2. The first term on the right-hand side corresponds to the probability that base $j+1$ is not paired. The second term corresponds to the probability associated to all possible nested and independent pairings between bases k and $j+1$ and where

$$w_{k,j+1} = \exp(-\beta \epsilon_{k,j+1}) \tag{10}$$

is the corresponding statistical weight. The base paring energy $\epsilon_{k,j+1} = \epsilon_0 + \epsilon h_k h_{j+1}$, according to Eq. (2), implicitly includes an entropic term ϵ_0^S defined as $\beta\epsilon_0 = \beta\epsilon_0^U - \epsilon_0^S$, where ϵ_0^U is the internal energy contribution. In what follows, the entropic term is obtained from $\beta\epsilon_0$ in the high-temperature limit $\beta \to 0$.

The recursive equation (9) allows one to compute the exact partition function $Z_N = Z_{1,N}$ without pseudoknots in $O(N^3)$ time, starting with the boundary conditions $Z_{i,i} = Z_{i,i-1} = 1$, $\forall i$ [3].

In [26], the authors have shown how to take into account loops of length n with a statistical weight [36],

$$v_n = n^{-c}, \tag{11}$$

where c is the *loop exponent*. In previous work [34], we have used a more complex form of the loop free energy which included bending for small loops (which, in fact, is dominated by the electrostatic contribution to the bending rigidity for not too large salt concentration) and the loop initiation free energy, which takes care of the stacking energy. The analytical calculation we perform in the present paper, however, is not tractable for this more complicated form and therefore we use the approximate simple form of Eq. (11).

Let us consider the restricted partition function $Z_{i,j}^M$ of a polymer going from monomer i to monomer j with $M \leqslant j - i$ non-nested (i.e., unlooped) links; see Fig. 1. Using the boundary conditions $Z_{i,i}^M = \delta_M$, $Z_{i,i-1}^M = \delta_{M+1}$, and $Z_{i,j}^{-1} = \delta_{j+1-i}$, the recursive equation for the partition function can be written as [26]

$$\widetilde{Z}_{i,j+1}^{M+1} = \widetilde{Z}_{i,j}^{M} + \sum_{k=i}^{j} w_{k,j+1} \widetilde{Z}_{i,k-1}^{M} \widetilde{Z}_{k,j+1}^{0}. \tag{12}$$

Contrary to Eq. (9), this recursive equation, illustrated in Fig. 3, requires a longer computational time that scales as $O(N^4)$ with the base number. Here, $\widetilde{Z}_{i,j}^M \equiv Z_{i,j}^M / u_M$ is the partition function rescaled by the statistical weight of M nonlooped links u_M, and

$$\widetilde{Z}_{k,j+1}^{0} = \sum_{l=-1}^{j-k-1} \widetilde{Z}_{k+1,j}^{l} v_{l+2} \tag{13}$$

is the partition function of a substrand that is terminated by a helix.

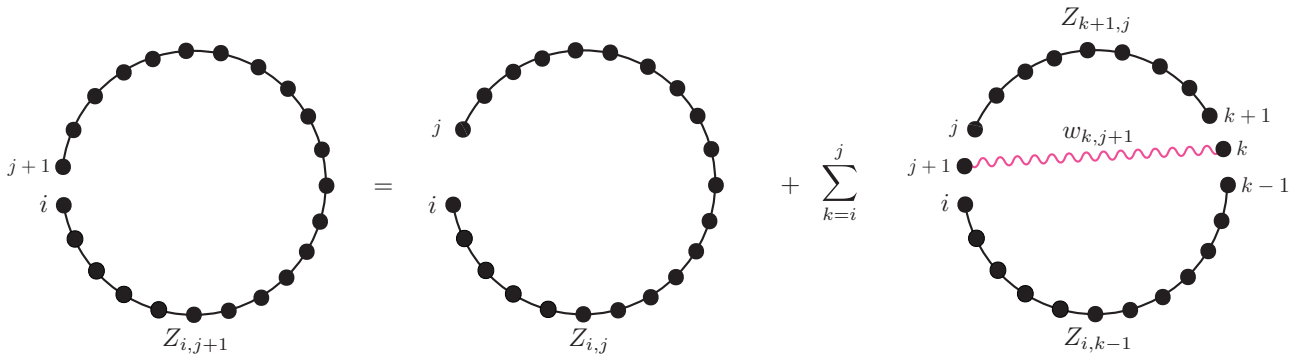


FIG. 2. Hierarchical recursive scheme for the partition function given by Eq. (9) without loop entropy. The subchain partition function from base i to base $j+1$ is the sum of the partition function from base i to base j and the partition functions of all base pairings formed between base $j+1$ and base $k \in (i, j)$ [3].

C. Helicity degree without loops

A key quantity to characterize the RNA conformation is the helicity degree defined as [3,13]

$$\theta = \frac{2}{N}\sum_{i<j}\overline{\langle s_{i,j}\rangle} = \frac{2}{N}\overline{\left\langle \sum_{i<j} s_{i,j}\right\rangle} = \frac{2}{N}\overline{\langle |\mathcal{S}|\rangle}, \tag{14}$$

where $\langle\cdot\rangle$ denotes the thermal average over the canonical ensemble and

$$|\mathcal{S}| \equiv \sum_{i<j} s_{i,j} \tag{15}$$

is the number of paired bases in the structure $\mathcal{S}$. Since $\overline{\langle |\mathcal{S}|\rangle} \in [0, N/2]$, the helicity degree is a function of the temperature in the interval [0,1] with $\theta = 1$ if every base is paired, corresponding to the native state, and $\theta = 0$ if no base is paired. Therefore, the helicity is a convenient measure of the order of the RNA conformation. The helicity can also be expressed in terms of the free energy by noting that

$$\begin{aligned}\theta &= \frac{2}{N}\overline{\frac{1}{Z_N(h)}\sum_{\{\mathcal{S}\}} e^{-\beta\epsilon_0|\mathcal{S}|}e^{-\beta\epsilon\sum_{i<j}s_{i,j}h_ih_j}|\mathcal{S}|}\\ &= \frac{2}{N}\overline{\left[-\frac{\partial}{\partial(\beta\epsilon_0)}\right]\ln\sum_{\{\mathcal{S}\}} e^{-\beta\epsilon_0|\mathcal{S}|}e^{-\beta\epsilon\sum_{i<j}s_{i,j}h_ih_j}}\\ &= 2\frac{\partial}{\partial(\beta\epsilon_0)}\beta\overline{f(h)}.\end{aligned} \tag{16}$$

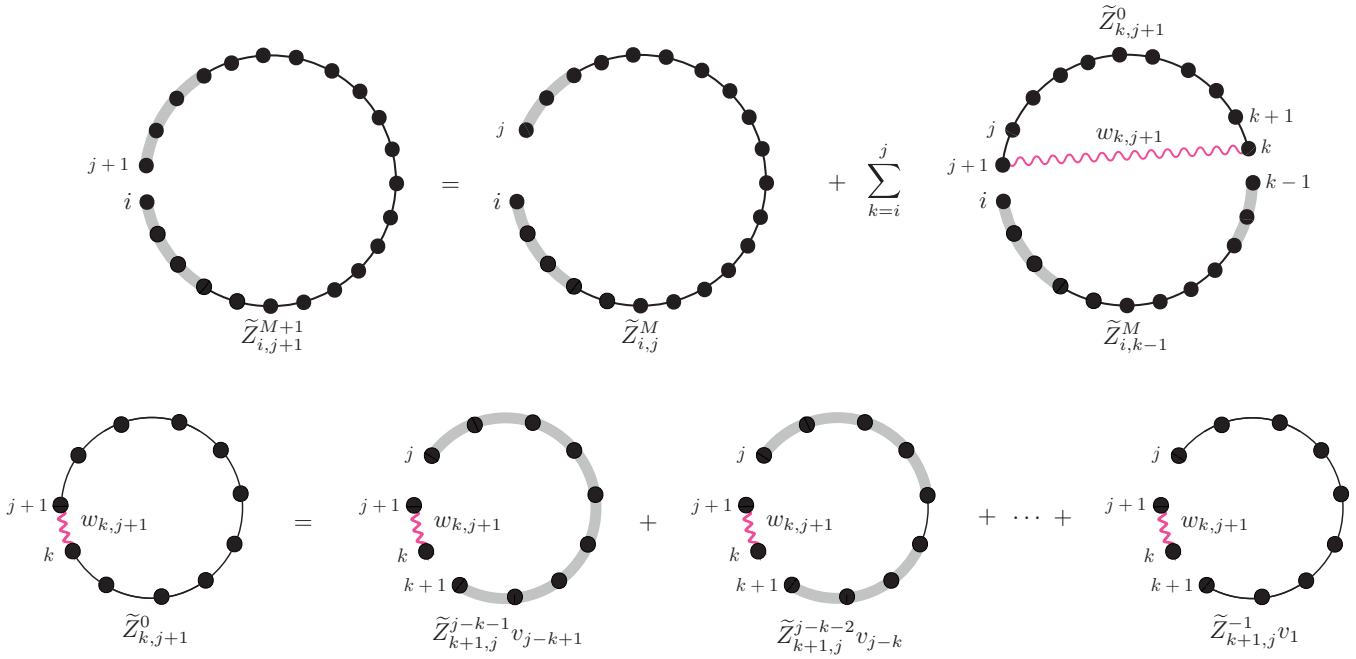
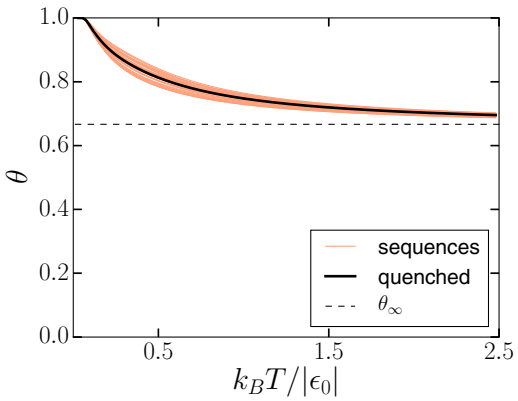
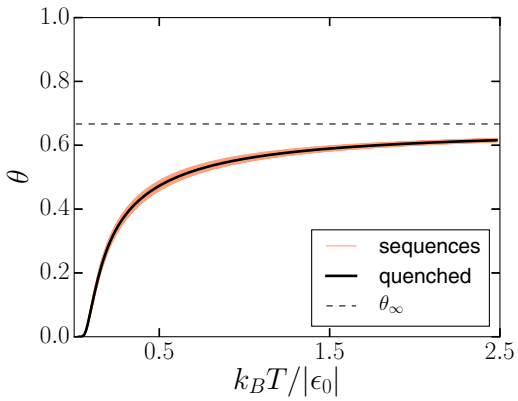


FIG. 3. Recursion scheme for the partition function given by Eq. (12) including loops. The partition function of a RNA sequence ranging from i to $j+1$ with $M+1$ non-nested backbones (thick gray lines) is computed from the sequence ranging from i to j with M non-nested backbones by adding a base at position $j+1$ and considering all possible pairings with a base $k \in (i, j)$ [26]. Each of these pairings defines a structure which has zero non-nested backbones, i.e., an arbitrary substrand that is terminated by a helix. The explicit diagrams for the latter are shown in the lower panel and are obtained by considering the sum over all non-nested backbones with associated statistical weight given by Eq. (11).



(a)



(b)

FIG. 4. Helicity degree for (a) $\epsilon_0 = -1$ and (b) $\epsilon_0 = +1$ with $\epsilon = 0.5|\epsilon_0|$ in the absence of loop entropy with probability of U base occurrence $p = 0.75$. Each red curve corresponds to the helicity of a single RNA sequence realization of the disorder. The quenched average (black line) is obtained from the exact computation of the partition function for 30 random sequences of length $N = 50$. The black dashed line denotes the high-temperature limit θ_∞ given by Eq. (25).

In the absence of loops, the helicity can be computed as

$$\theta = \frac{2}{N}\overline{\sum_{i<j} P_{i,j}}, \tag{17}$$

where $P_{i,j} = \langle s_{i,j}\rangle$ is the probability of base-pair formation between nucleotides i and j [14]. The latter is obtained from the partition function in the absence of loop entropy, given by Eq. (9), as [37]

$$P_{i,j} = e^{-\beta\epsilon_{i,j}}\frac{Z^{\rm int}_{i,j}Z^{\rm ext}_{i,j}}{Z_{1,N}}, \tag{18}$$

were $Z^{\rm int}_{i,j} = Z_{i+1,j-1}$ is the partition function of the internal sequence $(i+1,\ldots,j-1)$, while $Z^{\rm ext}_{i,j}$ is the partition function of the external sequence $(1,\ldots,i-1,j+1,\ldots,N)$ that can be computed by considering the duplicated sequence $(1,\ldots,N,N+1,\ldots,2N)$ as $Z^{\rm ext}_{i,j} = Z_{j+1,N+i-1}$.

The average of the helicity degree, obtained from numerical exact enumeration over 30 disordered sequences as a function of the temperature, is shown in Figs. 4(a) and 4(b) for attractive ($\epsilon_0 < 0$) and repulsive ($\epsilon_0 > 0$) background energy, respectively. It is seen that the helicity in the zero-temperature limit depends on the value of the energy parameters ϵ_0 and ϵ, as will be discussed in full detail further below.

III. HOMOPOLYMER RESULTS

Most of the known results on RNA folding properties were obtained for the special case of homopolymers, which we shall review in this section. De Gennes derived an expression for the canonical partition function based on a singularity analysis of the generating function [38]. By setting $\epsilon = 0$ in Eq. (1), the pairing energy becomes site independent, i.e., $\epsilon_{i,j} = \epsilon_0$, $\forall(i,j)$, thus making the energy of the structure $\mathcal{S}$ depending only on the number of paired bases $|\mathcal{S}|$ defined by Eq. (15).

A. Folded RNA without loops

First, let us consider the simpler no-loop scenario. Since for homopolymers the partition function is translational invariant, the restricted partition function for a sequence of length $j-i+1$ can be written as $Z_{i,j} = Q_{j-i+1}$, which depends only on the difference between position i and position j. It is useful to introduce the z-transform of Q_N,

$$Q(z) = \sum_{N=1}^{\infty} z^{-N} Q_N. \tag{19}$$

The canonical partition function

$$Z_N = Q_N = \sum_{\{\mathcal{S}\}} w^{|\mathcal{S}|}, \tag{20}$$

where $|\mathcal{S}|$ is defined by Eq. (15), is obtained by back transforming the positive root of the equation for $Q(z)$. In the limit of large N, a saddle-point approximation yields the scaling [14]

$$Z_N = \sum_{\{\mathcal{S}\}} w^{|\mathcal{S}|} \sim \xi(w) N^{\alpha-1} z_0^{-N}, \tag{21}$$

where $\alpha = -1/2$ and $z_0 = 1/(1+2\sqrt{w})$, while $\xi(w)$ is a scaling function that depends on the homopolymer statistical weight,

$$w = \exp(-\beta\epsilon_0). \tag{22}$$

In this description, the free energy assumes the same scaling for all temperatures with universal prefactor 3/2 characteristic of the folded state [38], and thus no phase transition takes place.

From Eq. (21), it is possible to express the helicity degree given by Eq. (14) as a function of the statistical weight w in a very simple form. By writing the homopolymer partition function as $Z_N = \sum_{\{\mathcal{S}\}} e^{|\mathcal{S}|\ln w}$, and using Eq. (14), we obtain

$$\theta = \frac{2}{N}\frac{1}{Z_N}\frac{\partial}{\partial \ln w} Z_N = \frac{2}{N}\frac{\partial \ln Z_N}{\partial \ln w}. \tag{23}$$

Finally, using $w\partial/\partial w = \partial/\partial \ln w$, and the scaling form found by de Gennes in Eq. (21) for $N \gg 1$, the helicity in the folded

phase takes the form

$$\begin{aligned}\theta &= \frac{2}{N} w \frac{\partial \ln Z_N}{\partial w} \\ &= \frac{2}{N} w \frac{\partial}{\partial w}\left[\ln \xi(w) - \frac{3}{2}\ln N + N \ln(1+2\sqrt{w})\right] \\ &\approx \frac{2\sqrt{w}}{1+2\sqrt{w}}. \end{aligned} \tag{24}$$

In this case, the helicity, as shown in Fig. 4, asymptotically approaches a constant value in the high-temperature limit,

$$\theta_\infty = \lim_{T\to\infty} \theta = \lim_{\beta\to 0} \frac{2e^{-\beta\epsilon_0/2}}{1+2e^{-\beta\epsilon_0/2}} = \frac{2}{3}, \tag{25}$$

where the last equality follows from $\beta\epsilon_0 = \beta\epsilon_0^U - \epsilon_0^S$ with a vanishing entropic term ϵ_0^S. We note that the occurrence of a finite fraction of bound base pairs at high temperatures follows if the binding energy does not contain an explicit entropic contribution. A finite entropic cost of base pairing, $\epsilon_0^S < 0$, decreases the fraction of bound base pairs at high temperature.

B. RNA folding with loops

Let us now consider the case of a finite loop entropy. As in the no-loop scenario, for homopolymers, we set $w = e^{-\beta\epsilon_0}$ for all pairs and consider the translationally invariant partition function $\widetilde{Q}_N^M$ with $M \in [-1, N]$ non-nested backbones of a chain of N links. In the absence of external forces [39,40], the canonical partition function including loops is obtained by summing over all non-nested backbones as $Z_N^{\rm loop} = \sum_{M=0}^{\infty} \widetilde{Q}_N^M$. The grand-canonical partition function follows as

$$\mathcal{Z}^{\rm loop}(z) = \sum_{N=0}^{\infty} z^N Z_N^{\rm loop} = \sum_{N=0}^{\infty}\sum_{M=0}^{\infty} z^N \widetilde{Q}_N^M, \tag{26}$$

where z is the fugacity. By performing the double sum $\sum_{N=0}^{\infty} z^N \sum_{M=0}^{\infty}$ on both sides of the recursive equation and rearranging indices, one obtains [31]

$$\mathcal{Z}^{\rm loop}(z) = \frac{\kappa(w,z)}{1 - z\kappa(w,z)}, \tag{27}$$

where $\kappa(w,z)$ is the grand-canonical partition function of RNA structures with zero non-nested backbones, i.e., structures which consist of just one nucleotide or structures where the terminal bases are paired. For $|z\kappa| < 1$, the grand-canonical partition function given by Eq. (27) can be expanded into a geometric series as $\mathcal{Z}^{\rm loop} = \sum_{M=0}^{\infty} z^M \kappa^{M+1}$. Comparing the coefficients of the power series with Eq. (26) leads to the implicit equation for $\kappa(w,z)$,

$$\kappa(w,z) - 1 = \frac{w}{\kappa}\mathrm{Li}[c, z\kappa(w,z)], \tag{28}$$

where $\mathrm{Li}(c,x) \equiv \sum_{n=1}^{\infty} x^n n^{-c}$ is the polylogarithm [41]. This relation yields the first constitutive equation for homopolymers with loop entropy.

The canonical partition function obtained from Eq. (27) takes the general form of Eq. (21), but with α and z_0 not determined univocally. In fact, contrary to the no-loop scenario, now the grand-canonical partition function features two relevant singularities. These are the simple pole $z_0 = z_p$, where the denominator of Eq. (27) vanishes, and the branch point $z_0 = z_b$ of the function $\kappa(w,z)$, characteristic of the unfolded and folded phase, respectively.

For $z < z_b$, at least one real solution of Eq. (28) exists, while exactly at $z = z_b$, the two solutions merge and the slope of the right-hand side of Eq. (28) at the tangent point z_b equals unity [31]. By imposing this condition, Eq. (28) yields

$$\kappa_b^2 = w[\mathrm{Li}(c-1, z_b\kappa_b) - \mathrm{Li}(c, z_b\kappa_b)], \tag{29}$$

where we use the short notation $\kappa_b \equiv \kappa(w, z_b)$. This relation together with Eq. (28) univocally determines the branch singularity z_b, if it exists. By a first-order expansion of $\kappa(w,z)$ near the branch point, the canonical partition function can be shown to scale as [31]

$$Z_N^{\rm loop} \sim \xi_b(w) N^{-3/2} {z_b}^{-N}, \tag{30}$$

which demonstrates that the free energy scales logarithmically in N with universal prefactor $3/2$, in agreement with the no-loop scenario given by Eq. (21). The partition function given by Eq. (30) therefore describes homopolymeric RNA including loop entropy contributions in the folded phase. From Eq. (30), the helicity degree given by Eq. (23) follows as $\theta_b \approx (2w/N)(\partial \ln z_b^{-N}/\partial w) = (-2w/z_b)(\partial z_b/\partial w)$, which gives

$$\theta_b \approx \frac{2\mathrm{Li}(c, z_b\kappa_b)}{\mathrm{Li}(c-1, z_b\kappa_b)}. \tag{31}$$

Conversely, the simple pole z_p of Eq. (27) is determined, together with Eq. (28), by

$$z_p\kappa_p = 1, \tag{32}$$

where $\kappa_p \equiv \kappa(w, z_p)$. By inserting Eq. (32) into Eq. (28), the explicit expression for the pole singularity is obtained,

$$z_p = 2[1 + \sqrt{1 + 4w\zeta(c)}]^{-1}, \tag{33}$$

where $\zeta(c) = \mathrm{Li}(c,1) = \sum_{n=1}^{\infty} n^{-c}$ is the Riemann zeta function. In this phase, the partition function scales as [31]

$$Z_N^{\rm loop} \sim \xi_p(w) {z_p}^{-N}, \tag{34}$$

which in contrast to the branch point does not lead to a logarithmic N contribution in the free energy. Equation (34) describes the thermodynamics of homopolymers above the melting critical point. In this phase, the helicity degree $\theta_p \approx (-2w/z_p)\partial z_p/\partial w$ takes the explicit form

$$\theta_p \approx 1 - \frac{1}{\sqrt{1 + 4w\zeta(c)}}. \tag{35}$$

A phase transition is possible only if a critical fugacity z_m and a critical base-pairing weight w_m exist such that $z_m = z_b(w_m) = z_p(w_m)$. At the critical point, the three constitutive equations (28), (29), and (32) have to hold simultaneously. Using Eqs. (32) and (33), a closed-form expression can be given for the critical weight as a function of the loop exponent [31],

$$w_m(c) = \frac{\zeta(c-1) - \zeta(c)}{[\zeta(c-1) - 2\zeta(c)]^2}. \tag{36}$$

This equation defines the critical line in the phase diagram for homopolymer RNA and also determines an upper bound for

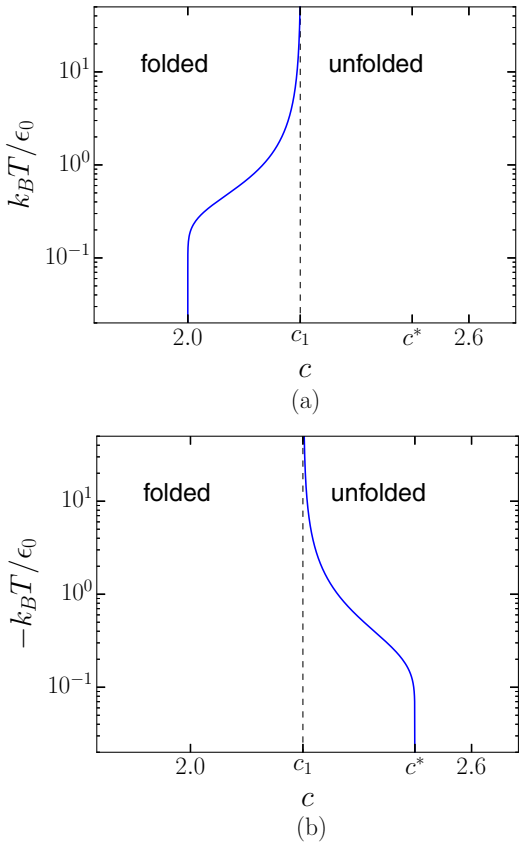


FIG. 5. Phase diagram of homopolymeric RNA in the $T - c$ plane featuring the unfolded and folded phase for (a) repulsive ($\epsilon_0 > 0$) and (b) attractive ($\epsilon_0 < 0$) base-pair interaction energy. The critical lines, obtained by solving $w_m(c) = w = e^{-\beta\epsilon_0}$, diverge for $c = c_1$ for both attractive and repulsive interaction energy. For $c \leqslant 2$, the molecule is always folded and at $c^* \approx 2.479$ the critical weight $w_m(c)$ diverges, so that for $c > c^*$ no folded phase can exist.

the range of the loop exponent c in which a phase transition can occur. The latter is given by the universal value $c^* \approx 2.479$ [31] for which the denominator of Eq. (36) vanishes,

$$0 = \zeta(c^* - 1) - 2\zeta(c^*). \tag{37}$$

As the statistical weight at the critical point, $w_m(c^*)$, diverges, the melting temperature is zero, and thus the molecule is always in the unfolded state. In addition, the derivative with respect to κ of the right-hand side of Eq. (28) converges only for $c > 2$, which sets the lower bound of the critical region. This means that only for $2 < c < c^*$ can a phase transition occur between the folded and the unfolded phases. The corresponding phase diagram for attractive and repulsive interaction is obtained by solving $w_m(c) = \exp(-\beta\epsilon_0)$ and is displayed in Fig. 5.

We define an additional universal value for the loop exponent that will play a crucial role in the following. Depending on the nature of the interaction, repulsive if $\epsilon_0 > 0$ and attractive otherwise, the behavior of the statistical weight of base-pair formation $w = \exp(-\beta\epsilon_0)$ can be smaller than or greater than one. We define the value $c_1 \approx 2.241$ from the equation

$$0 = \zeta(c_1 - 1) - \zeta(c_1) - [\zeta(c_1 - 1) - 2\zeta(c_1)]^2, \tag{38}$$

which determines the statistical weight at the melting point with $w_m(c_1) = 1$. We note that a nonvanishing entropic contribution $\epsilon_0^S \neq 0$ in the base-pairing energy ϵ_0 would change the universal value of c_1 accordingly by imposing the condition $w_m(c_1) = e^{\epsilon_0^S}$ instead of $w_m(c_1) = 1$. As we will show next, a necessary condition for cold melting is that $c > c_1$ is satisfied.

IV. CONSTRAINED ANNEALING WITH LOOP ENTROPY

A. Outline of the method

Instead of the replica approach used in spin-glass theory [21], to compute the average over the disorder of the free energy given by Eq. (7), we use the *constrained annealing approximation* [33]. Here we closely follow the approach outlined in [13]. A related approach for the constrained annealed ensemble was first presented for RNA with quenched disorder and no loops in [42]. The basic idea of the method is to perform an annealed average,

$$f^a = -\frac{1}{\beta N} \ln \overline{Z_N(h)}, \tag{39}$$

in which the disorder variables h_i are free to evolve with the dynamical degrees of freedom $s_{i,j}$ in the partition function (5), but with the additional requirement that $\{h\}$ are coupled to appropriate constraints $\{\mu\}$, which assume the form of Lagrange multipliers. The values of the constraints $\{\mu\}$ that for $N \gg 1$ maximize the thermodynamic potential,

$$f^{ca}(\mu) = -\frac{1}{\beta N} \ln Z_N^{ca}(\mu), \tag{40}$$

are those that select the realizations with a correct value of the disorder intensive variables and, at the same time, minimize the difference between the quenched free energy (7) and the annealed free energy (39).

The constrained annealing free energy $f^{ca}(\mu)$ improves the lower bound estimation for the quenched free energy based on Jensen's inequality [43] given by f^a, so that

$$\overline{f(h)} \geqslant f^{ca}(\mu) \geqslant f^a \quad \forall\mu. \tag{41}$$

The constrained annealing partition function [33],

$$Z_N^{ca}(\mu) = \overline{Z_N(h) e^{-N\mu\alpha(h)}}, \tag{42}$$

is defined in terms of a function $\alpha(h)$, that we define as

$$\alpha(h) \equiv \frac{1}{N} \sum_{i=1}^{N} [h_i - (2p - 1)], \tag{43}$$

so that its average vanishes. Using

$$\bar{h} \equiv \sum_{\{h\}} \mathcal{P}(h) h = \sum_{h=\pm 1} \rho(h) h = 2p - 1, \tag{44}$$

it follows immediately that

$$\alpha(h) \xrightarrow[N\to\infty]{} \overline{\alpha(h)} = 0. \tag{45}$$

In the following, we show that since the model considered here is separable [44] as the disorder in the Hamiltonian (1) depends on sites and not on links, each term can be averaged independently. To see this, we write the partition function

given by Eq. (5) as

$$Z_N(h) = \sum_{\{\mathcal{S}\}} e^{-\beta\epsilon_0|\mathcal{S}|} \prod_{k<l}^{N} e^{-\beta\epsilon s_{k,l} h_k h_l}. \tag{46}$$

The constrained annealing partition function (42) then reads

$$Z_N^{ca}(\mu) = e^{N\mu(2p-1)} \sum_{\{\mathcal{S}\}} e^{-\beta\epsilon_0|\mathcal{S}|} \Pi(\mu), \tag{47}$$

with

$$\Pi(\mu) \equiv \sum_{\{h\}} \prod_{i=1}^{N} \rho(h_i) e^{-\mu h_i} \prod_{k<l}^{N} e^{-\beta\epsilon s_{k,l} h_k h_l}. \tag{48}$$

In this product, we have contributions that are different from unity only when $s_{k,l} = 1$ and each base can only participate in, at most, one base pair. Thus, in the average over the disorder, we get a product of $|\mathcal{S}| = \sum_{i<j} s_{i,j}$ times the factor $e^{-\beta\epsilon h_k h_l}$ and, since h_i are mutually independent, each disorder term can be averaged independently. Performing the average in Eq. (48), we obtain

$$\begin{aligned}\Pi(\mu) &= \left[\sum_{h=\pm1} \rho(h) e^{-\mu h}\right]^{N-2|\mathcal{S}|} \\ &\quad \times \left[\sum_{h,h'=\pm1} \rho(h)\rho(h') e^{-\mu h} e^{-\mu h'} e^{-\beta\epsilon h h'}\right]^{|\mathcal{S}|} \\ &= \Omega^N(\mu) \left[\frac{\Upsilon(\mu)}{\Omega^2(\mu)}\right]^{|\mathcal{S}|},\end{aligned} \tag{49}$$

where we have defined the two auxiliary quantities,

$$\Omega(\mu) = pe^{-\mu} + (1-p)e^{\mu},$$
$$\Upsilon(\mu) = e^{-\beta\epsilon}[p^2 e^{-2\mu} + (1-p)^2 e^{2\mu}] + e^{\beta\epsilon} 2p(1-p).$$

The disorder average yields the new interaction energy,

$$\bar{\epsilon}(\mu) \equiv -\frac{1}{\beta} \ln \frac{\Upsilon(\mu)}{\Omega^2(\mu)}, \tag{50}$$

where all the information relative to the disorder is included in the variational parameter μ. As a consequence, Eq. (47) reduces to

$$Z_N^{ca}(\mu) = e^{N\mu(2p-1)} \Omega^N(\mu) Z_N^{\text{hom}}(\mu). \tag{51}$$

Here,

$$Z_N^{\text{hom}}(\mu) = \sum_{\{\mathcal{S}\}} [w^{ca}(\mu)]^{|\mathcal{S}|} \tag{52}$$

is the partition function of homopolymers of the form (20) with statistical weight

$$w^{ca}(\mu) = \exp[-\beta\epsilon^{ca}(\mu)], \tag{53}$$

with constant interaction energy between monomers,

$$\epsilon^{ca}(\mu) = \epsilon_0 + \bar{\epsilon}(\mu). \tag{54}$$

The maximization with respect to μ of the reduced free energy,

$$\beta f^{ca}(\mu) = \ln \frac{e^{-\mu(2p-1)}}{\Omega(\mu)} - \frac{1}{N} \ln Z_N^{\text{hom}}(\mu), \tag{55}$$

is achieved by imposing

$$\left.\frac{\partial}{\partial\mu} f^{ca}(\mu)\right|_{\mu=\tilde{\mu}} = 0. \tag{56}$$

This condition yields the value $\tilde{\mu}(\beta\epsilon_0, \beta\epsilon)$ for which

$$f^{ca}(\tilde{\mu}) = \max_{\mu} f^{ca}(\mu) \approx \overline{f(h)}, \tag{57}$$

while the lower bound given by the annealed free energy is obtained simply by setting $\mu = 0$ in Eq. (51).

B. Critical behavior: Cold melting

The behavior of the specific heat and the helicity degree for the two-letter model with quenched disorder has been previously studied with the constrained annealing method [13] by solving Eq. (56) and using the folded-phase partition function given by Eq. (21) for Z_N^{hom} in (55). The comparison between the quenched free energy, computed by exact enumeration of the disorder realizations and the constrained annealing average, showed remarkable agreement. A double-peak structure in the specific heat C_V appears for $0.5 < p < 1$ and $\epsilon_0 \neq 0$. Interestingly, if the energy parameters satisfy the condition $\epsilon > |\epsilon_0|$, the low-temperature peak is more pronounced than the high-temperature peak and the helicity degree θ decreases with lowering temperature.

Here we provide a physical interpretation of the results obtained in [13] and account for loop entropy by using the folded and the unfolded scaling expressions for the partition function, given by Eqs. (30) and (34), with corresponding free energies

$$\beta f_{[b,p]}^{ca}(\mu) \approx \ln \frac{e^{-\mu(2p-1)}}{\Omega(\mu)} + \ln z_{[b,p]}(\mu). \tag{58}$$

Here the lower index in square brackets distinguishes between the branch and the pole singularity associated with the folded and unfolded phases, respectively. While z_b can only be computed by numerically solving Eqs. (28) and (29) simultaneously, the pole singularity z_p is explicitly given by Eq. (33) for $w = w^{ca}(\mu)$. A phase transition in the disordered model can then be described in terms of the two singularities $z_{[b,p]}$ of the homopolymeric partition function. In particular, the unusual decrease in paired bases found in [13] when lowering the temperature can formally be interpreted as a thermal phase transition to the unfolded state.

To proceed, Eq. (56) yields the equation

$$\begin{aligned}2p-1 &= \left\{\frac{\partial \ln \Omega(\mu)}{\partial\mu} + \frac{1}{NZ_N^{\text{hom}}(\mu)}\right. \\ &\quad \left.\times \sum_{\{\mathcal{S}\}} |\mathcal{S}|[w^{ca}(\mu)]^{|\mathcal{S}|-1} \frac{\partial}{\partial\mu} w^{ca}(\mu)\right\}_{\mu=\tilde{\mu}} \\ &= \left[\frac{\partial \ln \Omega(\mu)}{\partial\mu} + \frac{\langle|\mathcal{S}|\rangle}{N} \frac{\partial \ln w^{ca}(\mu)}{\partial\mu}\right]_{\mu=\tilde{\mu}} \\ &= \left[\frac{\partial \ln \Omega(\mu)}{\partial\mu} + \frac{\theta^{ca}(\mu)}{2} \frac{\partial}{\partial\mu} \ln \frac{\Upsilon(\mu)}{\Omega^2(\mu)}\right]_{\mu=\tilde{\mu}},\end{aligned} \tag{59}$$

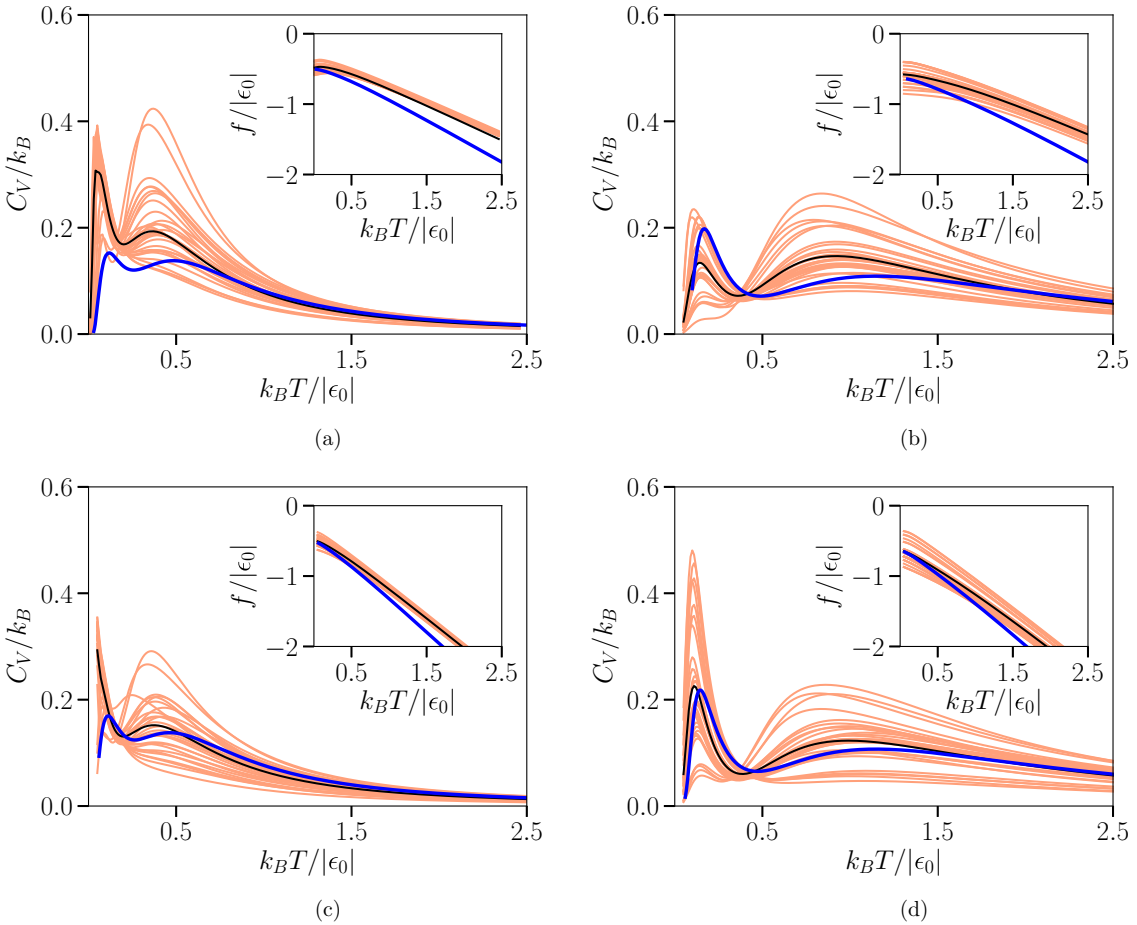


FIG. 6. Specific heat with loops in the folded phase (a),(b) with $c = c_{\rm RW} = 1.5$ and (c),(d) without loops (corresponding to $c = 0$). The quenched average (black curve) is numerically obtained from averaging over 30 random sequences (red curves) with $N = 50$, $p = 0.75$, and interaction energy (a),(c) $\epsilon = 0.5|\epsilon_0|$ and (b),(d) $\epsilon = 1.5|\epsilon_0|$. The analytical result using the constrained annealing method is shown in blue. Here, the partition function of each sequence is computed with the loop recursive equation (12) and the constrained annealing free energy is $f_b^{ca}(\widetilde{\mu})$ defined by Eq. (58). Insets: The corresponding free energies from numerics and the constrained annealing method from which the specific-heat curves in the main panels are obtained.

where we have used Eqs. (23), (50), and (54), and where

$$\begin{aligned}\theta^{ca}(\mu) &= 2\frac{\partial}{\partial(\beta\epsilon_0)}\ln\left[\frac{1}{1+2\sqrt{w^{ca}(\mu)}}\right]\\ &= \frac{-2}{1+2\sqrt{w^{ca}(\mu)}}\left[\frac{1}{\sqrt{w^{ca}(\mu)}}\frac{\partial w^{ca}(\mu)}{\partial(\beta\epsilon_0)}\right]\\ &= \frac{2\sqrt{w^{ca}(\mu)}}{1+2\sqrt{w^{ca}(\mu)}}.\end{aligned} \tag{60}$$

Once the solution $\widetilde{\mu}(\beta\epsilon_0, \beta\epsilon)$ of Eq. (59) is known, the folded-phase singularity $z_b(\widetilde{\mu})$ that maximizes the constrained annealing free energy can be obtained by solving the following system of equations:

$$\begin{aligned}\widetilde{\kappa_b^{ca}}\left(\widetilde{\kappa_b^{ca}}-1\right) &= w^{ca}\mathrm{Li}\left(c, z_b\widetilde{\kappa_b^{ca}}\right)\\ \left(\widetilde{\kappa_b^{ca}}\right)^2 &= w^{ca}\left[\mathrm{Li}\left(c-1, z_b\widetilde{\kappa_b^{ca}}\right) - \mathrm{Li}\left(c, z_b\widetilde{\kappa_b^{ca}}\right)\right],\end{aligned} \tag{61}$$

where we use the notation $\widetilde{\kappa_b^{ca}} \equiv \kappa[w^{ca}(\widetilde{\mu}), z_b(\widetilde{\mu})]$ and where all variables are evaluated at $\mu = \widetilde{\mu}$. In the unfolded phase, $z_p(\widetilde{\mu})$ is instead determined by the explicit expression

$$z_p(\widetilde{\mu}) = 2[1+\sqrt{1+4w^{ca}(\widetilde{\mu})\zeta(c)}]^{-1}. \tag{62}$$

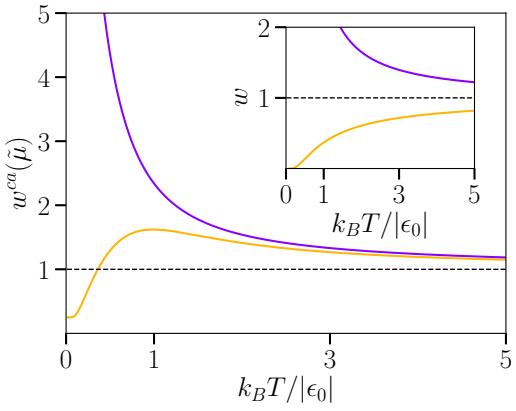


FIG. 7. Statistical weight in the constrained annealing approximation in the competitive regime ($\Lambda = 2/3$) with $\epsilon = 1.5|\epsilon_0|$ (orange) and in the noncompetitive regime ($\Lambda = 2$) with $\epsilon = 0.5|\epsilon_0|$ (violet) as a function of the temperature at $p = 0.75$. Inset: Homopolymer weight $w = e^{-\beta\epsilon_0}$ with $\epsilon_0 > 0$ (orange) and $\epsilon_0 < 0$ (violet), for repulsive and attractive background interaction, respectively.

The disorder average of the free energy in both phases (58) together with the system of Eqs. (61) and (62) constitute the main result of this paper.

At this point, we have all the necessary ingredients to describe the melting phase transition of disordered RNA molecules with loop entropy. First, we note that the critical weight $w_m(c)$ defined by Eq. (36) is a monotonically increasing function of the loop exponent c in the interval $2 < c < c^*$, which defines the critical region for homopolymers. For $w > w_m$, the molecule is always folded, governed by z_b, and unfolded otherwise, governed by z_p.

In Figs. 6(a) and 6(b), we compare quenched and constrained annealing averages in the folded phase for loop exponent $c = c_{\rm RW} = 1.5$, which qualitatively reproduce the behavior for $c = 0$ shown in Figs. 6(c) and 6(d), with the characteristic double-peak structure in the specific heat. The latter is computed from the free energy as

$$C_V^{ca}(\widetilde{\mu}) = -T\frac{\partial^2 f^{ca}(\widetilde{\mu})}{\partial T^2}. \tag{63}$$

The quenched and constrained annealing free energies show good agreement for low temperatures, with the exception of the first few data points at $T \approx 0$ in the no-loop noncompetitive scenario [Fig. 6(c)]. The deviation between numerics and the constrained annealing results is roughly linear in T and thus does not affect the specific heat significantly.

In the following, we distinguish the two energetic regimes, depending on whether or not quenched disorder significantly affects RNA folding. Since $\widetilde{\mu}$ depends implicitly on $\beta = (k_B T)^{-1}$, $\beta\epsilon_0$, and $\beta\epsilon$, we have two possible regimes depending on the value of the base-paring energy ratio,

$$\Lambda \equiv \frac{|\epsilon_0|}{\epsilon}. \tag{64}$$

If $\Lambda > 1$, there is no competition between the favorable UA and unfavorable AA/UU pairings, meaning that they are either both attractive or repulsive [see Eq. (2)]. We will call this the *noncompetitive regime*. Instead, if $\Lambda < 1$, the effect of quenched disorder becomes relevant and there is an energetic competition between favorable and unfavorable pairings. This case we refer to as the *competitive regime*.

By setting $\epsilon_0 < 0$, in the competitive regime $\epsilon_0 + \epsilon$ is positive. In this case, $w^{ca}(\widetilde{\mu})$ exhibits a global maximum as well as a global minimum at low temperatures, independently of the values of p, ϵ_0, and ϵ; see Fig. 7. From the temperature dependence of the statistical weight for homopolymers, $w = \exp(-\beta\epsilon_0)$, shown in the inset of Fig. 7, substantially different behavior is found for the disordered model in the competitive regime (orange curve). When $\Lambda < 1$, the competition between favorable and unfavorable base pairs results in a global maximum in $w^{ca}(\widetilde{\mu})$, for all probability values in the range $0.5 < p < 1.0$. As we show next, the presence of a global maximum in the statistical weight of base pairings is closely related to the behavior of the helicity degree.

From the temperature dependence of the helicity degree θ in the competitive and noncompetitive regimes, we find a behavior that is similar to that of the statistical weight of base pairings; see insets in Fig. 8. However, only in the competitive regime where $\Lambda < 1$ [Fig. 8(b)], the helicity degree drops when lowering the temperature and displays a

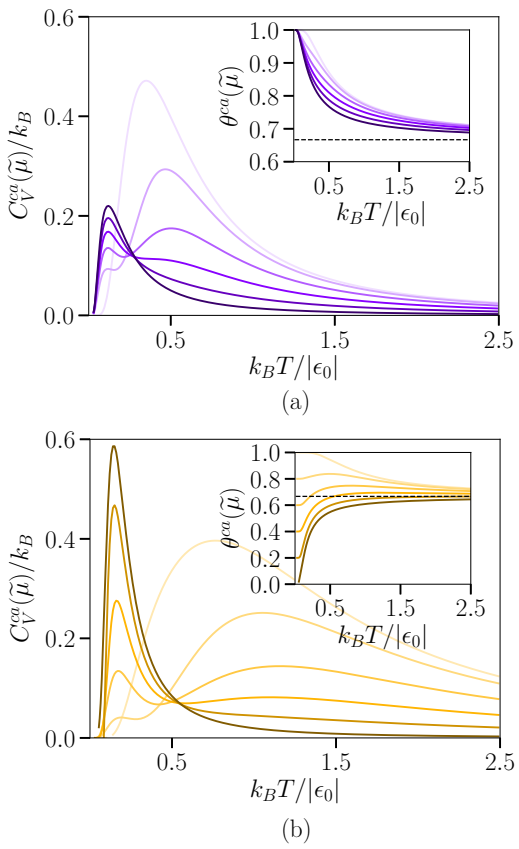


FIG. 8. Specific heat for $c = c_{\rm RW} = 1.5$ and different probabilities, from light to dark, $p = 0.5$, $p = 0.6$, $p = 0.7$, $p = 0.8$, $p = 0.9$, and $p = 1.0$, in the constrained annealing approximation in the (a) noncompetitive regime ($\Lambda = 2$) with $\epsilon = 0.5|\epsilon_0|$ and (b) competitive regime ($\Lambda = 2/3$) with $\epsilon = 1.5|\epsilon_0|$. Insets: Corresponding helicity degrees. All curves converge to the asymptotic value of Eq. (25) in the high-temperature molten phase.

global maximum for all $0.5 < p < 1.0$. In Fig. 8, we show that the specific heat C_V always displays two peaks, corresponding to hot and cold melting, respectively. This result confirms that cold melting appears only in the competitive regime where the low-temperature peak of C_V is more pronounced.

Finally, it is interesting to compare the helicity degree in the RNA structures with disorder with the fraction η^- of Watson-Crick (i.e., AU) and η^+ of non Watson-Crick (i.e., AA and UU) base pairs. To numerically estimate the fraction of favorable η^- and unfavorable base pairs η^+, we define the auxiliary quantity $\eta = \frac{2}{N}\overline{\langle \sum_{i<j} h_i h_j s_{i,j}\rangle}$, so that

$$\eta^{\pm} = \tfrac{1}{2}(\theta \pm \eta). \tag{65}$$

Performing the same derivation as in Eq. (16), in the constrained annealing approximation, we find $\eta(\mu) \approx 2\frac{\partial}{\partial(\beta\epsilon)}\beta f^{ca}(\mu)$, which gives

$$\eta(\mu) \approx \theta(\mu)\left[1 - \frac{4e^{\beta\epsilon}p(1-p)}{\Upsilon(\mu)}\right]. \tag{66}$$

The temperature dependence of $\eta^{\pm}$, shown in Fig. 9, gives additional insight into the mechanism behind the low-temperature phase transition. In the noncompetitive regime

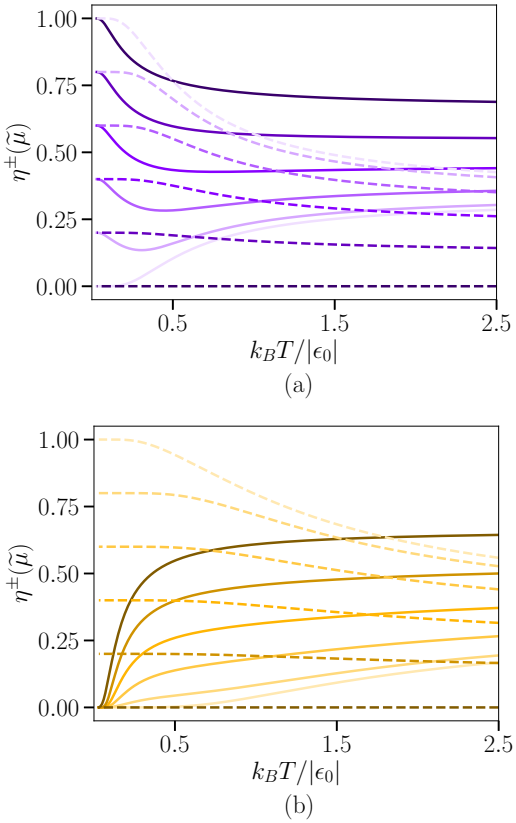


FIG. 9. Fraction η^- of favorable (dashed line) and η^+ of unfavorable (solid line) base pairs for $c = c_{\rm RW} = 1.5$ and different probabilities, from light to dark, $p = 0.5$, $p = 0.6$, $p = 0.7$, $p = 0.8$, $p = 0.9$, and $p = 1.0$, in the constrained annealing approximation in the (a) noncompetitive regime ($\Lambda = 2$) with $\epsilon = 0.5|\epsilon_0|$ and (b) competitive regime ($\Lambda = 2/3$) with $\epsilon = 1.5|\epsilon_0|$.

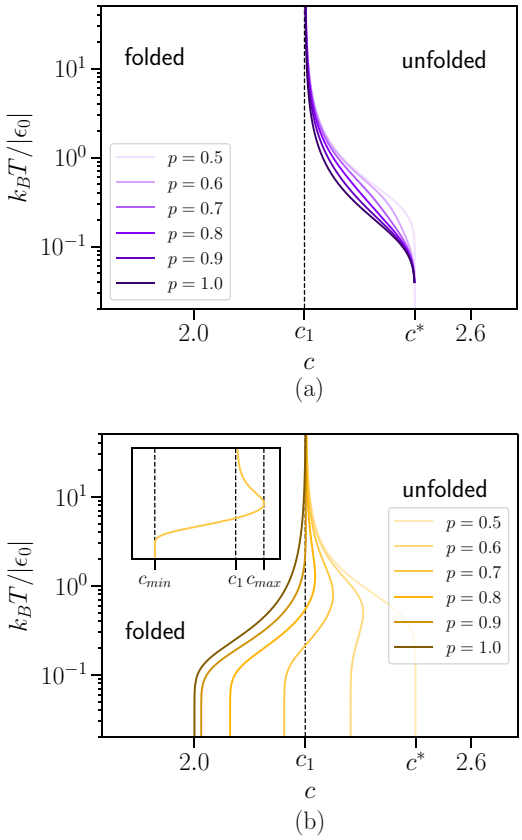


FIG. 10. (a) Phase diagram in the noncompetitive regime ($\Lambda = 2$) with $\epsilon = 0.5|\epsilon_0|$ and $\epsilon_0 < 0$. The critical line that defines hot melting is defined in the range $c_1 < c < c^*$ [see Eqs. (37) and (38)] so that for $c < c_1$, the RNA molecule is always folded. (b) Phase diagram in the competitive regime ($\Lambda = 2/3$) with $\epsilon = 1.5|\epsilon_0|$ and $\epsilon_0 < 0$. For $c < c_{\rm min}$, the molecule is always folded since $w_m(c) < w^{ca}(T)$, $\forall T$. For $c_{\rm min} < c < c_1$, there is only hot melting, while for $c > c_{\rm max}$, the molecule is always unfolded since $w_m(c) > w^{ca}[T, \widetilde{\mu}(T)]$, $\forall T$. In the range $c_1 < c < c_{\rm max}$, both hot and cold melting take place. In both panels, the color of each curve corresponds to a different value of the probability p, from light to dark, $p = 0.5$, $p = 0.6$, $p = 0.7$, $p = 0.8$, $p = 0.9$, and $p = 1.0$. Inset: Critical line for $p = 0.75$. Since $0.75 > p*$ in the range $c_1 < c < c_{\rm max}$, the double intersection between $w_m(c)$ and $w^{ca}[T, \widetilde{\mu}(T)]$ introduces the additional cold-melting phase transition.

[Fig. 9(a)], the behavior of $\eta^{\pm}$ follows the one found for the helicity in Fig. 8(a), with $\eta^{\pm}$ increasing at low temperature. We see that only in the competitive regime [Fig. 9(b)] does the fraction of non-Watson-Crick contacts η^+ (solid lines) abruptly decrease at low temperatures for all values of the probability p, roughly coinciding with the low-temperature peak of the specific heat found in Fig. 8(b). Thus, in this regime, the base-pairing mechanism at low temperatures between identical nucleotides AA and UU is suppressed by the disorder-induced competition between those pairings and the favorable pairings. Instead, Watson-Crick contacts η^- in the competitive regime monotonically grow when lowering the temperature in a similar way as in the noncompetitive regime.

C. Global phase diagram

In order to fully understand the peculiar behavior of the specific heat and the helicity degree in the competitive regime, it is instructive to consider the loop exponent values c_1, defined by Eq. (38), and the two additional values $c_{\rm max}$ and $c_{\rm min}$ defined, respectively, as

$$w_m(c_{\rm max}) \equiv \max_T w^{ca}[T, \widetilde{\mu}(T)], \tag{67}$$

$$w_m(c_{\rm min}) \equiv w^{ca}[0, \widetilde{\mu}(0)]. \tag{68}$$

Note that since w^{ca} is an implicit function of p, ϵ_0, and ϵ, so are the loop exponents $c_{\rm max}$ and $c_{\rm min}$.

Depending on whether we are in the competitive regime ($\Lambda < 1$) or not ($\Lambda > 1$) and depending on the value of the loop exponent c, there are three possibilities. There can be two, one, or no intersection between $w^{ca}[T, \widetilde{\mu}(T)]$ and $w_m(c)$. The last two cases correspond to the well-known high-temperature denaturation of RNA molecules and to the stable folded phase, respectively. If two intersections are present, we obtain more complex behavior with an additional cold-

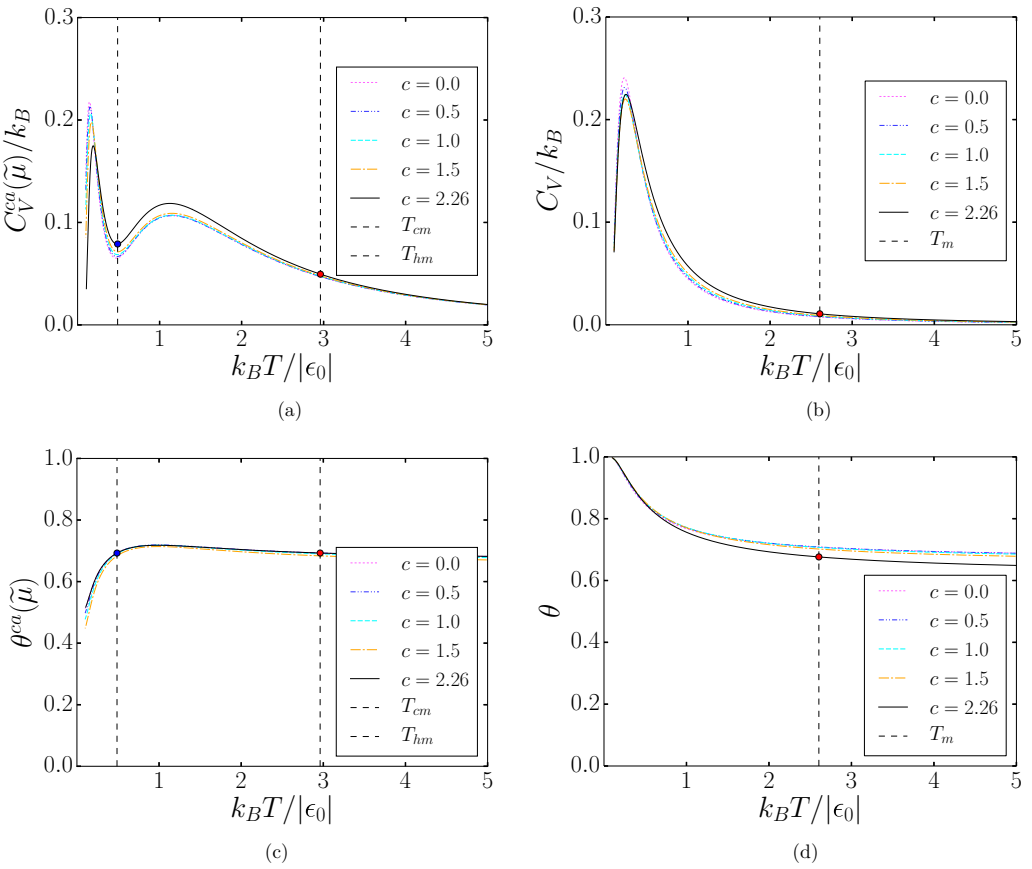


FIG. 11. Specific heat for the disordered model using the constrained annealing approximation with $p = 0.75$ and $\epsilon = 1.5|\epsilon_0|$ in (a) the competitive regime with $\Lambda = 2/3$ and (b) for a homopolymer with $\epsilon_0 < 0$. In both cases, the different curves correspond to different values of the loop exponent c. In the disordered model, the two critical temperatures $T_{cm} \approx 0.488|\epsilon_0|/k_B$ (blue circle) and $T_{hm} \approx 2.961|\epsilon_0|/k_B$ (red circle) correspond to the value of the loop exponent $c = 2.26$, while for the homopolymer there is only a single melting transition at $T_m \approx 2.605|\epsilon_0|/k_B$. In (c) and (d), the corresponding helicity degrees are shown.

melting phase transition in the low-temperature regime. In the latter case, if $p > p_1$, where p_1 is defined by $w_m[c_{\min}(p_1)] = 1$, cold melting appears for $c_1 < c < c_{\max}$, where $c_{\max}$ depends on p, ϵ_0, and ϵ, while $c_1 \approx 2.241$ is the universal value defined by Eq. (38). On the contrary, if $p < p_1$, one has $c_{\min}(p) > c_1$ and cold melting appears for $c_{\min} < c < c_{\max}$. In this case, the homopolymer melting weight $w_m(c)$ intersects $w^{ca}[T, \widetilde{\mu}(T)]$ at two different temperatures T_{cm} and T_{hm}. Cold melting is a proper thermodynamic phase transition and separates two phases that are characterized by different singularities of the grand-canonical partition function given by Eq. (27). However, this can only happen in the competitive regime ($\Lambda < 1$), when $w^{ca}[T, \widetilde{\mu}(T)]$ displays a global maximum. By contrast, in the noncompetitive regime, only hot melting takes place.

In Fig. 10, we present the global phase diagram for disordered RNA in the $T - c$ plane at different probabilities p with the critical lines separating the folded and unfolded phases for the noncompetitive [Fig. 10(a)] and the competitive [Fig. 10(b)] regimes. Each line is obtained by numerically solving the equation $w^{ca}[T, \widetilde{\mu}(T)] = w_m(c)$, which yields the melting temperature as a function of the loop exponent. Comparing these results to those obtained within a replica approach in [14], where in the no-disorder limit the glass-phase critical temperature vanishes, we similarly find that in the limiting case of no disorder, $\epsilon_{ij} = \epsilon_0$, $\forall(i, j)$, the cold-melting transition disappears.

The c dependence in the folded phase of the specific heat is very subtle for both the disordered model, where there is a two-peak structure, and the homopolymeric model, where C_V features only one peak; see Figs. 11(a) and 11(b). A similar weak dependence on c is found for the helicity degree [see Figs. 11(c) and 11(d)], where for the folded phase we use θ_b defined by Eq. (31) with $\kappa_b = \widetilde{\kappa_b^{ca}}$ obtained by solving Eq. (61). In the unfolded phase, θ_p is simply obtained from Eq. (35) with $w = w^{ca}(\widetilde{\mu})$. The cold- and the hot-melting temperatures are obtained numerically for $c = 2.26$, $p = 0.75$, and $\epsilon = 1.5|\epsilon_0|$ as $T_{cm} \approx 0.488|\epsilon_0|/k_B$ and $T_{hm} \approx 2.961|\epsilon_0|/k_B$, respectively. We note that the critical point for the cold-melting transition lies almost exactly in the valley formed by the two peaks of the specific heat.

V. CONCLUSIONS

In this paper, we have studied the critical behavior of RNA secondary structures including both quenched sequence disorder and loop entropy. In previous work based on the

two-letter model without loop entropy, it was shown that the heat capacity exhibits a peculiar two-peak structure in the case of energetic competition between favorable and unfavorable base pairs [13], indicating hot as well as cold melting. For homopolymers, on the other hand, a finite loop entropy leads to a genuine thermal phase transition between the folded and unfolded states in a well-defined range of the loop exponent c [26,31]. Here, we combine the relevant features of these two models and consider a two-letter RNA model with quenched randomness and in the presence of loop entropy. As the central result, we show that if there is an energetic competition between favorable and unfavorable base pairs, in a specific range of the loop exponent c, two different phase transitions occur at two different critical temperatures T_{hm} and T_{cm}, resulting in a folded state that is only stable at intermediate temperatures. Most importantly, the results obtained here are not of limited theoretical interest, but presumably are related to the experimental observation of cold melting of RNA [4,8]. While there is no direct physical interpretation of the loop exponent range within which both hot and cold melting occur, we have provided an interpretation of this phenomenon in terms of the behavior of the statistical weight of base-pair formation w that displays a global maximum at intermediate temperatures. By combining loop entropy with quenched disorder in our model, we are able to reproduce the cold-melting phenomenon and reveal that it is a genuine thermal phase transition.

A natural extension of the method outlined here consists in including all four RNA nucleotides uracil (U), adenine (A), guanine (G), and cytosine (C). This can be done simply by adding an additional disorder term in the Hamiltonian [33]. Then the constrained annealing average can be performed in a similar fashion, leading to qualitatively comparable results for the important thermodynamic quantities studied here; see [13]. The main difference from the two-letter model is that four-letter sequences do not always admit perfect matching, as first established by Valba and collaborators [45], and the limiting helicity degree at zero temperature can be below the value $\theta(T=0) = 2\min(p, 1-p)$ found in the inset in Fig. 8(b). While this might be relevant, to some extent, for investigating the glass transition, we stress that we find qualitatively the same behavior in the helicity degree with a global maximum at intermediate temperatures as found for the four-letter model [13], which is what ultimately explains, in our framework, the cold-melting transition.

Our results suggest that the cold-melting transition is continuous because it is triggered by the same conformational effect as found for homopolymers [31]. In particular, we argue that the cold-melting transition is of the order of n, where n is determined by $(c-2)^{-1} - 1 < n < (c-2)^{-1}$. It remains a challenge for future work to investigate the connection between cold denaturation found in the disordered model and the glass transition [14,15,46,47].

[1] R. F. Gesteland, T. R. Cech, and J. F. Atkins, *The RNA World* (Cold Spring Harbor Laboratory Press, New York, 2005).
[2] I. Tinoco, O. C. Uhlenbeck, and M. D. Levine, Nature (London) **230**, 362 (1971).
[3] J. McCaskill, Biopolymers **29**, 1105 (1990).
[4] P. L. Privalov, Crit. Rev. Biochem. Mol. Biol. **25**, 281 (1990).
[5] P. De Los Rios and G. Caldarelli, Phys. Rev. E **63**, 031802 (2001).
[6] P. L. Privalov, Y. V. Griko, S. Yu. Venyaminov, and V. P. Kutyshenko, J. Mol. Biol. **190**, 487 (1986).
[7] N. C. Pace and C. Tanford, Biochemistry **7**, 198 (1968).
[8] P. J. Mikulecky and A. L. Feig, J. Am. Chem. Soc. **124**, 890 (2002).
[9] P. J. Mikulecky and A. L. Feig, Nucleic Acids Res. **32**, 3967 (2004).
[10] A. V. Finkelstein, *Protein Physics* (Academic Press, London, 2002).
[11] F. Sedlmeier, D. Horinek, and R. R. Netz, J. Chem. Phys. **134**, 055105 (2011).
[12] P. G. Higgs, Phys. Rev. Lett. **76**, 704 (1996).
[13] G. N. Hayrapetyan, F. Iannelli, J. Lekscha, V. F. Morozov, R. R. Netz, and Y. Sh. Mamasakhlisov, Phys. Rev. Lett. **113**, 068101 (2014).
[14] R. Bundschuh and T. Hwa, Phys. Rev. E **65**, 031903 (2002).
[15] A. Pagnani, G. Parisi, and F. Ricci-Tersenghi, Phys. Rev. Lett. **84**, 2026 (2000).
[16] R. Bundschuh and T. Hwa, Phys. Rev. Lett. **83**, 1479 (1999).
[17] K. A. Dill, Biochemistry **24**, 1501 (1985).
[18] F. Krzakala, M. Mézard, and M. Müller, Europhys. Lett. **57**, 752 (2002).
[19] R. Bundschuh and T. Hwa, Europhys. Lett. **59**, 903 (2002).
[20] R. Nussinov and A. B. Jacobson, Proc. Natl. Acad. Sci. USA **77**, 6309 (1980).
[21] M. Mezard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1988).
[22] M. Bon, G. Vernizzi, H. Orland, and A. Zee, J. Mol. Biol. **379**, 900 (2008).
[23] G. Vernizzi, P. Ribeca, H. Orland, and A. Zee, Phys. Rev. E **73**, 031902 (2006).
[24] M. Pillsbury, H. Orland, and A. Zee, Phys. Rev. E **72**, 011911 (2005).
[25] H. Orland and A. Zee, Nucl. Phys. B **620**, 456 (2002).
[26] T. R. Einert, H. Orland, and R. R. Netz, Eur. Phys. J. E **34**, 55 (2011).
[27] P-G. de Gennes, *Scaling Concepts in Polymer Physics* (Cornell University Press, Ithaca, 1979).
[28] B. Duplantier, Phys. Rev. Lett. **57**, 941 (1986).
[29] M. Müller, Phys. Rev. E **67**, 021914 (2003).
[30] Y. Kafri, D. Mukamel, and L. Peliti, Phys. Rev. Lett. **85**, 4988 (2000).
[31] T. R. Einert, P. Näger, H. Orland, and R. R. Netz, Phys. Rev. Lett. **101**, 048103 (2008).
[32] M. V. Tamm and S. K. Nechaev, Phys. Rev. E **75**, 031904 (2007).
[33] M. Serva and G. Paladin, Phys. Rev. Lett. **70**, 105 (1993).
[34] T. R. Einert and R. R. Netz, Biophys. J. **100**, 2745 (2011).
[35] D. J. Amit, H. Gutfreund, and H. Sompolinsky, Phys. Rev. A **32**, 1007 (1985).
[36] D. Poland and H. A. Scheraga, J. Chem. Phys. **45**, 1464 (1965).

[37] C. Monthus and T. Garel, Phys. Rev. E **75**, 031103 (2007).
[38] P-G. de Gennes, Biopolymers **6**, 715 (1968).
[39] M. Müller, F. Krzakala, and M. Mézard, Eur. Phys. J. E **9**, 67 (2002).
[40] U. Gerland, R. Bundschuh, and T. Hwa, Biophys J. **81**, 1324 (2001).
[41] G. E. Andrews, R. Askey, and R. Roy, *Special Functions* (Cambridge University Press, Cambridge, 1999).
[42] T. Liu and R. Bundschuh, Phys. Rev. E **72**, 061905 (2005).
[43] M. Mézard and A. Montanari, *Information, Physics, and Computation* (Oxford University Press, Oxford, 2009).
[44] J. Bascle, T. Garel, and H. Orland, J. Phys. I France **3**, 259 (1992).
[45] O. V. Valba, M. V. Tamm, and S. K. Nechaev, Phys. Rev. Lett. **109**, 018102 (2012).
[46] M. Lässig and K. J. Wiese, Phys. Rev. Lett. **96**, 228101 (2006).
[47] F. David and K. J. Wiese, Phys. Rev. Lett. **98**, 128102 (2007).