RNA folding in the presence of counterions

Yevgeni Sh. Mamasakhlisov,[1,2] Shura Hayryan,[1] V. F. Morozov,[2] and Chin-Kun Hu[1,3,*]
[1]*Institute of Physics, Academia Sinica, Nankang, Taipei 11529, Taiwan*
[2]*Department of Molecular Physics, Yerevan State University, 1 Al. Manougian str., Yerevan 375025, Armenia*
[3]*Center for Nonlinear and Complex Systems and Department of Physics, Chung-Yuan Christian University, Chungli 320, Taiwan*
(Received 29 September 2006; revised manuscript received 1 March 2007; published 11 June 2007)

We present a general thermodynamic picture of the folding of RNA-like heteropolymer based on the basic physical principles. The Hamiltonian of the model includes all characteristic interactions explicitly. A particular attention is paid to the electrostatic interactions whose role in the RNA folding is known to be crucial. In this paper we study RNA folding with the full Hamiltonian and describe the spin-glass behavior on the level of tertiary structure. We show that formation of the stable tertiary structure is possible in the random RNA-like molecule. By including into the model the nonspecific interactions of the RNA molecule with counterions, we derive the logarithmic dependencies of the melting and freezing temperatures on the ion concentration, which is consistent with experimental data [R. Shiman and D. E. Draper, J. Mol. Biol. **302**, 79 (2000)]. We also infer that the large RNA folds slower than the hierarchical model predicts, which was observed in the experiments.

DOI: 10.1103/PhysRevE.75.061907 PACS number(s): 87.10.+e, 87.15.Aa, 87.15.Cc

I. INTRODUCTION

The RNA molecule plays a remarkably versatile role in cellular processes. Its functions vary from serving as a primer in DNA replication to the catalysis of many important molecular reactions. For some viruses RNA is also a depository of genetic information. Correspondingly, many different aspects of RNA behavior are being studied intensively: from mechanical properties to drug binding propensities.

Following one of nature's most ubiquitous principles, RNA's biological function is closely related to its structure. RNA molecules involved in the replication process utilize their nucleotide sequence (the primary structure) to code for proteins. For other functions of RNA the specific folded spatial structures are crucial. The formation of the specific compact folds of RNA is a result of complex physical interactions of different intensity and scale present in the RNA-solvent system. Understanding the thermodynamics and kinetics of these processes constitute one of the most important research problems of the contemporary biophysics. Because of the extremely complex nature of the problem it is impossible to construct a complete theoretical model to investigate the RNA folding. Here we study the general thermodynamic behavior of the RNA molecule by constructing a Hamiltonian which contains all relevant physical interactions. Before describing the model let us recall some details of structural organization of RNA.

Chemically RNA molecule is a heteropolymer consisting of four types of naturally existing nitrogenous bases: A (Adenine), G (Guanine), U (Uracil), and C (Cytosine). The base pairs (A-U) and (G-C) are called complementary or Watson-Crick type pairs. These pairs connect with each other by hydrogen bonds and, when stacked together, they make a major contribution to the stability of spatial structures of the polymer. The typical topology forms of hydrogen bonds are shown in Fig. 1. Each monomer (base) in RNA carries a negative charge and hence the electrostatic interactions (both inter molecular and intramolecular) play an important role in the stabilization of the structure. Three levels of structures are distinguished in the RNA molecule: primary structure, the sequence of nitrogenous bases along the chain; secondary structure, formed by contacts of the bases; and the tertiary structure which is composed of the secondary structure elements.

According to the basic hypothesis the biologically active RNA structure is encoded in the nucleotide sequence [1]. Thermodynamically this structure corresponds to minimum of the free energy. The stability of the folded conformation depends on many factors among which perhaps the most important one is counterions in RNA environments [2]. It is believed that without neutralizing the RNA backbone charges by counterions the collapsed compact structures are not possible at all because the strong repulsion drives the molecule to the rodlike configuration.

In the recent decade, RNA structures have been intensively studied [3]. However, the main focus of these studies is on the secondary structure formation while the tertiary structure formation remains less understood. The ultimate aim of the RNA folding investigations is to discover the relationship between nucleotide sequence and the biologically functional tertiary structure.

It is known that the folding processes of proteins and RNA share many common features, but the existing experimental data show that in contrast to proteins, long RNA mol-

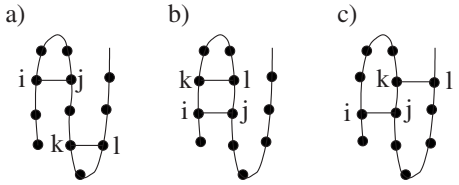

FIG. 1. Three typical pairing types. (a) Bonds (i,j) and (k,l) are independent; (b) bonds (i,j) and (k,l) are nested; (c) bonds (i,j) and (k,l) create a pseudoknot.

*Corresponding author. Electronic address: huck@phys.sinica.edu.tw

ecules exhibit "slow and inaccurate" folding [4,5] accompanied by formation of thermodynamically stable intermediates, usually understood as misfolding. Thus, long RNA molecules seem much more similar to random heteropolymer than the protein. Thus, modelling a RNA molecule by a random heteropolymer sounds more reasonable than the protein.

The basic goal of our present research is to study the folding thermodynamics of RNA with a full Hamiltonian which includes all characteristic interactions present in the dissolved RNA molecule. The question we ask is: what kind of fundamental information on the RNA thermodynamics can be derived from general physical principles.

The secondary structure of RNA is stabilized by strong hydrogen bonds and the "stacking" while the tertiary folds are governed by relatively weak interactions between the secondary structure elements. This separation of the energy scales for RNA secondary and tertiary structures allows one to investigate the formation and thermodynamics of these structures separately with some degree of approximation. Most algorithms for theoretical study of RNA folding are based on the concept of hierarchical mechanism of folding [6], according to which the secondary structure is formed first and is much more stable than the tertiary structure. The three-dimensional structure is formed through packing of the elements of the secondary structure, and alters the later only minimally. However, there are certain indications that the large RNA molecules fold slower than the hierarchical mechanism predicts [4]. This means that separating the folding process into two successive "independent" processes may appear to be an oversimplification.

In the present paper we examine the behavior of a RNA-like molecule without assuming any preassigned order of the folding, and considering any type of possible structures, including pseudoknots as shown in Fig. 1(c). The Hamiltonian of the model includes an explicit realistic electrostatic term. Using a replica approach [7,8], we derive the logarithmic dependencies of the melting and freezing temperatures on the ion concentration, which is consistent with experimental data [9].

This paper is organized as follow. In Sec. II, we define the Hamiltonian for a RNA sequence. In Sec. III, we use the Hamiltonian of Sec. II to calculate the free energy averaged over all possible RNA sequences. We also use the free energy to derive coupled equations for physical variables. In Sec. IV, we use the results of Sec. III to derive equations for freezing and melting temperatures. In Sec. V, we use the results of Sec. IV to discuss the dependence of freezing and melting temperatures on the ion concentration. We also discuss the folding time problem and the role of repulsion between parts of the RNA in the structure stabilization of RNA. In Appendix A, we present a detailed derivation of the free energy given in Sec. III. Appendix B contains calculations for the four-letter nucleotide sequences.

II. THE MODEL

We describe a RNA molecule with N monomers by a random Gaussian chain in which N electrically charged

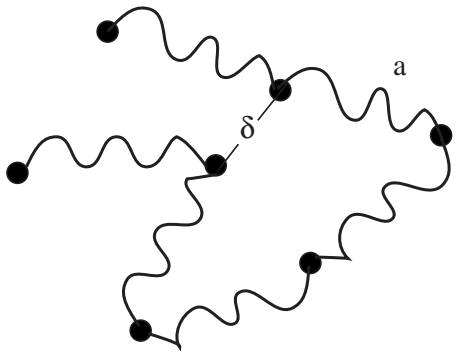

FIG. 2. Schematic description of a random Gaussian chain with $N=7$ beads.

"beads," representing nucleotides, are connected with elastic "springs," (Fig. 2). The monomers are enumerated from 1 to N. The structure of the molecule is described by introducing two $N\times N$ matrices: (1) The complementarity matrix $\hat{M}=\|M_{ij}\|$, where $M_{ij}=1$, if the ith and the jth bases are complementary (Watson-Crick type) and $M_{ij}=0$, otherwise; (2) the contact matrix $\hat{C}=\|C_{ij}\|$, where $C_{ij}=1$, if the ith and the jth bases are hydrogen bonded, and $C_{ij}=0$, otherwise. Thus, the base pair (i,j) contributes to the secondary or tertiary structure only if $M_{ij}C_{ij}=1$. The spatial configuration of the chain is defined by the set $\mathbf{r}\equiv\{\mathbf{r}_i;i=1,2,\cdots,N\}$ of the radius vectors of the monomers. Hereafter the bold roman letters indicate space vectors. The Hamiltonian of the model reads

$$H\{\mathbf{r},\hat{C},\hat{M}\}=H_{\mathrm{el}}(\mathbf{r})+H_{\mathrm{rep}}(\mathbf{r})-\epsilon\sum_{i<j=2}^{N}M_{ij}C_{ij}+\sum_{i<j=2}^{N}M_{ij}C_{ij}\frac{3T(\mathbf{r}_i-\mathbf{r}_j)^2}{2\delta^2},\tag{1}$$

where T is the temperature, δ is the linear size of the Watson-Crick base pair, $\epsilon>0$ is the base pair stabilization energy, and it is assumed to be the same for the base pairs (G-C) and (A-U). The energy of "stacking" is not included into the model explicitly. Note that some other models of RNA secondary structure consider only "stacking" and neglect the hydrogen bond energy [10]. We assume that the complementary pairing is of saturated nature, i.e., each base can be paired with only one other base.

In (1), the term

$$H_{\mathrm{el}}(\mathbf{r})=\frac{3T}{2a^2}\sum_{i=1}^{N-1}(\mathbf{r}_i-\mathbf{r}_{i+1})^2\tag{2}$$

describes the energy due to polymeric elasticity with a being the length of the Kuhn segment of the chain; the term $H_{\mathrm{rep}}(\mathbf{r})$ describes the role of the electrostatic repulsion. In the real physiological aqueous solutions with solute counterions the electrostatic interactions is short ranged due to the screening. In this case the Coulomb interactions can be replaced by an excluded volume pseudopotential [11,12] $v_C\delta(\mathbf{r})$, where $v_C=4\pi l_B f^2 r_D^2$ is the electrostatic excluded volume. Here $l_B=e^2/(\varepsilon T)$ is the Bjerrum length, f is the fractional charge on

a monomer, $r_D=(\varepsilon T/4\pi Ie^2)^{1/2}$ is the Debye radius, ε is the dielectric constant, e is the electron charge, I is the ionic strength of the solution, and $\delta(\mathbf{r})$ is the Dirac's function. The counterions condensation is not the subject of the present paper and, thus $f=1$. With this approximation the repulsion term $H_{\rm rep}$ can be represented as [13,14]

$$H_{\rm rep}=2\pi l_B f^2 r_D^2 \sum_{i\neq j}\delta(\mathbf{r}_i-\mathbf{r}_j). \tag{3}$$

In the case of spatially uniform monomer density ρ one obtains

$$\beta H_{\rm rep}(\mathbf{r})=2\pi l_B f^2 r_D^2 V\rho^2=2\pi l_B f^2 r_D^2 N\rho, \tag{4}$$

where V is the volume occupied by macromolecule and $\rho=N/V$.

The third term in (1) describes the energy of saturated hydrogen-bond interactions. To some degree of approximation we assume that the energy of base stacking is also incorporated in this term implicitly. The justification is that separating the hydrogen-bond and the stacking interactions does not affect the basic results.

The last sum in (1) is the energy of the elastic deformations of the hydrogen bonds. Like the first term, it is calculated in the approximation of the elastic spring.

Some of the parameters introduced above have been considered in [15,16]. The persistence length for the one-strand DNA is estimated as ~20 Å which well agrees with results obtained in [17]. However, in the case of single-stranded nucleic acids one-strand RNA is more flexible, with a persistence length of ~8–9 Å [18–20] in comparison to one-strand DNA with a persistence length of ~15–30 Å [17]. Thus, in the case of one-strand RNA the persistence length can be estimated as $\tilde{l}\approx 8$ Å. Here the length of the Kuhn segment is $a\approx 2\tilde{l}$. We have introduced the parameter δ as a linear size of the Watson-crick base pair. More accurately, we understand it as an equilibrium distance between the antiparallel chains in the double-stranded regions. Thus, δ is of the order of the diameter of double-strand RNA (about 20 Å) and, consequently, one can use $\delta\sim 10$ Å for approximate estimates. The energy of the base pair stabilization is estimated as $\beta\epsilon\approx 3$ ($T\approx 300$ K) [15,16].

III. THE PARTITION FUNCTION AND THE FREE ENERGY

The partition function is a central quantity in thermodynamics. It is defined as the weighted sum over all configurational states. For a given RNA sequence with Hamiltonian $H\{\mathbf{r},\hat{C},\hat{M}\}$ defined by Eq. (1), we carry out integration over all possible $\mathbf{r}$ and summation over all possible $\hat{C}$ to obtain the partition function $Z(\hat{M})$. We assume that the RNA sequence is random, i.e., the entries of the matrix $\hat{M}$ are independent random variables. For the sake of simplicity, here we describe only the results obtained for two-letter sequences (A and U or G and C) from which the entries of the matrix M_{ij} take values 1 or 0 with equal probabilities. The real polynucleotides contain four types of nitrogenous bases and a question arises about the validity of our results. We have carried out the corresponding calculations for the four-letter sequences in Appendix B and found that the whole picture of discussed phenomena remains the same. More accurately speaking, the coefficients in the expansion over order parameter are changed but none of the terms changes its sign from positive to negative and vice versa. And this means that the general behavior of the model is the same.

To average the free energy over quenched disorder of the entries of $\hat{M}$, we use a replica approach [7,8], which requires averaging the partition function of n replicas of the system over the distribution of the $\hat{M}$ entries. After averaging and carrying out the summation over all entries of the matrix $\hat{C}$, we obtain for the n-replica reduced free energy (see Appendix A for details)

$$\beta\mathcal{F}(\rho,[q])=-S[q]+n\beta H_{\rm rep}(\rho)-\max_{\eta} G([\eta,\rho,q]). \tag{5}$$

Here we denoted by q the set of the equilibrium two-replica (a and b) overlap parameters,

$$q_{ab}(\mathbf{x},\mathbf{y})=\left\langle \sum_i \delta(\mathbf{x}-\mathbf{r}_i^a)\delta(\mathbf{y}-\mathbf{r}_i^b)\right\rangle, \tag{6}$$

where $\langle\ldots\rangle$ means thermal averaging, η is some auxiliary parameter appearing due to the calculation method, and $S[q]$ is the entropy of the system of n ideal polymer chains under the restriction $\Sigma_i\delta(\mathbf{x}-\mathbf{r}_i^a)\delta(\mathbf{y}-\mathbf{r}_i^b)=q_{ab}(\mathbf{x},\mathbf{y})$. The function $G([\eta,\rho,q])$ in (5) is given by the equation

$$G([\eta,\rho,q])=-\frac{Nn\eta^2}{2\rho\delta^3}+Nn\ln(1+\eta e^{\beta\epsilon/2}) + Nn\frac{\delta^6}{2(e^{-\beta\epsilon/2}+\eta)^4}\mathcal{M}[q]. \tag{7}$$

The quantity $\mathcal{M}[q]=1/(Nn)\Sigma_{a\neq b}\int dxdx' q_{ab}^2(\mathbf{x},\mathbf{x}')$ describes the contribution into the free energy from the inter-replica overlapping. The value of the parameter η at which $G([\eta,\rho,q])$ takes its maximum, satisfies the equation

$$1=\frac{\eta(\eta+e^{-\beta\epsilon/2})}{\rho\delta^3}+\frac{2\delta^6\mathcal{M}[q]}{(\eta+e^{-\beta\epsilon/2})^4}. \tag{8}$$

The equilibrium density of monomers is obtained by minimizing the free energy with respect to ρ, with constraint $\rho V=N$,

$$0=\frac{\partial}{\partial\rho}\frac{\beta\mathcal{F}(\rho,[q])}{Nn}=2\pi l_B f^2 r_D^2-\frac{\eta^2}{2\delta^3\rho^2}. \tag{9}$$

From Eqs. (8) and (9), we find

$$\frac{2\delta^6\mathcal{M}[q]}{(e^{-\beta\epsilon/2}+\eta)^4}=1-\frac{\eta e^{-\beta\epsilon/2}}{\rho\delta^3}-4\pi\rho l_B f^2 r_D^2. \tag{10}$$

In Eq. (5) the free energy $\mathcal{F}$ is expressed as a functional of the order parameter q. By minimizing $\mathcal{F}$ with respect to q for some n we determine the overlaps between the folds of different pure states (replicas) [7,8,21]. In our case the pure states are characterized by a set of spatial coordinates $\{\mathbf{r}_i^a\}$. In

the mean-field approximation the 2-replicas overlap parameter reads [21,22]

$$q_{ab}(\mathbf{x},\mathbf{y}) \simeq \frac{\rho}{R^3}\phi\left(\frac{\mathbf{x}-\mathbf{y}}{R}\right), \tag{11}$$

where R is the replica overlap scale associated with the difference between the positions $\mathbf{x}$ and $\mathbf{y}$ of the monomers in the replicas a and b, respectively.

The free energy can be minimized over the family of functions ϕ defined by Eq. (11). Like in the case of random heteropolymers without saturated interactions [23], we obtain $\mathcal{F}\sim -A_1/R^2+A_2/R^3$, where the contribution $-1/R^2$ is due to the entropy loss for the chain confined in a tube with diameter R [24], the term $\mathcal{M}[q]$ gives the contribution which scales as $1/R^3$, and A_1 and A_2 are some positive numbers. The free energy has two minima: at $R\simeq\delta$ and at $R=\infty$. The first one corresponds to two replicas, which coincide at the microscopic level δ, while the second one describes replicas which do not overlap at all. It is assumed that here we can use the one-step replica breaking scheme [23,24] according to which the n replicas are divided into groups. Each group consists of x replicas with identical folds, and the replicas belonging to different groups have completely different folds. Then the free energy can be estimated from the following equation [21,24]:

$$\frac{\beta\mathcal{F}(\rho,[q])}{Nn} \equiv \beta f(\rho,x) \simeq (1-1/x)S - g(\eta,\rho,x), \tag{12}$$

where $S=3\ln(a/\delta)$ is the entropy loss per monomer due to the chain freezing, $g(\eta,\rho,x)\equiv G([\eta,\rho,q])/(Nn)$, and $\mathcal{M}[q]\simeq\rho\delta^{-3}(x-1)$.

The free energy minimum over the parameter x are found from the equation

$$x^2 = \frac{2S(\eta+e^{-\beta\epsilon/2})^4}{\rho\delta^3}. \tag{13}$$

There exists some characteristic, so-called "freezing" temperature above which ($T\geqslant T_{\mathrm{fr}}$) the replicas are symmetric (all conformations are different), and hence, $x=1$. In this region Eqs. (8), (10), and (13) are reduced to

$$1 = \frac{\eta(\eta+z)}{\sigma},$$

$$\omega\sigma = 1-\frac{\eta z}{\sigma},$$

$$1 = \frac{2S}{\sigma}(\eta+z)^4, \tag{14}$$

where $z=\exp(-\beta\epsilon/2)$, $\sigma=\rho\delta^3\leqslant 1$ is the volume fraction of the RNA base pairs and $\omega=4\pi l_B f^2 r_D^2\delta^{-3}$. By solving this set of equations one finds the equilibrium values of η, σ, and z which minimize the free energy.

IV. FREEZING AND MELTING TRANSITIONS

Below the freezing temperature the configurational space of the molecule is drastically reduced. Instead of exponentially large numbers of conformations, now the state of the molecule is dominated by only a few low-energy conformations. The molecule is found in the frozen state [23,25]. This phase is defined by the condition $x\leqslant 1$. In the replica symmetric phase the number of available conformations is $\sim e^N$, and $x=1$. In order to estimate the freezing temperature T_{fr} we first calculate the volume fraction σ of base pairs.

From the first two equations of the set (14), we get easily

$$\sigma = \eta\omega^{-1/2},$$

$$\eta = \omega^{-1/2} - z, \tag{15}$$

which imply

$$\sigma = \omega^{-1} - z\omega^{-1/2}. \tag{16}$$

From the last equation of the set (14) and Eqs. (15) and (16), we obtain the following expression for the freezing temperature:

$$z_{\mathrm{fr}} = \exp[-\epsilon/(2T_{\mathrm{fr}})] = \frac{1}{\sqrt{\omega}}\left(1-\frac{2S}{\omega}\right). \tag{17}$$

It is seen from Eq. (17) that the freezing temperature T_{fr} decreases with growing entropy. For a given value of entropy S, the freezing temperature takes its maximum at $\omega=6S$.

Below the freezing transition the RNA-like macromolecule exhibits a glassy behavior (just like the random heteropolymers without saturated interactions) which is characterized by slow relaxation [26] and a few dominant folds. This might be the reason for the experimentally observed "slow and inaccurate" folding of large RNA [4,5] which was understood as misfolding.

A convenient quantitative order parameter for describing the stability of the RNA secondary structure is the degree of helicity which is defined as the average fraction of the hydrogen-bonded base pairs and can be represented as

$$\theta = \left\langle \sum_{i<j} M_{ij}C_{ij} \right\rangle \Big/ (N/2). \tag{18}$$

The average is calculated through the following transformations:

$$\left\langle \sum_{i<j} M_{ij}C_{ij} \right\rangle = \frac{\partial}{\partial(\beta\epsilon)} \ln Z. \tag{19}$$

Using the known relation $\ln Z=-\beta\mathcal{F}$ and Eq. (12), we obtain

$$\left\langle \sum_{i<j} M_{ij}C_{ij} \right\rangle = N\frac{\partial g}{\partial(\beta\epsilon)}. \tag{20}$$

Since in the replica symmetric phase $\mathcal{M}[q]=0$ [in the view of $\mathcal{M}[q]\sim(x-1)$ and $x=1$], then we can write

$$g = -\frac{\eta^2}{2\sigma} + \ln[1+\eta\exp(\beta\epsilon/2)]. \tag{21}$$

Combining (15) with (21), we obtain

$$\theta = 1 - z\sqrt{\omega}/2. \tag{22}$$

Another important quantity characterizing the stability of the secondary structure is the so-called "melting tempera-

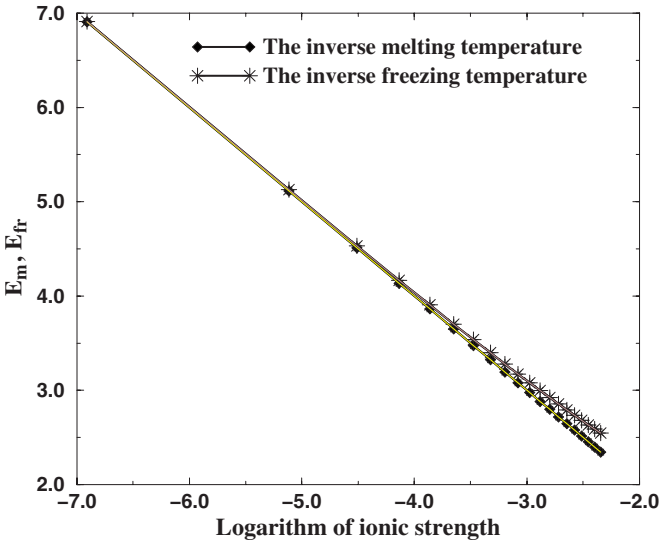

FIG. 3. (Color online) The dependencies of the inverse freezing temperature $E_{\mathrm{fr}}=\epsilon/T_{\mathrm{fr}}$ (asterisks) and the inverse melting temperature $E_m=\epsilon/T_m$ (diamonds) on logarithm of the reduced ionic strength $\mu=\ln(I\delta^3)$. The range of values $\mu\sim-3;-1$ corresponds to the range $10^{-2}-1$ M for the ionic strength I.

ture" T_m, defined as the point, where the bounded and nonbounded states of base pairs have an equal probability, and $\theta=1/2$. By substituting this value in (22), we find

$$T_m=\frac{\epsilon}{\ln\omega}. \tag{23}$$

Combining with Eq. (17), we obtain $T_m>T_{\mathrm{fr}}$. The dependencies of T_{fr} and T_m on salt concentration are shown in Fig. 3.

V. DISCUSSION

In this work we assume a quenched disorder for the RNA sequence. A rightful question arises about the general differences between this model and the model with annealed sequence. Note that at temperatures $T>T_{\mathrm{fr}}$ the replica parameter $x=1$. This means that the free energies for quenched and annealed systems are the same. At the end of the preceding section we obtained that the melting temperature for this model is higher than the freezing temperature: $T_m>T_{\mathrm{fr}}$. With this we can speculate that annealed disorder and quenched disorder give the same melting transition and we can expect that the logarithmic dependence of the melting temperature on the ionic strength (Fig. 3) holds also for the annealed system. On the other hand, we mentioned that quenched disorder is a key condition for attaining a freezing transition.

Now we provide some approximate quantitative estimations which follow from the described model. It is easy to see that the value of the electrostatic parameter ω scales with δ as $\omega=f^2/I\delta^3$. The fractional charge f is supposed to be non-neutralized and $f=1$. It was mentioned above that the value $\delta\sim10$ Å can be used. Hence, under the normal conditions ($T\approx300$ K and $\varepsilon\approx80$) and for experimentally used salt ionic strength interval $I\sim(0.01-1)$ M, one obtains $1/\omega\sim(10^{-3}-10^{-1})$.

The σ parameter has the meaning of the volume fraction of the RNA base pairs. It is obvious from Eq. (16) that σ takes the physically significant positive values only when $1/\omega>z^2$. This expression is easily transformed as $I>z^2/\delta^3$. Taking into account that $\beta\epsilon\approx3$ [15,16], one obtains that $z^2/\delta^3\approx5\times10^{-5}$ Å^{-3} and the necessary condition for the $\sigma>0$ is $I>I^*$ ($I^*\approx0.1$ M). At the lower ionic strengths it would be invalid to use this model for the mean-field approximation of σ.

The freezing temperature T_{fr} takes positive values only if $\frac{1}{\omega}<\frac{1}{2S}$. Taking into account the estimation $a\approx16$ Å [18–20] and $\delta\approx10$ Å we obtain $S\approx3\ln1.6$. Thus, $I<1/(2S\delta^3)$ only if $I<I^{**}$, where $I^{**}\approx3.5\times10^{-4}$ Å$^{-3}\approx0.6$ M.

The estimations derived above show clearly that the tertiary structure formation is possible only for the ionic strength in the interval $I^*<I<I^{**}$, and this is in a good agreement with the experimental data [9].

The important point of the proposed theory is the "strong screening approximation" [see Eq. (3)], where the screened electrostatic potential is substituted by the short-range pseudopotential $v_C\delta(\mathbf{r})$. As it was shown in [12], this approximation is valid for the concentration of the added salt higher than the critical value which satisfies the equation

$$\frac{\delta}{r_D}\sim\left(\frac{l_B}{\delta}\right)^{1/3}. \tag{24}$$

At room temperatures ($T\approx300$ K) the Bjerrum length is $l_B\approx0.7$ nm and the Debye screening length is $r_D\sim1$ nm. Thus, the "strong screening" approximation is valid at ionic strengths $I>0.13$ M and our results are valid in the interval of ionic strength $I^*<I<I^{**}$.

If the solvent contains only monovalent cations, e.g., Na^+ or NH_4^+, then the electrostatic parameter ω can be estimated as $\omega\propto1/c$, where c is proportional to the concentration of the counterions. This means that the inverse melting temperature $1/T_m$ decreases with the logarithm of monovalent counterions concentration, which has been observed experimentally [9]. In particular, it has been shown in [9] that in the presence of NH_4^+ counterions the inverse melting temperature $1/T_m$ decreases proportionally with $\ln[NH_4^+]$, where $[NH_4^+]$ is the molar concentration of the NH_4^+ ions. Thus the proposed theory is in qualitative agreement with the experimental data.

As mentioned above, below the freezing transition the RNA-like macromolecule exhibits a glassy behavior characterized by a few dominate folds just like the conventional random heteropolymers. The experiments on the folding of the large RNA molecules such as *Tetrahymena* ribozyme [27], show that the real folding rates are considerably slower than it follows from the hierarchical mechanism. It is believed that in large RNA molecules conformational rearrangements are coupled to strong interactions between the parts of the RNA molecule which can create kinetic traps and slow down the folding. This glassylike behavior is consistent with the replica symmetry breaking and frozen phase formation, predicted in our model. The possibility of the spin-glass-like behavior of the RNA molecule was, particularly predicted numerically in [28,29]. However these authors

used the recursive model which does not consider pseudoknots and describes only the secondary structure formation.

The more detailed comparison is difficult because our model predicts spin-glass-like behavior for the tertiary structure. However, the RNA secondary structure formation is closely related to interactions between the base pairs, which stand apart along the macromolecule and the tertiary structure freezing may be a master to secondary structure folding (see, e.g., [30]). In the framework of present model by formation of the stable base pair (i,j) as an element of the secondary structure, the relative location of the i and j bases is fixed. Then the spatial structure freezing is accompanied by the secondary structure freezing and the different low-energy structures in the frozen phase correspond to the stable tertiary and secondary structures.

The frozen (glassy) phase formation is typical for the random heteropolymers at low temperatures [21,23]. The characteristic feature of the RNA folding is the role of the balance between the chain entropy and electrostatic repulsion. As it is easy to see from Eqs. (16), (17), (22), and (23), the electrostatic excluded volume v_C and the effective volume of the base pair $\sim\delta^3$ appear in the thermodynamic characteristics only as the ratio $\omega\sim v_C/\delta^3$ and, consequently have a competitive influence on the thermodynamics of the model under consideration. It is obvious from Eq. (17) that RNA can undergo a freezing transition only if $\omega>2S$, i.e., only at sufficiently strong electrostatic repulsions. Thus unlike proteins the repulsion between the parts of the RNA plays a constructive role in the structure stabilization. Apparently, the reason of the qualitative difference between the role, playing by the repulsive interactions in the RNA and protein molecules is the saturated nature of the hydrogen bonds in RNA, while the protein structures are formed basically by the ordinary van der Waals interactions. The possible reason seems to be the following. In the absence of the electrostatic repulsion a local collapsed structure may occur in which three and more bases are colliding. At the same time, only two complementary bases can form a stable pair. The contacts between more than two bases can only increase the scale of fluctuations of the three-dimensional (3D) structure and create an additional overlapping between different stable folds of the RNA molecule. Large scale fluctuations themselves indicate instability. Thus, the electrostatic repulsions facilitate in formation of the stable specific 3D structures by preventing large fluctuations.

It follows from the conditions $\sigma_{\rm fr}=2S/\omega^2\ll 1$ that at the freezing transition point the RNA molecule is not compact and attains a loose, low-density structure. This can be understood as follows. At the temperatures $T>T_{\rm fr}$ the RNA macromolecule does not have any stable tertiary structure and can be presented as the set of double-strand stems, loops, and other elements of the secondary structure fluctuating in the space. The thickness of the double-strand stem is $\sim\delta$, but the spatial arrangement of the stems is governed by the electrostatic repulsion $\sim v_C$. That is why $\sigma_{\rm fr}\propto(\delta^3/v_C)^2$.

We can summarize that the proposed model of the large RNA macromolecule describes correctly at least qualitatively some experimentally observed features of the RNA folding, such as the misfolded states which are slowing down the long RNA's folding and the dependence of the melting temperature on the monovalent salt concentration.

ACKNOWLEDGMENTS

This work was supported by the National Science Council in Taiwan under Grant No. NSC 95-2112-M 001-008, National Center for Theoretical Sciences in Taiwan, and Academia Sinica (Taiwan) under Grant Nos. AS-92-TP-A09 and AS-95-TP-A07.

APPENDIX A: DERIVATION OF THE FREE ENERGY

Here we give a detail derivation of the equation for the free energy of the Hamiltonian $H\{\mathbf{r},\hat{C},\hat{M}\}$ defined in Eq. (1). The conformational partition function of a given sequence has the following form:

$$Z(\hat{M})=\sum_{\hat{C}}^{\rm sat}\int D\mathbf{r}\,\exp[-\beta H\{\mathbf{r},\hat{C},\hat{M}\}], \tag{A1}$$

where $\beta=1/T$, $D\mathbf{r}=\prod_{i=1}^{N}d\mathbf{r}_i$ and

$$\sum_{\hat{C}}^{\rm sat}=\prod_{i<j=2}^{N}\sum_{C_{ij}\in\{0,1\}}\prod_{k=1}^{N}\Theta\left(\alpha+1-\sum_{l=1}^{N}M_{kl}C_{kl}\right). \tag{A2}$$

Here $\Theta(x)$ is the Heaviside step function, $0<\alpha<1$ is some auxiliary number which is not present in the final result. Its role is to take into account the restriction $\Sigma_{l=1}^{N}M_{kl}C_{kl}\leqslant 1$ (for any k) which reflects the saturated nature of the complementary pairing. In further calculations it is convenient to use the integral representation of the Heaviside step function

$$\Theta(x)=\frac{1}{2\pi\imath}\int_{\gamma-\imath\infty}^{\gamma+\imath\infty}ds s^{-1}\exp(sx), \tag{A3}$$

for arbitrary $\gamma>0$. After taking sum over all arrangements of the secondary structure matrix $\hat{C}$ the partition function is transformed into

$$Z(\hat{M})=\int Ds\exp\left((\alpha+1)\sum_{k=1}^{N}s_k\right)\int D\mathbf{r}\exp[-\beta(H_{\rm el}+H_{\rm rep})]\times\exp\left[\sum_{i<j=2}^{N}M_{ij}\ln\left(\frac{1}{2}+\frac{1}{2}\exp(-s_i-s_j)\Phi(\mathbf{r}_i-\mathbf{r}_j)\right)\right], \tag{A4}$$

where $\int Ds=\prod_{l=1}^{N}\frac{1}{2\pi\imath}\int_{\gamma-\imath\infty}^{\gamma+\imath\infty}ds_l s_l^{-1}$ and $\Phi(\mathbf{r})=\exp(\beta\epsilon-3\mathbf{r}^2/2\delta^2)$. Since the value $\gamma>0$ is arbitrary then one can choose $\gamma\gg 1$ which corresponds to $|\exp(-s_i)|\ll 1$. Then the logarithm in Eq. (A4) can be expanded up to the fourth order of $\exp(-s_i)$,

$$Z(\hat{M}) = \int Ds \exp\left((\alpha+1)\sum_{k=1}^{N} s_k\right) \int D\mathbf{r} \exp[-\beta(H_{\rm el}+H_{\rm rep})] \times \exp\left[\sum_{i<j=2}^{N} M_{ij}\left(-\ln 2 + \exp(-s_i-s_j)\Phi(\mathbf{r}_i-\mathbf{r}_j) - \frac{1}{2}\exp[-2(s_i+s_j)]\Phi^2(\mathbf{r}_i-\mathbf{r}_j)\right)\right]. \quad \text{(A5)}$$

According to the definition of the model, the elements of the matrix $\hat{M}$ are independent random variables taking values 0 and 1 with equal probabilities. To average the free energy over quenched disorder of the entries of $\hat{M}$, we use a replica approach [7,8], which requires averaging the partition function of n replicas of the system over the distribution of the $\hat{M}$ entries. The averaged n-replica partition function reads

$$\langle Z^n(\hat{M})\rangle_{\rm av} = \int D\{s\}\exp\left((\alpha+1)\sum_{a=1}^{n}\sum_{k=1}^{N} s_k^a\right) \times \int D\mathbf{r}^a \exp\left(-\beta\sum_{a=1}^{n}[H_{\rm el}(\mathbf{r}^a)+H_{\rm rep}(\mathbf{r}^a)]\right) \times \left\langle \exp\left[\sum_{i<j=2}^{N} M_{ij}\sum_{a=1}^{n}\left(-\ln 2+\exp(-s_i^a-s_j^a) \times\Phi(\cdot) - \frac{1}{2}\exp[-2(s_i^a+s_j^a)]\Phi^2(\cdot)\right)\right]\right\rangle_{\rm av}, \quad \text{(A6)}$$

where a is the replica index, $\int D\{s\}=\prod_{a=1}^{n}\int Ds^a$, $D\{r\}=\prod_{a=1}^{n}\int Dr^a$, and $\Phi(\cdot)=\Phi(\mathbf{r}_i^a-\mathbf{r}_j^a)$.

After carrying out an averaging over disorder and expanding the effective Hamiltonian to the fourth order of the variables $\exp(-s_i^a)$ we obtain

$$\langle Z^n(\hat{M})\rangle_{\rm av} \simeq \int D\{s\}\exp\left((\alpha+1)\sum_{a=1}^{n}\sum_{k=1}^{N} s_k^a\right) \times \int D\mathbf{r}^a \exp\left(-\beta\sum_{a=1}^{n}[H_{\rm el}(\mathbf{r}^a)+H_{\rm rep}(\mathbf{r}^a)]\right) \times \exp\left[\sum_{i<j=2}^{N}\left(d_n\sum_{a=1}^{n}\exp(-s_i^a-s_j^a)\Phi(\mathbf{r}_\mathbf{i}^a-\mathbf{r}_\mathbf{j}^a) - \frac{a_n}{2}\sum_{a=1}^{n}\exp[-2(s_i^a+s_j^a)]\Phi^2(\mathbf{r}_\mathbf{i}^a-\mathbf{r}_\mathbf{j}^a) + f_n\sum_{a,b=1}^{n}\Phi(\mathbf{r}_\mathbf{i}^a-\mathbf{r}_\mathbf{j}^a)\Phi(\mathbf{r}_\mathbf{i}^b-\mathbf{r}_\mathbf{j}^b) \times\exp(-s_i^a-s_j^a-s_i^b-s_j^b)\right)\right], \quad \text{(A7)}$$

where $d_n=2^{-n}/(1+2^{-n})$ and $f_n=2^{-(n+1)}[1-2^{-n}/(1+2^{-n})]/(1+2^{-n})$. The last expression may be rewritten in terms of some order parameters $\hat{Q}$, $\hat{n}$ and $\hat{m}$ arising from the replica technique

$$\langle Z^n(\hat{M})\rangle_{\rm av} \simeq \int D\{s\}\exp(W_0\{s\}) \times \int D\{\mathbf{r}\}\exp\left(-\beta\sum_{a=1}^{n}[H_{\rm el}(\mathbf{r}^a)+H_{\rm rep}(\mathbf{r}^a)]\right) \times\exp\left(\mathcal{A}\sum_{a=1}^{n}\int d\mathbf{x}\hat{m}_a^2(\mathbf{x}) - \mathcal{B}\sum_{a=1}^{n}\int d\mathbf{x}\hat{n}_a^2(\mathbf{x}) + \mathcal{C}\sum_{a,b=1}^{n}\int d\mathbf{x}d\mathbf{x}'\hat{Q}_{ab}^2(\mathbf{x},\mathbf{x}')\right), \quad \text{(A8)}$$

where $\hat{Q}_{ab}(\mathbf{x},\mathbf{x}')=\Sigma_i\delta(\mathbf{x}-\mathbf{r}_\mathbf{i}^a)\delta(\mathbf{x}'-\mathbf{r}_\mathbf{i}^b)e^{-s_i^a-s_i^b}$, $\hat{m}_a(\mathbf{x})=\Sigma_i\delta(\mathbf{x}-\mathbf{r}_\mathbf{i}^a)e^{-s_i^a}$ and $\hat{n}_a(\mathbf{x})=\Sigma_i\delta(\mathbf{x}-\mathbf{r}_\mathbf{i}^a)e^{-2s_i^a}$. The $\hat{m}_a$ and $\hat{n}_a$ are densitylike quantities and $\hat{Q}_{ab}$ is the replica-overlap parameter. Here we have also introduced the notations

$$\mathcal{A} = d_n\exp(\beta\epsilon)\left(\frac{2\pi\delta^2}{3}\right)^{3/2},$$

$$\mathcal{B} = \frac{d_n\exp(2\beta\epsilon)}{2}\left(\frac{\pi\delta^2}{3}\right)^{3/2},$$

$$\mathcal{C} = f_n\exp(2\beta\epsilon)\left(\frac{2\pi\delta^2}{3}\right)^{3},$$

$$W_0\{s\} = (\alpha+1)\sum_{a=1}^{n}\sum_{i=1}^{N} s_i^a - \frac{d_n e^{\beta\epsilon}}{2}\sum_{a,i}\exp(-2s_i^a) + \frac{d_n e^{2\beta\epsilon}}{4}\sum_{a,i}\exp(-4s_i^a) - \frac{f_n e^{2\beta\epsilon}}{2} \times\sum_{a,b}\sum_{i}\exp[-2(s_i^a+s_i^b)]. \quad \text{(A9)}$$

It is easy to linearize the expression (A8) by using the Hubbard-Stratonovich transformation. Then one obtains

$$\langle Z^n(\hat{M})\rangle_{\rm av} \simeq \int D\{\varphi\}D\{\eta\}D\{\chi\}\exp\left(-\frac{1}{2\mathcal{A}}\sum_{a}\int d\mathbf{x}\,\eta_a^2(\mathbf{x}) - \frac{1}{2\mathcal{B}}\sum_{a,b}\int d\mathbf{x}\chi_a^2(\mathbf{x}) - \frac{1}{2\mathcal{C}}\sum_{a}\int d\mathbf{x}d\mathbf{x}'\varphi_{ab}(\mathbf{x},\mathbf{x}')^2\right) \times\int D\{\mathbf{r}\}\exp\left(-\beta\sum_{a=1}^{n}[H_{\rm el}(\mathbf{r}^a) + H_{\rm rep}(\mathbf{r}^a)]\right)\int D\{s\}\exp(W_0\{s\})$$

$$\times \exp\left(\sum_{a=1}^{n}\int d\mathbf{x}\hat{m}_a(\mathbf{x})\,\eta_a(\mathbf{x}) + \imath\sum_{a=1}^{n}\int d\mathbf{x}\hat{n}_a(\mathbf{x})\chi_a(\mathbf{x}) + \sum_{a,b=1}^{n}\int d\mathbf{x}d\mathbf{x}'\hat{Q}_{ab}(\mathbf{x},\mathbf{x}')\varphi_{ab}(\mathbf{x},\mathbf{x}')\right). \quad \text{(A10)}$$

This transformation introduces the auxiliary fields φ_{ab}, η_a, and χ_a. These fields satisfy the saddle-point equations [22], i.e.,

$$\eta_a(\mathbf{x}) = \langle \hat{m}_a(\mathbf{x})\rangle,$$

$$\chi_a(\mathbf{x}) = \langle \hat{n}_a(\mathbf{x})\rangle,$$

$$\varphi_{ab}(\mathbf{x},\mathbf{x}') = \langle \hat{Q}_{ab}(\mathbf{x},\mathbf{x}')\rangle, \quad \text{(A11)}$$

where $\langle\cdots\rangle$ means thermal averaging. Now it is possible to factorized integration over the variables $\{s\}$ by the individual monomers. After subsequent expansion over the small parameters $\exp(-s_i^a)$ we have

$$\int D\{s\}\exp(W_0\{s\})\exp\left(\sum_{a=1}^{n}\int d\mathbf{x}\hat{m}_a(\mathbf{x})\,\eta_a(\mathbf{x}) + \imath\sum_{a=1}^{n}\int d\mathbf{x}\hat{n}_a(\mathbf{x})\chi_a(\mathbf{x}) + \sum_{a,b=1}^{n}\int d\mathbf{x}d\mathbf{x}'\hat{Q}_{ab}(\mathbf{x},\mathbf{x}')\varphi_{ab}(\mathbf{x},\mathbf{x}')\right)$$
$$\simeq \prod_{i=1}^{N}\int D\{p\}\exp\left(-\frac{1}{4}\sum_{a,b}\varphi_{ab}(\mathbf{r}_i^a,\mathbf{r}_i^b)^{-1}p_a p_b\right) \times \prod_{a=1}^{n}[1+\eta_a(\mathbf{r}_i^a)+p_a]. \quad \text{(A12)}$$

One can see that

$$\int Dp_a \exp\left(-\frac{1}{4}\sum_{ab}\varphi_{ab}(\mathbf{r}_i^a,\mathbf{r}_i^b)^{-1}p_a p_b\right)\prod_{a=1}^{n}[1+\eta_a(\mathbf{r}_i^a)+p_a] = \sum_{k=0}^{[n/2]}\sum_{\{a,b,\ldots\}}(2\varphi)^{(k)}(1+\eta)^{(n-2k)}, \quad \text{(A13)}$$

where $[\cdots]$ in the upper limit of the summation means the next lowest integer and the summation $\Sigma_{\{a,b,\ldots\}}$ is carried out over all possible different pairs of replicas, $(\varphi)^{(k)}$ is the product of k corresponding terms $\varphi_{ab}(\mathbf{r}_\mathbf{i}^a,\mathbf{r}_\mathbf{i}^b)$, and $(1+\eta)^{(n-2k)}$ has a similar meaning.

Using these notations the n-replica partition function can be rewritten as

$$\langle Z^n(\hat{M})\rangle_{\mathrm{av}} \simeq \int D\{\mathbf{r}\}\exp\left(-\beta\sum_{a=1}^{n}[H_{\mathrm{el}}(\mathbf{r}^a)+H_{\mathrm{rep}}(\mathbf{r}^a)]\right) \times \int D\{\varphi\}D\{\eta\}\exp\left(-\frac{1}{2\mathcal{A}}\sum_a\int d\mathbf{x}\,\eta_a^2(\mathbf{x}) - \frac{1}{2\mathcal{C}}\times\sum_{a,b}\int d\mathbf{x}d\mathbf{x}'\varphi_{ab}^2(\mathbf{x},\mathbf{x}') + \int d\mathbf{X}\hat{q}_{1\ldots n}(\mathbf{X})L(\mathbf{X};\{\varphi,\eta\})\right), \quad \text{(A14)}$$

where $L(\mathbf{X};\{\varphi,\eta\}) = \ln\Sigma_{k=0}^{[n/2]}\Sigma_{\{a,b,\ldots\}}(2\varphi)^{(k)}(1+\eta)^{(n-2k)}$, $\mathbf{X}=(\mathbf{x_1},\ldots,\mathbf{x_n})$ and $\hat{q}_{1\cdots n}(\mathbf{X})=\Sigma_i\Pi_a\delta(\mathbf{x}^a-\mathbf{r}_\mathbf{i}^a)$ is the n-replica overlap parameter.

In the mean-field approximation the μ-replicas overlap parameter may be presented as [21,22]

$$\langle\hat{q}_{abc\ldots}(\mathbf{x}_a,\mathbf{x}_b,\mathbf{x}_c,\ldots)\rangle \simeq \frac{\rho}{R^{3(\mu-1)}}\phi\left(\frac{\mathbf{x}_a-\mathbf{x}_b}{R},\frac{\mathbf{x}_a-\mathbf{x}_b}{R},\ldots\right), \quad \text{(A15)}$$

where R is the replica overlap scale associated with the difference between the positions $\mathbf{x}_a,\mathbf{x}_b,\ldots$ of the monomers in the replicas $a,b,\ldots$, respectively. Here ρ is the density of monomers, which is supposed to be constant over the volume occupied by the system. The replica overlap parameter is subject to the constraint $\int d\mathbf{x}_a d\mathbf{x}_b d\mathbf{x}_c\cdots\langle\hat{q}_{abc\ldots}(\mathbf{x}_a,\mathbf{x}_b,\mathbf{x}_c,\ldots)\rangle=\rho$. It is easy to show from equations (A11) and (A15) that in the mean-field approximation $|\varphi^{(1)}/(1+\eta)^{(2)}|\simeq 1/R^3$. Taking into account this estimation we expand the $L(\mathbf{X};\{\varphi,\eta\})$ function as

$$L(\mathbf{X};\{\varphi,\eta\}) \simeq \sum_a \ln[1+\eta_a(\mathbf{x}_a)] + 2\sum_{a<b}\frac{\varphi_{ab}(\mathbf{x}_a,\mathbf{x}_b)}{[1+\eta_a(\mathbf{x}_a)][1+\eta_b(\mathbf{x}_b)]}. \quad \text{(A16)}$$

The higher-order terms do not influence the three-dimensional structure freezing. The contribution to the free energy from these terms scales as $\geqslant 3$ powers of $1/R$. From equations (A14) and (A16) the mean-field approximation of the n-replica partition function can be obtained

$$\langle Z^n(\hat{M})\rangle_{\mathrm{av}} \simeq \exp\left(S[q]-\beta\sum_a H_{\mathrm{rep}}(\rho_a)\right)\exp(-\beta F_0[\rho,q]), \quad \text{(A17)}$$

where $\rho_a=\rho(a=1,\cdot,n)$ is the replica-symmetric equilibrium density of monomers, $q_{ab}(\mathbf{x},\mathbf{y})=\langle\Sigma_i\delta(\mathbf{x}-\mathbf{r}_\mathbf{i}^a)\delta(\mathbf{y}-\mathbf{r}_\mathbf{i}^b)\rangle$ is the equilibrium two-replica overlap parameter, $S[q]$ is the entropy of the n ideal polymeric chains under restriction $\Sigma_i\delta(\mathbf{x}-\mathbf{r}_\mathbf{i}^a)\delta(\mathbf{y}-\mathbf{r}_\mathbf{i}^b)=q_{ab}(\mathbf{x},\mathbf{y})$, and

$$\exp(-\beta\mathcal{F}_0[\rho,q]) \simeq \int D\{\eta\}\exp\left(-\frac{1}{2\delta^3}\sum_a \int d\mathbf{x}\,\eta_a^2(\mathbf{x}) + \sum_a \int d\mathbf{x}\rho_a(\mathbf{x})\ln[1+\eta_a(\mathbf{x})e^{\beta\epsilon/2}] + \frac{1}{2\delta^3} \times \sum_{a\neq b}\int d\mathbf{x}d\mathbf{x}'\frac{q_{ab}^2(\mathbf{x},\mathbf{x}')}{(e^{-\beta\epsilon/2}+\eta_a)^2(e^{-\beta\epsilon/2}+\eta_b)^2}\right). \tag{A18}$$

The $\eta_a(\mathbf{x})$ fields are replica symmetric. Besides, let us assume that the system is spatially homogeneous $\eta_a(\mathbf{x})=\eta$. Thus we obtain for the n-replica reduced free energy

$$\beta\mathcal{F}(\rho,[q]) = -S[q] + n\beta H_{\mathrm{rep}}(\rho) - \max_{\eta} G([\eta,\rho,q]), \tag{A19}$$

where

$$G(\eta,[\rho,q]) = -\frac{Nn\eta^2}{2\rho\delta^3} + Nn\ln(1+\eta e^{\beta\epsilon/2}) + \frac{Nn\delta^6}{2(e^{-\beta\epsilon/2}+\eta)^4}\mathcal{M}[q]. \tag{A20}$$

Here $\mathcal{M}[q]=\frac{1}{Nn}\Sigma_{a\neq b}\int d\mathbf{x}d\mathbf{x}'q_{ab}^2(\mathbf{x},\mathbf{x}')$ describes the contribution into the free energy from the inter-replica overlapping. Equations (A19) and (A20) are Eqs. (5) and (7), respectively, in the main paper.

APPENDIX B: RNA MODEL WITH FOUR-LETTER SEQUENCES

Let us consider the free energy of the RNA molecule with four-letter sequence. It is easy to see that the difference between two- and four-letter sequences appears after the averaging over disorder of Eq. (A6). In the case of the four-letter sequence the probability distribution for each entry M_{ij} has the form $p(M_{ij})=q\delta_{Kr}(M_{ij}-1)+(1-q)\delta_{Kr}(M_{ij})$, where $q=1/4$. The two-letter sequence corresponds to $q=1/2$. After averaging over disorder, Eq. (A6) transforms into

$$\langle Z^n(\hat{M})\rangle_{\mathrm{av}} \simeq \int D\{s\}\exp\left((\alpha+1)\sum_{a=1}^{n}\sum_{k=1}^{N}s_k^a\right) \times \int D\mathbf{r}^a\exp\left(-\beta\sum_{a=1}^{n}[H_{\mathrm{el}}(\mathbf{r}^a)+H_{\mathrm{rep}}(\mathbf{r}^a)]\right) \times \exp\left(\sum_{i<j=2}^{N}\ln\left\{1+c\exp\left[\sum_{a=1}^{n}\left(-\ln 2 + \exp(-s_i^a-s_j^a)\Phi(\cdot) - \frac{1}{2}\exp[-2(s_i^a+s_j^a)]\Phi^2(\cdot)\right)\right]\right\}\right), \tag{B1}$$

where $c=\frac{q}{1-q}$. After subsequent expansion of the logarithm in the last equation up to the fourth order of the variables $\exp(-s_i^a)$ we obtain

$$\langle Z^n(\hat{M})\rangle_{\mathrm{av}} \simeq \int D\{s\}\exp\left((\alpha+1)\sum_{a=1}^{n}\sum_{k=1}^{N}s_k^a\right) \times \int D\mathbf{r}^a\exp\left(-\beta\sum_{a=1}^{n}[H_{\mathrm{el}}(\mathbf{r}^a)+H_{\mathrm{rep}}(\mathbf{r}^a)]\right) \times \exp\left[\sum_{i<j=2}^{N}\left(\tilde{d}_n\sum_{a=1}^{n}\exp(-s_i^a-s_j^a)\Phi(\mathbf{r_i}^a-\mathbf{r_j}^a) - \frac{\tilde{d}_n}{2}\sum_{a=1}^{n}\exp[-2(s_i^a+s_j^a)]\Phi^2(\mathbf{r_i}^a-\mathbf{r_j}^a) + \tilde{f}_n\sum_{a,b=1}^{n}\Phi(\mathbf{r_i}^a-\mathbf{r_j}^a)\Phi(\mathbf{r_i}^b-\mathbf{r_j}^b) \times\exp(-s_i^a-s_j^a-s_i^b-s_j^b)\right)\right], \tag{B2}$$

where $\tilde{d}_n=c2^{-n}/(1+c2^{-n})$ and $\tilde{f}_n=c2^{-(n+1)}[1-c2^{-n}/(1+c2^{-n})]/(1+c2^{-n})$. Thus, for the four-letter sequence the expansion over the order parameters (A8) has the same form with the coefficients $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ obtained by substitution $d_n\rightarrow\tilde{d}_n$ and $f_n\rightarrow\tilde{f}_n$. It is easy to see that the coefficients $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ have the same sign and order of magnitude for the two- and four-letter sequences and, thus the free energy exhibits qualitatively the same behavior.

[1] J. M. Burke *et al.*, in *Catalytic RNA*, edited by F. Eckshtein and David M. J. Lilley (Springer-Verlag, Berlin, 1996), p. 130.

[2] J. R. Morrow, in *Metal Ions in Biological Systems*, edited by A. Sigel and H. Sigel (Marcel Dekker, New York, 1996), Vol. 33, p. 561.

[3] R. Bundschuh and T. Hwa, Phys. Rev. Lett. **83**, 1479 (1999); Phys. Rev. E **65**, 031903 (2002); F. Krzakala, M. Mezard, and M. Müller, Europhys. Lett. **57**, 752 (2002); M. Müller, Phys. Rev. E **67**, 021914 (2003); V. Guttal and R. Bundschuh, Phys. Rev. Lett. **96**, 018105 (2006); M. Lassig and K. J. Wiese, *ibid.* **96**, 228101 (2006).

[4] D. K. Treiber and J. R. Williamson, Curr. Opin. Struct. Biol. **9**, 339 (1999).

[5] O. C. Uhlenbeck, RNA **1**, 4 (1995).

[6] I. Tinoco, Jr. and C. Bustamante, J. Mol. Biol. **293**, 271 (1999).

[7] M. Mezard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).
[8] K. Binder and A. P. Young, Rev. Mod. Phys. **58**, 801 (1986).
[9] R. Shiman and D. E. Draper, J. Mol. Biol. **302**, 79–91 (2000).
[10] S. Cao and S.-J. Chen, RNA **11**, 1884 (2005).
[11] A. R. Khokhlov and K. A. Khachaturian, Polymer **23**, 1742 (1982).
[12] P. G. Higgs and E. Raphael, J. Phys. I **1**, 1 (1991).
[13] M. Ulner and B. Jönsson, Macromolecules **29**, 6645 (1996).
[14] C.-Y. Shew and A. Yethiraj, J. Chem. Phys. **110**, 676 (1999).
[15] B. Maier, D. Bensimon, and V. Croquette, Proc. Natl. Acad. Sci. U.S.A. **97**, 1202 (2000).
[16] A. Montanari and M. Mezard, Phys. Rev. Lett. **86**, 2178 (2001).
[17] M. C. Murphy, I. Rasnik, W. Cheng, T. M. Lohman, and T. Ha, Biophys. J. **86**, 2530 (2004).
[18] M. N. Lambert, E. Vocker, S. Blumberg, S. Redemann, A. Gajrai, J.-C. Meiners, and N. G. Walter, Biophys. J. **90**, 3672 (2006).
[19] Y. Seol, G. M. Skinner, and K. Visscher, Phys. Rev. Lett. **93**, 118102 (2004).
[20] F. Vanzi, Y. Takagi, H. Shuman, B. S. Cooperman, and Y. E. Goldman, Biophys. J. **89**, 1909 (2005).
[21] C. D. Sfatos, A. M. Gutin, and E. I. Shakhnovich, Phys. Rev. E **48**, 465 (1993).
[22] G. Z. Archontis and E. I. Shakhnovich, Phys. Rev. E **49**, 3109 (1994).
[23] V. S. Pande, A. Y. Grosberg, and T. Tanaka, Rev. Mod. Phys. **72**, 259 (2000).
[24] E. I. Shakhnovich and A. M. Gutin, Biophys. Chem. **34**, 187 (1989).
[25] E. I. Shakhnovich and A. M. Gutin, Nature (London) **346**, 773 (1990).
[26] A similar example is the spin glass, in which the relaxation at low temperatures is slow power-law decay, see, e.g., C. Dasgupta, S.-K. Ma, and C.-K. Hu, Phys. Rev. B **20**, 3837 (1979) rather than the much faster exponential decay in the Ising model, see, e.g., C.-K. Hu, Phys. Rev. B **29**, 5103 (1984) and F.-G. Wang and C.-K. Hu, Phys. Rev. E **56**, 2310 (1997).
[27] D. W. Celander and T. R. Cech, Science **251**, 401 (1991).
[28] P. G. Higgs, Phys. Rev. Lett. **76**, 704 (1996).
[29] A. Pagnani, G. Parisi, and F. Ricci-Tersenghi, Phys. Rev. Lett. **84**, 2026 (2000).
[30] D. Thirumalai, Proc. Natl. Acad. Sci. U.S.A. **95**, 11506 (1998).