Dependence of RNA secondary structure on the energy model

Bernd Burghardt* and Alexander K. Hartmann†
Institut für Theoretische Physik, Universität Göttingen, Friedrich-Hund-Platz 1, D-37077 Göttingen, Germany
(Received 1 September 2004; revised manuscript received 23 November 2004; published 25 February 2005)

We analyze a microscopic RNA model, which includes two widely used models as limiting cases; namely, it contains terms for bond as well as for stacking energies. We numerically investigate possible changes in the qualitative and quantitative behavior while going from one model to the other; in particular, we test whether a transition occurs when continuously moving from one model to the other. For this we calculate various thermodynamic quantities, at both zero temperature and finite temperatures. All calculations can be done efficiently in polynomial time by a dynamic programming algorithm. We do not find a sign for the transition between the models, but the critical exponent ν of the correlation length, describing the phase transition in all models to an ordered low-temperature phase, seems to depend continuously on the model. Finally, we apply the ε-coupling method to study low-energy excitations. The exponent θ describing the energy scaling of the excitations seems to depend not much on the energy model.

DOI: 10.1103/PhysRevE.71.021913 PACS number(s): 87.15.Aa, 64.60.Fr

I. INTRODUCTION

RNA plays an important role in the biochemistry of all living systems [1,2]. Similar to DNA, it is a linear-chain molecule built from four types of bases: i.e., adenine (A), cytosine (C), guanine (G), and uracil (U). It does not only transmit pure genetic information, but, e.g., works as a catalyst. While for the former the primary structure—i.e., the sequence of bases—is relevant, for the latter kind of higher-order structures—i.e., secondary and tertiary structures—is relevant.

Like in the double helix of DNA, in RNA complementary bases can build hydrogen bonds between each other. Compared to DNA, where the bonds are built between two different strands, RNA builds bonds between bases of the same RNA strand. The information, which bases of the strand are paired, gives the secondary structure, and the spatial structure is called the tertiary structure. The tertiary structure is stabilized by a much weaker interaction than the secondary structure. This leads to a separation of energy scales between secondary and tertiary structures, and gives justification to neglect the latter [3]. Therefore we deal here with the secondary structure only.

One crucial point in calculating the secondary structure is the energy model used: On the one hand, if one aims to get minimum structures close to the experimentally observed one, one uses energy models that take into account structure elements [4–7]—e.g., hairpin loops. On the other hand, if one is interested in the qualitative behavior, one uses models as simple as possible that keep the general behavior—e.g., only one kind of base [8] or using energies depending only on the number and on the type of paired bases [9–12]. Here we will consider only models with the latter kind of interaction energy. In recent years several authors examined this kind of model with regard to the thermodynamic behavior—i.e., searching for phase transitions and describing the type of phase involved [10–13]. Liu and Bundschuh [8,14] recently discussed whether native RNA is already in the regime of the thermodynamic limit or finite-size effects have to be taken into account. In this paper we numerically investigate a hybrid model of two well-known energy models [9–12,15]: i.e., a *pair energy model*, where only base pairs are considered regardless of their neighborhood, and a *stacking energy model*, where only consecutive paired bases—i.e., forming a stack—give an energy contribution. It has been claimed that the stacking energy is more relevant than just the pair energy in real RNA [16]. Our model contains terms for *both* types of interactions and allows one to move continuously from one model to the other. We are interested in whether the two limiting models are qualitatively different—in particular, whether a phase transition occurs, when moving from one model to the other.

The paper is organized as follows. In Sec. II, we define our model; i.e., we formally define secondary structures and introduce our energy model. In Sec. II B, we explain how to calculate the partition function with a dynamic programming algorithm. In Sec. III, we introduce the observables which we investigate in the following Sec. IV. While in Secs. IV B and IV C we do finite-temperature calculations, in Sec. IV D we use the ε-coupling method at zero temperature.

II. MODEL

Because RNA molecules are linear chains of bases, they can be described as a (quenched) sequence $\mathcal{R}=(r_i)_{i=1,\ldots,L}$ of bases $r_i \in \{\mathrm{A},\mathrm{C},\mathrm{G},\mathrm{U}\}$, where L is the length of the sequence. Within this single-stranded molecule some bases can pair and build a secondary structure. Typically Watson-Crick base pairs—i.e., A-U and C-G—have the strongest affinity to each other; they are also called complementary base pairs. Each base can be paired at most once. For a given sequence $\mathcal{R}$ of bases the secondary structure can be described by a set $\mathcal{S}$ of pairs (i,j) (with the convention $1 \leqslant i < j \leqslant L$), meaning that bases r_i and r_j are paired. For convenience of notation

*Electronic address: burghard@theorie.physik.uni-goettingen.de
†Electronic address: hartmann@theorie.physik.uni-goettingen.de

we further define a matrix $(S_{i,j})_{i,j=1,\ldots,L}$ with $S_{i,j}=1$ if $(i,j) \in \mathcal{S}$ and $S_{i,j}=0$ otherwise. Two restriction are used: (i) Here we exclude so-called *pseudoknots*, which means, for any (i,j), $(i',j') \in \mathcal{S}$, either $i<j<i'<j'$ or $i<i'<j'<j$ must hold; i.e., we follow the notion of pseudoknots being more an element of the tertiary structure [16].

(ii) Between two paired bases a minimum distance is required: $|j-i| \geqslant s$ is required, granting some flexibility of the molecule (here $s=2$).

A. Energy models

Every secondary structure $\mathcal{S}$ is assigned a certain energy $E(\mathcal{S})$; note that this energy in general depends on the $\mathcal{R}$ as well, so it is more precise to write $E(\mathcal{S},\mathcal{R})$, but we assume that the structure also includes the information about the sequence. With such an energy model it is possible to calculate the canonical partition function Z of a given sequence $\mathcal{R}$ by summing over all possible structures $Z=\sum_{\mathcal{S}} e^{-\beta E(\mathcal{S})}$, but it is computationally more efficient to compute it by using the partition functions of the subsequences—i.e., by a dynamic programming approach.

Motivated by the observation that the secondary structure is due to building of numerous base pairs where every pair of bases is bound by hydrogen bonds, one assigns each pair (i,j) a certain energy $e(r_i,r_j)$ depending only on the kind of bases. The total energy is the sum over all pairs,

$$E_p(\mathcal{S}) = \sum_{(i,j)\in\mathcal{S}} e(r_i,r_j); \tag{1}$$

e.g., by choosing $e(r,r')=+\infty$ for noncomplementary bases r, r' pairings of this kind are suppressed.

Another possible model is to assign an energy E_s to a pair $(i,j) \in \mathcal{S}$ iff also $(i+1,j-1) \in \mathcal{S}$. This *stacking energy* can be motivated by the fact that a single pairing gives some gain in the binding energy, but also reduces the entropy of the molecule, because through this additional binding it loses some flexibility. Formally the total energy of a structure can be written as

$$E_s(\mathcal{S}) = \sum_{(i,j)\in\mathcal{S}} E_s S_{i+1,j-1}, \tag{2}$$

assuming that for every pair $(i,j) \in \mathcal{S}$ the bases r_i and r_j are complementary bases. The total number t of consecutive base pairs is called the *stacking size*. Single base pairs are not considered as stacks; therefore, $t \geqslant 2$ for any stack.

Both types of energy models are discussed in the literature [9–12,15], but to our knowledge, only Müller *et al.* [17,18] have discussed a model with both stacking and pair energy contributions. However, the above authors do not discuss the dependence of their results on the energy parameters E_p and E_s. Instead they use a fixed set of parameters to study the unzipping behavior of RNA, which is in the focus of their work. Here, we examine the sum of both models—stacking energy and pair energy:

$$E(\mathcal{S}) := E_p(\mathcal{S}) + E_s(\mathcal{S}), \tag{3}$$

where the parameters E_s and $e(r,r')$ can be adjusted freely, including both models discussed above. In particular we are interested in varying the parameters and studying whether the models are qualitatively different. Here we use

$$e(r,r') = \begin{cases} E_p & \text{if } r \text{ and } r' \text{ are complementary bases,} \\ +\infty & \text{otherwise,} \end{cases} \tag{4}$$

with a pair energy $E_p \leqslant 0$ independent of the kind of bases.

Due to the simple structure of the energy model—e.g., the energies depend not on the position of the bases within the sequence or whether the paired bases include some structure elements like hairpins—the ground state is highly degenerated [12,13].

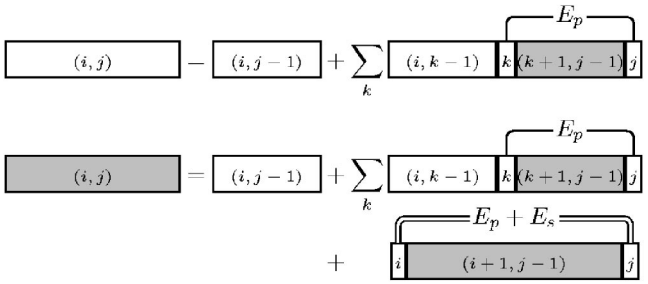

FIG. 1. Schematical explanation of Eqs. (5) and (6); e.g., white boxes represent Z, gray boxes $\hat{Z}$.

B. Calculating the partition function

Due to the fact that pseudoknots are excluded from our model (see Sec. II A), the calculation of the partition function can be done recursively. The algorithm is similar to that of Nussinov *et al.* [19,20]. The algorithms calculate the elements of two upper-triangular matrices $(Z_{i,j})_{1\leqslant i\leqslant j\leqslant L}$ and $(\hat{Z}_{i,j})_{1\leqslant i\leqslant j\leqslant L}$, where $Z_{i,j}$ is the partition function of the subsequence from base r_i to r_j under the boundary condition that bases r_{i-1} and r_{j+1} are not paired, and $\hat{Z}_{i,j}$ the partition function under the boundary condition that bases r_{i-1} and r_{j+1} be complementary; $\hat{Z}_{i,j}$ is only used as an auxiliary matrix. Then $Z_{i,j}$ can be computed from the partition functions Z and $\hat{Z}$ of smaller subsequences in the following way (remember that s denotes the minimum distance between two bases of a pair):

$$Z_{i,j} = Z_{i,j-1} + \sum_{k=i}^{j-s-1} Z_{i,k-1} \mathrm{e}^{-e(r_k,r_j)/k_\mathrm{B}T} \hat{Z}_{k+1,j-1},$$

$$\hat{Z}_{i,j} = Z_{i,j-1} + \mathrm{e}^{-(e(r_i,r_j)+E_s)/k_\mathrm{B}T} \hat{Z}_{i+1,j-1} + \sum_{k=i+1}^{j-s-1} Z_{i,k-1} \mathrm{e}^{-e(r_k,r_j)/k_\mathrm{B}T} \hat{Z}_{k+1,j-1},$$

$$Z_{i,j} = \hat{Z}_{i,j} = 1 \quad \text{for } i \geqslant j,$$

$$\hat{Z}_{i,j} = 0 \quad \text{for } 0 < j-i < s-2, \tag{5}$$

which is schematically explained in Fig. 1. Because both matrices depend on each other, they must be calculated si-

multaneously, starting along the diagonal and continuing along the off diagonals. The calculation of the partition function can be done in $O(L^3)$ steps, where L is the length of the sequence. The partition function Z of the entire sequence is $Z_{1,L}$, but also the other matrix elements are useful for generating ensembles of structures according to the Boltzmann distribution (see Sec. III).

A similar algorithm can be used to calculate the ground-state energy:

$$N_{i,j} = \min\{N_{i,j-1}, \min_{k=i}^{j-s-1}[N_{i,k-1} + e(r_k, r_j)] + \hat{N}_{k+1,j-1}\},$$

$$\hat{N}_{i,j} = \min\{N_{i,j-1}, e(r_i, r_j) + E_s + \hat{N}_{i+1,j-1}, \min_{k=i+1}^{j-s-1}[N_{i,k-1} + e(r_k, r_j) + \hat{N}_{k+1,j-1}]\},$$

$$N_{i,j} = \hat{N}_{i,j} = 0 \quad \text{for } i \geqslant j,$$

$$\hat{N}_{i,j} = +\infty \quad \text{for } 0 < j - i < s - 2. \tag{6}$$

In comparison to Eq. (5) additions are replaced by min operations, multiplications by additions, and the exponentials of the energies by the energies themselves.

III. OBSERVABLES

After calculating the partition function Z for a given random sequence, we want to measure some quantities to compare the members of the ensemble. In principle most quantities could be calculated by a similar dynamic programming algorithm introduced above; e.g., the calculation of quantities which are derivatives of the partition function with respect to temperature, external fields, etc., can also be calculated with $O(L^3)$ effort; this can be easily seen from Eq. (5). For other quantities—e.g., the distribution of the overlaps $P(q)$, which we we study here—the running time behavior would be of higher order (than three) in the sequence length [9]. For this reason we use a different method, where an ensemble $\mathcal{E}$ of structures is generated due to its Boltzmann weight. The procedure to build a sequence is essentially a backtracking algorithm: Starting with the entire sequence a partner for an outermost base—e.g., base L—is chosen with the appropriate probability—e.g., base k—after this the procedure is applied to the subsequences 1 to $k-1$ and $k+1$ to L. If base L is chosen not to be paired at all, one uses the sequence shortened by base L. Considering a subsequence from base k to l, the probability $p_{i,j;k,l}$ that bases i and j ($k \leqslant i < j \leqslant l$) are paired is given by

$$p_{i,j;k,l} = \begin{cases} Z_{k,l}^{-1} Z_{k,i-1} e^{-(e(r_k,r_j)+E_0)/k_B T} \hat{Z}_{j+1,l} & \text{bases } i-1 \text{ and } j+1 \text{ paired,} \\ Z_{k,l}^{-1} Z_{k,i-1} e^{-e(r_k,r_j)/k_B T} \hat{Z}_{j+1,l} & \text{bases } i-1 \text{ and } j+1 \text{ unpaired.} \end{cases} \tag{7}$$

For each member of this ensemble $\mathcal{E}$ the quantity of interest X is calculated and the average $\langle X \rangle = (1/|\mathcal{E}|)\Sigma_{\mathcal{S}} X(\mathcal{S})$ is used as an approximation to the expectation value of the Gibbs-Boltzmann ensemble; for large enough ensembles $\mathcal{E}$, this average approaches the true expectation value. In Sec. IV all finite-temperature observables are calculated by an ensemble generated according to Eq. (9).

A simple observable is the energy E and its fluctuations $(\Delta E)^2$, the latter is directly connected with the specific heat $c_V = (\Delta E)^2 / L k_B T^2$.

Of particular importance is the overlap q between two structures $\mathcal{S}$ and $\mathcal{S}'$,

$$q(\mathcal{S}, \mathcal{S}') := \frac{2}{L} \sum_{1 \leqslant i < j \leqslant L} S_{i,j} S'_{i,j}, \tag{8}$$

which is the number of bases paired to the same base in both structures normalized such that q lies always between 0 and 1. This is a measure of how similar two structures are. Note, however, that with this definition the overlap of one structure with itself is $q(\mathcal{S}, \mathcal{S}) \lneqq 1$ unless all bases are paired, where it is equal to 1. A definition of q where also bases unpaired in both structures are counted is used in [9], resulting in an overlap definition that is normalized; however, this similarity measurement has the drawback that the fewer bases that are paired, the more two structures become similar. We further remark that for any two structures $\mathcal{S}, \mathcal{S}'$ the Cauchy-Schwarz inequality $[q(\mathcal{S}, \mathcal{S}')]^2 \leqslant q(\mathcal{S}, \mathcal{S}) q(\mathcal{S}', \mathcal{S}')$ holds. With this quantity q two ensembles $\mathcal{E}$ and $\mathcal{E}'$ can be compared—e.g., looking at the distribution of $q(\mathcal{S}, \mathcal{S}')$ of all $\mathcal{S} \in \mathcal{E}, \mathcal{S}' \in \mathcal{E}'$.

The ensemble averages $\langle \cdot \rangle$ in general depend on the chosen sequence; therefore, a further averaging over several (random) sequences is required. This sequence average is denoted by $[\cdot]$. We again approximate this average by summing over a finite set of sequences.

Because of this two stage averaging, it is probably preferable instead of looking at $[\langle q \rangle]$ directly to use functions of the first and higher moments of q—e.g., the Binder cumulant [21–23], which is defined by

$$B := \frac{1}{2}\left(3 - \frac{[\langle q^4 \rangle]}{[\langle q^2 \rangle^2]}\right), \tag{9}$$

where $\langle q^n \rangle$ is either the average over all pairs of one ensemble,

$$q_{\mathcal{E}}^{n} := \frac{1}{|\mathcal{E}|(|\mathcal{E}|-1)} \sum_{\substack{\mathcal{S},\mathcal{S}' \in \mathcal{E} \\ \mathcal{S} \neq \mathcal{S}'}} q^{n}(\mathcal{S},\mathcal{S}'), \tag{10}$$

or the average over all pairs of two ensembles,

$$q_{\mathcal{E},\mathcal{E}'}^{n} := \frac{1}{|\mathcal{E}||\mathcal{E}'|} \sum_{\mathcal{S} \in \mathcal{E}, \mathcal{S}' \in \mathcal{E}'} q^{n}(\mathcal{S},\mathcal{S}'). \tag{11}$$

The ensembles $\mathcal{E}$ and $\mathcal{E}'$ are finite collections of structures belonging to the same sequence—e.g., generated according to Eq. (7). Note, however, that in general the probabilities $p_{i,j;k,l}$ will be different—e.g., different temperatures T or energy parameter (E_p, E_s). The latter choice [Eq. (11)] is appropriate when one is looking for a change in the behavior of the models when one varies the parameters in comparison to a reference model, while the former [Eq. (10)] is the better choice, if an external parameter—e.g., the temperature—is varied.

$q_{\mathcal{E}}$ is a measure of how similar the structures within ensemble $\mathcal{E}$ are, while $q_{\mathcal{E},\mathcal{E}'}$ gives in addition a measure how many structures both ensembles have in common. Typically, e.g., in spin-glass studies, one examines the self-overlap $q_{\mathcal{E}}$, but we believe that the cross overlap $q_{\mathcal{E},\mathcal{E}'}$ is more sensitive for determining a parameter driven phase transition which might occur in our case and therefore examine both quantities.

The Binder cumulant B vanishes at high temperatures, while for low temperatures it approaches a finite value in the thermodynamic limit.

A similar quantity—first introduced in [24]—has been used in [12]:

$$A := \frac{[\chi_{\mathcal{R}}^{2}] - [\chi_{\mathcal{R}}]^{2}}{[\chi_{\mathcal{R}}]^{2}}, \tag{12}$$

where

$$\chi_{\mathcal{R}} := L(\langle q^{2} \rangle - \langle q \rangle^{2}) \tag{13}$$

is the variance of the q distribution. This parameter A measures how the probability distribution of q varies from sequence to sequence. A value close to zero indicates a self-averaging behavior. A similar physical content is represented by another quantity introduced in [25]:

$$G := \frac{[\chi_{\mathcal{R}}^{2}] - [\chi_{\mathcal{R}}]^{2}}{[L^{2}\langle (q - \langle q \rangle)^{4} \rangle] - [\chi_{\mathcal{R}}]^{2}}. \tag{14}$$

The general strategy to find a phase transition with respect to the energy parameters is to examine various quantities—e.g., the above introduced A, B, G—to find any of this indicating a phase transition. It is quite possible that one quantity shows a phase transition, while the others do not; see e.g., the ongoing discussion on whether Heisenberg spin glasses exhibit just chiral ordering without traditional spin-glass order [26–29] or both [30,31], where the first group of authors argues that the "right" observable shows a phase transition while others might show no sign at all.

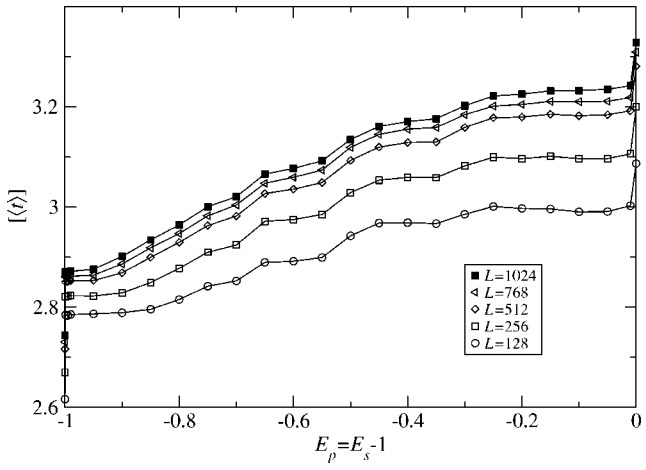

FIG. 2. Average stacking size as a function of energy parameter E_p for different system sizes ($T=0$). At $E_p=-1$—i.e., $E_s=0$ and $E_p=0$—the curves are discontinuous. Missing error bars are of the size of the symbols or smaller and omitted for legibility. Calculated data points are indicated by symbols. Lines are drawn to guide the eye.

IV. NUMERICAL RESULTS

In order to find some possible differences in the behavior of the energy model, Eq. (3), for different parameters E_p, E_s, as introduced in Eqs. (4) and (2), respectively, we numerically calculated the quantities mentioned in Sec. III above. In all our examples we averaged over randomly generated sequences $\mathcal{R}=(r_i)$, where the probability for a specific base r_i at position i is $\frac{1}{4}$ for all base types independent of the other bases $r_{j\neq i}$. The size of the sequence varied from $L=128$ up to $L=1024$; for the disorder average 2000, up to 6000 random sequences were used. Pairing of bases is only allowed for complementary bases and the minimum distance between two bases was chosen as $s=2$.

In Secs. IV A and IV B, we vary continuously the energy parameters between the two extreme cases $(E_p=0, E_s=-1)$ and $(E_p=-1, E_s=0)$. In Sec. IV A we shortly discuss the averages of the stacking size t and of the q for different energy parameters. In Sec. IV B we examine the cross overlap $q_{\mathcal{E},\mathcal{E}^{\mathrm{ref}}}$ [see Eq. (11)] between a reference ensemble $\mathcal{E}^{\mathrm{ref}}$ generated for fixed energy parameters and ensembles generated for different energy parameters. In the following Sec. IV C we look at the temperature variation of various quantities without using any reference ensemble to estimate some critical parameters. In the last Sec. IV D we apply the ε-coupling method at $T=0$ to determine the critical exponent θ describing the behavior of low-lying excitations.

A. Basic observables

In Sec. II A, where we introduced the energy model, we opposed the pair energy to the stacking energy model. In Fig. 2 we show the average size t of a stacking as a function of the energy parameter E_p at temperature $T=0$. We keep $E_p+E_s=-1$ constant to fix the overall energy scale. For all fixed energy parameters the average stacking size t increases with increasing system size, which is expected as with increasing

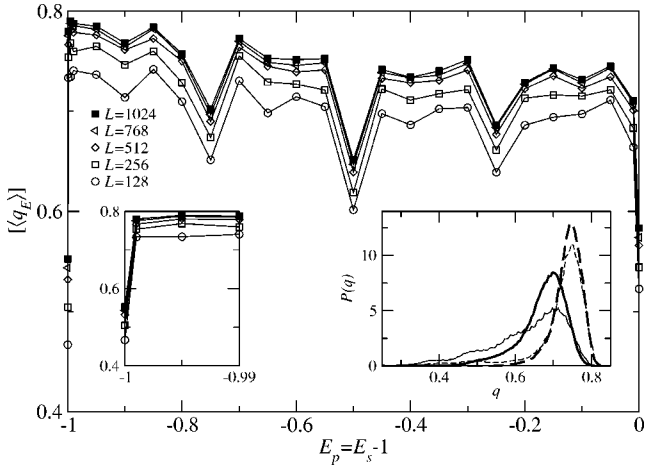

FIG. 3. Average overlap $q_{\mathcal{E}}$ as a function of the energy parameter E_p for different system sizes at $T=0$. The local minima at $E_p=-0.75,-0.5,-0.25$ are due to the commensurability of E_p and E_s and indicate a broad distribution of the ground states in configuration space. The left inset is an enlargement to show the discontinuity. The right inset is the q distribution for $E_s=E_p=-0.5$ (solid lines) and $E_s=-0.49$, $E_p=-0.51$ (dashed lines) for sequence lengths $L=512$ (thin lines) and $L=1024$ (thick lines).

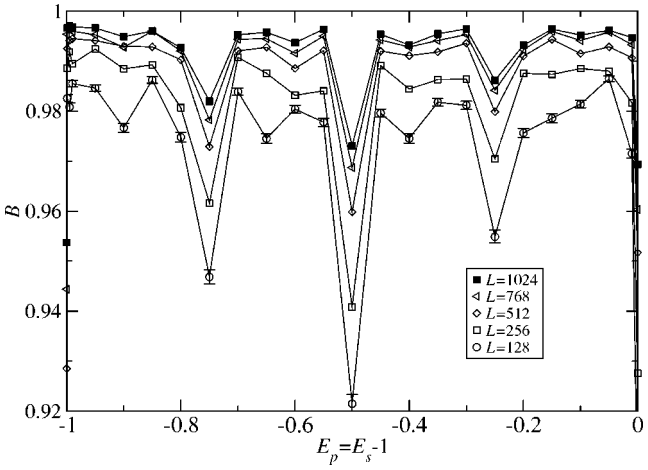

FIG. 4. The Binder cumulant B of Eq. (9) with $q=q_{\mathcal{E}}$ [see Eq. (10)] at temperature $T=0$. The curves for different system sizes do not cross, and therefore B gives no hint of a phase transition. Missing error bars are of the size of the symbols or smaller and omitted for legibility. Calculated data points are indicated by symbols. Lines are drawn to guide the eye.

length the probability for longer stacks increases. Also as expected is the overall increase in the average stacking size with the energy parameter E_s, because the stacking energy prefers to build stacks. The large increase of t while changing E_s from a zero to a nonzero value can be explained as follows: The ground state for $E_p=-1.0, E_s=0.0$ is highly degenerated, while changing to a nonzero E_s only those states stay ground states which have a high stacking contribution to the energy. This selection increases the average of t. A similar argument applies at the other end, where E_p changes from a nonzero value to zero.

Similarly the overlap q jumps to a larger value when changing from $E_p=0$ to a nonzero value (Fig. 3). In addition, at positions where E_p/E_s are fractions with small numerator and denominator—e.g., $\frac{1}{1}$ or $\frac{1}{3}$—for this energy parameter the ground states in configuration space are more broadly distributed than for slightly different parameters. This can be seen in the right inset of Fig. 3, where the q distribution in the symmetric case ($E_s=E_p=-0.5$) is broader than in the slightly asymmetric case ($E_s=-0.49, E_p=-0.51$). The width of this minimum in the main plot, as well as that in Figs. 4 and 5, is below $\Delta E=0.001$, as one can estimate from the left inset of Fig. 3.

For both the average stacking size and the average overlap, the behavior changes smoothly when moving from one model to the other, with the exception of the highly degenerate points, where we can observe the jumps in the overlap q. Hence, there is no sign of a transition in between the two extremal models. To confirm this, we next study the Binder cumulant.

B. Binder cumulant

Since we introduced in Eq. (3) a whole class of energy models depending on the pair energy E_p and the stacking energy E_s, we examined the behavior of the Binder cumulant depending on these two energy parameters and the sequence length L. Second-order phase transitions are characterized by crossing of the Binder cumulant for different system sizes at the transition point. We examined two slightly different Binder cumulants in the hope that—in the sense of the discussion at the end of Sec. III—at least one of them shows such a crossing.

In Fig. 4 the Binder cumulant, Eq. (9), is shown at $T=0$ using the "self-overlap," Eq. (10). Again, the energy parameters E_p and E_s are varied such that always $E_p+E_s=-1$ holds. The value of B increases with increasing system size for E_p; i.e., the curves do not cross each other and therefore give no evidence of a phase transition. The local minima are—as the minima in Fig. 3—due to the commensurability of the energy parameters E_p and E_s.

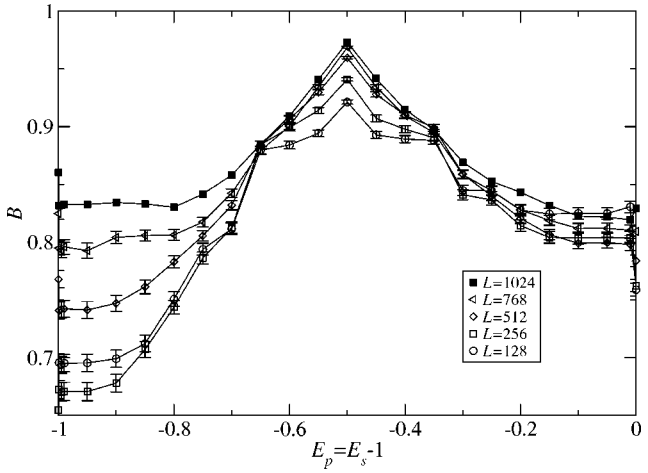

FIG. 5. The Binder cumulant B of Eq. (9) with $q=q_{\mathcal{E},\mathcal{E}'}$ [see Eq. (11)] at temperature $T=0$. Missing error bars are of the size of the symbols or smaller and omitted for legibility. Calculated data points are indicated by symbols. Lines are drawn to guide the eye.

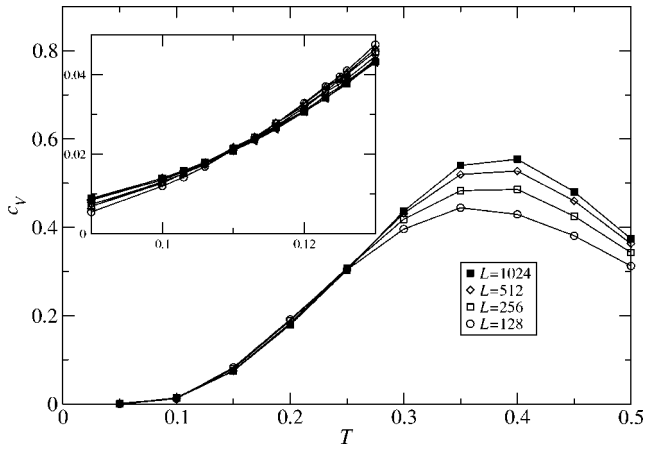

FIG. 6. Specific heat c_V as a function of temperature for parameters $E_s=-1.0$ and $E_p=0.0$. The curves for different system sizes crosses at $T\approx 0.25$ and $T\approx 0.11$. The inset is an enlargement of the main plot. Missing error bars are of the size of the symbols or smaller and omitted for legibility. Calculated data points are indicated by symbols. Lines are drawn to guide the eye.

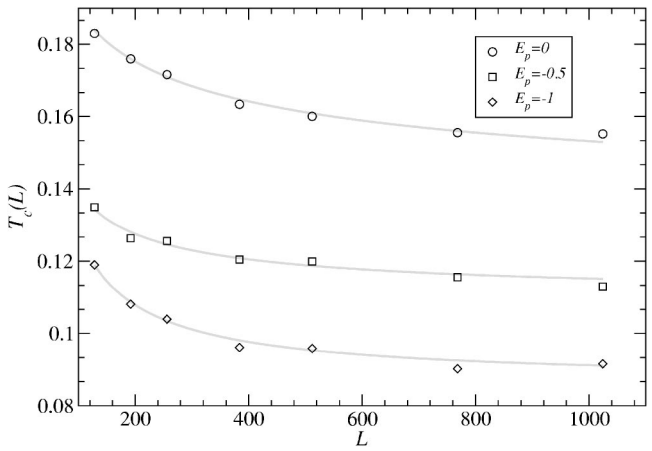

FIG. 7. The position of the maxima of $\chi_{\mathcal{R}}$ for different energy parameter. The curves are fitted to the form $T_c(L)=T_c+aL^{-1/\nu}$ and gives the critical parameters of Table I.

Further we used a reference ensemble generated for energy parameters $E_p=E_s=-0.5$ and used the "cross overlap" of Eq. (11). First, the values of B at $E_p=E_s=-0.5$ coincide with the values shown in Fig. 4. Similar to the observation above, B roughly increases with the system size, although the separation of the curves is not as clear as in Fig. 4, especially in the range from $E_p=-0.4$ to $E_p=0.0$, where the curves coincide within the error bars. To summarize, the dependence of the Binder cumulant on the energy parameters does not indicate a phase transition.

C. Temperature dependence of the energy models

Another possible method to discriminate between the different energy parameters is to examine the temperature dependence of some quantities, especially the behavior at a critical temperature. In Ref. [12] it was shown for a similar model that below a critical temperature, almost all sequences fold to a compact structure, but for most sequences not into a single structure. They point out that this kind of low-temperature behavior is well known from spin glass and other disordered systems. In Fig. 6 we plotted the specific heat for different system sizes as a function of temperature for $E_s=-1.0$ and $E_p=0.0$. All curves cross each other close to $T=0.11$ and $T=0.25$, which is evidence for a phase transition at this temperature region. The data for other energy parameters (E_s,E_p) look similar, but the curves cross at different temperatures.

To determine the critical behavior quantitatively we investigated the width $\chi_{\mathcal{R}}$ of the overlap distribution. We found that for all examined energy parameters and sequence lengths $\chi_{\mathcal{R}}$ as a function of temperature has a maximum. We show this maximum position in Fig. 7. We assume that the maximum position follows the form $T_c(L)=T_c+aL^{-1/\nu}$ and fit the data to this form to get the critical parameters (see Fig. 7). The results for three different pairs of energy parameters are shown in Table I. The critical exponent ν for the parameter pair $(E_s=0,E_p=-1)$ is clearly different from that of the parameter pair $(E_s=-1,E_p=0)$, showing that the quantitative behavior changes when varying from one limiting case to the other. On the other hand, for the intermediate parameter set the critical exponent is consistent with both within the error range.

In Fig. 8 we present the parameter G of Eq. (14) for the energy parameters $(E_s,E_p)=(0,-1)$, $(-0.5,-0.5)$, and $(-1,0)$. As known for other spin-glass-like models [25] the value of G approaches zero with increasing temperature. The above authors find a limiting value of $\frac{1}{3}$ for decreasing temperature, which has been shown to be the exact result under some more restricted conditions [32]. For two of our energy parameters, $(E_s,E_p)=(0,-1)$ and $(-0.5,-0.5)$, the data are compatible with the same limiting value, while in the third case, which has the lowest critical temperature (Table I), it is not clear whether the values approach $\frac{1}{3}$ or not.

Finally, we also found (not shown) that the behavior of A of Eq. (12) is in agreement with this observation and with the results for a two-letter RNA model [12]. For all three cases studied here, the width of the q distribution varies only slightly from realization to realization at high temperatures, while for low temperatures the self-averaging behavior disappears.

D. ε-coupling method

Previously the *ε-coupling method* has been used [11,33] to investigate low-energy excitations of RNA secondary

TABLE I. Critical parameter of a $\chi_{\mathcal{R}}$ maximum fit. Comparing the two limiting cases $(E_s,E_p)=(-1,0)$ and $(E_s,E_p)=(0,-1)$ the parameters ν and T_c are different and indicate a quantitative change. The last column belongs to the ε-coupling method in Sec. IV D.

Energy model		$1/\nu$	T_c	θ
$E_s=0$	$E_p=-1$	0.93(15)	0.086(3)	0.229(38)
$E_s=-0.5$	$E_p=-0.5$	0.70(36)	0.109(7)	0.237(50)
$E_s=-1$	$E_p=0$	0.36(17)	0.125(21)	0.194(67)

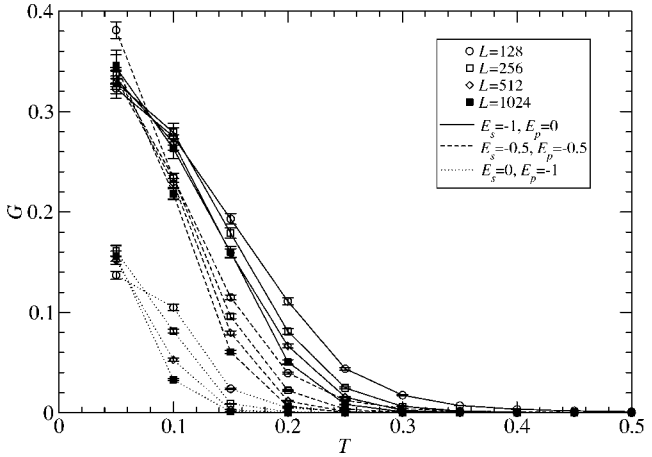

FIG. 8. The parameter G [Eq. (14)] as a function of temperature for different energy parameters $(E_s,E_p)=(0,-1)$, $(-0.5,-0.5)$, and $(-1,0)$. For low temperatures the first two cases approach the theoretical expected value $G(T\to 0)=\frac{1}{3}$, while in the third case—the one with the lowest critical temperature—this is not clear.

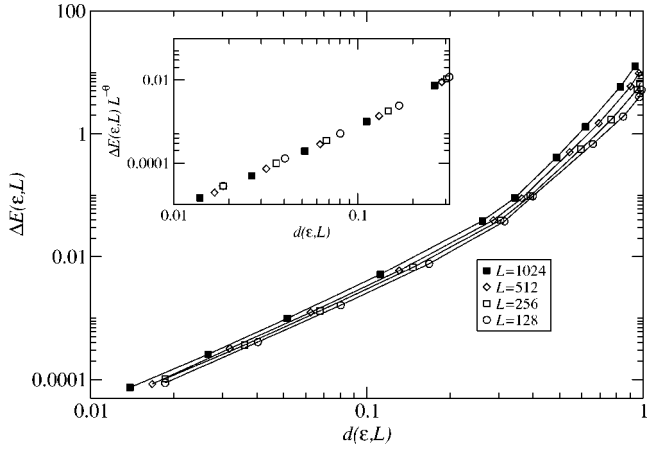

FIG. 9. ε-coupling results for $E_s=-0.5$, $E_p=-0.5$. The energy difference $\Delta E(\mathcal{S})$ between the original ground-state structure $\mathcal{S}_0$ and the ground-state structure $\mathcal{S}_\varepsilon$ of the disturbed model is plotted as a function of the distance between this structures. The inset is a scaled plot of the data with $\varepsilon<1$ of the main plot ($\theta=0.24$; see Table I). Missing error bars are of the size of the symbols or smaller and omitted for legibility. Calculated data points are indicated by symbols. Lines are drawn to guide the eye.

structures. The basic idea is to add another term to the energy model in Eq. (3), which depends on the ground state of the original problem: Assume $\mathcal{S}_0$ is the unique ground state of $E(\mathcal{S})$; then, a new energy function is defined as

$$E_\varepsilon(\mathcal{S}) = E(\mathcal{S}) + \varepsilon q(\mathcal{S},\mathcal{S}_0), \tag{15}$$

with $\varepsilon>0$. The additional term penalizes structures similar to $\mathcal{S}_0$, where $\varepsilon q(\mathcal{S},\mathcal{S}_0)$ is largest for $\mathcal{S}=\mathcal{S}_0$. In general the ground state $\mathcal{S}_\varepsilon$ of the new energy model E_ε will be different from $\mathcal{S}_0$. The difference $\Delta E(\varepsilon):=E(\mathcal{S}_\varepsilon)-E(\mathcal{S}_0)$ is an increasing function of ε and $\Delta E(\varepsilon)\leqslant\varepsilon$ holds. The latter implies that $\mathcal{S}_\varepsilon$ is for small enough ε a low-lying excitation of the original energy model and has the smallest overlap with $\mathcal{S}_0$.

The average distance $d(\varepsilon,L):=1-q(\mathcal{S}_\varepsilon,\mathcal{S}_0)$ between the new and old ground states scales as $d(\varepsilon,L)\propto\varepsilon L^{-\theta}$, ε constant, while the average energy difference scales as $\Delta E(d,L)\propto L^\theta$, d constant, with the critical exponents θ. For details see Refs. [11,33].

The assumption of a unique ground state used above does not hold for our model used so far: in general the ground state is highly degenerated because only two energy parameters (E_s and E_p) are used and many structures will have the same energy. The degeneracy of the ground state renders the ε-coupling method as described above almost useless. Since in natural RNA the contributions to the energy are more complex, different structure will never be degenerated. This justifies to change the energy model slightly by adding a random energy $\eta_{i,j}$ to the pair energies introduced in Eq. (1): $e(r_i,r_j)\to e(r_i,r_j)+\eta_{i,j}$. There are several possibilities to choose the distributions of the η's (see [11]); here, we use identical, independently distributed Gaussian random numbers with zero mean $\langle\eta\rangle=0$ and variance $\langle\eta^2\rangle=\eta_0^2/L$ with $\eta_0=0.1$ (the quasidegenerate (QD) model in Ref. [11]).

The raw result for different values of $\varepsilon\in[0.01,100]$ is shown in Fig. 9. A scaling plot of the data for $\varepsilon<1$ according the scaling form $\Delta EL^{-\theta}=f(d)$ is shown in the inset of Fig. 9. The scaling parameters θ leading to the best data collapse for different energy parameters are shown in the rightmost column of Table I. They are equal within the error margins and thus do not give us a further hint of a quantitative different behavior for different energy parameters. These difficulties in doing a good scaling of the data in the QD model were already pointed out [11]. However, for a different model using a scaling function with finite-size corrections a critical exponent $\theta=0.23\pm0.05$ was obtained [33], which is close to our findings.

V. SUMMARY

We have introduced a RNA secondary structure model which continuously interpolates between two well-known models incorporating both bond and stacking energies. The model does not allow for pseudoknots; nor does it take different entropic terms for different structural elements into account. The simplicity of the model enables us to present a recursive formula for the partition function, which allows a fast exact evaluation for each realization of the disorder.

We sought an answer to the question whether there is a qualitative difference between the behavior induced by either bond or stacking energies. In particular we are interested whether there is any phase transition of the thermodynamic behavior when moving continuously from the pure bond to the pure stacking model. The way we look at the model to find such a transition is to consider it as a statistical mechanics model exhibiting quenched disorder through the sequences. We measure typical quantities like distributions of overlaps and derived quantities which are based on higher moments of these distributions. It was known already that the pure bond-energy model exhibits at low temperatures a complex ground-state landscape and has some features with spin glasses in common. In particular, this phase exhibits a

lack of self-averaging for the distribution of overlaps. To search for a transition, we considered both zero as well as finite temperatures.

Zero-temperature results give no evidence for a parameter-driven phase transition apart from trivial crossovers—e.g., discontinuities of t at points with $E_s=0$ or $E_p=0$. The curves of the Binder cumulant do not cross at any point, which would be an indication of a phase transition. Similar, the critical exponent θ derived from the ε-coupling method seems to be independent of the energy parameters E_s and E_p, and therefore gives no evidence for a quantitative difference in the thermodynamic properties. But as stated in [11], the determination of the critical exponent is rather difficult in the quasidegenerated case studied here.

Hence, at low temperatures a somehow complex ordered phase, with many similarities to spin glasses, is present for all parameter combinations in our model. However, the finite-temperature results show a quantitative dependence on the energy parameters, when studying the transition from the high-temperature phase to the low-temperature ordered phase, at fixed values for the energy parameters. The critical exponent ν for the correlation length seems to depend on the energy model and may vary continuously while going from one limiting model to the other.

Wrapping up, there seems to be no strong qualitative difference between the models with just bond and just stacking energies. For all considered values of the energy parameters, there exists an ordered low-temperature glassylike phase, characterized by the lack of self-averaging of the distribution of overlaps. Just the quantitative behavior seems to depend somehow on the energy contributions; i.e., a crossover may occur. It is still possible that these results are due to the simplicity of our model, since it does not allow for pseudoknots; nor does it give different entropic weights to different structural elements like hairpins, bulges, etc. When including entropic contributions, we do not expect a fundamental change in behavior, because this approach changes only the relative weights given to the different structural elements. On the other hand, the inclusion of pseudoknots might really have a strong qualitative effect. Unfortunately, the case of pseudoknots has a much higher computational complexity. Hence, only small systems can be studied. Nevertheless, preliminary results [34] indicate that the presence of pseudoknots makes no strong difference in the low-temperature behavior, at least for the case where only bond energies are considered.

As a last point, it might be worthwhile not to use random sequences but sequences which show up in biological systems—e.g., such as ribozymes. These special selected sequences might have some special properties random sequences in general do not have, e.g., a unique ground state, which in turn might influence the quantities discussed in this paper.

ACKNOWLEDGMENTS

The authors have obtained financial support from the Volkswagenstiftung (Germany) within the program "Nachwuchsgruppen an Universitäten." The simulations were performed at the Paderborn Center for Parallel Computing in Germany and on a workstation cluster at the Institut für Theoretische Physik, Universität Göttingen, Germany. We thank E. Yewande for helpful remarks.

[1] *The RNA World*, 2nd ed., edited by R. F. Gesteland, T. R. Cech, and J. F. Atkins (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1999).
[2] P. G. Higgs, Q. Rev. Biophys. **33**, 199 (2000).
[3] R. Bundschuh and T. Hwa, Phys. Rev. Lett. **83**, 1479 (1999).
[4] M. Zuker, Science **244**, 48 (1989).
[5] J. S. McCaskill, Biopolymers **29**, 1105 (1990).
[6] I. L. Hofacker, W. Fontana, P. F. Stadler, L. S. Bonhoeffer, M. Tacker, and P. Schuster, Monatsh. Chem. **125**, 167 (1994).
[7] R. Lyngsø, M. Zuker, and C. N. S. Pedersen, Bioinformatics **15**, 440 (1999).
[8] T. Liu and R. Bundschuh, Phys. Rev. E **69**, 061912 (2004).
[9] P. G. Higgs, Phys. Rev. Lett. **76**, 704 (1996).
[10] R. Bundschuh and T. Hwa, Phys. Rev. E **65**, 031903 (2002).
[11] E. Marinari, A. Pagnani, and F. Ricci-Tersenghi, Phys. Rev. E **65**, 041919 (2002).
[12] A. Pagnani, G. Parisi, and F. Ricci-Tersenghi, Phys. Rev. Lett. **84**, 2026 (2000).
[13] A. K. Hartmann, Phys. Rev. Lett. **86**, 1382 (2001).
[14] T. Liu and R. Bundschuh, e-print physics/0304108.
[15] R. Mukhopadhyay, E. Emberly, C. Tang, and N. S. Wingreen, Phys. Rev. E **68**, 041904 (2003).
[16] I. Tinoco, Jr. and C. Bustamante, J. Mol. Biol. **293**, 271 (1999).
[17] M. Müller, F. Krzakala, and M. Mézard, Eur. Phys. J. E **9**, 67 (2003).
[18] M. Müller, Phys. Rev. E **67**, 021914 (2003).
[19] R. Nussinov, G. Pieczenik, J. R. Griggs, and D. J. Kleitman, SIAM (Soc. Ind. Appl. Math.) J. Appl. Math. **35**, 68 (1978).
[20] R. Durbin, S. R. Eddy, A. Krogh, and G. Mitchison, *Biological Sequence Analysis* (Cambridge University Press, Cambridge, England, 1998).
[21] K. Binder, Z. Phys. B: Condens. Matter **43**, 119 (1981).
[22] R. N. Bhatt and A. P. Young, Phys. Rev. Lett. **54**, 924 (1985).
[23] R. N. Bhatt and A. P. Young, Phys. Rev. B **37**, 5606 (1988).
[24] E. Marinari, C. Naitza, F. Zuliani, G. Parisi, M. Picco, and F. Ritort, Phys. Rev. Lett. **82**, 5175 (1999).
[25] E. Marinari, C. Naitza, F. Zuliani, G. Parisi, M. Picco, and F. Ritort, Phys. Rev. Lett. **81**, 1698 (1998).
[26] H. Kawamura, Phys. Rev. Lett. **80**, 5421 (1998).
[27] K. Hukushima and H. Kawamura, Phys. Rev. E **61**, R1008 (2000).
[28] M. Matsumoto, K. Hukushima, and H. Takayama, Phys. Rev. B **66**, 104404 (2002).
[29] D. Imagawa and H. Kawamura, Phys. Rev. B **70**, 144412 (2004).

[30] F. Matsubara, T. Shirakura, and S. Endoh, Phys. Rev. B **64**, 092412 (2001).
[31] L. W. Lee and A. P. Young, Phys. Rev. Lett. **90**, 227203 (2003).
[32] F. Ritort and M. Sales, J. Phys. A **33**, 6505 (2000).
[33] F. Krzakala, M. Mézard, and M. Müller, Europhys. Lett. **57**, 752 (2002).
[34] A. Morales Gallardo, B. Burghardt, and A. K. Hartmann (private communication).