

## Statistical mechanics of RNA folding: A lattice approach

P. Leoni<sup>1</sup> and C. Vanderzande<sup>1,2</sup>

<sup>1</sup>Departement WNI, Limburgs Universitair Centrum, 3590 Diepenbeek, Belgium

<sup>2</sup>Instituut voor Theoretische Fysica, Katholieke Universiteit Leuven, Celestijnenlaan 200D, 3001 Heverlee, Belgium

(Received 31 March 2003; published 11 November 2003)

We propose a lattice model for the secondary structure of RNA based on a self-interacting two-tolerant trail. Self-avoidance and pseudoknots are taken into account. We investigate a simple version of the model in which the native state of RNA consists of just one hairpin. Using exact arguments and Monte Carlo simulations we determine the phase diagram for this case. We show that the denaturation transition is first order and can either occur directly or through an intermediate molten phase.

DOI: 10.1103/PhysRevE.68.051904

PACS number(s): 87.15.Aa, 05.40.-a, 36.20.Ey

### I. INTRODUCTION

In recent years, the diversity of roles played by RNA in cellular processes has become increasingly clear [1]. RNA is a heteropolymer which is built from four types of monomers (nucleotides) [2] and as for proteins, a major issue is the prediction of the tertiary (folded) structure for a given sequence of nucleotides [3]. In some respects, this question is easier to answer for RNA than for proteins. First, the four nucleotides are chemically more similar than the 20 amino acids that form proteins. Second, it has been argued [3] that folding in RNA is hierarchical in the sense that the energy scales involved in secondary structure elements are larger than those of tertiary structure. Therefore much attention, also in the physics literature [4–11], has been devoted to predicting the secondary structure of RNA.

In most of the existing models, self-avoidance has not been fully taken into account. If one also neglects certain types of monomer-monomer interactions such as those occurring in pseudoknots and kissing hairpins [2], it becomes possible to calculate secondary structures with a recursive algorithm whose complexity only grows as the third power of the number of monomers  $L$  [5]. Despite its success, this approach has several drawbacks. First, the neglect of self-avoidance leads to unphysical properties for the radius of gyration [7], especially at low temperatures. Second, the neglect of part of the physically relevant interactions is often justified *a posteriori* from the observation that (for example) pseudoknots are not very common in real RNA. However, it would be more attractive to have a model which allows one to *predict* the rate of occurrence of these structural elements. A first attempt along these lines, but neglecting self-avoidance, was made in Ref. [12].

In this paper, we propose a lattice model for RNA which can take into account both self-avoidance and most of the physically relevant interactions. To get a first insight into the properties of the model, we study here in detail a simplified version, which is a lattice variant of the model studied in Ref. [6]. By using rigorous and numerical techniques well known from the study of other lattice models of polymers [13], we find that RNA can exist in three phases. In the present work we study the properties of these phases and of the transitions between them. In a forthcoming publication,

we will investigate questions such as the probability of formation of pseudoknots.

This paper is organized as follows. In Sec. II we present our lattice model and show how one can include several aspects of real RNA in a natural way. We also introduce the simplified version of the model. In Sec. III we present a number of semiexact results from which the phase diagram for this case can be obtained. In Sec. IV we give the results of an extensive Monte Carlo study of this phase diagram in two dimensions. Finally, in Sec. V, we present our conclusions.

### II. AN INTERACTING TWO-TOLERANT TRAIL MODEL OF RNA

RNA is a heteropolymer whose primary structure consists of a sequence of ribonucleotide bases. Of these there are four types, which are denoted by the symbols  $C$ ,  $G$ ,  $U$ , and  $A$ . The particular sequence of these bases determines the three-dimensional structure. Here we are mainly interested in the secondary structure which depends on base-base interactions, of which pairing and stacking are the most important.

We make a lattice model for RNA starting from a two-tolerant trail [14]. This is a lattice random walk that can visit each edge of the lattice at most twice. The walk has  $L$  steps, each of which corresponds to a monomer of RNA. Doubly visited edges correspond to bonded base pairs. In molecules such as DNA and RNA, complementary base pairs bind through hydrogen bonds. Once two bases are bonded, no further hydrogen bonds can be made with other bases and it is this restriction that is taken into account by the two-tolerance of the walk. In Fig. 1 we show a typical two-tolerant trail of 500 steps. When step  $i$  and  $j$  are on the same edge, we say that they are *bonded*. The two-tolerant trail consists of sets of connected bonded steps (corresponding to helices in real RNA) alternating with sets of singly visited edges (corresponding to loops and bulges) [2]. With each  $L$ -step two-tolerant trail  $\mathcal{T}$  we associate a set  $S_{\mathcal{T}}$  whose elements are the bonded steps of the trail:  $S_{\mathcal{T}} = \{(i,j): i,j \text{ are bonded}\}$ . In most of the existing theoretical work on RNA the following restriction is put on the bonded steps: when both  $(i,j)$  and  $(i',j')$  are bonded, one only allows the combinations  $i < j < i' < j'$  and  $i < i' < j' < j$ . The situation in which  $i < i' < j < j'$  will be referred to as a

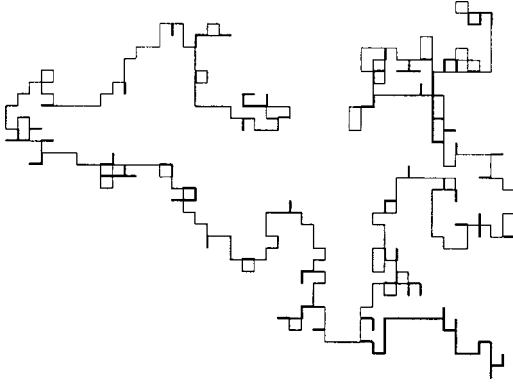


FIG. 1. A two-tolerant trail of 500 steps. Doubly visited edges are represented by double lines.

pseudoknot, even though in the biological literature this class of structures is further divided into several types [2]. In our model, pseudoknots are treated on the same footing as the other types of bonds.

Next, we associate to each doubly visited edge a pairing energy  $\varepsilon_{i,j}$  which depends on the nature of the bases  $i$  and  $j$  present on the edge. Below we will propose a simple form for  $\varepsilon_{i,j}$  inspired by the work of Bundschuh and Hwa [6] and investigate the phase diagram of the resulting model.

Before doing that we comment here on how other physically relevant interactions can be introduced in a natural way in our model. The most important of these is the stacking energy. This is an interaction between neighboring bonded base pairs and it can be modeled by associating an energy  $\sigma_{i,i+1,j-1,j}$  with each element of  $S_T$ . The value of this stacking energy depends on the four nucleotides forming the stack. When the pair  $(i+1,j-1)$  is not bonded we put  $\sigma_{i,i+1,j-1,j}=0$ .

Because of the semiflexible nature of nucleic acids, it is also necessary to add a bending rigidity for which we assume the form  $\kappa(\vec{n}_i - \vec{n}_{i+1})^2$  with  $\kappa \geq 0$  [15]. Here  $\vec{n}_i$  is the unit vector in the direction of the  $i$ th step. The total energy associated with the two-tolerant trail  $T$  thus becomes

$$E_T = \kappa \sum_{i=1}^{L-1} (\vec{n}_i - \vec{n}_{i+1})^2 + \sum_{(i,j) \in S_T} (\varepsilon_{i,j} + \sigma_{i,i+1,j-1,j}). \quad (1)$$

The thermodynamic properties of the RNA model can then be determined from the partition sum

$$Z_L = \sum_T \exp(-\beta E_T), \quad (2)$$

where the sum is over all  $L$ -step two-tolerant trails and  $\beta$  is the inverse temperature  $\beta = 1/k_B T$ . Moreover, if  $\mathcal{A}$  is a particular set of bonded monomer pairs we can determine the probability  $p_{\mathcal{A},L}$  that this set occurs by calculating the ratio

$$p_{\mathcal{A},L} = \frac{Z_{\mathcal{A},L}}{Z_L}, \quad (3)$$

where  $Z_{\mathcal{A},L}$  is given by an expression similar to Eq. (2) with the sum restricted to trails  $T$  for which  $\mathcal{A} \subset S_T$ . In this way, it

is possible to calculate, for example, the probability that a pseudoknot appears as a function of temperature.

Within this model we can calculate the thermodynamic properties of a RNA molecule with a given sequence of bases by taking from the literature the estimated values of the relevant pairing and stacking energies. Clearly, such a calculation can only be performed numerically. Here, we follow another route with the main purpose of gaining a more detailed insight into a simplified version of our model. Such an approach can give useful information against which the results of a full numerical calculation can be understood. In the simplified model we neglect the stacking energies and the bending rigidity. Moreover, we modify the nucleotide dependence of the pairing energies. A common approach in the physics literature is to take it as random [8–11]. In this way, the study of RNA can be linked to that of random systems such as spin glasses. However, clear differences between random and real RNA have been pointed out [16]. The latter has an evolutionarily evolved, correlated sequence of bases, which is such that most often the ground state secondary structure is less degenerated than that of random RNA. An expression for the pairing energy which has this feature and which lends itself to detailed analysis was introduced in Ref. [6]. Following that work, we take

$$\varepsilon_{i,j} = \varepsilon_0 + \tilde{\varepsilon} \delta_{i+j,L+1} \quad (4)$$

(where  $\delta$  is the Kronecker-delta and from now on we take  $L$  even). In the physical region of interest  $\varepsilon_0 < 0$  and  $\tilde{\varepsilon} < 0$ . The second term in Eq. (4) favors the formation of just one hairpin. This structure thus corresponds to the native state of our model. With all these simplifications, Eq. (1) becomes

$$E_T = \varepsilon_0 I + \tilde{\varepsilon} N, \quad (5)$$

where  $I$  is the total number of bonded base pairs and  $N$  the total number of *native* interactions (i.e., those pairs  $i,j$  for which  $\delta_{i+j,L+1} = 1$ ). Finally, we introduce  $q = \exp(-\beta \varepsilon_0)$  and  $\tilde{q} = \exp(-\beta \tilde{\varepsilon})$ , so that the partition sum (2) becomes

$$Z_L(q, \tilde{q}) = \sum_T q^I \tilde{q}^N. \quad (6)$$

In the rest of this paper we study the phase diagram of the model defined by Eq. (6).

### III. THE PHASE DIAGRAM

The two-tolerant trail was originally introduced as a simple model for the coil-globule transition of homopolymers [17]. The authors of that work studied the two-tolerant trails with attractive self-interactions, which corresponds to our model for  $\tilde{q} = 1$ . A closer investigation [18] revealed however that in the low-temperature phase, the universal properties of the trail are not those of a collapsed globular polymer, but coincide with those of branched polymers (BP's). In the high-temperature phase it was found [18] that the universality class of two-tolerant trails is that of the self-avoiding walk (SAW) [13]. These results were based on ex-



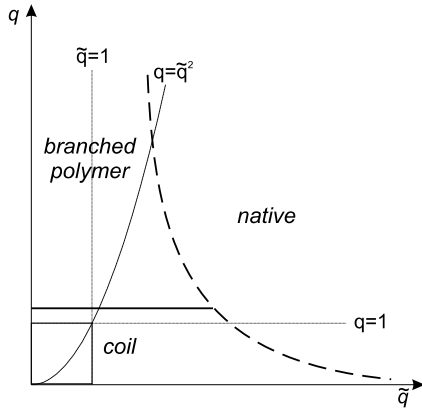


FIG. 3. Schematic phase diagram of our model. The thick full line is a line of second-order transitions, whereas along the broken line the transition is first order. The thin full lines indicate the three paths along which we investigated the model numerically.

$$n(q, \tilde{q}) = 2 \lim_{L \rightarrow \infty} \frac{\langle N \rangle}{L} = 2 \lim_{L \rightarrow \infty} \frac{\sum_T N q^L \tilde{q}^N}{LZ(q, \tilde{q})}. \quad (14)$$

The factor 2 ensures that  $n(q, \tilde{q}) = 1$  when the two-tolerant trail consists of one hairpin. In terms of the free energy this density equals

$$n(q, \tilde{q}) = 2 \frac{\partial f(q, \tilde{q})}{\partial \tilde{q}}. \quad (15)$$

Thus, from our earlier results on free energy we find that  $n(q, \tilde{q}) = 0$  for  $q < \tilde{q}_c(q)$  while this density becomes strictly positive above  $\tilde{q}_c(q)$ . We therefore interpret the phase transition at  $\tilde{q}_c(q)$  as a denaturation transition, i.e., a transition into (or out of) the native state. Also note that this transition exists for every value of  $q$ .

As discussed in the beginning of this section, numerical evidence shows that there is also a transition between a self-avoiding walk regime and a BP one along the line  $\tilde{q} = 1$ , at some critical value  $q_c(1)$ . Clearly, since the free energy does not depend on  $\tilde{q}$  for  $\tilde{q} < \tilde{q}_c(q)$ , the SAW-BP transition has to be also present for some range of  $\tilde{q}$  values, and moreover neither the location of the transition point nor its critical properties can depend on  $\tilde{q}$ .

We therefore arrive at the phase diagram shown in Fig. 3. There are three phases. At low  $q$  (or  $|\varepsilon_0|$ ) and  $\tilde{q}$  (or  $|\tilde{\varepsilon}|$ ), RNA is denaturated and behaves as a coil in the universality class of the self-avoiding walk. For  $q > q_c(1)$ , and for  $\tilde{q}$  sufficiently small, there is a collapse into a branched polymer or “molten” phase. Finally for  $\tilde{q}$  sufficiently large, we are in the native, hairpin phase. The location and the nature of the phase transitions drawn in Fig. 3 follow from the numerical data presented in the following section. For the transition into the native state we were also guided by the bounds derived above.

The arguments we have given are quite general and we therefore believe that this phase diagram is correct, independent of dimension. While the denaturation transition is always present, it is possible that the SAW-BP transition disappears above some critical dimension. As an example of this we mention that in the model of Ref. [6] where pseudoknots and self-avoidance are neglected, and which thus can be seen as a mean-field model, no evidence for the coil phase is found, and only the transition between the native and the molten phase is present. It is interesting to remark that in that work it was also found that in the molten phase, RNA has the properties of (mean-field) branched polymers, as is the case here.

#### IV. NUMERICAL RESULTS

In order to get a more quantitative insight into the phase diagram, we have investigated our model in the  $(q, \tilde{q})$  plane with extensive Monte Carlo simulations. Further numerical insight was obtained from an exact enumeration of our model for  $L \leq 18$ . In these numerical approaches, we worked on a square lattice. This is not a severe approximation when one is only interested in secondary structures. Indeed, it represents already a significant step forward in comparison with the mean-field approaches used so far. Going to three dimensions could change some of the critical exponents of the model, but here we are mostly interested in qualitative features of the phase diagram.

For the simulations, we used the pivot algorithm [24], a well-known technique that generates a Markov chain in the set of all allowed walks of a certain type. The pivot algorithm was originally introduced for self-avoiding walks. In Ref. [25] we investigate the extension of this method to non-interacting two-tolerant trails, discussing such aspects as ergodicity of the algorithm, acceptance ratio, autocorrelation times, and so on.

Here we have to take into account interactions and do this by adding a standard Metropolis step to the algorithm. Moreover, we implement the program within a multiple Markov chain (MMC) approach [26]. In this method several Markov chains at different temperatures are run in parallel and at regular intervals an attempt is made to switch two trails between Markov chains at different temperatures. Such an attempt is accepted with a probability that is a trivial extension of that of the Metropolis algorithm. The MMC approach has the advantage that it allows better sampling at lower temperatures, where a standard algorithm may easily get stuck near a metastable state. Besides the pivot moves, we also found it useful to add some local moves to enhance the performance of the algorithm. Typically, at each temperature we performed  $10^9$  Monte Carlo steps resulting in  $(0.5-1.0) \times 10^6$  independent configurations over which averages are calculated.

With this approach, we investigated our model along three lines. For these, we chose  $q = 1$ ,  $\tilde{q} = 1$ , and  $q = \tilde{q}^2$ . We now discuss the results.



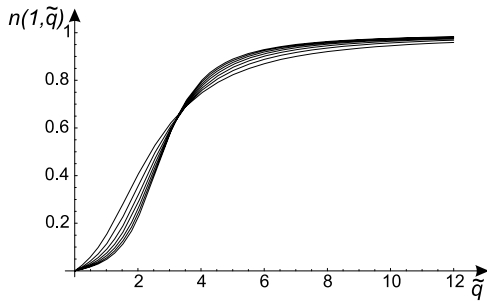


FIG. 4. The number of native contacts as a function of  $\tilde{q}$  along the line  $q=1$  for different  $L$  values, as obtained from exact enumerations.

### A. The line $q=1$

On the basis of the phase diagram obtained in the preceding section, we expect that along this line there is only the denaturation transition. Evidence for this transition can most easily be found by looking at the density of native contacts. In Fig. 4 we show our results for  $n(1, \tilde{q})$  for two-tolerant trails of different  $L$  as obtained from the exact enumeration data. We clearly recognize the behavior predicted in Sec. III, dressed with finite size roundings. From the intersections of the curves for different  $L$  values a first estimate for the location of the critical point can be made. Moreover, since the transition into the native state shares some properties with the adsorption transition of a polymer onto a surface, we expect that right at the critical point the density of contacts scales as  $n \sim L^{\varphi_n - 1}$ , where  $\varphi_n$  is a crossover exponent. From the exact enumeration data shown in Fig. 4, we estimate  $\varphi_n \approx 0.88$ . However, a study of the same quantity with the Monte Carlo approach shows that this estimate is still strongly affected by finite size effects. For example, the value for  $\varphi_n$  tends to increase to a value close to 1. This suggests that the actual value of  $\varphi_n$  is 1, which would be the case for a first-order transition. To verify this idea we made histograms for our data for  $n$  at different  $\tilde{q}$  values. These are indeed consistent with a first-order transition. As an example, we show in Fig. 5 such a histogram at the transition point. There is clear evidence for two peaks, one near  $n=0$ , the other around  $n \approx 0.65$ . In fact, it turns out that from these

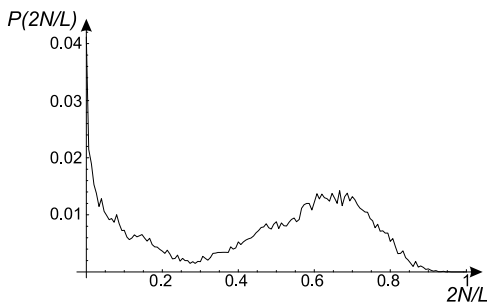


FIG. 5. Histogram for the density of native contacts at the denaturation transition for  $L=300$  ( $q=1$ ). The vertical axis indicates the fraction of simulated configurations with a given number of native contacts.

histograms, a rather sharp estimate for  $\tilde{q}_c$  can be obtained, with the result  $\tilde{q}_c = 4.20 \pm 0.02$ .

The fact that the denaturation transition is first order can also be understood with hindsight from a comparison with known results from DNA models. Indeed, in recent years there has been quite some interest in understanding the nature of the denaturation transition for that biopolymer. Almost all existing evidence now shows that this transition is first order, both in  $d=2$  and  $d=3$  [27–29]. When within our model we divide the two-tolerant trail into two halves, we can see them as the two strands of a DNA molecule whose starting point is halfway on the two-tolerant trail. The native interactions, which are the only ones appearing along the line  $q=1$ , can thus be interpreted as interactions between homologous bases on the two strands of the DNA. In this way, our model with  $q=1$  becomes in a sense the dual of a recently studied lattice model of DNA [27,29], and we can therefore expect both models to have a similar critical behavior.

### B. The line $\tilde{q}=1$

The phase transition between the coil and branched polymer regime along the line  $\tilde{q}=1$  is more difficult to analyze. There is no obvious order parameter characterizing this transition, since the average number of bonded base pairs,  $\langle I \rangle$ , is extensive on both sides of the transition.

We therefore investigated two other quantities. First, we looked at the specific heat, which for  $\tilde{q}=1$  equals

$$C_L(q) = \frac{1}{L} \left[ \frac{\sum_T I^2 q^I}{\sum_T q^I} - \left( \frac{\sum_T I q^I}{\sum_T q^I} \right)^2 \right]. \quad (16)$$

While both the exact enumerations and the Monte Carlo simulations show that the specific heat has a peak that slowly grows with  $L$ , it is difficult to obtain reliable estimates for the location of the critical point and for the crossover exponent from these data. Nevertheless, the small value of the specific heat exponent shows that the crossover exponent associated with the SAW-branched polymer transition is much smaller than 1, evidence that this transition is second order.

Second, it is possible to get information on the SAW-BP transition from the ratio  $Y(q,1)$  of the average squared end-to-end distance over the average squared radius of gyration. It is well known that this is a universal quantity so we expect its behavior to be stepwise as a function of  $q$ , at least for  $L \rightarrow \infty$ . In fact, since for large  $q$  the two-tolerant trail visits each edge twice (see Fig. 2), the end-to-end distance approaches zero and hence  $Y(q,1)$  goes to zero for large  $L$ . On the other hand, we verified recently that for noninteracting two-tolerant trails  $Y(1,1) = 7.1235 \pm 0.001$  [25], fully consistent with the value for the SAW. Hence along the line  $\tilde{q}=1$ ,  $Y(q,1)$  should assume this value below  $q_c(1)$ , and then drop to zero. In Fig. 6 we present our data for  $Y(q,1)$ , which have the expected behavior, dressed with finite size round-

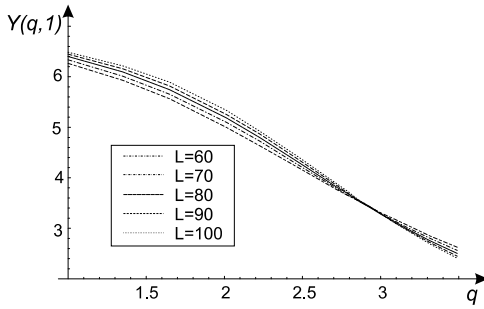


FIG. 6. Plot of  $Y(q,1)$  (see text) for different  $L$  values ( $L=60$ , dotted line;  $L=70$ , thick dashed line;  $L=80$ , full line;  $L=90$ , dash-dotted line; and  $L=100$ , thin dashed line).

ing. From these results we are able to obtain the most accurate estimate for  $q_c(1)$  which equals  $2.91 \pm 0.08$ .

In Fig. 7 we show data for the squared radius of gyration  $R_G^2(L)$  as a function of  $L$  in the BP phase. From this we can obtain the geometric exponent  $\nu$  since

$$R_G^2(L) \sim L^{2\nu}. \tag{17}$$

From a fit of the exact enumeration and Monte Carlo data in the BP phase we find  $\nu \approx 0.55$ . This is still far from the best known value for two-dimensional branched polymers which is  $\nu = 0.64075 \pm 0.00015$  [30]. This difference is probably due to strong corrections to scaling. Indeed, this also happens for the noninteracting situation where very long trails can be simulated, up to  $L=7500$ . From them we obtain  $\nu = 0.749 \pm 0.001$ , as should be expected for a walk in the SAW universality class. However, this exponent is only recovered for  $L > 200$ . We expect that the  $\nu$  exponent for branched polymers will show up if one studies longer trails at temperatures sufficiently below the transition. But that regime is difficult to probe with our Monte Carlo technique.

### C. The line $q = \tilde{q}^2$

From the phase diagram shown in Fig. 3, from the available estimates for  $q_c(1)$ , and from Eq. (13) we expect that along this line two phase transitions will be encountered. The first one is the SAW-BP transition, which can be analyzed most suitably from an investigation of the ratio  $Y(q, \tilde{q})$ . The

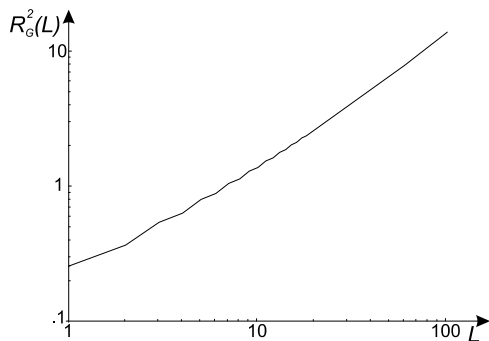


FIG. 7. Plot of  $R_G^2(L)$  vs  $L$  in the BP phase at  $q = 3.49$  and  $\tilde{q} = 1$ .

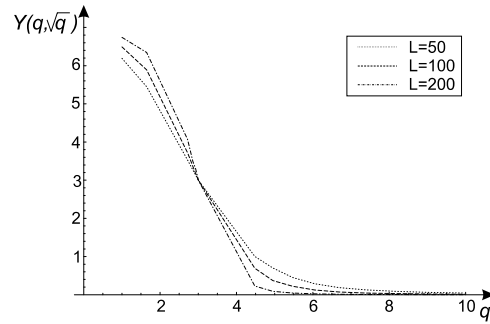


FIG. 8. Plot of  $Y(q, \sqrt{q})$  for different  $L$  values ( $L=50$ , dotted line;  $L=100$ , dashed line; and  $L=200$ , dashed-dotted line).

BP-native transition, on the other hand, should show itself through a study of the density of native contacts.

In Fig. 8 we show our results for  $Y(q, \sqrt{q})$ . As was the case before, the data for different  $L$  values intersect, yielding the estimate  $q_c = 3.00 \pm 0.06$ . This result is within the numerical accuracy the same as that for  $\tilde{q} = 1$ , as was predicted in Sec. III.

From a study of the density of native contacts, we conclude that the transition between the branched polymer and the native phase is also first order. Since in this case it is a transition between two rather dense phases, it is more difficult to obtain a reliable estimate for  $\tilde{q}_c$ . We find  $\tilde{q}_c = 2.10 \pm 0.15$ . Figure 9 shows a histogram for the number of native contacts at this point.

## V. CONCLUSIONS

In this paper, we have introduced a lattice model, based on a two-tolerant trail that seems well suited to investigate the secondary structure of RNA. We have found that for a simple version of the interaction energies, there are three phases and we have investigated in detail the transitions between these phases with a Monte Carlo method.

We believe that the structure of the phase diagram that we found here is not particular for the choice of the interaction energy (4) but would be similar also for other choices of  $\epsilon_{i,j}$  that break the homogeneity of the interaction energies. Indeed, it is known that, at least in mean-field theory [10], the inclusion of a random part in the interaction energy gives rise

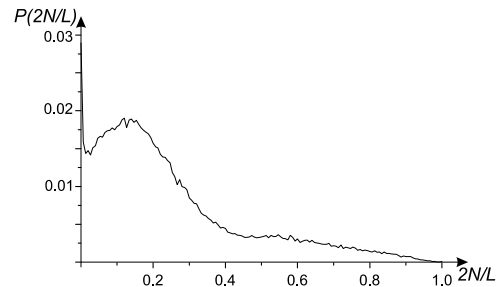


FIG. 9. Histogram for the number of native contacts at the BP-native transition along the line  $q = \tilde{q}^2$ . The data are for  $L=200$  and  $\tilde{q} = 2.10$ . The vertical axis indicates the fraction of simulated configurations with a given number of native contacts.

to the appearance at low temperature of a spin-glass-like phase. This phase then plays the role of the native one. It therefore seems quite possible that taking into account self-avoidance and for rather general choices of nonhomogeneous interaction energies, one recovers the three phases found here.

Upon lowering the temperature at fixed values of the interaction energies  $\varepsilon_{i,j}$  there are thus two possible scenarios. Either one goes directly into the native regime, or one goes through an intermediate molten phase. We believe that the first-order transition line approaches the line  $\tilde{q}=1$  when  $q \rightarrow \infty$ , although we could not prove this and simulations in this regime are difficult with the pivot algorithm. If this belief is correct it seems that the molten phase, where a description in terms of homogeneous interactions is correct, can never be the stable one at very low temperatures.

The coil or SAW phase was not present in simpler, mean-field models of RNA such as those studied in Refs. [6–8,10]. Moreover, our model does not show the divergence of the radius of gyration with  $L$  which occurs in the low-temperature phase of simpler models [7]. Within our model, the  $\nu$  exponent remains at its SAW value in the native phase. This is of course a peculiarity of the simple interaction which we choose and will probably not hold anymore when a more realistic version for the bonding energies  $\varepsilon_{i,j}$  is chosen. Moreover, if one wants to determine the true value of the exponent  $\nu$  it is necessary to introduce all relevant interactions, also those considered to be only relevant for tertiary structure. But clearly the precise value of critical exponents is not a major concern from a biochemical point of view.

We believe that the main interest of our model lies in further applications that go beyond the work presented here and that our results can be used as a starting point for studies which are of more interest from the point of view of molecular biophysics. First, still within the context of the simplified model we are investigating the probability of occurrence of pseudoknots. It is intuitively obvious that in the hairpin phase the occurrence of pseudoknots will be severely sup-

pressed. But one would be interested in obtaining a more quantitative insight into this issue.

Another application for which we think that our model can be useful is an investigation of the elastic properties of RNA. These have been measured recently using micromanipulation techniques [31]. The theoretical study of elastic properties of biomolecules is usually performed within simple continuum models such as the wormlike chain (WLC) [32]. This is certainly a very good approach when effects of self-avoidance can be neglected. For noninteracting models of polymers, it is known [33] that as soon as an infinitesimal force is applied, the polymer becomes stretched and in such a regime it can be expected that it can be described by the WLC or in terms of directed polymers. However, for homopolymers below the  $\theta$  point it has been established that they undergo a transition to a stretched phase only for forces greater than a critical force  $F_c > 0$  [34,35]. Then, in the whole region where the forces are below this threshold, effects of self-avoidance are of importance. We expect that a similar scenario might hold within the low-temperature phases (BP and native) of RNA. That self-avoidance can be of importance in stretching experiments on single stranded DNA (which is much like RNA) and proteins was also recently pointed out in Ref. [36].

Finally, the scenario which we have found here for the denaturation transition could be quite general and also be of relevance for proteins. Also, in that case it is possible that depending on the ratios of relevant interactions, the denaturation transition takes place immediately, or through an intermediate molten phase. Indeed, models of proteins showing this kind of behavior have been studied in the past (see, e.g., Refs. [37,38]).

#### ACKNOWLEDGMENT

We would like to thank R. Kawai for the use of his computer cluster. P.L. thanks the FWO-Vlaanderen for financial support.

- 
- [1] For a review, see the articles in *Nature*, (London) **418**, 213 (2002), special issue on RNA, edited by R. Turner.
  - [2] B. Alberts *et al.*, *Molecular Biology of the Cell* (Garland Science, New York, 2002).
  - [3] I. Tinoco and C. Bustamante, *J. Mol. Biol.* **293**, 271 (1999).
  - [4] P.G. Higgs, *Q. Rev. Biophys.* **33**, 199 (2000).
  - [5] J.S. McCaskill, *Biopolymers* **29**, 1105 (1990).
  - [6] R. Bundschuh and T. Hwa, *Phys. Rev. Lett.* **83**, 1479 (1999).
  - [7] M. Müller, *Phys. Rev. E* **67**, 021914 (2003).
  - [8] A. Pagnani, G. Parisi, and F. Ricci-Tersenghi, *Phys. Rev. Lett.* **84**, 2026 (2000).
  - [9] R. Bundschuh and T. Hwa, *Europhys. Lett.* **59**, 903 (2002).
  - [10] R. Bundschuh and T. Hwa, *Phys. Rev. E* **65**, 031903 (2002).
  - [11] F. Krzakala, M. Mézard, and M. Müller, *Europhys. Lett.* **57**, 752 (2002).
  - [12] H. Orland and A. Zee, *Nucl. Phys. B: Field Theory Stat. Phys.* **620**[FS], 456 (2002); M. Pilsbury, H. Orland, and A. Zee, e-print physics/0207110.
  - [13] C. Vanderzande, *Lattice Models of Polymers* (Cambridge University Press, Cambridge, UK, 1998).
  - [14] This idea has been introduced independently by M. Baiesi, E. Orlandini, and A.L. Stella, *Phys. Rev. Lett.* (to be published), e-print cond-mat/0303226.
  - [15] To make the model even more realistic, one could take the value of the bending rigidity higher when both the neighboring bonds are doubly visited. This would correspond to a different bending rigidity for the single and for the double stranded parts of RNA.
  - [16] P.G. Higgs, *J. Phys. I* **3**, 43 (1993).
  - [17] J.A. Marqusee and J.M. Deutch, *J. Chem. Phys.* **75**, 5179 (1981).
  - [18] E. Orlandini, F. Seno, A.L. Stella, and M.C. Tesi, *Phys. Rev. Lett.* **68**, 488 (1992).
  - [19] P. Leoni, Ph.D. thesis, Limburgs Universitair Centrum, 2003 (unpublished).

- [20] N. Madras and G. Slade, *The Self-Avoiding Walk* (Birkhäuser, Boston, 1993).
- [21] K. De'Bell and T. Lookman, *Rev. Mod. Phys.* **65**, 87 (1993).
- [22] H.A. Lim and H. Meirovitch, *Phys. Rev. A* **39**, 4176 (1989).
- [23] D.S. Gaunt and H. Ruskin, *J. Phys. A* **11**, 1369 (1978).
- [24] N. Madras and A. Sokal, *J. Stat. Phys.* **50**, 109 (1988).
- [25] P. Leoni and C. Vanderzande (unpublished).
- [26] M.C. Tesi, E.J. Janse van Rensburg, E. Orlandini, and S.G. Whittington, *J. Stat. Phys.* **82**, 155 (1996).
- [27] M.S. Causo, B. Coluzzi, and P. Grassberger, *Phys. Rev. E* **62**, 3958 (2000).
- [28] Y. Kafri, D. Mukamel, and L. Peliti, *Phys. Rev. Lett.* **85**, 4988 (2000).
- [29] E. Carlon, E. Orlandini, and A.L. Stella, *Phys. Rev. Lett.* **88**, 198101 (2002).
- [30] B. Derrida and D. Stauffer, *J. Phys. (Paris)* **46**, 1623 (1985).
- [31] J. Liphardt *et al.*, *Science* **292**, 733 (2001).
- [32] S. Cocco, J.F. Marko, and R. Monasson, *CR Physique* **3**, 1 (2002).
- [33] P.G. de Gennes, *Scaling Concepts in Polymer Physics* (Cornell University Press, Ithaca, 1979).
- [34] P. Grassberger and H.P. Hsu, *Phys. Rev. E* **65**, 031807 (2002).
- [35] D. Marenduzzo, A. Maritan, A. Rosa, and F. Seno, *Phys. Rev. Lett.* **90**, 088301 (2003).
- [36] T.R. Strick, M.-N. Dessinges, G. Charvin, N.H. Dekker, J.-F. Allemand, D. Bensimon, and V. Croquette, *Rep. Prog. Phys.* **66**, 1 (2003).
- [37] S. Doniach, T. Garel, and H. Orland, *J. Chem. Phys.* **105**, 1601 (1996).
- [38] J. Chahine, H. Nymeyer, V.B.P. Leite, N.D. Socci, and J.N. Onuchic, *Phys. Rev. Lett.* **88**, 168101 (2002).