

Reentrant Melting of RNA with Quenched Sequence Randomness

G. N. Hayrapetyan,¹ F. Iannelli,² J. Lekscha,² V. F. Morozov,¹ R. R. Netz,² and Y. Sh. Mamasakhlisov¹

¹*Department of Molecular Physics, Yerevan State University, 1 Alex Manouagian Street, Yerevan 0025, Armenia*

²*Fachbereich Physik, Freie Universität Berlin, 14195 Berlin, Germany, EU*

(Received 12 August 2013; published 5 August 2014)

The effect of quenched sequence disorder on the thermodynamics of RNA secondary structure formation is investigated for two- and four-letter alphabet models using the constrained annealing approach, from which the temperature behavior of the free energy, specific heat, and helicity is analytically obtained. For competing base pairing energies, the calculations reveal reentrant melting at low temperatures, in excellent agreement with numerical results. Our results suggest an additional mechanism for the experimental phenomenon of RNA cold denaturation.

DOI: 10.1103/PhysRevLett.113.068101

PACS numbers: 87.14.gn, 87.15.ad, 87.15.bd

RNA plays a central role in all living systems. In addition to transmitting genetic information, RNA participates actively in a variety of cellular processes [1]. RNA chains consist of four different bases A, C, G, and U that can form double-helical structures consisting of stacks of A-U or G-C Watson-Crick base pairs; thereby, the base sequence controls the three-dimensional RNA structure. RNA sequence is, thus, subject to evolutionary selection as the structure is important for proper functioning [2]. Experimentally, RNA secondary structure stability is probed in denaturation studies, where RNA chains are unfolded by change of temperature or denaturant concentration. In the standard scenario, RNA secondary structures melt at high temperature in order to maximize the conformational chain entropy [3]. But RNA melting at low temperatures was also observed, a finding that was controversially discussed and still is only partially understood [4,5]. For proteins, cold unfolding is argued to be a consequence of the nonmonotonic temperature dependence of the hydrophobic effect [6–8], and for RNA, it is not clear whether the same mechanism applies [5]. In this Letter, we show that quenched sequence disorder alone can weaken RNA secondary structure stability at low temperatures, which thus offers an alternative mechanism for the experimental phenomenon of RNA cold melting.

Because of the challenging theoretical nature of the RNA folding problem, it has been extensively investigated with emphasis on the effects of pseudoknots [9], chain loops [10], externally applied pulling forces [11], and most relevant in the present context, sequence disorder [12–16]. The key point of the treatment of RNA sequence disorder is to correctly distinguish between the base pairing degrees of freedom, which are annealed and averaged over in the partition function, and the sequence degrees of freedom, which are quenched and averaged on the level of the free energy. Previously, replica and numerical methods were applied to quenched disordered RNA models. Here, we use the constrained annealing approach [17–20], which we first

validate by comparison with exact enumeration methods [21,22] for finite-length RNA chains for one selected parameter value. We demonstrate that the constrained annealing approach offers a surprisingly accurate description of the RNA folding problem with quenched disorder. We next analyze the thermodynamics of an RNA model with sequence disorder in the complete parameter space spanned by the base-pair energy and sequence distribution parameters. We find the temperature-dependent specific heat to exhibit two broad maxima, which is paralleled by a helicity decrease at both low and high temperatures only for competing base pairing energies (favorable for unequal and unfavorable for equal bases).

We study RNA sequences using two different alphabets: two letter, composed of only A and U bases, and four letter, composed of A, U, G, and C bases. The Hamiltonian for an RNA chain with N bases reads

$$\mathcal{H}(M, \{h_N\}) = \sum_{1 \leq i < j \leq N} m_{ij} \epsilon_{ij}, \quad (1)$$

where M is the $N \times N$ base-pairing matrix with $m_{ij} = 1$ if bases i and j are paired and $m_{ij} = 0$ otherwise. For the two-letter alphabet, the interaction

$$\epsilon_{ij} = \epsilon_0 + \epsilon h_i h_j \quad (2)$$

contains a constant ϵ_0 and a term dependent on the sequence variables $h_i \pm 1$ that describes the nucleotide type; $h_i = +1$ corresponds to A, and $h_i = -1$ corresponds to U. For the four-letter alphabet, we have

$$\epsilon_{ij} = \epsilon_0 + \epsilon h_i h_j + \epsilon' g_i g_j, \quad (3)$$

with an additional degree of freedom $g_i = \pm 1$. The partition function reads

$$Z_N(\{h_N, g_N\}) = \sum_M' \exp[-\beta \mathcal{H}(M, \{h_N, g_N\})], \quad (4)$$

where $\beta = 1/k_B T$ and the sum is taken over all realizations of M that have at most one nonzero entry in each row and column and that contain no pseudoknots. The first restriction reflects that each base can at most pair with one other base, while the absence of pseudoknots means that for any two base pairs (i, j) and (k, l) with $i < j, k < l$, and $i < k$ we have either $i < k < l < j$ or $i < j < k < l$ [12]. The base sequence distribution for the two-letter alphabet factorizes as

$$\mathcal{P}_2(\{h_N\}) = \prod_{i=1}^N \rho(h_i, q) \quad (5)$$

with $\rho(h_i) = q\delta(h_i - 1) + (1 - q)\delta(h_i + 1)$ where q and $1 - q$ are the probabilities to have $h_i \pm 1$, respectively. For the four-letter alphabet, the base sequence distribution factorizes as $\mathcal{P}_4(\{h_N\}, \{g_N\}) = \prod_{i=1}^N \rho(h_i, q)\rho(g_i, p)$. Without pseudoknots, the partition function of an RNA subchain $Z_{i,j}$ between nucleotides i and j can be recursively calculated as [21,22]

$$Z_{i,j} = Z_{i,j-1} + \sum_{k=i}^{j-1} Z_{i,k-1} Q_{kj} Z_{k+1,j-1}, \quad (6)$$

where $Q_{kj} = \exp(-\beta\epsilon_{kj})$ is the statistical weight of a base pair between nucleotides k and j and $Z_N = Z_{1,N}$.

We first focus on the two-letter alphabet case. The free energy per nucleotide for a given sequence $\{h_N\}$ is defined as $\beta f(\{h_N\}) = -N^{-1} \ln Z_N(\{h_N\})$. Self-averaging means that $f(\{h_N\})$ in the thermodynamic limit $N \rightarrow \infty$ becomes independent of $\{h_N\}$ and equal to the sequence-averaged free energy [23]

$$\beta \bar{f} = -\frac{1}{N} \overline{\ln Z_N(\{h_N\})}, \quad (7)$$

where $\bar{\mathcal{O}}$ denotes the average over the sequence distribution $\mathcal{P}_2(\{h_N\})$ in Eq. (5). This fact is exploited by constructing the constrained annealed free energy

$$\beta f_{\text{ca}}(\mu) = -\frac{1}{N} \ln Z_N^{\text{ca}} = -\frac{1}{N} \overline{\ln Z_N(\{h_N\}) e^{-N\mu\alpha(\{h_N\})}}, \quad (8)$$

where $\alpha(\{h_N\})$ is an appropriately defined intensive sequence-dependent quantity that self-averages to zero. According to Ref. [19], f_{ca} satisfies the conditions

$$\bar{f} \geq f_{\text{ca}}(\mu) \geq f_a, \quad (9)$$

where $\beta f_a = -N^{-1} \overline{\ln Z_N(\{h_N\})}$ is the reduced annealed free energy. Thus, $f_{\text{ca}}(\mu)$ improves the annealing approximation and constitutes a strict lower bound of the quenched free energy \bar{f} . The best estimate follows from maximization of f_{ca} with respect to the Lagrange multiplier μ ,

$$f_{\text{ca}}^* = f_{\text{ca}}(\mu^*) = \max_{\mu} f_{\text{ca}}(\mu). \quad (10)$$

The simplest constraint imposed on the quenched variables $\{h_N\}$ uses the linear variable $\alpha(\{h_N\}) = \frac{1}{N} \sum_{i=1}^N [h_i - (2q - 1)]$. After some algebra (see the Supplemental Material [24]), Z_N^{ca} defined in Eq. (8) follows as

$$Z_N^{\text{ca}} = e^{N\mu(2q-1)} \Omega^N Z_N^{\text{hom}}(\epsilon_{\text{ca}}), \quad (11)$$

where $Z_N^{\text{hom}}(\epsilon_{\text{ca}})$ is the partition function of a homogeneous RNA chain obtained from Eq. (4) by setting $\epsilon_{ij} = \epsilon_{\text{ca}}$. Here, $\epsilon_{\text{ca}} = \epsilon_0 - (1/\beta) \ln(W/\Omega^2)$, $\Omega = qe^{-\mu} + (1 - q)e^{\mu}$, and $W = e^{-\beta\epsilon} [q^2 e^{-2\mu} + (1 - q)^2 e^{2\mu}] + 2q(1 - q)e^{\beta\epsilon}$. Since Z_N^{hom} for $N \gg 1$ to leading order scales as [14] $Z_N^{\text{hom}} \sim N^{-3/2} (1 + 2\sqrt{Q_{\text{ca}}})^N$, where $Q_{\text{ca}} = e^{-\beta\epsilon_{\text{ca}}}$, the variational free energy becomes

$$\frac{f_{\text{ca}}(\mu)}{k_B T} = -\mu(2q - 1) - \ln \Omega(\mu) - \ln(1 + 2\sqrt{Q_{\text{ca}}}). \quad (12)$$

Maximization of this expression with respect to μ yields μ^* and, thereby, the optimal constrained annealed free energy f_{ca}^* .

In Fig. 1(a) we compare f_{ca}^* (thick blue line) for $q = 0.75$ with exact enumeration results for $N = 50$ bases using the recursion algorithm Eq. (6), where red lines show 30 distinct random sequence realizations and the dotted line shows the corresponding quenched average \bar{f} according to Eq. (7). The overall agreement between the numerical and constrained annealing approaches is very good. The energy parameters $\epsilon_0 < 0$ and $\epsilon = 1.5|\epsilon_0|$ are chosen such that the UA base-pair energy $\epsilon_0 - \epsilon = -2.5|\epsilon_0|$ is favorable but the UU and AA base-pair energies $\epsilon_0 + \epsilon = 0.5|\epsilon_0|$ are unfavorable, yielding the particularly interesting competitive scenario.

A more critical comparison is possible with the specific heat, defined as $C_V = -T\partial^2 f/\partial T^2$, shown in Fig. 1(b). The constrained annealing result (blue line) shows an interesting double-peak structure, which is in perfect agreement with the quenched average numerical result (dotted line). The spread in the individual sequence realizations (red lines) is substantial, showing that the $N = 50$ chain is still far from the self-averaging limit. The structural significance of two peaks in the heat capacity can be assessed by the helicity order parameter θ

$$\theta = \frac{2}{N} \left\langle \sqrt{\sum_{i < j} m_{ij}} \right\rangle, \quad (13)$$

where $\langle \mathcal{O} \rangle$ denotes the thermal average over all permitted base pairing matrices M as in Eq. (4). $\theta = 1$ corresponds to a situation where all bases are paired, and $\theta = 0$ corresponds to the situation where no base is paired. In the

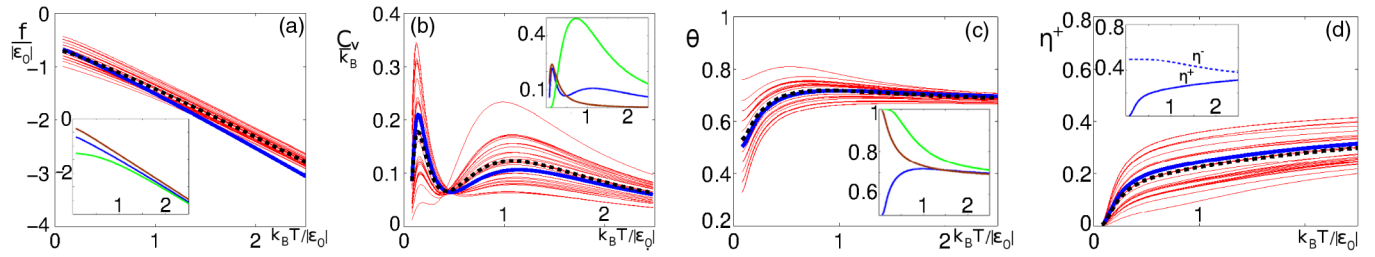


FIG. 1 (color online). (a) Free energy f , (b) specific heat C_v , (c) helicity degree θ , and (d) fraction of unfavorable base pairs η^+ vs rescaled temperature $k_B T/|\epsilon_0|$ for parameters $\epsilon_0 < 0$, $\epsilon = 1.5|\epsilon_0|$ and $q = 0.75$. Thin red lines denote numerical results for 30 different base sequence realizations for an $N = 50$ RNA chain, and the thick dashed black line gives the corresponding quenched average. The thick blue line gives the constrained annealing result in the thermodynamic limit $N \rightarrow \infty$. In the insets (a)–(c), constrained annealing (blue line), annealed (green line), and homopolymeric [brown line, with a homogeneous base pairing energy of $\bar{\epsilon} = \epsilon_0 + \epsilon(2q - 1)^2 = -5/8|\epsilon_0|$] cases are compared. The inset in (d) compares the fraction of unfavorable (blue solid line) and favorable (blue broken line) base pairs η^+ and η^- from the constrained annealing approach.

constrained annealing approximation, θ is given by the previously derived expression for a homogeneous RNA chain [14] $\theta = 2\sqrt{Q_{ca}}/(1 + 2\sqrt{Q_{ca}})$. In the numerics, θ is calculated from the probability of base pair formation between nucleotides i and j , $\langle m_{ij} \rangle = Q_{ij}Z_{i+1,j-1}Z_{j+1,N+i-1}/Z_{1,N}$ [14], where the partition function $Z_{j+1,N+i-1}$ is calculated from Eq. (6) for a periodically replicated sequence. In Fig. 1(c), we compare θ obtained by the constrained annealing approach and by exact enumeration (same color coding as before), again giving excellent agreement. The helicity exhibits a distinct maximum at intermediate temperatures, indicating the presence of melting at both high and low temperatures.

Insight into the mechanism of the low-temperature melting can be gained by considering the fractions η^+ and η^- of energetically unfavorable (i.e., AA and UU) and favorable (i.e., UA) base pairs, defined as

$$\eta^\pm = \frac{2}{N} \left\langle \sum_{i < j} \delta(h_i h_j \mp 1) m_{ij} \right\rangle, \quad (14)$$

which follow straightforwardly by suitable derivatives of the free energy (Supplemental Material [24]). By definition, $\theta = \eta^+ + \eta^-$. In Fig. 1(d), we compare η^+ from the constrained annealing and numerical methods. It is seen that the fraction of unfavorable contacts quite abruptly decreases at low temperature. In the inset of Fig. 1(d) the constrained annealing results for η^+ and η^- demonstrate that favorable contacts η^- monotonically grow with decreasing temperature over a broad range roughly paralleling the high-temperature peak in C_v , while the abrupt fall of η^+ roughly coincides with the low-temperature peak of C_v . The antagonistic temperature behavior of η^+ and η^- , thus, not only explains the nonmonotonicity of θ in Fig. 1(c) but also the double-peak structure of C_v in Fig. 1(b).

That the double-peak structure of C_v is really a consequence of quenched randomness is demonstrated in the inset of Fig. 1(b), where we compare the constrained

annealing result (blue line) with the annealed case (green line) and the homogeneous case of an RNA chain with an averaged base pairing energy of $\bar{\epsilon} = \epsilon_0 + \epsilon h_i h_j = \epsilon_0 + \epsilon(2q - 1)^2 = -5/8|\epsilon_0|$ (brown line). The annealed calculation only produces a single broad high-temperature peak, while the homogeneous calculation reproduces accurately the low-temperature peak. The reason is the absence of competition between favorable and unfavorable base pairing, leading to a monotonic increase in helicity θ with sinking temperature for the annealed and homogeneous scenarios, see Fig. 1(c) inset.

After having validated the constrained annealing approach for one parameter value, we next use it to explore the global RNA phase behavior. Figure 2 shows (a) helicity θ and (b) heat capacity C_v for the previously used fixed energy parameters $\epsilon_0 < 0$, $\epsilon = 3|\epsilon_0|/2$, and for various probabilities $1/2 \leq q \leq 1$ (note that the system is symmetric around $q = 1/2$, so the range $q < 1/2$ gives identical results). Except the homopolymeric case $q = 1$ and the symmetric case $q = 1/2$, all systems show a double peak in C_v and a decrease of θ at both high and low temperatures. The homopolymeric limit for $q = 1$ has been discussed before; the case $q = 1/2$ is more interesting: Minimization of Eq. (12) for $q = 1/2$ yields $\mu^* = 0$ corresponding to an annealed average in Eq. (8), which

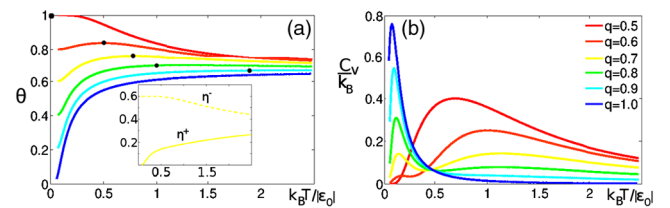


FIG. 2 (color online). Temperature dependence of (a) helicity θ and (b) specific heat C_v for the competitive scenario $\epsilon_0 + \epsilon > 0$ ($\epsilon_0 < 0$, $\epsilon = 3|\epsilon_0|/2$) for various base probabilities q . The black dots in (a) indicate the maxima. The inset (a) compares the fraction of unfavorable (η^+ , yellow solid line) and favorable (η^- , yellow broken line) base pairs for $q = 0.7$.

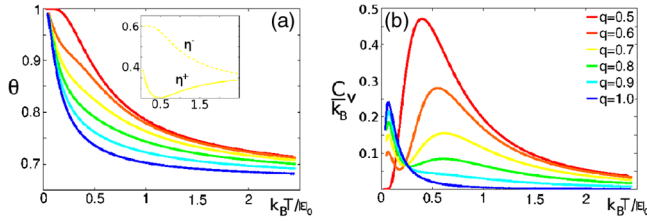


FIG. 3 (color online). Temperature dependence of (a) helicity θ and (b) specific heat C_v for the noncompetitive scenario $\epsilon_0 + \epsilon < 0$ ($\epsilon_0 < 0$, $\epsilon = 2|\epsilon_0|/3$) for various q . The inset (a) compares the fraction of less favorable (η^+ , solid line) and more favorable (η^- , broken line) base pairs for $q = 0.7$.

as we showed already in Fig. 1, has a single peak in C_v and monotonic θ . This is in harmony with the numerical work on the same model by Pagniani *et al.*, who only considered the symmetric situation $q = 1/2$ [13]. We note that for all probabilities $q \neq 1/2$, the number of possible unfavorable base pairs UU and AA, $q^2 + (1 - q)^2$, is larger than the number of favorable base pairs, UA and AU, $2q(1 - q)$; thus, at high temperature UU and AA base pairs are for $q \neq 1/2$ statistically favored. This explains why the symmetry line $q = 1/2$ is special, since only on this line do favorable and unfavorable base pairs have the same statistical weight. We conclude that the double peak in C_v and nonmonotonicity of θ are obtained for all q except the special cases $q = 0, 1/2, 1$.

Whereas the energy parameters considered so far of $\epsilon_0 < 0$, $\epsilon = 3|\epsilon_0|/2$ give rise to competition between favorable UA base pairs (energy $\epsilon_0 - \epsilon < 0$) and unfavorable AA/UU base pairs (energy $\epsilon_0 + \epsilon > 0$), we next consider the noncompetitive situation for which both $\epsilon_0 - \epsilon < 0$ and $\epsilon_0 + \epsilon < 0$. The results in Fig. 3 for $\epsilon_0 < 0$, $\epsilon = 2|\epsilon_0|/3$ for various q values demonstrate that the heat capacity C_v still displays two maxima but that the cold melting has disappeared and the helicity θ monotonically increases with lowering temperature.

While the low-temperature peak of the heat capacity C_v in the competitive case is caused by the gradual disappearance of unfavorable base pairs η^+ with lowering temperature [see the inset of Fig. 2(a)], the low-temperature C_v peak in the noncompetitive case is produced by a low-temperature increase of less-favorable base pairs η^+ [see the inset of Fig. 3(a)].

Figure 4 summarizes the RNA folding behavior in a phase diagram as a function of the sequence probability q and the base pairing energy $\epsilon_0 + \epsilon$ between identical bases UU and AA (we assume the UA and AU base pairing energy to be favorable, i.e., $\epsilon_0 - \epsilon < 0$). For the competitive situation $\epsilon_0 + \epsilon > 0$ (upper half), two peaks in C_v are present and both high- and low-temperature melting occurs as indicated by a maximum in the helicity at intermediate temperatures. In the noncompetitive situation $\epsilon_0 + \epsilon < 0$ (lower half), we still find two peaks in C_v but only high-temperature melting takes place and the helicity increases

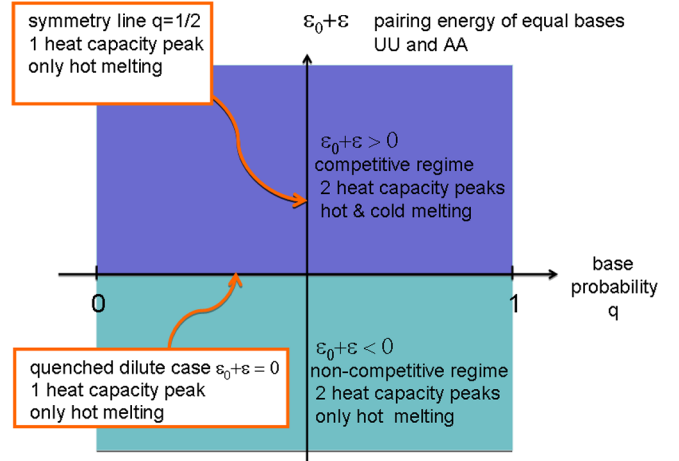


FIG. 4 (color online). Phase diagram of RNA thermal folding behavior as a function of base probability q and base-pair energy between similar nucleotides $\epsilon_0 + \epsilon$. Except for the homopolymeric scenarios for $q = 0$ and $q = 1$, the annealed scenario for $q = 1/2$, and the quenched dilute case $\epsilon_0 + \epsilon = 0$, the heat capacity shows two peaks. For $\epsilon_0 + \epsilon > 0$ (upper half), both high- and low-temperature melting occurs, while for $\epsilon_0 + \epsilon < 0$ (lower half) only high-temperature melting takes place.

monotonically towards $\theta = 1$ as the temperature decreases. The lines $q = 0$ and $q = 1$ correspond to the standard homopolymer case, and the symmetry line $q = 1/2$ is within the constrained annealing approach mapped onto the annealed scenario, which shows phase behavior similar to the homopolymeric case. The line $\epsilon_0 + \epsilon = 0$ corresponds to the quenched dilute case, where only hot melting takes place and the specific heat exhibits only one peak.

Similar behavior is observed for the four-letter alphabet. In Figs. 5(a) and 5(b), we compare the helicity degree θ and the specific heat from the constrained annealing calculation with numerical results (red and blue lines) for $q = p = 0.75$. We see that two heat-capacity peaks are also obtained for the more realistic four-letter alphabet, so our results for the low-temperature RNA melting are robust. The mechanism for the experimental phenomenon of cold RNA denaturation [4,5] has been controversially discussed in terms of a positive specific heat difference between the

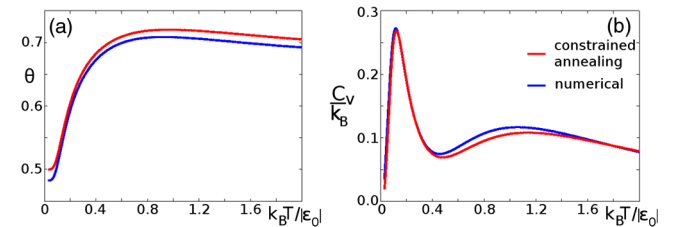


FIG. 5 (color online). Temperature dependence of helicity degree (a) and specific heat (b) for the four-letter alphabet with the energy parameters $\epsilon = 3|\epsilon_0|/2$ and $\epsilon' = -|\epsilon_0|/10$ at the point $p = q = 0.75$.

denatured and native states due to the hydrophobic effect [6,7], electrostatics [4,5], and the competition between inter- and intramolecular hydrogen bonds [26] and constitutes a long-standing puzzle in biopolymer basic research. Our results show that quenched randomness itself weakens the secondary structure formation tendency for nonsymmetric sequence distributions (i.e., $q \neq 1/2$) with competitive base pairing energies at low temperature and leads to reentrant behavior; it thus constitutes an additional mechanism for cold melting. Future experiments with disordered RNA sequences would be highly desirable in order to check our predictions. In the presence of loop entropies, the melting crossovers obtained in the present Letter are expected to turn into real thermodynamic phase transitions in the $N \rightarrow \infty$ limit [10], which will be addressed in future work.

This research was supported by the Volkswagen Foundation under Grant “Equilibrium and non-equilibrium behavior of single- and double-stranded biological molecules,” DFG under Grant No. NE810/8 and the ANSEF Foundation condmatth Grant No. 3087.

-
- [1] *The RNA World*, Cold Spring Harbor Monograph Series 2, edited by R. F. Gesteland and J. F. Atkins (Cold Spring Harbor Laboratory Press, New York, 1993).
 - [2] P. G. Higgs, *J. Phys. I (France)* **3**, 43 (1993).
 - [3] I. Tinoco, Jr. and C. Bustamante, *J. Mol. Biol.* **293**, 271 (1999).
 - [4] P. J. Mikulecky and A. L. Feig, *J. Am. Chem. Soc.* **124**, 890 (2002).
 - [5] P. J. Mikulecky and A. L. Feig, *Nucleic Acids Res.* **32**, 3967 (2004).

- [6] P. L. Privalov, *Crit. Rev. Biochem. Mol. Biol.* **25**, 281 (1990).
- [7] A. V. Finkelstein, *Protein Physics* (Academic Press, New York, 2002).
- [8] F. Sedlmeier, D. Horinek, and R. R. Netz, *J. Chem. Phys.* **134**, 055105 (2011).
- [9] G. Vernizzi, H. Orland, and A. Zee, *Phys. Rev. Lett.* **94**, 168103 (2005).
- [10] T. R. Einert, P. Nager, H. Orland, and R. R. Netz, *Phys. Rev. Lett.* **101**, 048103 (2008).
- [11] M. Manosas and F. Ritort, *Biophys. J.* **88**, 3224 (2005).
- [12] P. G. Higgs, *Phys. Rev. Lett.* **76**, 704 (1996).
- [13] A. Pagnani, G. Parisi, and F. Ricci-Tersenghi, *Phys. Rev. Lett.* **84**, 2026 (2000).
- [14] R. Bundschuh and T. Hwa, *Phys. Rev. E* **65**, 031903 (2002).
- [15] F. Krzakala, M. Mezard, and M. Mueller, *Europhys. Lett.* **57**, 752 (2002).
- [16] M. Lassig and K. J. Wiese, *Phys. Rev. Lett.* **96**, 228101 (2006).
- [17] T. Morita, *J. Math. Phys. (N.Y.)* **5**, 1401 (1964).
- [18] R. Kühn, *Markov Proc. Relat. Fields* **10**, 523 (2004).
- [19] M. Serva and G. Paladin, *Phys. Rev. Lett.* **70**, 105 (1993).
- [20] T. Liu and R. Bundschuh, *Phys. Rev. E* **72**, 061905 (2005).
- [21] M. Zuker and P. Stiegler, *Nucleic Acids Res.* **9**, 133 (1981).
- [22] J. S. McCaskill, *Biopolymers* **29**, 1105 (1990).
- [23] M. Mezard, G. Parisi, and M. Virasoro, *Spin Glass Theory And Beyond* (World Scientific, Singapore, 1987).
- [24] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.113.068101>, which includes Ref. [25].
- [25] P. G. Higgs, *Q. Rev. Biophys.* **33**, 199 (2000).
- [26] A. V. Badasyan, Sh. A. Tonoyan, Y. Sh. Mamasakhlisov, A. Giacometti, A. S. Benight, and V. F. Morozov, *Phys. Rev. E* **83**, 051903 (2011).