Behavior of random RNA secondary structures near the glass transition

William D. Baez,[1] Kay Jörg Wiese,[2] and Ralf Bundschuh[3]
[1]*Department of Physics, The Ohio State University, Columbus, Ohio 43210*
[2]*CNRS-Laboratoire de Physique Théorique de l'Ecole Normale Supérieure, 24 rue Lhomond, 75005 Paris, France*
[3]*Department of Physics, The Ohio State University, Columbus, Ohio 43210, USA*

(Received 8 August 2018; revised manuscript received 14 January 2019; published 20 February 2019)

RNA forms elaborate secondary structures through intramolecular base pairing. These structures perform critical biological functions within each cell. Due to the availability of a polynomial algorithm to calculate the partition function over these structures, they are also a suitable system for the statistical physics of disordered systems. In this model, below the denaturation temperature, random RNA secondary structures exist in one of two phases: a strongly disordered, low-temperature glass phase and a weakly disordered, high-temperature molten phase. The probability of two bases to pair decays with their distance with an exponent $3/2$ in the molten phase and about $4/3$ in the glass phase. Inspired by previous results from a renormalized field theory of the glass transition separating the two phases, we numerically study this transition. We introduce distinct order parameters for each phase that both vanish at the critical point. We finally explore the driving mechanism behind this transition.

DOI: 10.1103/PhysRevE.99.022415

I. INTRODUCTION

Heteropolymer folding is a fundamental biophysical process all living systems rely on. It is of medical relevance, since, e.g., misfolded proteins and nucleic acids are strongly implicated in the development of several neurological disorders, such as Alzheimer's, Parkinson's, and Lou Gehrig's disease [1–6]. While general questions of heteropolymer folding can be addressed both in proteins and in ribonucleic acids (RNA), RNA is the simpler of the two systems as it is composed of only four monomers (as opposed to 20 for proteins), the nucleotides adenine (A), cytosine (C), guanine (G), and uracil (U). These nucleotides have a tendency to form Watson-Crick (A-U, C-G) base pairs. To form these pairs, the RNA strand folds back onto itself, which leads to the creation of RNA secondary structures. From a statistical physics standpoint, heteropolymer folding presents a challenging task for the physics of disordered systems. In particular, RNA secondary structures are one of the few disordered systems for which one can calculate the partition function in polynomial time [7].

Previous studies show that RNA secondary structures exist within one of two well-identified phases. Above a critical temperature T_c, the system is in a phase where sequence disorder does not play a significant role. This simplifying assumption allows one to model any random base sequence or a sequence with random pairing energies, $\epsilon_{i,j}$, as a homopolymer. The defining trait for this phase is that the partition function of long RNA molecules is dominated by an exponentially large number of secondary structures with energies differing only at the order of $O(k_B T)$. This phase is denoted the *molten* phase.

Beginning in 1968, de Gennes [8], while studying the folded homopolymer, laid the theoretical foundation for our understanding of the molten phase by showing that the probability of two bases to pair scales as $p(\ell) \sim \ell^{-\alpha}$, where ℓ is the distance between any two bases labeled i and $j = i + \ell$ for $1 \leqslant i < j \leqslant N$, and that $\alpha = 3/2$. In later studies [9,10] it was demonstrated analytically that sequence disorder, when introduced perturbatively, is irrelevant for small disorder or high temperatures. It was also confirmed numerically that the pairing probability of a four-letter representation of the RNA heteropolymer model matches the scaling behavior of de Gennes' earlier prediction at high enough temperature [9,10].

At low temperatures sequence disorder can no longer be ignored. This *glass* (or "frozen") phase is characterized by the existence of a small number of low-energy secondary structures. In the glass phase the scaling exponent of the pairing probabilities was found numerically to be $\alpha \approx 1.34$ [11]. Properties of the glass phase have been extensively studied [9–19]. While the existence of the glass phase is well established and the glass phase itself has been characterized extensively, there is yet no clear signature and characterization of the expected phase transition between the molten and the glass phase.

Through the use of a renormalized field theory [13], Lässig and Wiese (LW) showed analytically that the freezing transition between the molten and the glass phase is of at least second order. They proved that their model is renormalizable at 1-loop order. They further argued that the value of the exponent α at the transition is the same as in the glass phase. The field theory was later refined by David and Wiese [14,15], showing that it is renormalizable to all orders in perturbation theory. At 2-loop order [14,15] it then predicts that $\alpha = 3/2$ in the molten phase and $\alpha \approx 1.36$ at the transition and in the glass phase.

However, questions concerning the location of this critical point, the existence and behavior of a suitable order parameter, and the transition's driving mechanism remained open. In the present study, through the use of numerical simulations and

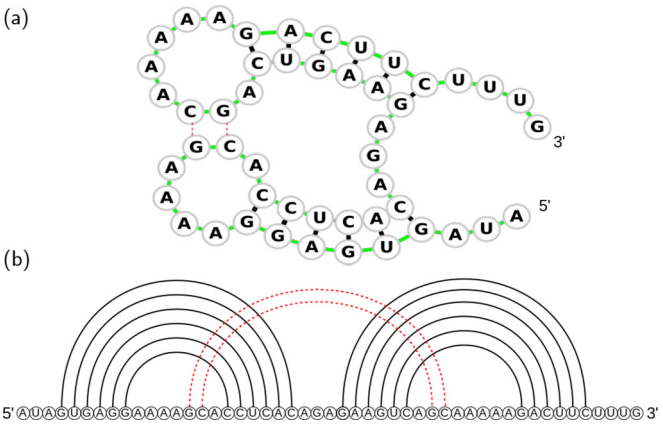

FIG. 1. Diagrammatic and abstract representations of an RNA secondary structure. (a) The backbone of the molecule is represented by the solid bright line while the solid black lines stand for the hydrogen-bonded base pairs where each base is depicted as a circle. The shape of the backbone is such that stems of stacked base pairs and the loops connecting or terminating them can be clearly seen. Stems form double-helical structures similar to that of DNA. (b) The same structure as in (a) but with the backbone of the molecule stretched out turning the solid black lines representing each base pair into arches. The base pairs represented by the red dashed lines in (a) and (b) create a pseudoknot visible in (b) as a crossing of the arches. We exclude such pairings in our definition of a secondary structure. Pseudoknots, such as these, do not contribute much to the total energy and are often deemed part of an RNA molecule's tertiary structure.

analytical tools inspired by the field theory, we study two order parameters which both vanish at the transition. One of them is non-zero in the molten phase, the other in the frozen phase. This allows us to precisely locate the critical point. Measuring pairing probability distributions, we explore the mechanism driving this transition.

This paper is organized as follows: In Sec. II we define the model. Section III details our efforts to locate the critical temperature by establishing two order parameters. Section IV probes the transition mechanism. In Sec. V we discuss our results.

II. RNA SECONDARY-STRUCTURE MODEL

A. RNA secondary structures

Each RNA molecule is described by its sequence of N bases $b_1 \dots b_N$, its *primary structure*. Given a molecule, an RNA *secondary structure*, such as the one shown in Fig. 1, can then be defined by a list of ordered pairs (i, j) with $i < j$ representing the base pairs formed by the molecule. Following previous studies in the field of RNA secondary structure reviewed in Ref. [20], we require each base to pair with no more than one other base and we only consider structures that exclude *pseudoknots*. The latter is accomplished by requiring that any two base pairs (i, j) and (k, l) with $i < k$ satisfy either $i < j < k < l$ or $i < k < l < j$ (see Fig. 1). Configurations violating this rule are termed *pseudoknots*.

While pseudoknots are biologically relevant [21,22], the no pseudoknot constraint is necessary to make both analytical and numerical calculations feasible. The error introduced by the no pseudoknot constraint is limited due to their relatively infrequent occurrence in real folded RNA [23]. We consider any pseudoknots and base triples as parts of the *tertiary structure* of the molecule, i.e., its three-dimensional conformation, given the secondary structure. Importantly, we also ignore any other constraints on the secondary strutures due to the tertiary structure, such as, e.g., excluded-volume constraints, in our model.

B. Energy model

To fully describe the statistical physics of RNA secondary structures we have to complement the definition of valid secondary structures by an energy function that assigns an energy to every structure. While very sophisticated energy models are available [24], which allow detailed quantitative descriptions of the folding of actual RNA molecules, in this paper we follow previous studies [9,25] that focus on the universal aspects of RNA secondary structure formation and adopt a simplified RNA energy model. Specifically, we consider contributions from each base pair, and associate with the pair of bases labeled i and j an interaction energy, $\epsilon_{i,j}$. The total energy of a structure S is defined as

$$E[S] = \sum_{(i,j)\in S} \epsilon_{i,j}. \tag{1}$$

In realistic energy models used in actual secondary-structure prediction of biologically relevant RNA molecules [24], the interaction energies capture several physically distinct effects: the enthalpic contributions of hydrogen bonding between the two bases of the pair, stacking between the aromatic rings of two consecutive bases along the strand, and (screened) Coulomb repulsion between the negatively charged phosphate groups of the backbones. In addition, these interaction "energies" are in reality *free* energy differences between a paired and an unpaired state. They include the difference in entropy between two unpaired bases freely fluctuating in three-dimensional space and a base pair constrained in its three-dimensional fluctuations due to the requirement to remain paired and due to the greater stiffness of double-stranded over single-stranded RNA. The enthalpic and entropic contributions nearly balance. At room or body temperature, the enthaltpic contributions slightly outweigh the enthalpic contributions and RNA forms secondary structures. At temperatures of around 80 °C (depending on the sequence) the entropic contributions outweigh the enthalpic contributions, the interaction free energies become positive, and the RNA denaturates. Thus, in principle, the interaction (free) energies $\epsilon_{i,j}$ depend on temperature and the identities of the bases b_i and b_j rendering the $\epsilon_{i,j}$ temperature dependent random variables with discrete values and a complicated correlation structure if the RNA sequences are chosen randomly. Since we are here interested in universal properties of phase transitions, we further simplify our model [9] by choosing these interaction energies as independent Gaussian random variables taken from the distribution

$$\rho(\epsilon) = \frac{1}{\sqrt{2\pi D}} e^{-\frac{\epsilon^2}{2D}} \tag{2}$$

with mean energy zero and variance D. A mean energy of zero is still low enough to guarantee that most bases engage in base pairs and thus to avoid the denatured phase that we are not interested in. Also, while we ignore the explicit temperature dependence of the realistic interaction free energies, the universal properties of the phase transition should not depend on which parameter is tuned to cross the transition.

Since $\sqrt{D}$ is the only energy scale in this model, we set $D = 1$ in our simulations, but report all results in units scaled by $\sqrt{D}$. Finally note that this kind of uncorrelated Gaussian-disorder model is expected to have less severe finite-size effects [11] than other choices of disorder models.

C. Partition function

Once an energy has been assigned to each secondary structure S, the partition function is defined as

$$Z(N) = \sum_{S\in\Omega(N)} e^{-\beta E[S]}, \tag{3}$$

where $\Omega(N)$ is the set of allowed secondary structures of N bases, and $\beta = 1/k_{\mathrm{B}}T$. In the absence of pseudoknots, the partition function can be studied by considering substrands of the total sequence from base i to base j. The restricted partition function $Z_{i,j}$ for these substrands obeys the recursive equation [7]

$$Z_{i,j} = Z_{i,j-1} + \sum_{k=i}^{j-1} Z_{i,k-1} e^{-\beta\epsilon_{k,j}} Z_{k+1,j-1}, \tag{4}$$

in which by convention $Z_{i,i-1} = 1$. In this recursion, the right-hand side involves only restricted partition functions for shorter substrands than the one on the left-hand side. Thus, the total partition function, $Z(N) = Z_{1,N}$, can be calculated by progressing from the shortest substrands to longer ones in $\mathrm{O}(N^3)$ time.

The analogous recursion equation [26]

$$E_{i,j} = \min_{i\leqslant k<j} [E_{i,j-1}, E_{i,k-1} + \epsilon_{k,j} + E_{k+1,j-1}] \tag{5}$$

holds for the energy $E_{i,j}$ of the lowest-energy structure on the substrand of the total sequence from base i to base j, where $E_{i,i-1} = 0$ by convention. This equation can be used in conjunction with a backtracking mechanism [26] to calculate the (zero-temperature) ground-state secondary structure for any disorder configuration $\{\epsilon_{i,j}\}_{i,j}$ in $\mathrm{O}(N^3)$ time.

D. Observables

From the partition function of RNA secondary structures, we can calculate a variety of physical observables that characterize the structural ensemble. Primarily, we study the pairing probability, $p_{i,j}$, for a given base pair (i, j). This probability can be obtained as

$$p_{i,j} \equiv \frac{e^{-\beta\epsilon_{i,j}} Z_{i+1,j-1} Z_{j+1,i-1}}{Z_{1,N}}, \tag{6}$$

where $Z_{i+1,j-1}$ can be calculated from Eq. (4) and $Z_{j+1,i-1}$ is the partition function of the sequence $b_{j+1}b_{j+2}...b_N b_1...b_{i-2}b_{i-1}$. This last quantity can be obtained when the recursion Eq. (4) is applied to a duplicated sequence $b_1...b_N b_1...b_N$ and calculated as $Z_{j+1,N+i-1}$.

We are specifically interested in the dependence of moments of the ensemble averaged base-pairing probability on the distance $|i - j| = \ell$ between the two bases. Within the molten phase, and for large ℓ, this quantity has a power-law dependence

$$\langle p(\ell)^n \rangle \equiv \left\langle p_{i,i+\ell}^n \right\rangle \sim \left[\frac{\ell(N-\ell)}{N}\right]^{-\alpha_n}, \tag{7}$$

where the brackets stand for the ensemble average over the random base-pairing energies $\{\epsilon_{i,j}\}_{i,j}$ and α_n is a critical exponent. The parameter n allows probing different moments of the distribution of the base pairing probabilities $p(\ell)$. Note that the form of Eq. (7) has a practical advantage over the asymptotic behavior $\ell^{-\alpha_n}$, as it allows us to use larger ℓ in extrapolations.

The logarithms of each pairing probability,

$$\Delta F_{i,j} = -k_{\mathrm{B}}T \ln(p_{i,j}), \tag{8}$$

have been interpreted as the "pinching free energy" [9], which is the free-energy difference between a pinch of the monomers labeled i and j, and the unperturbed, or unpinched, state. In Sec. III, as in previous studies [9], we will show how we can use the free energy of the largest possible pinch,

$$\Delta F(N) \equiv \Delta F_{1,\frac{N}{2}+1}, \tag{9}$$

to obtain an estimate of the critical temperature. Here monomers 1 and $\frac{N}{2}+1$ are treated as representatives of all splits of a molecule of N bases into two equal pieces of length $\frac{N}{2}-1$.

E. Review of main results of the Lässig-Wiese field theory

Our numerical analysis of the glass transition is inspired by a field-theoretical calculation by Lässig and Wiese [13], and its refinements in Refs. [14,15]. This theory starts from the Gaussian-disorder model defined in Sec. II B. It introduces replicas (copies) of the RNA molecule and shows that the disorder-averaged partition function of the replicated system can be written in terms of two relevant operators. The first is the contact field $\Phi_\alpha(i, j)$ that is 1 if the bases labeled i and j in replica α are paired and 0 otherwise. The second is the overlap field $\Psi_{\alpha,\beta}(i, j)$ between two replicas α and β defined as $\Psi_{\alpha,\beta}(i, j) \equiv \Phi_\alpha(i, j)\Phi_\beta(i, j)$, i.e., $\Psi_{\alpha,\beta}(i, j)$ is 1 if the bases labeled i and j are paired in *both* replicas α and β and 0 otherwise. The scaling dimensions of these operators, i.e., the exponents with which their thermal and disorder-averaged expectation values depend on the distance $|j - i|$ between the two bases, are called ρ and θ, respectively.

Lässig and Wiese treat the disorder in the model perturbatively, prove renormalizability of the theory at first order (later extended to renormalizability to all orders by David and Wiese [14,15]), calculate the β function, find a critical point that they identify with the glass transition, and calculate the critical exponents ρ^* and θ^* (the scaling dimensions of Φ and Ψ) at this transition. Importantly, they find expressions for the two critical exponents as a function of the scaling dimension of the disorder and the number of replicas. In the relevant limit of zero replicas, the expression for the exponent θ^* (the

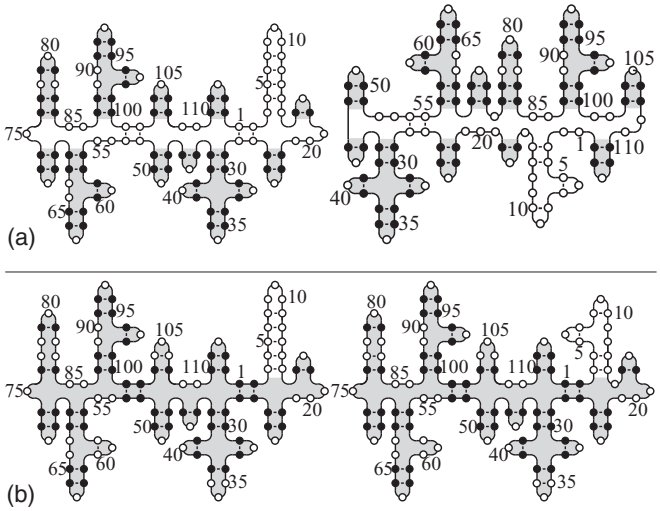

FIG. 2. Secondary structures of a random RNA molecule at distant times as presented in Ref. [13]. The pairing overlap is defined by the common base pairings between the left and right configuration (corresponding bases are shown in black). (a) Above T_c, the molecule contains conserved subfolds on scales up to the correlation length ξ (indicated by shading) and is molten on large scales. (b) Below T_c, the molecule is "locked" into its minimum-energy structure on all scales, up to rare fluctuations (unshaded).

scaling dimension of Ψ at the transition) becomes smaller than the expression for the exponent ρ^* (the scaling dimension of Φ at the transition) as the scaling dimension of the disorder approaches its physical value. Since $\Psi_{\alpha,\beta}(i,j) \leqslant \Phi_\alpha(i,j)$, it is impossible that the probability for two bases to be paired in multiple replicas decays slower than the probability for the same two bases to be paired in a single replica. Thus, Lässig and Wiese conclude that rather than crossing each other, the two exponents must become identical, i.e., that $\theta^* = \rho^*$. They interpret this physically as the different replicas "locking" with each other, so that at the transition (and below) there is no difference between looking at the pairing behavior of one or multiple replicas. This results in the picture of small-scale locked regions in a background of an overall molten structure above the transition temperature, and an overall locked structure with small-scale molten regions below the transition temperature as indicated in Fig. 2. The length scale ξ of these small-scale regions diverges as $\xi \sim |T - T_c|^{-\nu^*}$ with $\nu^* = 1/(2-\theta^*)$ as the phase transition is approached from either side. The actual values of the exponents from the 2-loop calculation by David and Wiese [14,15] are

$$\theta^* = \rho^* \approx 1.36 \tag{10}$$

and thus

$$\nu^* \approx 1.56. \tag{11}$$

They are very close to the numbers of $\theta^* = \rho^* \approx 11/8 = 1.375$ and $\nu^* \approx 8/5 = 1.6$ provided by Lässig and Wiese in their original one-loop calculation [13].

To connect the results of the Lässig-Wiese theory to the observables described in Sec. II D, we have to realize that the pairing probability $p(\ell)$ is the thermal average of the contact field $\Phi_\alpha(i, i+\ell)$. Since ρ is the scaling dimension of the thermal *and* ensemble averaged contact field $\Phi_\alpha(i, i+\ell)$, it describes the scaling of the disorder average of the pairing probability $p(\ell)$, i.e., the quantity described by $n = 1$ in Eq. (7). Thus,

$$\alpha_1 = \rho. \tag{12}$$

Similarly, the thermal average of the overlap field $\Psi_{\alpha,\beta}(i, i+\ell)$ equals the square of the pairing probability $p(\ell)$, since the bases labeled i and $i+\ell$ have to independently pair in the two replicas. Since θ is the scaling dimension of the thermal and ensemble averaged overlap field $\Psi_{\alpha,\beta}(i, i+\ell)$, it describes the scaling of the disorder average of the square of the pairing probability $p(\ell)$, i.e., the quantity described by $n = 2$ in Eq. (7). Thus,

$$\alpha_2 = \theta. \tag{13}$$

F. Numerical approach

To investigate the RNA secondary-structure glass transition, we use RNA molecules of lengths within the range of $500 \leqslant N \leqslant 4000$. For each length, many independent realizations of the random base-pairing energies $\epsilon_{i,j}$ are chosen, and the partition function of all allowable secondary structures is calculated for each realization of the random variables using Eq. (4). Then, the observables described above are calculated for each realization of the random variables and averaged over those realizations (as well as over the starting base i in case of the base-pairing probability). In the aim to keep the numerical effort manageable we averaged over 20 000 samples for $N = 500$, over 10 000 samples for $N \in 750, 1000, 1500$, over 5000 samples for $N = 2000$, and over 1000 samples for $N = 4000$. We varied the temperature in a range of $0.1 \leqslant k_B T/\sqrt{D} \leqslant 1.0$ to capture the behavior deep within each phase and close to the phase transition, which we will find to be located at $k_B T/\sqrt{D} \approx 0.53$. Note that the smaller number of samples for large systems is partially compensated by a factor of N in the statistics; e.g., the number of starting points for base pairs at a given distance grows as N, and we average over these starting points. As a result, our statistics are almost independent of size at the chosen numbers of samples.

III. SCALING OF THE CONTACT AND OVERLAP OBSERVABLES

To understand the model's behavior at the transition, we must know its precise location. Previous studies [9,17,27] using different disorder models found that for the system sizes achievable in numerical simulations, thermodynamic signatures of the phase transition are quite weak, making an exact localization of the critical temperature T_c challenging. In this section we will identify two order parameters inspired by the Lässig-Wiese field theory to obtain the precise location of the critical temperature T_c.

A. Pairing-probability exponents alone do not specify transition temperature

In principle, the critical exponents α_n alone should allow us to locate the phase-transition temperature T_c. In the molten phase $p(\ell)$ is independent of disorder and thus

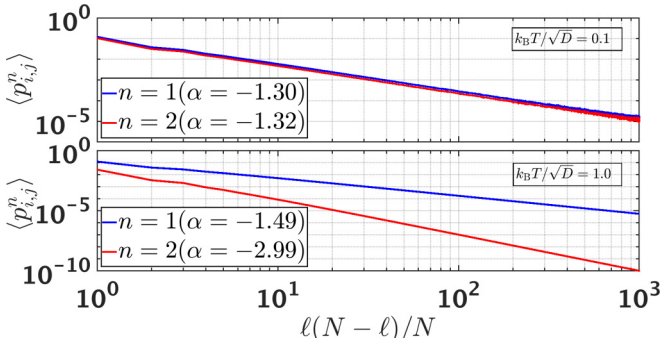

FIG. 3. Average pairing probability versus substrand length. The upper figure shows how the pairing probability scales as Eq. (7) in the glass phase at the lowest temperature considered, $k_BT/\sqrt{D} = 0.1$. As mentioned in the text, within this phase the critical exponent, $\alpha_n^{(g)}$, is insensitive to the value of n. The lower figure shows how the pairing probability scales as Eq. (7) in the molten phase at the highest temperature considered, $k_BT/\sqrt{D} = 1.0$, but demonstrates how the value of $\alpha_n^{(m)}$ directly depends on the value of n.

$\langle p(\ell)^n \rangle \approx \langle p(\ell) \rangle^n$, i.e.,

$$\alpha_n^{(m)} = n\alpha_1^{(m)} = n\frac{3}{2}, \tag{14}$$

where we identify exponents in the molten phase with a superscript (m). In the glass phase and at the transition, the Lässig-Wiese theory predicts [see Eqs. (10), (12), and (13)]

$$\alpha_n^* = \alpha_1^* \approx 1.36 \quad \text{and} \quad \alpha_n^{(g)} = \alpha_1^{(g)} \approx 1.36 \tag{15}$$

independent of n, where we denote exponents at the phase transition with a superscript asterisk and exponents in the glass phase with a superscript (g).

To verify this, we calculated the disorder-averaged base-pairing probabilities $\langle p(\ell)^n \rangle$ numerically for different temperatures at $N = 4000$ and fitted them to Eq. (7) to obtain α_n. Indeed, Fig. 3 shows that the expected power laws hold, for values of $n = 1, 2$, both at low temperature ($T = 0.1$) and at high temperature ($T = 1$).

However, Fig. 4 shows the behavior of α_1 and α_2 across the transition, which we show later is at $T_c = 0.53\sqrt{D}/k_B$. One can see that the apparent values of the exponents change gradually from their value in the glass phase to their value in the molten phase as the system undergoes the phase transition. The steepness of this change only increases slightly as the system size is increased from $N = 1000$ to $N = 4000$. Thus, these exponents cannot be used to precisely locate the critical temperature or to obtain the values of the exponents at the transition.

This gradual change in exponents is consistent with the weakness of the transition observed in earlier work by Pagnani *et al.* [17]. There, it was found that the specific heat has a very broad "feature" at the glass transition, and that the second derivative of the specific heat is the first derivative akin to a divergence, albeit with large finite-size effects. For this reason, we did not try to identify the location of the phase transition using derivatives of the free energy.

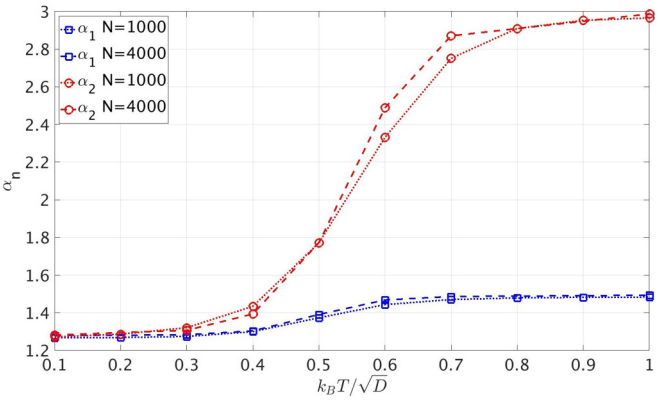

FIG. 4. The exponents α_1 and α_2 as a function of temperature. As expected, α_n becomes independent of n at a value of $\alpha_n^{(g)} \approx 1.36$ at very low temperatures and approaches $\alpha_n^{(m)} = n\frac{3}{2}$ at high temperatures. However, the changes in apparent α_n are so gradual and the steepness increases only so slightly as the system size is quadrupled that this quantity can neither be used to determine the critical temperature nor the exponents at the transition.

B. Coarse estimate of phase-transition temperature from pinch free energies

To obtain a first estimate of the transition temperature, we follow [9] and consider the disorder-averaged pinch free energy $\langle \Delta F \rangle$. As seen in the inset of Fig. 5, this quantity has a logarithmic dependence on the sequence length N at both low and high temperatures [9]. We thus fit it to the linear form

$$\langle \Delta F(N) \rangle = a(T)\ln(N) + c(T). \tag{16}$$

for sequence lengths in the range $500 \leqslant N \leqslant 2000$. The resulting prefactor $a(T)$ is shown in the main part of Fig. 5. As expected [8], in the molten phase $a(T) \approx \frac{3}{2}k_BT$ resulting in a linear dependence of the prefactor on temperature. However, for lower temperatures, the prefactor is non-monotonic in temperature and thus lends itself as an estimator of the critical

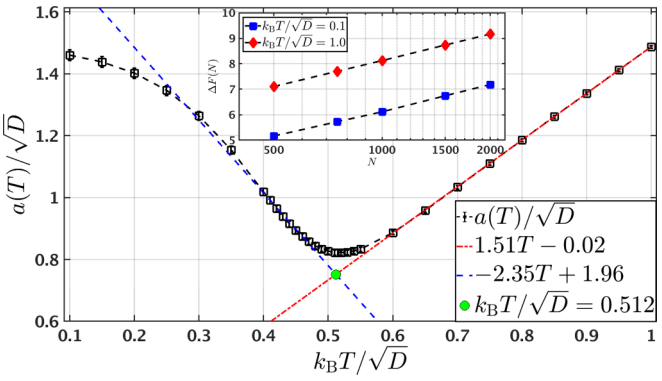

FIG. 5. Pinch free energy prefactor $a(T)$ versus temperature. The inset shows the disorder-averaged pinch free energy versus $\ln(N)$ in both the glass phase ($k_BT = 0.1\sqrt{D}$) and the molten phase ($k_BT = 1.0\sqrt{D}$). This allows us to extract the prefactors of these logarithmic behaviors, which are shown as a function of temperature in the main graph. The temperature at which the prefactor changes slope is an estimate of the phase transition temperature T_c.

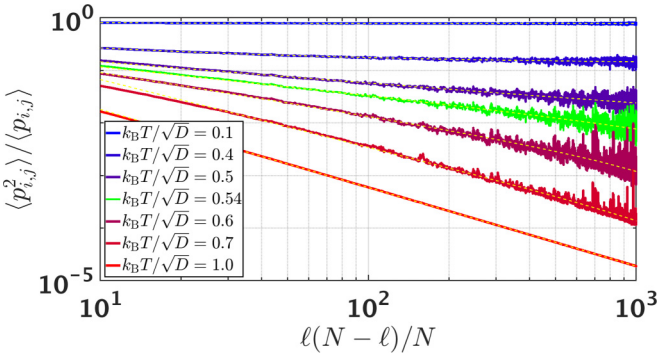

FIG. 6. High-temperature pairing ratio $\langle p(\ell)^2\rangle/\langle p(\ell)\rangle$ versus substrand length. The curves show the behavior of this new observable at several temperatures in the vicinity of T_c and for temperatures deep within the glass phase ($k_BT/\sqrt{D} = 0.1$) and molten phase ($k_BT/\sqrt{D} = 1.0$). The dashed yellow lines are fits to Eq. (17).

temperature. The intersection of linear fits on either side of the minimum yields $T_c \approx 0.51\sqrt{D}/k_B$.

C. Order parameters for the transition, and a more precise estimation of the phase-transition temperature

Using this value of T_c as a guide, we limited the range over which we performed our simulations to $0.4 \leqslant k_BT/\sqrt{D} \leqslant 0.7$ for sequence lengths up to $N = 4000$. Since the scaling of the base-pairing probability $\langle p(\ell)^n\rangle$ by itself does not demonstrate a clear indication of when a finite system of length N approaches T_c, we defined two additional observables that allow us to extract when the system approaches T_c from either side. The first is the ratio $\langle p(\ell)^2\rangle/\langle p(\ell)\rangle$, which we would expect to be constant for large ℓ in the glass phase due to $\alpha_2 = \alpha_1$ and revert back to a power law similar to Eq. (7) with an exponent of 3/2 in the molten phase. Conversely, the second new observable, $\langle p(\ell)\rangle^2/\langle p(\ell)^2\rangle$ we expect to be constant in the high-temperature regime for large ℓ and then decay with the power law of Eq. (7) and an exponent of approximately 1.36 at low temperatures. Figures 6 and 7 show these observables for several temperatures, both above, below, and close to the critical temperature at $N = 4000$. As

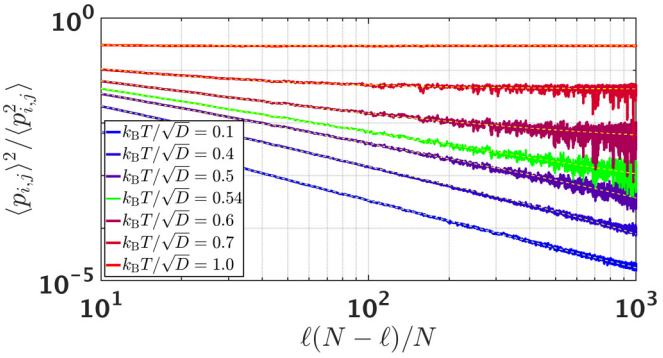

FIG. 7. Low-temperature pairing ratio $\langle p(\ell)\rangle^2/\langle p(\ell)^2\rangle$ versus substrand length. Shown is the behavior of this new observable at several temperatures in the vicinity of T_c and for temperatures deep within the glass phase ($k_BT/\sqrt{D} = 0.1$) and molten phase ($k_BT/\sqrt{D} = 1.0$). The dashed yellow lines are fits to Eq. (18).

indicated with dashed yellow lines, we fit these curves to the forms

$$\frac{\langle p(\ell)^2\rangle}{\langle p(\ell)\rangle} = A_g\left\{\left[\frac{\ell(N-\ell)}{N}\right]^{-\omega^{(g)}} + \Delta_g\right\}, \tag{17}$$

$$\frac{\langle p(\ell)\rangle^2}{\langle p(\ell)^2\rangle} = A_m\left\{\left[\frac{\ell(N-\ell)}{N}\right]^{-\omega^{(m)}} + \Delta_m\right\}, \tag{18}$$

with offsets Δ_g and Δ_m, exponents $\omega^{(g)}$ and $\omega^{(m)}$, and global prefactors A_g and A_m. We identify the offsets Δ_g and Δ_m as the order parameters of the system.

It may seem surprising to have *two* order parameters, one for each phase. It is not commonly known that such observables can be constructed quite generally. In the Appendix we illustrate this fact for a two-dimensional Ising model, where it is analytically known that there is only one phase transition. As that example shows, an order parameter is a quantity that vanishes in one phase, is nonzero in the other one, and can be used to determine the critical temperature; it may, or may not, be associated with a specific symmetry. The two quantities Δ_g and Δ_m we constructed each vanish in one of the phases and are nonzero in the other one, as they probe specific properties of their respective phases. While we cannot exclude that there are two consecutive phase transitions at very close temperatures, we have no indication for this hypothesis. The example of the Ising model suggests that there is likely only one transition, even though it is mathematically possible to have $\langle p(\ell)\rangle^2 < \langle p(\ell)^2\rangle < \langle p(\ell)\rangle$, which would make both quantities vanish simultaneously, and possibly for more than a single temperature.

The order parameters Δ_g and Δ_m scale close to the transition as $\xi^{-\omega}$, where ξ is a characteristic length scale, diverging at the transition as $\xi \sim |T - T_c|^{-\nu^*}$. This implies that Δ_g and Δ_m should depend on temperature T as

$$\Delta_g \sim |T - T_c|^{\omega^{(g)}\nu^*}, \tag{19a}$$

$$\Delta_m \sim |T - T_c|^{\omega^{(m)}\nu^*}. \tag{19b}$$

To determine the two exponents, we vary them over a range of $1 \leqslant (\omega^{(g)}\nu^*, \omega^{(m)}\nu^*) \leqslant 4$ for each system size N and fit a linear regression to $\Delta_g^{(\omega^{(g)}\nu^*)^{-1}}$ versus T, and $\Delta_m^{(\omega^{(m)}\nu^*)^{-1}}$ versus T for a small temperature range immediately below and above $k_BT/\sqrt{D} = 0.53$, respectively. We then choose for each system size the values of $\omega^{(g)}\nu^*$ and $\omega^{(m)}\nu^*$ that maximize the coefficient of determination R^2 of the linear fits. Last, we treat each system size as an independent realization (upon inspection the dependence of the optimal $\omega^{(g)}\nu^*$ and $\omega^{(m)}\nu^*$ on system size does not seem to have a systematic contribution) and determine the average and standard error of the mean of the two exponents. On the left side of the critical point ($T < T_c$), we find

$$\omega^{(g)}\nu^* \approx 2.12 \pm 0.08, \tag{20}$$

while on its right side ($T > T_c$)

$$\omega^{(m)}\nu^* \approx 2.43 \pm 0.06. \tag{21}$$

Figure 8 shows $\Delta_g^{\frac{1}{\omega^{(g)}\nu^*}}$ and $\Delta_m^{\frac{1}{\omega^{(m)}\nu^*}}$ with the such determined values for $\omega^{(g)}\nu^*$ and $\omega^{(m)}\nu^*$ as functions of temperature.

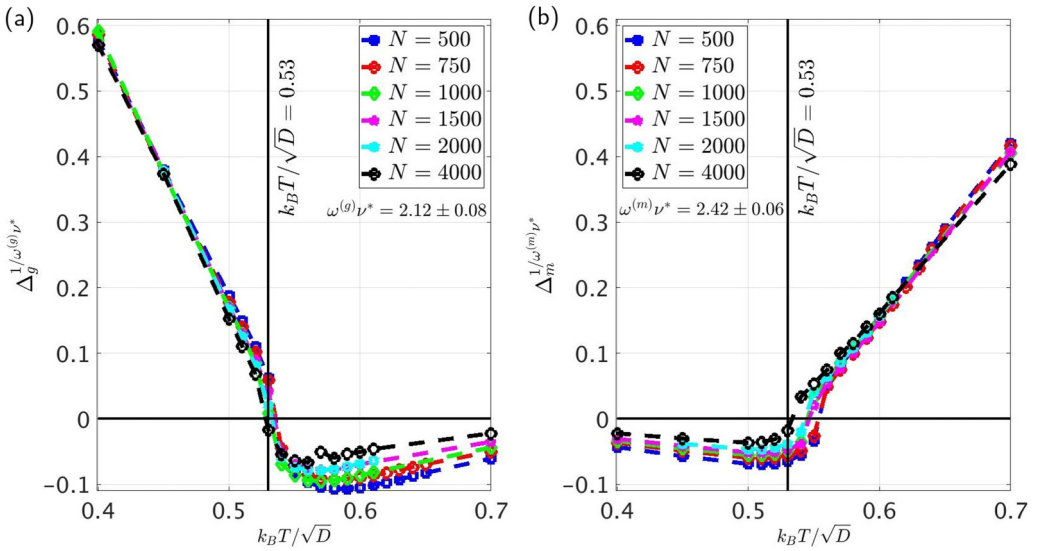

FIG. 8. Appropriate powers of the order parameters Δ_g and Δ_m versus temperature. (a) As Δ_g defined by Eq. (18) approaches $k_BT/\sqrt{D} \approx 0.53$ from below it drops below zero. (b) Similarly, as Δ_m defined by Eq. (17) approaches $k_BT/\sqrt{D} \approx 0.53$ from above it becomes negative. Note that to take the appropriate powers of the order parameters, the power is taken of their absolute value and the sign is restored afterwards. On the side of the transition, where the order parameters Δ_g and Δ_m are negative, they tend to zero with increasing system size N. This allows precise determination of the phase transition temperature $k_BT/\sqrt{D} \approx 0.53 \pm 0.01$.

One clearly sees that each quantity approaches zero roughly linearly on its respective side of $k_BT_c/\sqrt{D} = 0.53$ (Δ_g for $T < T_c$ and Δ_m for $T > T_c$). Both cross zero at $k_BT_c/\sqrt{D} = 0.53$ and become dependent on the system size N on the opposite side of $k_BT_c/\sqrt{D} = 0.53$, where they remain small and appear to tend toward zero with increasing system size N. We thus conclude that Δ_g and Δ_m are nonzero on their respective side of the phase transition and vanish on the opposite side, rendering them proper order parameters for the transition. We further conclude that the phase transition takes place at $k_BT_c/\sqrt{D} = 0.53 \pm 0.01$. To verify the consistency of the observed critical behavior, we note that the field theoretical predictions given in Eqs. (11) and (14), (15) read

$$\omega^{(g)}\nu^* = \left(2\alpha_1^{(g)} - \alpha_2^{(g)}\right)\nu^* = 1.36 \times 1.56 = 2.12 \qquad (22)$$

and

$$\omega^{(m)}\nu^* = \left(\alpha_2^{(m)} - \alpha_1^{(m)}\right)\nu^* = \frac{3}{2} \times 1.56 = 2.34. \qquad (23)$$

These values are in good agreement with our numerically determined values in Eqs. (20) and (21).

IV. TRANSITION MECHANISM

With the critical temperature in hand, we aim to understand the mechanism of the phase transition itself. Deep in the molten phase, base-pairing energies are irrelevant and as a consequence any base can pair with any other base. Thus, the thermally averaged base-pairing probability $p(\ell)$ is the same for any pair of bases with the same distance ℓ and decays with the distance ℓ as the power law introduced in Eq. (7). This probability is independent of the disorder realization. In contrast, at zero temperature the RNA molecule folds into the ground-state structure, which is typically nondegenerate within the Gaussian-disorder model. Thus, for a given realization of the disorder the thermally averaged pairing probability of a given base pair is either zero (if the base pair is not in the ground-state structure) or one (if the base pair is in the ground-state structure). Which base pairs have a zero probability and which have a probability of one changes with the disorder realization. Thus, the scaling behavior Eq. (7) at zero temperature arises only *after* ensemble averaging.

As discussed in Sec. II E, and illustrated in Fig. 2, the picture of the phase transition proposed in Ref. [13] is that as the phase transition is approached from below, the fraction of base pairs "locked" into a ground-state structure decreases by the appearance of molten regions. At the transition, the situation inverts and the "locked" base pairs become localized regions of decreasing size within an overall molten structure on the high-temperature side of the phase transition. These ideas are reflected in the order parameters introduced in Eqs. (17) and (18). Let us stress that while the expression "locked" used in Ref. [13] suggests that bases are paired with probability 1, this condition can be relaxed to mean that they are paired with probability $p > 0.5$, or even $p \geqslant 0.1$. This suffices to render the expectations Eqs. (17) and (18) nontrivial. We will understand "locked" in this sense from now on.

A. Base-pairing probability distribution

To directly probe the transition mechanism, we numerically calculate the entire *distribution* of base-pairing probabilities $P(p;\ell)$. We note that all moments of the disorder-averaged base-pairing probabilities can be reconstructed from these distributions as

$$\langle p(\ell)^n \rangle = \int_0^1 p^n P(p;\ell)\, dp. \qquad (24)$$

We obtain these distributions by explicitly calculating the base-pairing probabilities $p_{i,i+\ell}$ using Eq. (6) and then tabulating their frequencies averaged over all i and many real-

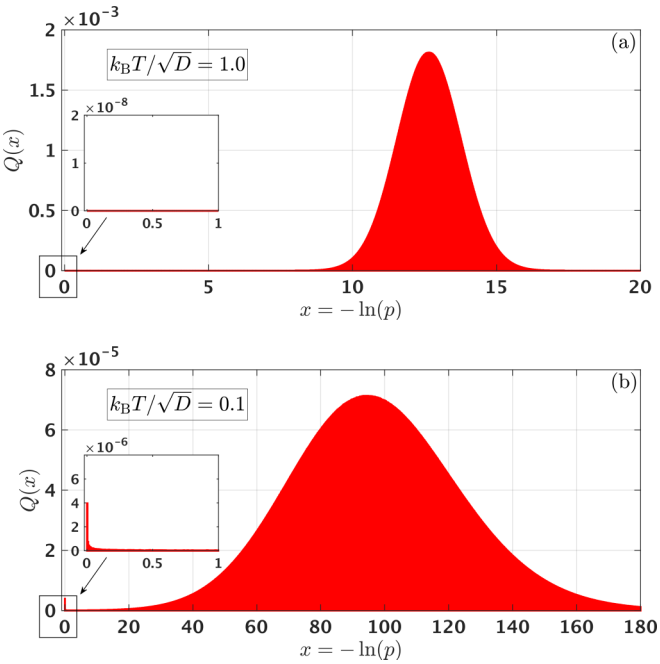

FIG. 9. Distributions of the logarithm of base-pairing probabilities $x = -\ln(p)$ for system size $N = 4000$ averaged over the final third of all ℓ's. (a) Deep in the molten phase ($k_BT/\sqrt{D} = 1.0$) the distribution of base-pairing probabilities has a distinct Gaussian-like form, which would go to a delta peak as $T \to \infty$ ($D \to 0$). The inset shows that no weight of the distribution is present close to $x \approx 0$ ($p \approx 1$). (b) Deep in the glass phase ($k_BT/\sqrt{D} = 0.1$) the base-pairing distribution still has a Gaussian-like form, albeit broader than the high-temperature distribution. As the inset shows there exists now a peak close to $x \approx 0$ ($p \approx 1$).

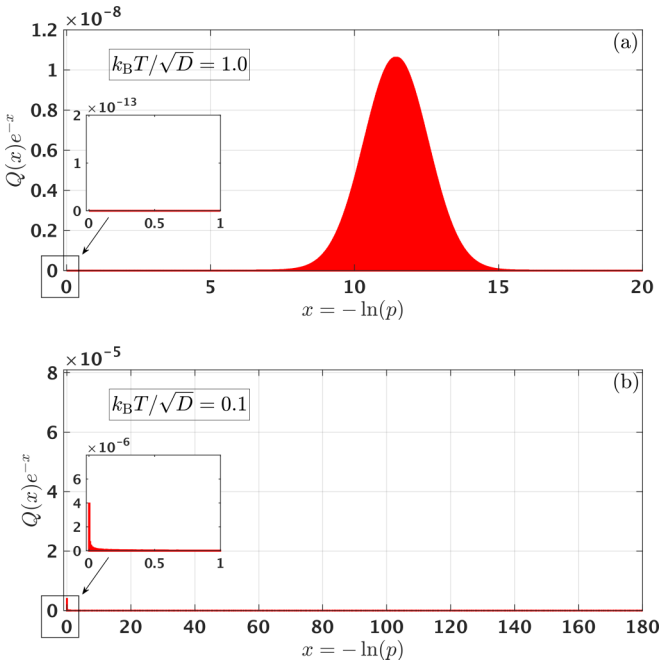

FIG. 10. Distributions of the logarithm of base-pairing probabilities $x = -\ln(p)$ for system size $N = 4000$ averaged over the final third of all ℓ's multiplied with e^{-x} for the same ranges as shown in Fig. 9, there without the factor of e^{-x}. (a) Deep in the molten phase ($k_BT/\sqrt{D} = 1.0$) the distribution of base-pairing probabilities has a distinct Gaussian-like form, which would go to a delta peak as $T \to \infty$ ($D \to 0$). (b) Deep in the glass phase ($k_BT/\sqrt{D} = 0.1$) the base-pairing distribution consists of only a peak at $x \approx 0$ ($p \approx 1$). Note the difference in scales between Figs. 9(a) and 10(a) while the scales in Figs. 9(b) and 10(b) are identical.

izations of the disorder. Moreover, due to the large range of base-pairing probabilities, instead of taking histograms of the $p_{i,i+\ell}$ themselves, we sample the distribution $Q(x;\ell)$, where $x = -\ln(p(\ell))$. The two distributions are connected to each other by $Q(x;\ell)\,dx = P(p;\ell)\,dp$ and thus moments of the disorder-averaged base-pairing probability can be obtained from $Q(x;\ell)$ as

$$\langle p(\ell)^n \rangle = \int_0^\infty e^{-nx} Q(x;\ell)\,dx. \tag{25}$$

To reduce the statistical noise, we further use

$$Q(x) := \int_{N/3}^{N/2} Q(x;\ell)\,d\ell, \tag{26}$$

where the range of integration is motivated by the fact that according to Eq. (7), the pairing probability $p(\ell)$ is a function of $\ell(N-\ell)/N$, and in the chosen range the latter does not vary by more than 10%.

Figure 9 shows that in both temperature regimes the base-pairing probability distribution $Q(x)$ has a Gaussian-like shape. Figure 9(a) is the distribution of $Q(x)$ at the highest temperature considered, $k_BT/\sqrt{D} = 1$, for a system of size $N = 4000$. The distribution is centered around $x \approx 13$ and we interpret it as a broadened version of what would be a δ-peak around the value of $\langle p(\ell)\rangle$ from Eq. (7) at infinite temperature or $D = 0$. Figure 9(b) shows the same distribution at the lowest temperature numerically accessible to us, $k_BT/\sqrt{D} = 0.1$. Here we see a broad Gaussian-like distribution again. However, it is by far wider than the high-temperature distribution and is located at $x \approx 95$, i.e., at much smaller base-pairing probabilities than its high-temperature counterpart. We interpret this Gaussian-like feature as the contributions from all the base pairs that are *not* part of the ground-state structure and thus would have exactly zero probability at zero temperature, while acquiring a finite, albeit very small ($x \gg 0$, $p \ll 1$), pairing probability at low temperatures. Importantly, as the inset of Fig. 9(b) shows, close to the origin ($x \approx 0$ or $p \approx 1$) there exists another peak. We interpret this prominent peak at $x \approx 0$ as the contributions from those base pairs that are "locked" in the ground-state structure and thus define the glass phase of this model. As the inset in Fig. 9(a) shows, no corresponding peak at $x \approx 0$ occurs deep in the molten phase.

As discussed above, when calculating the ensemble averaged pairing probability $\langle p(\ell)\rangle$ the distribution $Q(x;\ell)$ has to be integrated against e^{-x}. Thus, to appreciate where the weight contributing to the ensemble-averaged pairing probability $\langle p(\ell)\rangle$ is coming from in the two phases, Fig. 10 shows $e^{-x}Q(x)$. Deep in the molten phase ($k_BT/\sqrt{D} = 1.0$) multiplying with e^{-x} simply shifts the peak of the Gaussian-like distribution to somewhat smaller x (larger p). In contrast, deep in the glass phase ($k_BT/\sqrt{D} = 0.1$) the Gaussian peak at large x observed in the distribution $Q(x)$ alone becomes invisible

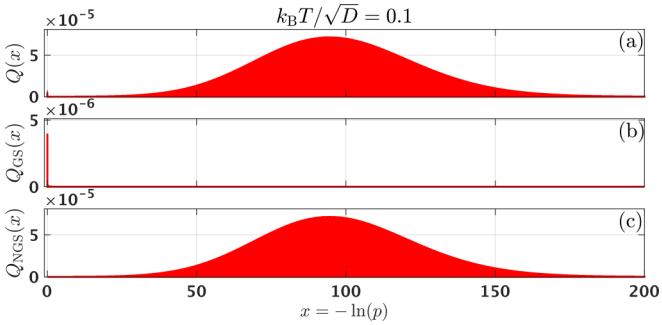

FIG. 11. Separation of the glass-phase ground and nonground-state distributions. (a) The total base-pairing distribution within the glass phase ($k_BT/\sqrt{D} = 0.1$). (b) The ground-state base-pairing distribution (note the difference in scale). This peak represents the frequency of those pairs "locked" in the ground-state structure and are the contributing weight seen in the inset of Fig. 9(b). (c) The non-ground-state base-pairing distribution. As mentioned in the text, the broad Gaussian-like peak seen in the low-temperature regime is a result of non-ground-state pairs that have a small, nonzero pairing probability at finite temperatures. These probabilities would be zero at $T = 0$.

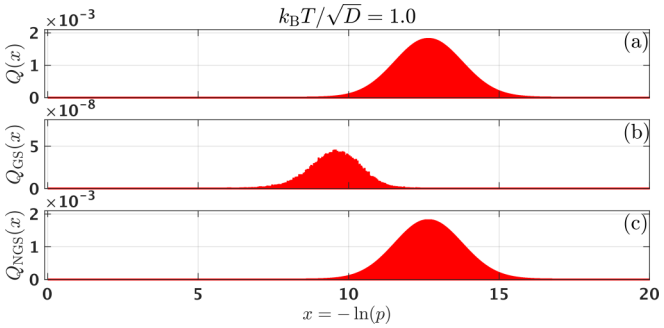

FIG. 12. Separation of the molten-phase ground and non-ground-state distributions. (a) The total base-pairing distribution within the molten phase ($k_BT/\sqrt{D} = 1.0$). (b) The ground-state base-pairing distribution (note the difference in scale). The lack of weight close to $x \approx 0$ is due to a significantly reduced pairing probability of ground-state base pairs at high temperatures leading to the absence of any feature in the inset of Fig. 9(a). (c) The non-ground-state pairs dominate the total pairing probability distribution.

and only the feature at $x \approx 0$ remains and is unaffected in amplitude by multiplication with e^{-x}. Thus, as suspected, in the glass phase only the small-x range contributes significantly to the ensemble-averaged pairing probability.

B. Locked base pairs at finite temperature stem from ground-state base pairs at zero temperature

To confirm that the peak at $x \approx 0$ in the low-temperature behavior is indeed generated from the base pairs that form the ground-state structure at $T = 0$, we explicitly separate the distribution $Q(x)$ into two sub-distributions. We use the approach outlined at the end of Sec. II C to calculate the ground-state structure for every disorder configuration and then tabulate the base-pairing probabilities across all ℓ for base pairs participating in the ground-state structure separately from those which do not. This allows us to split the entire base-pairing probability $Q(x)$ into a ground-state contribution $Q_{GS}(x)$ and a non-ground-state contribution $Q_{NGS}(x)$. Figures 11 and 12 show these ground-state and non-ground-state distributions at the lowest and highest temperatures considered. In Fig. 11(b) one can see that the ground-state pairs directly contribute to the distinct peak close to $x \approx 0$, which we interpret as those base pairs that are "locked" in the ground-state structures. This further confirms that it is the non-ground-state pairs that make up the bulk of the total distribution seen in Fig. 9(b), each of which would otherwise be zero at $T = 0$. Figure 12 then explains the absence of any weight in the inset of Fig. 9(a). The lack of ground-state pair contributions in the region $x \approx 0$ indicates that the system is no longer in the glass phase at this temperature. Since the transition is driven by the behavior of the ground-state base pairs alone, we focus on $Q_{GS}(x)$ from now on.

C. Locked base pairs disappear at the phase transition

Our next goal is to quantify how the weight of "locked" ground-state base pairs changes through the transition. Since a ground-state base pair remaining "locked" at a finite temperature does not mean that the pairing probability of this base pair is one, but rather that it is sizable in some way, we introduce a pragmatic way to characterize the disappearance of "locked" base pairs. Specifically, we divide the range $0 \leqslant x \leqslant 0.7$ (corresponding to $p \geqslant 0.5$) into an equal number of bins and ask which fraction of these bins receive at least one ground-state base pair with $N/3 \leqslant \ell \leqslant N/2$ when we sample 10 000 independent disorder configurations (for $N = 1000$). This quantity is shown in Fig. 13. While its detailed behavior is definitely dependent on the choice of the upper cutoff

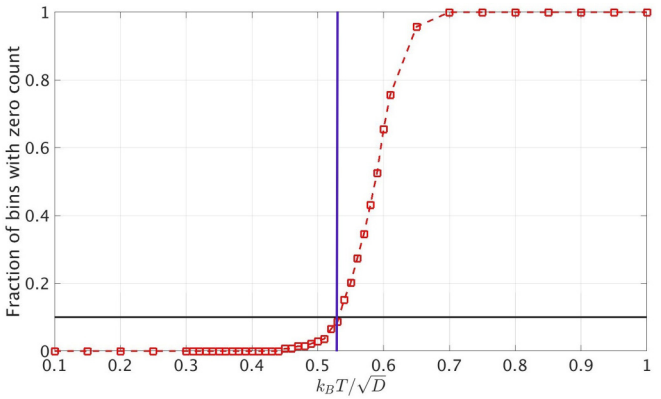

FIG. 13. Fraction of bins at $x \leqslant 0.7$ with zero counts as a function of temperature for $N = 1000$. We split the range $0 \leqslant x < 0.7$ (corresponding to $p \geqslant 0.5$) into 139 equal bins of width 0.005. A bin has zero counts when not a single ground-state base pair with $N/3 \leqslant \ell \leqslant N/2$ in 10 000 disorder configurations has a thermally averaged base pairing probability in the interval of x associated with the bin. The fraction of bins with zero counts dramatically increases from near zero to near one at the phase transition temperature. The horizontal solid line indicates where 10% of the bins have zero counts, the vertical line the location of T_c.

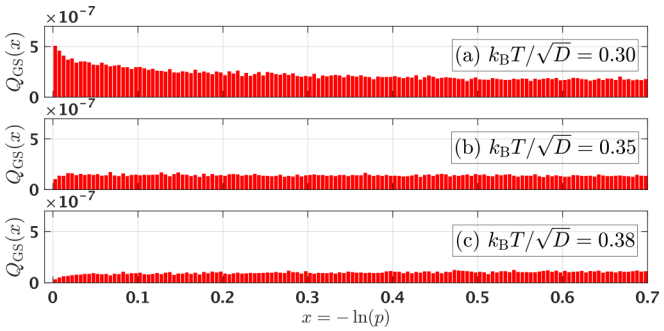

FIG. 14. Average ground-state base-pairing probabilities versus x for system size $N = 1000$. (a) At $k_BT/\sqrt{D} = 0.3$, and all temperatures below, the region close to $x \approx 0$ is the dominant peak. (b) As the temperature is increased, the height of these peaks all drop to roughly the same height as the neighboring peaks. (c) Above $k_BT/\sqrt{D} \approx 0.35$, the probability density $Q_{GS}(x)$ at $x \approx 0$ is actually lower than at moderate x.

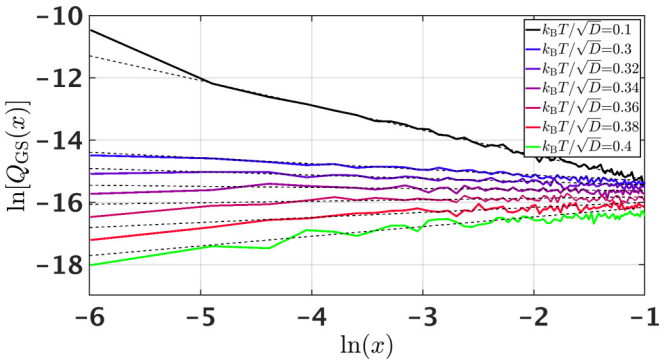

FIG. 15. Double logarithmic plot of the average ground base-pairing distributions versus x for system size $N = 1000$. At temperatures below the glass transition, the ground-state pairing distributions at moderate pairing probabilities are approximate power laws (black dashed fit lines, bins with zero counts ignored in the fits) whose exponents turn from negative to positive as the temperature increases.

on x, the width of the bins, and the number of samples, it is quite clear from the very strong temperature dependence shown in Fig. 13 that essentially all log-probabilities $x \leqslant 0.7$ are present for $k_BT/\sqrt{D} \leqslant 0.45$, while they are absent for $k_BT/\sqrt{D} \geqslant 0.7$. This switch should only moderately depend on our particular choices. It is consistent with the disappearance of locked base pairs at the phase transition.

D. Behavior of the base-pairing probability distribution at $x \approx 0$

As we have seen in Fig. 11, the base-pairing probability distribution is highly peaked at $x \approx 0$ deep in the glass phase, consistent with a picture in which a base pair that is part of the ground-state structure at zero temperature still has probability essentially 1 at $k_BT/\sqrt{D} = 0.1$. Since we have seen in Sec. IV C that the entire base-pairing probability at $x \leqslant 0.7$ ($p \geqslant 0.5$) disappears at the transition to the molten phase, it is of interest to look at the shape of the distribution as the transition is approached. Figure 14 shows these distributions at various temperatures below, but close to, the glass-transition temperature. The most striking feature is that they switch from a behavior where the bulk of the distribution is at $x \approx 0$ ($p = 1$) at low temperatures to a behavior where base-pairing probabilities with small x are still prevalent but probabilities at $x \approx 0$ ($p \approx 1$) itself are suppressed. An intermediate case of a distribution that is essentially independent of x over the entire range $0 \leqslant x \leqslant 0.7$ appears at $k_BT/\sqrt{D} \approx 0.35$.

When plotting the same data on a double logarithmic scale (Fig. 15) we find that over the range of $0 \leqslant x \leqslant 0.7$ the ground-state base-pairing distribution is well described by a power law,

$$Q_{GS}(x) \sim x^{-\gamma} , \tag{27}$$

that allows us to capture its behavior in terms of a single temperature-dependent exponent γ. Figure 15 shows that the slope of these fits (the black dotted lines) turns from negative to positive with increasing temperature. The temperature dependence of the exponent γ is shown in Fig. 16. For $T \to 0$ it approaches -1 consistent with a true δ-peak of the distribution at base-pairing probability one. The exponent crosses zero for $k_BT/\sqrt{D}$ between 0.34 and 0.36 and tends to one at the glass-transition temperature, where it becomes impossible to determine it, since too many bins in x receive zero counts (see above). We conclude that while the transition, at which high-probability base pairs disappear is at $k_BT/\sqrt{D} \approx 0.53$ as shown in Sec. III, already at a lower temperature of $k_BT/\sqrt{D} \approx 0.35$ the system transitions from an ensemble where the bulk of the pairing probability distribution is at $p = 1$ to an ensemble where the probability distribution has a power law with positive exponent at $x = 0$, i.e., where the probability density goes to zero for pairing probabilities of $p = 1$. We call this transition, close in spirit to the one proposed in Ref. [13], *secondary freezing*, stressing that we have not found any signature of a thermodynamic phase transition.

V. CONCLUSIONS

In this work, we numerically studied the glass-molten phase transition of RNA secondary structures using a Gaussian disorder model. With the guidance of previous studies of this system using renormalization-group theory [13–15]

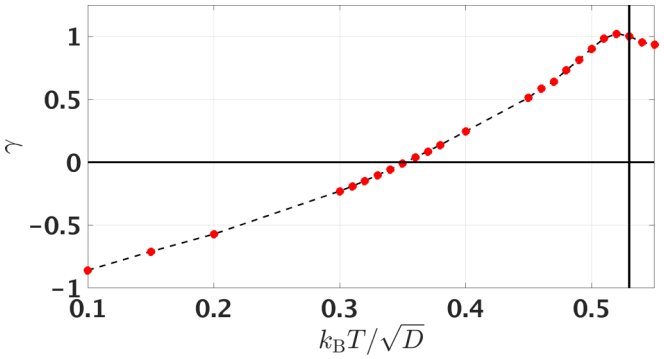

FIG. 16. γ versus temperature. When plotted against temperature, the slopes of the fit lines from Fig. 15 cross zero at a point between $k_BT = 0.34\sqrt{D}$ and $k_BT = 0.36\sqrt{D}$, which could be interpreted as the *secondary freezing* event mentioned in Ref. [13].

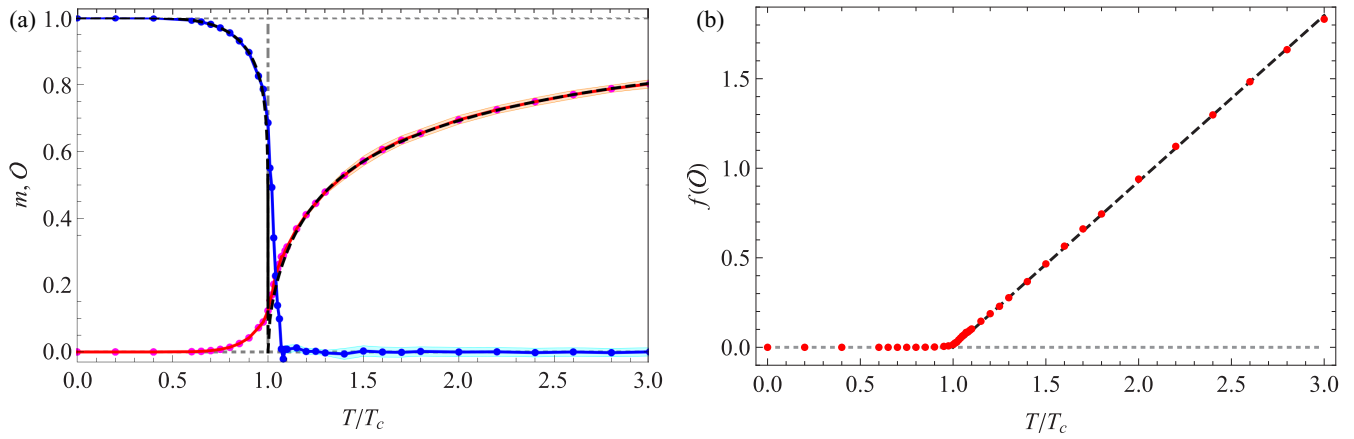

FIG. 17. (a) Simulation results for the two order parameters of the 2D Ising model at linear size $L = 150$: Nonvanishing for $T < T_c$ (blue in online version) the magnetization m; superimposed the analytical result of Yang [28] (black dashed). Nonvanishing for $T > T_c$ (red in online version), the high-temperature order-parameter $\mathcal{O}$ defined in Eq. (A1). The black dashed line is a fit of the data, starting with a square-root singularity in $T - T_c$ (exact form to be confirmed analytically). This can be verified in (b), where $f(x) = x^2 + 4x^6 + 10x^{20}$; the higher-order terms where chosen such that even for large T/T_c the data roughly lie on a straight line, namely $0.927(T/T_c - 1)$. This shows that $\mathcal{O}$ is an order parameter for the high-temperature phase. Simulations where performed for system size $L = 150$, using a total of $1000L^2$ updates per temperature in the region $1 \leqslant T \leqslant 1.3$, and $200L^2$ updates elsewhere, starting from a configuration annealed at the next smaller temperature. Statistical fluctuations are given by the shaded regions. The dominant error is systematic.

we determined the precise location of the transition by the introduction of two order parameters, one for the glass phase and one for the molten phase. In addition to precisely locating the critical temperature of this system, we confirmed the numerical values of the critical exponents at the phase transition predicted by the field theory.

We provide an explanation as to how this model transitions between its molten and glass states by studying the behavior of the base-pairing probability distributions. In particular, we show that the total base-pairing probability distribution can be broken into two sub-distributions composed of base pairs that are "locked" in the ground-state structures below the glass transition temperature, $Q_{\mathrm{GS}}(x;\ell)$, and non-ground-state base pairs, $Q_{\mathrm{NGS}}(x;\ell)$, that never develop a sizable pairing probability. Interestingly, we also identify a potential *secondary freezing* event [13] within the low temperature phase. At temperatures below this secondary freezing event, the most prevalent base pairing probability of ground-state base pairs remains one even at finite temperatures indicating that they are truly locked. Above this secondary freezing event but below the glass transition temperature, base pairs with significant pairing probabilities ($p \geqslant 0.5$) remain prevalent but the density of base pairs with pairing probability one itself vanishes, indicating that locking is maintained only in the weaker sense that locked base pairs are common to a finite fraction of the thermodynamic ensemble rather than to "all" structures. It would be interesting to study the implications of this secondary freezing event on the kinetics of RNA structures.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grants No. DMR-1410172 and No. DMR-1719316. K.W. thanks W. Janke for discussions and early simulations addressing the Lässig-Wiese phase transition scenario and for pointing out the importance of finite-size effects. We acknowledge useful discussions with F. David and M. Lässig.

APPENDIX: CONSTRUCTION OF TWO ORDER PARAMETERS IN THE TWO-DIMENSIONAL ISING MODEL

In this Appendix we construct a pair of order parameters vanishing on either side of a phase transition in a system known to have only a single phase transition. To this end, consider the Curie point in magnets, separating a ferromagnetic from a paramagnetic phase. In the ferromagnetic phase, the magnetization is an order parameter, vanishing when approaching the transition from below. This order parameter has a $\mathbb{Z}_2$ symmetry, which is spontaneously broken in the ordered phase. The connected correlation function in the ordered phase decays exponentially, defining a correlation length ξ_{ferro}, which diverges at the transition, $\xi_{\mathrm{ferro}} \sim |T - T_c|^{-\nu^*}$. While this correlation length does not show the $\mathbb{Z}_2$ symmetry breaking, its inverse could still be used as an order parameter.

In the high-temperature phase it is more difficult to find an order parameter. One idea is to use the (inverse) correlation length, which vanishes as $1/\xi_{\mathrm{para}} \sim |T - T_c|^{\nu^*}$. A better observable can be constructed as follows: Draw a line along one of the lattice directions, and note the numbers S of consecutive spins with the same sign. As an example, in a system of linear size $L = 10$ with periodic boundary conditions, the configuration "0111100010" on the chosen line translates into the four random variables $\{S_1, \ldots, S_4\} = \{4, 3, 1, 2\}$. Define

the observable

$$\mathcal{O} := \left(\frac{\langle S\rangle}{\langle S^2\rangle} - \frac{1}{L}\right) \times \left(\frac{1}{3} - \frac{1}{L}\right)^{-1}. \tag{A1}$$

At $T = 0$, all spins are aligned and S takes only one value, $S = L$. By construction $\mathcal{O}$ vanishes. For $T \to \infty$, the spins are independent, and the ratio $\langle S\rangle/\langle S^2\rangle \to 1/3$. Thus, $\mathcal{O}$ approaches 1 for $T \to \infty$. The behavior for intermediate temperatures is shown in Fig. 17 (left): The observable $\mathcal{O}$ vanishes (approximately) for $T \leqslant T_c$ and roughly grows as $\mathcal{O} \simeq 0.927\sqrt{T - T_c}$ for $T > T_c$. This is more clearly seen by plotting $\mathcal{O}^2$ (plus higher-order terms to bring the result closer to a straight line for larger T), see the right of Fig. 17. A small rounding around $T = T_c$ persists due to finite-size effects. Since there is only one phase transition in the 2D Ising model, this shows that one can construct two order parameters for this single transition.

[1] G. R. Bowman, V. A. Voelz, and V. S. Pande, Curr. Opin. Struct. Biol. **21**, 4 (2011).
[2] P. Wu, X. Zuo, H. Deng, X. Liu, L. Liu, and A. Ji, Brain Res. Bull. **97**, 69 (2013).
[3] J. P. Taylor, J. Hardy, and K. H. Fischbeck, Science **296**, 1991 (2002).
[4] E. G. Conlon and J. L. Manley, Genes Dev. **31**, 1509 (2017).
[5] F. U. Hartl, Annu. Rev. Biochem. **86**, 21 (2017).
[6] T. Ishiguro, N. Sato, M. Ueyama, N. Fujikake, C. Sellier, A. Kanegami, E. Tokuda, B. Zamiri, T. Gall-Duncan, M. Mirceta, Y. Furukawa, T. Yokota, K. Wada, J. P. Taylor, C. E. Pearson, N. Charlet-Berguerand, H. Mizusawa, Y. Nagai, and K. Ishikawa, Neuron **94**, 108 (2017).
[7] J. S. McCaskill, Biopolymers **29**, 1105 (1990).
[8] P. G. de Gennes, Biopolymers **6**, 715 (1968).
[9] R. Bundschuh and T. Hwa, Phys. Rev. E **65**, 031903 (2002).
[10] R. Bundschuh and T. Hwa, Europhys. Lett. **59**, 903 (2002).
[11] M. Müller, F. Krzakala, and M. Mézard, Eur. Phys. J. E **9**, 67 (2002).
[12] T. Liu and R. Bundschuh, Phys. Rev. E **69**, 061912 (2004).
[13] M. Lässig and K. J. Wiese, Phys. Rev. Lett. **96**, 228101 (2006).
[14] F. David and K. J. Wiese, Phys. Rev. Lett. **98**, 128102 (2007).
[15] F. David and K. J. Wiese, J. Stat. Mech. (2009) P10019.
[16] S. Hui and L.-H. Tang, Eur. Phys. J. B **53**, 77 (2006).
[17] A. Pagnani, G. Parisi, and F. Ricci-Tersenghi, Phys. Rev. Lett. **84**, 2026 (2000).
[18] E. Marinari, A. Pagnani, and F. Ricci-Tersenghi, Phys. Rev. E **65**, 041919 (2002).
[19] F. Krzakala, M. Mézard, and M. Müller, Europhys. Lett. **57**, 752 (2002).
[20] P. G. Higgs, Q. Rev. Biophys. **33**, 199 (2000).
[21] J.-L. Chen and C. W. Greider, Proc. Natl. Acad. Sci. USA **102**, 8080 (2005).
[22] D. W. Staple and S. E. Butcher, PLoS Biol. **3**, e213 (2005).
[23] I. Tinoco and C. Bustamante, J. Mol. Biol. **293**, 271 (1999).
[24] R. Lorenz, S. H. Bernhart, C. Honer Zu Siederdissen, H. Tafer, C. Flamm, P. F. Stadler, and I. L. Hofacker, Alg. Mol. Biol. **6**, 26 (2011).
[25] P. G. Higgs, Phys. Rev. Lett. **76**, 704 (1996).
[26] R. Nussinov and A. B. Jacobson, Proc. Natl. Acad. Sci. USA **77**, 6309 (1980).
[27] W. Janke and K. J. Wiese (unpublished).
[28] C. N. Yang, Phys. Rev. **85**, 808 (1952).