Exactness of the annealed and the replica symmetric approximations for random heteropolymers

Ugo Bastolla[1,2] and Peter Grassberger[1]
[1]*HLRZ, Forschungszentrum Jülich, D-52425 Jülich, Germany*
[2]*Max Planck Institute for Colloids and Interfaces, D-14424 Potsdam, Germany*
(Received 29 March 2000; published 14 February 2001)

We study a heteropolymer model with random contact interactions introduced some time ago as a simplified model for proteins. The model consists of self-avoiding walks on the simple cubic lattice, with contact interactions between nearest-neighbor pairs. For each pair, the interaction energy is an independent Gaussian variable with mean value B and variance Δ^2. For this model the annealed approximation is expected to become exact for low disorder, at sufficiently high dimension and in the thermodynamic limit. We show that corrections to the annealed approximation in the three-dimensional high-temperature phase are small, but do not vanish in the thermodynamic limit, and are in good agreement with our replica symmetric calculations. Such corrections derive from the fact that the overlap between two typical chains is nonzero. We explain why previous authors had come to the opposite conclusion, and discuss consequences for the thermodynamics of the model. Numerical results were obtained by simulating chains of length $N \leqslant 1400$ by means of the recent PERM algorithm, in the coil and molten globular phases, well above the freezing temperature.

DOI: 10.1103/PhysRevE.63.031901 PACS number(s): 87.15.Aa

I. INTRODUCTION

Apart from their extreme biological importance, proteins are also very interesting objects from the point of view of statistical mechanics. They possess a very well-defined native structure, which they are able to find in a short time among a potentially huge number of competing ones, and in spite of many metastable states. How proteins reconcile the stability of the native structure with the requirement that this structure is rapidly reached constitutes the essence of the fascinating and still open protein folding problem [1].

An interesting question is whether the property of folding is a generic property of randomly assembled polypeptidic chains, regardless of their biological function, or is a special property that has evolved through natural selection. This kind of question makes the protein folding problem a bridge between theoretical biology and the statistical mechanics of disordered systems. Motivated by this, numerous authors have studied simple models of random heteropolymers [2–18], see Ref. [19] for a review.

In the following, we shall discuss only the ''random bond model'' introduced independently by Garel and Orland [3] and by Shakhnovich and Gutin [4]. More precisely, in our numerical simulations we will study a lattice version of this model. Preliminary results of this paper have already been presented in Ref. [20]. A ''protein'' with $N+1$ ''amino acids'' is represented as a self-avoiding walk [21] of N steps on the simple cubic lattice. Each pair (i,j) of nonbonded monomers on nearest-neighbor lattice sites contributes to the total energy an amount given by an independent and identically distributed Gaussian variable B_{ij} with mean B' and variance Δ'^2. Formally, one defines the contact map of configuration $\mathcal{C}$, $\sigma(\mathcal{C})$, as the matrix of binary variables $\sigma_{ij} \in \{0,1\}$, with $i,j \in \{0, \ldots N\}$, such that

$$\sigma_{ij}(\mathcal{C}) = \begin{cases} 1, & \text{if } i \text{ and } j \text{ are in contact and nonbonded} \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

The energy of the model can then be written as

$$E(\mathcal{C},\{B\}) = \sum_{i<j} \sigma_{ij}(\mathcal{C}) B_{ij}, \tag{2}$$

with $\overline{B_{ij}} = B'$, $\overline{B_{ij}^2} - \overline{B_{ij}}^2 = \Delta'^2$. For a given realization of the interaction energies B_{ij} (representing a protein sequence in the biological analogy), the partition sum Z_N at temperature T can be formally computed as [23]

$$Z_N\{B_{ij}\} = \sum_{\mathcal{C}} \exp - E(\mathcal{C},\{B\})/k_B T, \tag{3}$$

where the sum over configurations $\mathcal{C}$ runs over all self-avoiding N-step walks. Obviously, the above expression depends only on the variables $\Delta = \Delta'/k_B T$ and $B = B'/k_B T$, i.e., we have a two-parameter phase diagram in the variables B and Δ. The main advantage in using B as one of the independent variables instead of T or $\beta = 1/k_B T$ is that we can pass continuously from positive (repulsive, hydrophilic) to negative (hydrophobic) B.

As usual with random models, we have to evaluate the quenched average of the free energy. This is a very difficult task, while it is rather easy to perform an annealed average over the disorder. For several models of random spin systems it is well known that such an annealed approximation becomes exact in the high-temperature phase, in the thermodynamic limit, and at sufficiently large dimension. The same is thought to be true for the present model. It was indeed predicted in Ref. [4] that the annealed approximation becomes exact in three dimensions when the chain length tends to infinity. For this to be true it is necessary that the overlap between two randomly chosen replicas with the same realization of disorder vanishes in the limit $N \to \infty$.

Numerical tests of this prediction have been made in the past for chains of length $\leqslant 36$, mostly by means of exact enumerations of maximally compact chains of length 27 [14,15]. These authors found deviations (replica overlap is

nonzero) which seemed to decrease with N. A similar result even for $d=2$ was found in Ref. [18], where exact enumeration of very short chains were used (up to $N=22$). But it is clear that tests with such short chains can hardly be significant. In the present paper we shall present Monte Carlo simulations for chains of length up to $N=1400$. These simulations are made with the PERM algorithm developed recently by one of us [25], and applied successfully to a number of different polymer problems [26–29].

We study the corrections to the annealed approximation using two different approaches. First, we compute them using the replica method and assuming replica symmetry, which is believed to hold for low disorder. Even if a full computation was not possible, the expected behavior was well confirmed by numerical simulations. Second, we notice that corrections to the annealed approximation in the weak disorder limit can be related exactly to the average overlap between pairs of homopolymers (without any disorder). We give strong theoretical arguments that this overlap does not vanish in the limit $N\to\infty$. We also calculate it by means of Monte Carlo simulations. Unlike in the previous case, these simulations do not involve the averaging over the disorder and thus can be applied to larger systems.

The two methods agree with each other and show that the corrections to the annealed approximation are small in $d=3$, but do not vanish in the thermodynamic limit. Deviations from the annealed approximation are larger in the coil (high-temperature) phase and very small in the collapsed (globular) phase.

The annealed approximation is presented in Sec. II and compared to results of Monte Carlo simulations. In order to explain the observed deviations, we study in Sec. III a scenario where the overlap is nonzero but replica symmetry is unbroken. We again compare theoretical predictions with simulation results. The relationship between the weak disorder limit and homopolymer overlap is discussed in Sec. IV. Additional thermodynamic considerations are presented in Sec. V, and our final conclusions are drawn in Sec. VI. The PERM algorithm used for the simulations is discussed in an appendix.

II. ANNEALED APPROXIMATION

In thermodynamic systems with quenched disorder we have to consider the average of the free-energy per monomer over individual realizations of disorder $\{B_{ij}\}$, which formally is given by

$$\begin{aligned} F_N(B,\Delta) &= -\frac{1}{\beta N}\,\overline{\ln[Z_N\{B_{ij}\}]} \\ &\equiv -\frac{1}{\beta N}\prod_{i<j}\int dB_{ij}\frac{\exp-(B_{ij}-B)^2/2\Delta^2}{\Delta\sqrt{2\pi}}\ln(Z_N\{B_{ij}\}). \end{aligned} \tag{4}$$

As for most random systems, this cannot be evaluated in closed form. Much easier to evaluate is the annealed approximation

$$F_{N,\mathrm{ann}}(B,\Delta) = -\frac{1}{N}\,\ln\,\overline{Z_N} \tag{5}$$

obtained by taking the disorder average before taking the log. Here the Gaussian integrals can be done explicitly, with the result

$$\overline{Z_N} = \sum_{\mathcal{C}}\,\exp-\left(B-\frac{1}{2}\Delta^2\right)\sum_{i<j}\,\sigma_{ij}(\mathcal{C}). \tag{6}$$

Since this is the partition sum for a homopolymer with pair energy

$$\tilde{B} = B-\tfrac{1}{2}\Delta^2, \tag{7}$$

we see that [4]

$$F_{N,\mathrm{ann}}(B,\Delta) = F_N(\tilde{B},0). \tag{8}$$

Therefore, all thermodynamic variables can be expressed in the annealed approximation in terms of an equivalent homopolymer with shifted interaction strength. This relationship is easiest for those observables whose definition does not involve a derivative with respect to temperature, such as the gyration and end-to-end radii, and the density of nonbonded nearest-neighbor contacts c. The latter is defined as the average number of nn contacts between nonconsecutive monomers divided by N. For these observables, we have

$$R_{N,\mathrm{ann}}(B,\Delta) = R_N(\tilde{B},0) \tag{9}$$

and

$$c_{\mathrm{ann}}(B,\Delta) = c(\tilde{B},0) \equiv \tilde{c}, \tag{10}$$

precisely as in Eq. (8).

For energy U and entropy S the relations are less simple, since these involve derivatives of the free energy with respect to T, which are changed into derivatives with respect to B and Δ by our convention of using $T=1$. For the energy per monomer it holds

$$U_{N,\mathrm{ann}}(B,\Delta) = \frac{B-\Delta^2}{\tilde{B}}U_N(\tilde{B},0) = (B-\Delta^2)\tilde{c}, \tag{11}$$

where we used the fact that the energy for homopolymers is $U_N(\tilde{B},0)=\tilde{c}\tilde{B}$. For the specific entropy $S_N(B,\Delta) = -(\partial/\partial T)F_N(B,\Delta,T)|_{T=1}$ we use $F_N(B,\Delta,T) = TF_N(B/T,\Delta/T,1)$ together with Eq. (8), and obtain

$$S_{N,\mathrm{ann}}(B,\Delta) = S_N(\tilde{B},0) - \frac{\Delta^2}{2\tilde{B}}U_N(\tilde{B},0). \tag{12}$$

The number of configurations with fixed c should increase as $\exp[Nf(c)]$ for large N, i.e., $f(c)$ is the entropy density in the fixed-N, fixed-c ensemble. For homopolymers, the ensemble

with fixed $\tilde{B}$ becomes equivalent to the fixed-c ensemble in the limit $N\to\infty$. Thus c becomes a nonfluctuating function of $\tilde{B}$, $c=c(\tilde{B})\equiv\tilde{c}$, and the above formula becomes simply

$$S_{N,\mathrm{ann}}(B,\Delta)=f(\tilde{c})-\frac{\Delta^2}{2}\tilde{c} \quad \text{for} \quad N\to\infty, \tag{13}$$

where $c(\hat{B})$ is the solution of the saddle-point equation

$$f'(\tilde{c})\equiv\left.\frac{\partial f(c)}{\partial c}\right|_{c=\tilde{c}}=\tilde{B}. \tag{14}$$

The condition for thermodynamic stability is that the second derivative of f should be negative, corresponding to F being minimal. This is equivalent to requiring that the specific heat is positive. In fact, the specific heat for a homopolymer is given by

$$C_V=B\frac{\partial c}{\partial T}=-B^2\left(\frac{\partial^2 f}{\partial c^2}\right)^{-1}, \tag{15}$$

which has been obtained by deriving both sides of Eq. (14) with respect to T.

Homopolymers with attraction between unbonded nearest neighbors show a collapse (''theta'') transition where the specific heat diverges in the limit $N\to\infty$. Thus we expect that the second derivative $\partial^2 f/\partial c^2$ vanishes at the theta point $c=c_\theta$ (the precise value of the transition point depends on the lattice considered).

The annealed approximation is supposed to be valid both above and below the theta transition. At very low temperatures and very large disorder, it has to break down since otherwise the entropy would become negative, according to Eq. (12). This signals another phase transition, the so-called freezing transition. We shall not discuss this regime in this paper, but will treat it in a forthcoming publication.

Since the theta point is a tricritical point [21,25], its upper critical dimension is $d=3$. Therefore, we expect that in three dimensions the ''swelling factor'' is constant,

$$\langle R^2\rangle/N\approx\text{const} \tag{16}$$

at the theta point, up to logarithmic corrections [30,25,31]. Here, R is any measure of the size of the polymer, such as the end-to-end distance or the gyration radius. We expect that this is still true for heteropolymers, as long as we are not yet in the frozen regime. While Eq. (16) gives the most precise numerical estimate of the collapse transition [with $B_\theta=-0.2690\pm0.0002$ Ref. [25]], estimates with similar precision can be obtained from the convexity of the free energy [28], and the volume dependence of the free energy in case of periodic boundary conditions [26].

The collapse line in the (B,Δ) plot obtained from simulating chains of length $N\leqslant1000$, is shown in Fig. 1. Here the solid line is the annealed approximation. We see that on this scale the annealed approximation seems perfect. But this is not quite true. Much more precise tests can be performed by comparing directly both sides of Eqs. (8) to (11). Typical plots obtained in this way are shown in Fig. 2. For each of these plots we used at least 300 realizations of disorder, and we got at least 10^3 independent configurations for each disorder realization. Similar plots were made also for several other values of B and Δ.

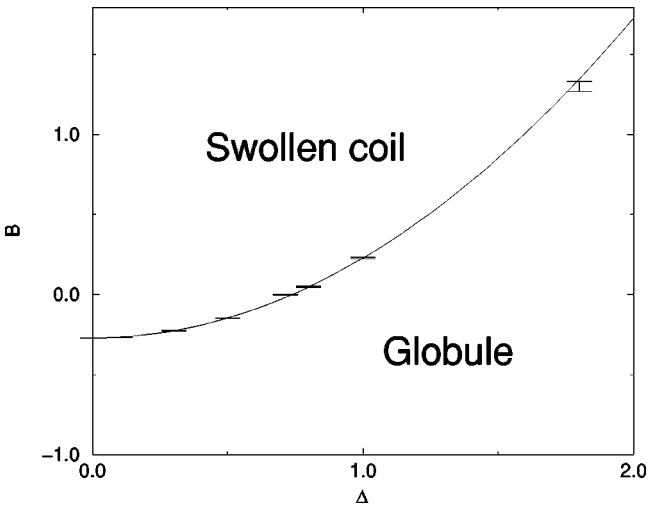

FIG. 1. Collapse transition line. The solid line is the annealed prediction, $B-\Delta^2/2=B_\theta=-0.269$. Numerical data are obtained by measuring the end-to-end distance. At $\Delta\geqslant2.5$ (not shown) there are significant deviations from the annealed approximation, that are most likely due to a direct freezing from the swollen phase.

In all these plots we see small but significant deviations. These deviations are present both in the collapsed (globular) and in the open (coil) phase. They depend only weakly on N. Therefore, even with our long chains and high statistics it is not clear whether they disappear for $N\to\infty$. Obviously, in order to proceed we need more refined theoretical predictions to compare with, and/or a more efficient way to do the disorder average.

Before we do this, we should point out that deviations from the annealed approximation were also found recently in a different model by Trovato *et al.* [22].

III. REPLICA SYMMETRIC APPROXIMATION

To go beyond the annealed approximation, we will use the replica trick

$$\overline{\ln Z_N}=\lim_{n\to0}\frac{\overline{Z_N^n-1}}{n}. \tag{17}$$

Alternatively, we could try a Taylor expansion

$$\overline{\ln(Z_N/\overline{Z_N})}=-\tfrac{1}{2}(\overline{Z_N^2}/\overline{Z_N}^2-1)+\ldots. \tag{18}$$

This expansion is most likely divergent. It is nevertheless useful since its first term gives already a good indication of the leading corrections. Also, it suggests the inequality

$$F_N(B,\Delta)\geqslant F_{N,\mathrm{ann}}(B,\Delta). \tag{19}$$

which can easily be derived exactly from convexity of the logarithm. The same inequality is expected to hold for U_N.

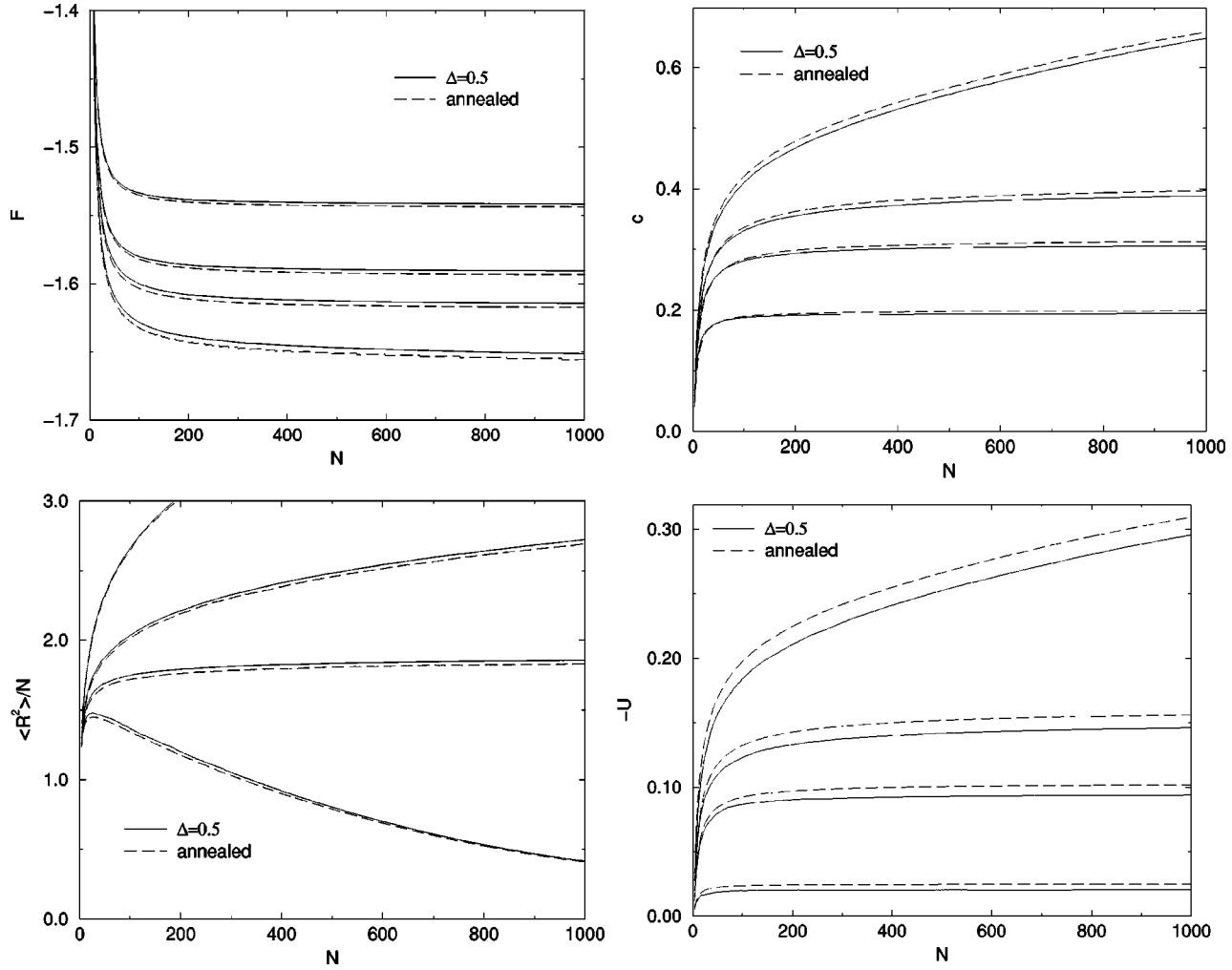

FIG. 2. Free energies F_N per monomer (top left), end-to-end swelling factors $\langle R^2\rangle/N$ (bottom left), density of contacts c (top right), and absolute values of energies per monomer U_N (bottom right) for four systems with $\Delta=0.5$, and with $\tilde{B}=0,-0.20,-0.269$, and -0.345. In the two top figures $|\tilde{B}|$ increases from top to bottom, in the bottom figures it increases from bottom to top. Full lines are from Monte Carlo simulations, dashed lines are predictions of the annealed approximation. In all panels, error bars are much smaller than the thickness of the lines. The values $\tilde{B}=0$ and -0.20 are in the swollen phase, -0.345 is in the collapsed phase, and -0.269 is on the theta line.

This inequality is equivalent to the existence of a negative correlation between the average energy and the partition function.

To use Eq. (17), we first have to evaluate disorder averages of Z_N^n for integer $n\geqslant 2$. These are performed similarly to the average over Z_N, except that the Gaussian integrals give rise formally to interactions between replicas [4],

$$\overline{Z^n}=\sum_{\mathcal{C}_1\cdots\mathcal{C}_n}\exp\left[-\tilde{B}\sum_{\alpha=1}^{n}\left(\sum_{i<j}\sigma_{ij}^{\alpha}\right)+\frac{\Delta^2}{2}\sum_{\alpha\neq\beta}\left(\sum_{i<j}\sigma_{ij}^{\alpha}\sigma_{ij}^{\beta}\right)\right]. \tag{20}$$

Here the Greek indices α and β refer to the different replicas, $\mathcal{C}_\alpha$ is a configuration of replica α, σ_{ij}^{α} is its contact map, and $\hat{B}$ is given by Eq. (7). The annealed approximation is equivalent to neglecting the two-replica term.

To proceed, we define the variables

$$c_\alpha=\frac{1}{N}\sum_{i<j}\sigma_{ij}^{\alpha},\quad q_{\alpha\beta}=\frac{1}{N\sqrt{c_\alpha c_\beta}}\sum_{i<j}\sigma_{ij}^{\alpha}\sigma_{ij}^{\beta}, \tag{21}$$

which are, respectively, the density of contacts for the contact map α and the overlap between two contact maps α and β. The overlap is a measure of similarity, and it is equal to one, if and only if the two contact maps coincide. Furthermore, we assume that, for large N, the number of configuration n tuples with $Nc_1,\ldots Nc_n$ contacts and mutual overlaps $\{q_{\alpha\beta}\}$ grows as

$$\exp\left[N\left(\sum_{\alpha=1}^{n}f(c_\alpha)-\sum_{k=2}^{n}\chi_k(\{c_\alpha\},\{q_{\alpha\beta}\})\right)\right]. \tag{22}$$

In other words, $\chi_k(\{c_\alpha\},\{q_{\alpha\beta}\})$ is the entropy loss per monomer when we impose that the replica $\mathcal{C}_k$ with density of

contacts c_α has overlaps $q_{1,k}\cdots q_{k-1,k}$ with the $k-1$ previous replicas. This quantity can be measured, for instance for $k=2$.

We can then write

$$\overline{Z^n}\approx \int d\{c_\alpha\}d\{q_{\alpha\beta}\}\exp\left\{N\left[\sum_{\alpha=1}^{n}[f(c_\alpha)-\tilde{B}c_\alpha]+\frac{\Delta^2}{2}\left(\sum_{\alpha\neq\beta}\sqrt{c_\alpha c_\beta}q_{\alpha\beta}-\sum_{k=2}^{n}\chi_k(\{c^\alpha\},\{q_{\alpha\beta}\})\right)\right]\right\}$$
$$\approx \exp[-NnF_n(\{c^\alpha\},\{q_{\alpha\beta}\})]. \quad (23)$$

Here, $F_n(\{c^\alpha\},\{q_{\alpha\beta}\})$ is the free energy per monomer in a system with n replicas. To evaluate it, we approximate the integrals over c_α and $q_{\alpha\beta}$ by their saddle points. We assume replica symmetry, which is expected to hold for low disorder: the saddle point is assumed to be given by $c_\alpha=c$ for all α and $q_{\alpha\beta}=q$ for all pairs $\alpha\neq\beta$.

Now, in order to obtain the correct free energy, we have to take the limit $n\to 0$. We obtain

$$F(B,\Delta)=-f(c)+\tilde{B}c+\tfrac{1}{2}\Delta^2 cq-\chi(c,q), \quad (24)$$

where $\chi(c,q)=-\lim_{n\to 0}\Sigma_{k=2}^{n}\chi_n(c,q)/n$ is the average entropy gain per replica due to the condition that the overlap among all replicas is equal to q. Note that this quantity is positive because, in the limit of vanishing n, the number of terms in the sum is -1. Finally, we have to compute the values of c and q at which F is evaluated by imposing two saddle-point conditions:

$$\frac{\partial f(c)}{\partial c}+\frac{\partial\chi(c,q)}{\partial c}=\tilde{B}+\frac{1}{2}\Delta^2 q, \quad (25)$$

$$\frac{1}{2}\Delta^2 c=\frac{\partial\chi(c,q)}{\partial q}. \quad (26)$$

For $\Delta=0$ (homopolymer) Eq. (26) just means that the value of the overlap is the most probable one for a given c, $q_0(c)$. Because of the normalization, it must be $\chi(c,q_0)=0$, thus the free energy of the homopolymer is just a special case of Eq. (24). Moreover, since $\chi(c,q_0)=0$ is an absolute minimum, also the derivative $\partial\chi/\partial c$ must vanish at that point, thus the saddle-point equation for c valid for the homopolymer is recovered for $\Delta=0$. It also follows from this argument that the second derivatives of $\chi(c,q)$ at the point $[c,q_0(c)]$ must be non-negative.

Notice that the free energy has to be maximized as a function of q because this variable refers to a space with a negative number of dimensions in the limit $n\to 0$. We thus get a first condition of thermodynamic stability $\partial^2\chi/\partial q^2>0$, which, from the above consideration, is expected to be fulfilled for Δ small enough. The situation is more complicated for the variable c. It enters both into the free energy of the replica interactions, which has to be maximized for $n\to 0$, and into the free energy of the homopolymer, which has to be minimized, at least for $\Delta=0$. We conjecture that the corresponding condition of thermodynamic stability is that the Hessian determinant of the free energy with respect to the variables c and q, $H(c,q)$, be nonpositive:

$$H(c,q)\equiv\left(\frac{\partial^2 f}{\partial c^2}+\frac{\partial^2\chi}{\partial c^2}\right)\frac{\partial^2\chi}{\partial q^2}-\left(\frac{\partial^2\chi}{\partial c\partial q}-\frac{1}{c}\frac{\partial\chi}{\partial q}\right)^2\leqslant 0. \quad (27)$$

For $\Delta=0$ we have $H(c,q)=(\partial^2 f/\partial c^2)(\partial^2\chi/\partial q^2)\leqslant 0$ as for homopolymers. In fact, at that point the first derivatives of $\chi(c,q)$ vanish. The Hessian determinant $\chi(c,q)$ vanishes also, because $\chi(c,q)$ stays constant at the value zero along the line $q=q_0(c)$. Thus both conditions of thermodynamic stability are fulfilled at Δ small enough.

The energy and entropy per monomer are obtained in the same way as in the annealed approximation. We find

$$U_N(B,\Delta)=[B-\Delta^2(1-q)]c, \quad (28)$$

$$S_N(B,\Delta)=f(c)+\chi(c,q)-\frac{\Delta^2}{2}c(1-q). \quad (29)$$

Although this is obtained from the saddle-point method, which is exact only for $N\to\infty$, we can use Eq. (28) to obtain effective overlaps $q'(B,\Delta,N)$, which tend to the saddle-point values $q(B,\Delta)$ in the thermodynamic limit. Results are shown in Fig. 3(a). Since corrections to the saddle point approximation are expected to be of order N^{-1}, while surface corrections in the compact phase are expected to be of order $N^{-1/3}$, the N behavior of $q(B,\Delta,N)$ for large systems should be dominated by the surface dependence of the overlap. Indeed, we measured the average overlap also using the definition Eq. (21), observing that its value and its finite-size behavior compare quite well with our numerical estimates based on Eq. (28). These measurements have been performed only at $\Delta=0$, and will be reported in the next section. In the present section we shall report the indirect measurement of the overlap through Eq. (28), which is much easier from the point of view of simulations.

We observed that $q'(B,\Delta,N)$ decreases with system size, but its asymptotic value seems to be finite in the random coil phase. This was confirmed by similar measurements at different values of Δ. Thus the annealed approximation does *not* hold in the random coil phase. The situation is more difficult for collapsed chains. In this case it cannot be excluded from Fig. 3(b) that the overlap asymptotically vanishes, and thus the annealed approximation becomes exact in the thermodynamic limit. However, simulations of systems of even larger size that will be presented in the next section show that this is not the case and that the corrections to the annealed approximation remain finite in the thermodynamic limit in both the coil and the collapsed phases, although in the latter phase they are rather small. As seen from Fig. 3(b), q does not depend very much on Δ in the collapsed phase and for large systems.

To obtain numerical estimates of χ, we subtract from Eq. (24) the analogous equation for homopolymers, and obtain

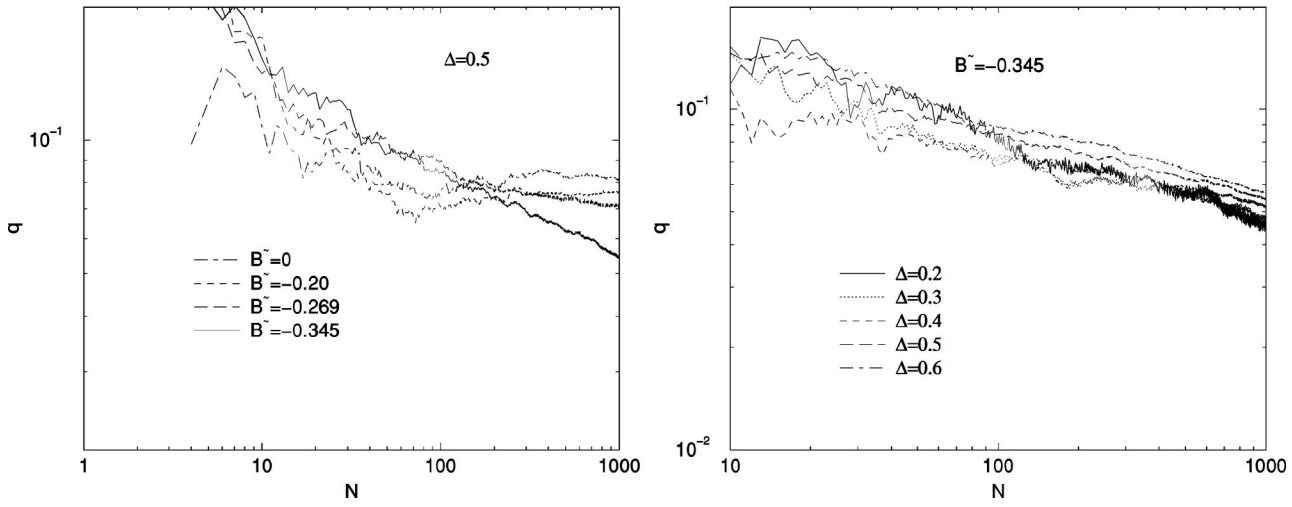

FIG. 3. Effective overlap $q'(B,\Delta,N)$ measured from the energy and the contact densities, as a function of system size for $\Delta=0.5$ and four different values of $=\tilde{B}$ (a) and for $\tilde{B}=-0.345$ and five different values of Δ (b).

$$F(B,\Delta)-F(\tilde{B},0)=(c-\tilde{c})\tilde{B}-[f(c)-f(\tilde{c})]+\frac{1}{2}\Delta^2 cq-\chi(c,q). \tag{30}$$

Expanding the difference $f(c)-f(\tilde{c})$ in the difference $c-\tilde{c}$ and using $f'(\tilde{c})=\tilde{B}$, we get

$$F(B,\Delta)-F(\tilde{B},0)=-\frac{1}{2}(c-\tilde{c})^2 f''(\tilde{c})+\frac{1}{2}\Delta^2 cq-\chi(c,q)+O[(c-\tilde{c})^2]. \tag{31}$$

This can be either evaluated directly, neglecting the last term. Alternatively, we can eliminate the term involving $f''(\tilde{c})$ by subtracting from Eq. (25) the analogous equation for homopolymers, which gives

$$(c-\tilde{c})f''(\tilde{c})=\frac{1}{2}\Delta^2 q-\frac{\partial\chi(c,q)}{\partial c}+O[(c-\tilde{c})^2]. \tag{32}$$

Combining this with Eq. (31) and neglecting terms $O[(c-\tilde{c})^2]$, we obtain finally

$$F(B,\Delta)-F(\tilde{B},0)=\frac{c+\tilde{c}}{4}\Delta^2 q-\frac{1}{2}[\chi(c,q)+\chi(\tilde{c},q)]+O[(c-\tilde{c})^2]. \tag{33}$$

As in the case of the overlap, we note that Eq. (33) defines an effective entropy $\chi(B,\Delta,N)$, which should tend to the leading correction to the annealed approximation [according to which, $\chi(B,\Delta)=0$ holds] in the thermodynamic limit.

In Fig. 4 we show numerical values for $[\chi(B,\Delta,N)+\chi(\tilde{B},0,N)]/2$ obtained in this way. We see that χ is very small but definitely not zero. Again we see that corrections to the annealed approximations are larger in the swollen phase than in the collapsed phase. Indeed, this time it seems that

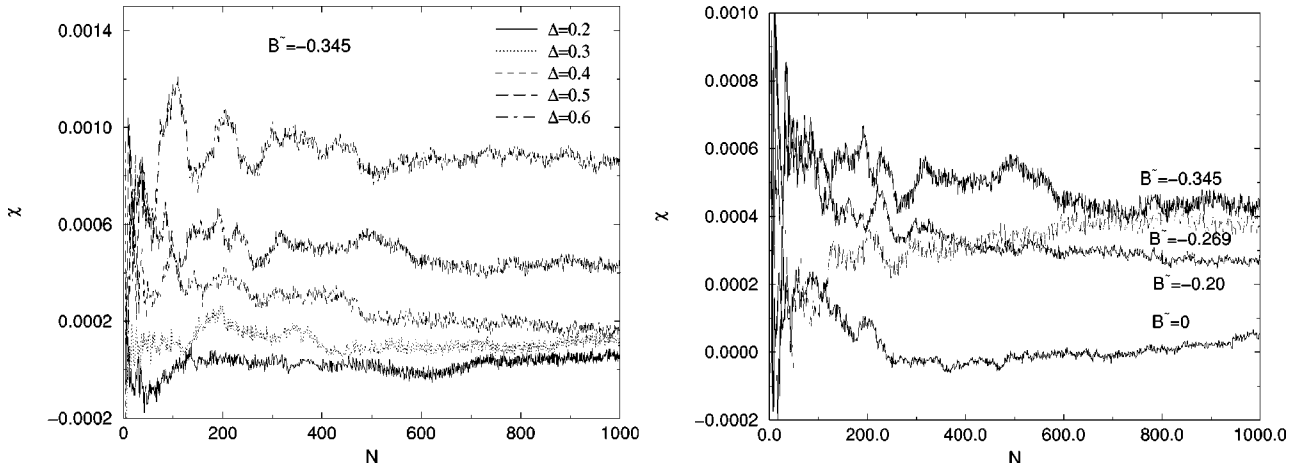

FIG. 4. $[\chi(B,\Delta,N)+\chi(\tilde{B},0,N)]/2$ measured from Eq. (33) as a function of system size for the same values of $\tilde{B}$ and Δ as in Fig. 3.

the deviations have a finite limit for $N\to\infty$ in both phases. This conclusion is supported by measurements at other values of Δ (not shown here). Basically the same conclusions are also drawn from Eq. (31), showing that $\chi(c,q)$ depends weakly on c and that Eq. (32) is very well satisfied.

IV. OVERLAP OF HOMOPOLYMERS

In this section we will discuss the overlap of contact matrices for homopolymer chains and their relation to corrections to the annealed approximation in the weak disorder limit. Consider the derivative of the free energy of a random heteropolymer with respect to Δ^2, at $\Delta=0$. Using Eqs. (18) and (20) and the results of Sec. II, we obtain

$$\left[\frac{\partial \overline{\ln Z_N}}{\partial \Delta^2}\right]_{\Delta=0} = 2\left[\frac{\partial \ln \overline{Z_N}}{\partial \Delta^2}\right]_{\Delta=0} - \frac{1}{2}\left[\frac{1}{\overline{Z_N^2}}\frac{\partial \overline{Z_N^2}}{\partial \Delta^2}\right]_{\Delta=0}$$

$$= \left[\frac{\partial \ln \overline{Z_N}}{\partial \Delta^2}\right]_{\Delta=0} + \frac{1}{2}\left\langle \sum_{i<j} \sigma_{ij}\sigma'_{ij}\right\rangle_{\Delta=0}, \tag{34}$$

where the angular brackets denote Boltzmann average over the ensemble of two replicas. Thus we have

$$\left[\frac{\partial}{\partial \Delta^2}(F_N - F_{N,\mathrm{ann}})\right]_{\Delta=0} = \frac{1}{2}\langle cq\rangle_{\Delta=0}. \tag{35}$$

This could have been obtained of course also within the replica symmetric approach, but the above derivation shows that it is indeed a rigorous result involving neither approximations nor unjustified assumptions.

Notice that this cannot be generalized to $\Delta\neq 0$, but in principle straightforward generalizations could be used to compute all higher derivatives

$$\left[\frac{\partial^k}{\partial \Delta^{2k}}(F_N - F_{N,\mathrm{ann}})\right]_{\Delta=0}, \quad k=1,2,3,\ldots. \tag{36}$$

Numerically, the right-hand side of Eq. (35) can be estimated by simulating pairs of chains simultaneously. For this we used a variant of the PERM algorithm where we add monomers alternatively to the first and to the second chain [32]. In this way we guarantee that both chains have exactly the same length (after having added an even number of monomers), and it is straightforward to estimate their overlap.

Results from such simulations with chains of length up to 1400 are shown in Fig. 5. These data agree nicely with extrapolations of the overlaps for $\Delta>0$ shown in the last section. They have much smaller statistical errors, since we do not have to average over any disorder explicitly. This makes the present method much faster and allows us to study larger systems.

The curve for $\tilde{B}=-0.2>B_\theta$ in Fig. 5 shows clearly that the annealed approximation does not become exact for $N\to\infty$ in the open coil phase. The same is true (although a bit less clear) exactly at the Θ point, as indicated by the curve

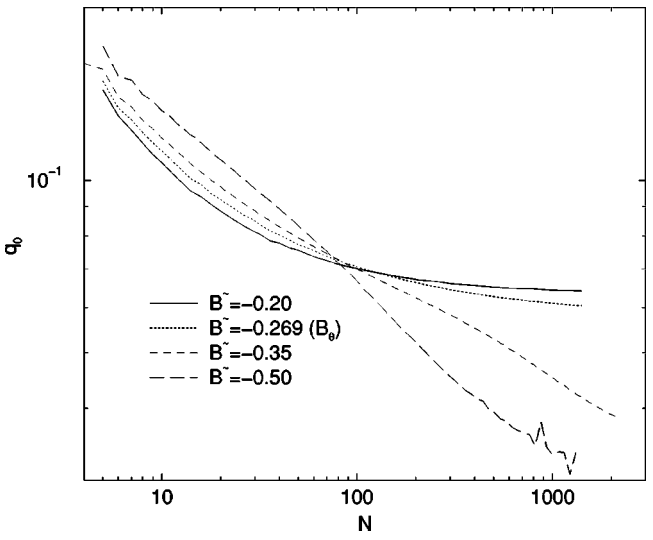

FIG. 5. Average overlap $q_0=\langle cq\rangle/\langle c\rangle$ between homopolymer chains of length N for different values of the monomer-monomer attraction B.

for $\tilde{B}=-0.269$. For the collapsed phase, the evidence is not so clear. The curves for $\tilde{B}=-0.35$ and for $\tilde{B}=-0.5$ both are much lower for large N and continue to decrease. Superficially, one might therefore conclude that the overlap disappears for $N\to\infty$. But both these curves show a distinct upward curvature for the largest values of N, indicating that the decrease will level off and q_0 tends to a finite constant for $N\to\infty$. To sustain this view, we show in Fig. 6 the plots of $\langle cq\rangle$ as a function of chain length N. In this case it is evident that the curves are going to a nonzero value. Since the fraction of contacts is limited (it holds $c\leqslant 2$ on the cubic lattice), also $q_0=\langle cq\rangle/\langle c\rangle$ should go to a nonzero value, and the decrease observed in Fig. 5 is just a consequence of the fact that the average fraction of contacts is increasing, approaching its stationary value.

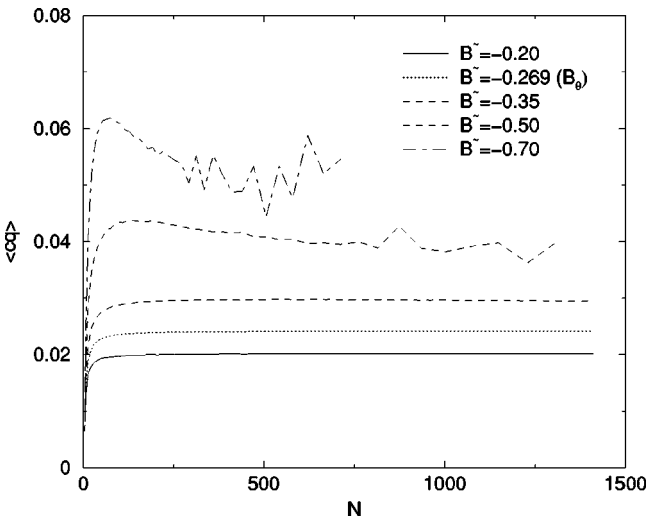

FIG. 6. Average fraction of common contacts $\langle cq\rangle$ between homopolymer chains of length N for different values of the monomer-monomer attraction B.

This conclusion is not, after all, very surprising. It can be backed by an argument, which could presumably, with some effort, be made rigorous. The density of contacts in a self-avoiding walk (corresponding to $B=0$) is dominated by short-range contacts, i.e., by contacts (i,j) with small $|i-j|$. The contribution of such a loop with fixed i and j depends weakly on the configuration of the chain far away from this monomer pair. Thus, if such a loop is present in one replica, it has a high chance to be present also in the other replica, even if the global structures of both replicas are entirely different.

In collapsed chains there are relatively more long-range contacts, which explains why the annealed approximation is better — but still not exact — in that regime. For random heteropolymers with strong disorder, and even more so for proteins, this argument suggests an increased overlap because of secondary structure. For instance, an alpha helix produces an array of contacts $(i,i+4)$, $(i+1,i+5)$, ..., where i labels the position of the amino acid along the protein chain. This enhances the overlap. Moreover, there is a finite probability that such contacts appear simultaneously even in the structures of two unrelated proteins. Thus the average overlap even in a large set of unrelated protein structure appears to attain a finite limit when the length of the chains increases [24].

V. THERMODYNAMICS IN THE REPLICA SYMMETRIC APPROXIMATION

In this section we study the predictions of the replica symmetric approximation on the behavior of thermodynamic variables with the external control parameters $\tilde{B}$ and Δ^2. We shall show the following. First, we shall justify that condition of thermodynamic stability is that the Hessian determinant of the free energy with respect to the parameters c and q, $H(c,q)$ given in Eq. (27), be nonpositive. Then we shall argue that, if $q_0(c)$, the average overlap of homopolymers with density of contacts c, is small, as it appears to be, also the corrections to the annealed approximation, given by $c-\tilde{c}(\tilde{B})$ and by q, will remain small [of order $q_0(c)$] for finite Δ. On very general grounds, we shall show that the density of contacts c is a decreasing function, while the overlap q is an increasing function, of the parameter $\tilde{B}$. If $\chi(c,q)$ depends only on the product cq, as assumed in previous works, then c is independent of Δ at fixed $\tilde{B}$, and it is exactly predicted by the annealed approximation. This is however at odds with simulations. Better assumptions on the functional form of $\chi(c,q)$ show that c is a decreasing function and q is an increasing function of Δ, as observed in simulations. At last, we discuss the possibility that the singularity in the specific heat at the theta point of the homopolymer may be smeared out by the disorder.

The two saddle-point equations for c and q cannot be explicitly solved without an explicit expression for the functions $f(c)$ and $\chi(c,q)$. Nevertheless, their qualitative behavior can be studied in more detail. Taking the derivatives of both Eqs. (25) and (26) with respect to the thermodynamic parameters $\tilde{B}$ and Δ we can compute the derivatives of c and q as

$$\left(\frac{\partial c}{\partial \tilde{B}}\right)_{\Delta}=\frac{\partial^2\chi/\partial q^2}{H(c,q)}\leqslant 0,$$

$$\left(\frac{\partial c}{\partial \Delta}\right)_{\tilde{B}}=\frac{-\Delta}{H(c,q)}\frac{\partial}{\partial q}\left(c\frac{\partial \chi}{\partial c}-q\frac{\partial \chi}{\partial q}\right), \tag{37}$$

$$\left(\frac{\partial q}{\partial \tilde{B}}\right)_{\Delta}=-\frac{\partial^2\chi/\partial c\,\partial q-(1/c)(\partial\chi/\partial q)}{H(c,q)},$$

$$\left(\frac{\partial q}{\partial \Delta}\right)_{\tilde{B}}=\frac{\Delta}{H(c,q)}\left[c\left(\frac{\partial^2 f}{\partial c^2}+\frac{\partial^2\chi}{\partial c^2}\right)-q\left(\frac{\partial^2\chi}{\partial c\,\partial q}-\frac{1}{c}\frac{\partial\chi}{\partial q}\right)\right],$$

where $H(c,q)$ is given by Eq. (27). From these, the specific heat can be computed as

$$\begin{aligned}C_v&=\partial U/\partial T\\&=[B-\Delta^2(1-q)]\left[(B-\Delta^2)\frac{\partial c}{\partial\tilde{B}}+\Delta\frac{\partial c}{\partial\Delta}\right]\\&\quad-\Delta^2 c\left[(1-q)+(B-\Delta^2)\frac{\partial q}{\partial\tilde{B}}+\Delta\frac{\partial q}{\partial\Delta}\right]\\&=-\frac{1}{H(c,q)}\left[\frac{\partial^2 F}{\partial c^2}(\Delta^2 c)^2-2\frac{\partial^2 F}{\partial c\,\partial q}(\Delta^2 c)[B-\Delta^2(1-q)]\right.\\&\quad\left.+\frac{\partial^2 F}{\partial q^2}[B-\Delta^2(1-q)]^2\right]+\Delta^2 c(1-q),\end{aligned} \tag{38}$$

where $F(c,q)$ is the free energy evaluated at the saddle point and $H(c,q)$ is its Hessian determinant. The three terms in the square brackets are a quadratic form whose determinant is expressed by $H(c,q)$. Since $\partial^2 F/\partial q^2$ is positive, they would give a negative contribution to the specific heat if $H(c,q)$ were positive. Thus, it is justified to require that $H(c,q)$ is negative as a condition for thermodynamic stability.

We now argue that the corrections to the annealed approximation remain small if the average overlap of homopolymers, $q_0(c)$, is small. In fact, the function $\chi(c,q)$ attains its absolute minimum value $\chi(c,q)=0$ along the line $q=q_0(c)$. Thus, assuming that $\chi(c,q)$ is an analytic function of q for $q>0$, it can be expressed in the form

$$\begin{aligned}\chi(c,q)&=\sum_{k=2}^{\infty}\frac{a_k(c)}{k!}[q-q_0(c)]^k\\&\equiv\sum_{k=2}^{\infty}\frac{A_k(c)}{k!}[Q-Q_0(c)]^k,\end{aligned} \tag{39}$$

with $a_2(c)>0$. The typical overlap $q_0(c)$ is small, and it is a decreasing function of c, or $q_0'(c)<0$ (the prime indicates derivative with respect to c). The coefficients $a_k(c)$ are expected to be quantities of order $[q_0(c)]^{-k+1}$, as it will be

argued later. We also introduce the notation $Q=cq$, $Q_0(c)=cq_0(c)$, and $A_k(c)=a_k(c)c^{-k}$. We can now develop the saddle-point equations for c close to the solution of the annealed approximation, $c=\tilde{c}(\tilde{B})$, given by $f'(\tilde{c})=\tilde{B}$:

$$\left(\frac{\partial^2 f}{c^2}+\frac{\Delta^2}{2}\frac{\partial^2 Q_0}{\partial c^2}\right)_{c=\tilde{c}}(c-\tilde{c})+\left(\frac{\Delta^2}{2}\frac{\partial Q_0}{\partial c}\right)_{c=\tilde{c}}+\left(\sum_{k=2}^{\infty}\frac{A_k'(c)}{k!}[Q-Q_0(c)]^k\right)_{c=\tilde{c}}=0,$$

$$\sum_{k=1}^{\infty}\frac{A_{k+1}(c)}{k!}[Q-Q_0(c)]^k=\frac{1}{2}\Delta^2. \tag{40}$$

From these expressions one sees that both $Q-Q_0(c)$ and $c-\tilde{c}$ are quantities of order $Q_0(c)$, thus corrections to the annealed approximation are finite but small for finite Δ.

We now compute the thermodynamic derivatives by developing Eq. (39) to the zeroth order in $\delta q=[q-q_0(c)]$ (this quantity must be positive for small Δ):

$$\left(\frac{\partial c}{\partial \tilde{B}}\right)_{\Delta}\approx\frac{a_2(c)}{H(c,q)}\leqslant 0,\quad \left(\frac{\partial c}{\partial \Delta}\right)_{\tilde{B}}\approx\frac{\Delta a_2(c)}{H(c,q)}\frac{\partial Q_0}{\partial c}, \tag{41}$$

$$\left(\frac{\partial q}{\partial \tilde{B}}\right)_{\Delta}\approx\frac{a_2(c)}{H(c,q)}\frac{\partial q_0}{\partial c}\geqslant 0,$$

$$\left(\frac{\partial q}{\partial \Delta}\right)_{\tilde{B}}\approx\frac{\Delta}{H(c,q)}\left[c\frac{\partial^2 f}{\partial c^2}+a_2(c)\frac{\partial q_0}{\partial c}\frac{\partial Q_0}{\partial c}\right] \tag{42}$$

$H(c,q)$ must be computed at the first order in δq, because the zeroth order term vanishes at the theta point $c=c_\theta$ at which $\partial^2 f/\partial c^2$ vanishes. The result reads

$$H(c,q)\approx\frac{\partial^2 f}{\partial c^2}\frac{\partial^2 \chi}{\partial q^2}-[Q-Q_0(c)]c[A_2(c)]^2\frac{\partial^2 Q_0}{\partial c^2}. \tag{43}$$

Thus we conclude that the density of contacts c decreases and the average overlap q increases with $\tilde{B}$. To proceed further, we first consider the simple case where the entropy $\chi(c,q)$ depends only on the product $Q=cq$ representing the number of conditions that we have to impose in order to fix an overlap q: $\chi(c,q)=\hat{\chi}(cq)$. This form was assumed in the work of Shakhnovich and Gutin [4]. In this case, Q_0 does not depend on c, thus from Eqs. (42) it follows that $c=c(\tilde{B})$ should not depend on Δ at fixed $\tilde{B}$ and assumes the value $\hat{c}$ predicted by the annealed approximation. This result can also be obtained directly from Eq. (37). However, our numerical results contradict this prediction. The overlap q, should be in this case, an increasing function of Δ, and its derivative with respect to $\tilde{B}$ should be proportional to $(\partial^2 f/\partial c^2)^{-1}$, which is expected to diverge at the theta point $c=c_\theta$. This fact can explain why the values of q are much smaller in the collapsed phase than in the coil phase.

A better theory for the function $\chi(c,q)$ was developed by Plotkin *et al.*, [13]. They computed the entropy loss for two replicas with density of contacts c being at overlap q, $\chi_2(c,q)$, in the mean-field approximation and in the collapsed phase. Although this can be different from $\chi(c,q)$, it is a good point for understanding its qualitative behavior. Unfortunately, we cannot use the formula obtained in Ref. [13] because of technical reasons and because it assumes that the homopolymer overlap $q_0(c)$ is zero in the thermodynamic limit, while our calculations show that this is not the case. We shall however use the fact that $\chi(c,q)$ is the sum of three contributions: the entropy loss due to imposing that Ncq contacts have to coincide, the loss due the fact that $Nc(1-q)$ contacts have to be different, and a combinatoric factor counting the number of different choices of Ncq contacts among Nc. The last term can be approximated by $\chi_{\rm mix}(c,q)=c[q\ln q+(1-q)\ln(1-q)]$, even if this is an overestimate, since not all of the combinations of different common contacts can be realized. Putting everything together we have

$$\chi(c,q)=\hat{\chi}(c,cq)+c[q\ln q+(1-q)\ln(1-q)]. \tag{44}$$

In the computation by Plotkin *et al.*, the mixed second derivative of $\hat{\chi}(c,cq)$ with respect to $Q=cq$ and c vanishes. This simplifies considerably formulas, and it will be assumed to hold for the rest of the paper. We shall thus introduce the notation $\hat{\chi}'(Q)$ to denote the derivative of $\hat{\chi}(c,Q)$ with respect to $Q=cq$ at fixed c. Comparing Eq. (44) to Eq. (39), we see that $\hat{\chi}''(Q)$ must be positive and that $\partial\hat{\chi}/\partial c=-(1-2q_0)\ln(1-q_0)$. We also see that $a_2(c)=c/[q_0(1-q_0)]+\hat{\chi}''(Q)$ is likely to be a quantity of order q_0^{-1}, as it has been assumed above. For the higher-order coefficients one finds $a_k=O(q_0^{-k+1})$, as anticipated. We have now to compute the derivatives of $Q_0(c)=cq_0(c)$:

$$\frac{\partial Q_0}{\partial c}=-\frac{\partial^2\chi/\partial c\,\partial Q}{\partial^2\chi/\partial Q^2}=\frac{q_0}{1+cq_0\hat{\chi}''(cq_0)(1-q_0)}\geqslant 0, \tag{45}$$

in qualitative agreement with Fig. 6. Inserting the above result in the formulas (42) we see that the density of contacts decreases with Δ at fixed $\tilde{B}$. This behavior is confirmed by our numerical results (see Fig. 7), which also show that the decrease is maximal for $\tilde{B}\approx B_\theta=-0.27$, as expected from the fact that $H[c,q_0(c)]$ vanishes at $c=c_\theta$. The overlap q increases with $\tilde{B}$ at fixed Δ, as expected from Eq. (42) (see Fig. 8), and increases with Δ at fixed $\tilde{B}$, as expected from Eq. (45) (see Fig. 8 again). It can thus be understood why the overlap decreases with system size: as the number N of monomers increases the importance of surface effects is reduced (as $N^{-1/3}$) and $c(N)$ increases, thus decreasing the value of q.

We now examine the condition of thermodynamic stability, $H(c,q)\leqslant 0$. As it was already observed, since at the point $[c,q_0(c)]$ both the gradient of $\chi(c,q)$ and its

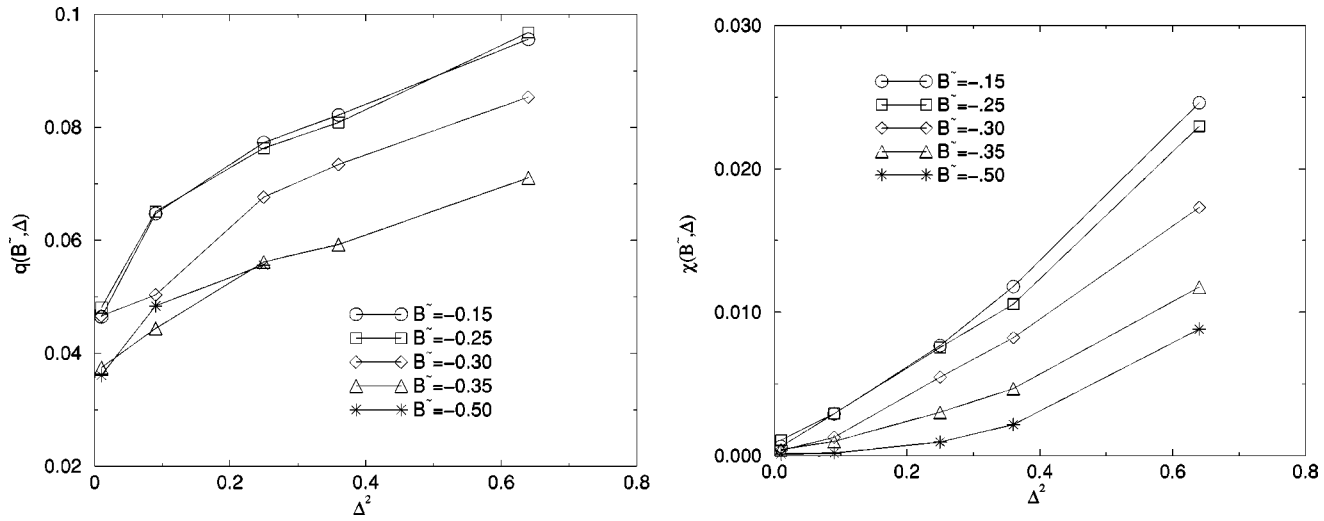

FIG. 7. Left: overlap q as a function of Δ^2 for different values of $\tilde{B}$. Right: entropy $[\chi(B,\Delta)+\chi(\tilde{B},0)]/2$ measured from Eq (33) as a function of Δ^2 for different values of $\tilde{B}$. In both cases, we use values for $N=800$, but other lengths give qualitatively the same behavior.

Hessian determinant vanish, we have $H[c,q_0(c)]=(\partial^2 f/\partial c^2)(\partial^2\chi/\partial q^2)\leqslant 0$. At the theta point this quantity vanishes, and $H(c_\theta,q)$ is given by the deviations from the annealed approximation. Three situations are possible: First, $H(c_\theta,q)$ can be positive at the leading order in $\delta q=q-q_0(c_\theta)$. In this case the thermodynamic stability would be violated around the theta point, but our simulations do not show anything peculiar in this region. Second, the leading order in δq can be negative. In this case, the specific heat would not diverge anymore at the theta point for finite Δ, but it would show a peak proportional to some negative power of δq. Thus the disorder would smear out the thermodynamic singularity at $c=c_\theta$, leaving unchanged the geometric characterization of the collapsed chains in terms of the gyration radius. It is rather difficult, if not impossible, to test this scenario by means of simulations. Third, $H(c_\theta,q)$ can vanish identically at $c=c_\theta$. It is easy to see that this condition, combined with the assumption that $\chi(c,q)$ is of the form (44), is fulfilled if and only if $\hat{\chi}'(cq)$ is of the form

$$\hat{\chi}'(cq)=\ln\left(\frac{1-cq/B}{cq/A}\right), \tag{46}$$

where $q\geqslant q_0(c)$, $A<0$, and $0<B<cq$ are two constants, and c is not too small so that the last inequality can be fulfilled. In this case, one finds $cq_0(c)=B(c-A)/(B-A)\in[0,c]$, and it is easy to check that all previous results are recovered, while $H(c,q)-H[c,q_0(c)]$ vanishes for all q and c, including the theta point.

Summarizing the discussion, we find that, if $\chi(c,q)$ is of the form (44), two possibilities are open: either $\hat{\chi}'(cq)$ is given by Eq. (46), in which case $H(c,q)\equiv H[c,q_0(c)]\leqslant 0$, or $\hat{\chi}'(cq)$ has a different form, in which case the specific heat is not anymore divergent at the theta point $c=c_\theta$. Unfortunately, we are not able to decide among these alternatives.

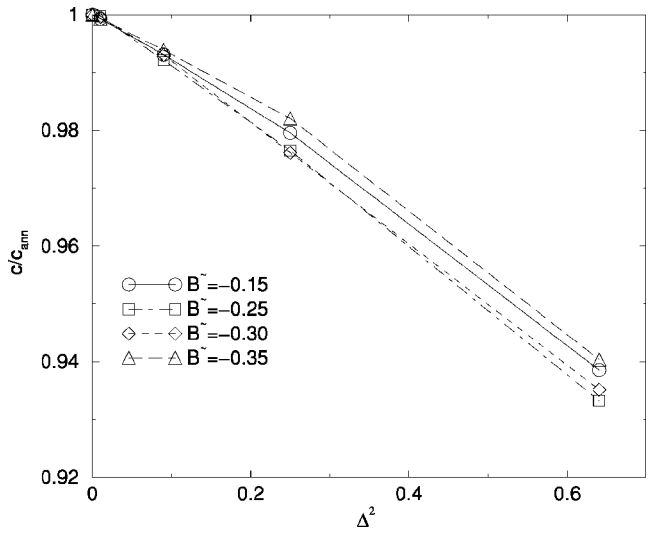

FIG. 8. Corrections to the density of contacts predicted by the annealed approximation as a function of Δ, for different values of $\tilde{B}$ and $N=800$. The deviations are negative, and they are maximal close to the theta point $\tilde{B}\approx -0.27$. The same pattern is observed for other lengths.

VI. DISCUSSION

We have shown that the annealed approximation is very good but not exact for a particular model of random heteropolymers, and we have given simple physical arguments for it. We have also computed the thermodynamics of the model using the replica symmetric approximation, and we have shown that such an approach can explain very well, at least qualitatively, the observed deviations from the annealed approximation in the high-temperature phase. The replica symmetric calculation also leaves open, surprisingly, the possibility that the disorder could cancel the thermodynamic singularity at the theta point. A numerical test of this possibility is very difficult, and it has been left out.

In our present paper we have not addressed the most interesting aspect of the model — the freezing of the system in a finite number of mesoscopic states. This transition should represent some features of the folding transition taking place for protein structures. Instead, we have studied the model at higher temperatures and at smaller disorder. This should however be of interest also in the context of the freezing, since it was conjectured [4] that freezing can be described in this model by the random energy model, for which a prerequisite is that the annealed approximation is exact.

In the present simulations we have studied chains of length up to $N=1400$. Deviations from the annealed approximation decrease fast for small N, which explains why studying very short chains has mislead several authors to the conclusion that these deviations vanish for $N\to\infty$. But in the high-T (open coil) phase this decrease clearly stops, and deviations are roughly independent of N for $N>100$. This is less clear in the collapsed phase. But also numerics, general arguments, and detailed calculations within a specific scenario with unbroken replica symmetry all indicate that these deviations will settle at a nonzero value for large N. This casts doubts on the validity of the random energy picture for protein folding.

There are of course a number of questions which are left open by the present paper. First of all, we have deliberately left out all questions related to freezing. Secondly, our treatment of Sec. III assumed that the overlap distribution is always dominated by a single peak. This is most likely not true in the frozen regime, and the distribution of overlaps is certainly a most interesting object. We shall address these questions in a forthcoming paper [35]. Finally, we have studied only one particular model, where contact energies are independent Gaussian variables. Several other models of random heteropolymers are studied in the recent literature [19], and several of them present very interesting open problems.

ACKNOWLEDGMENTS

We are indebted to many colleagues for useful discussions, in particular to H. Frauenkron, W. Nadler, H. Orland, E. Shakhnovich, A. Trovato, and M. Vendruscolo.

APPENDIX: THE PRUNED ENRICHED ROSENBLUTH METHOD

This method which was first described in detail in Ref. [25] is a chain growth method. It is based on Rosenbluth's idea of biased sampling [33], but it deviates from it by deleting (''pruning'') configurations with too low weight, and copying configurations with too large weight (''enrichment''). We remind the reader that the bias in the Rosenbluth method requires each configuration to carry a nontrivial weight that builds up gradually as monomer by monomer is added. In addition to this ''Rosenbluth factor,'' there is also a Boltzmann factor in the case of interacting polymers. The weights that control pruning/enrichment are the product of the two.

Both pruning and enrichment are done while the chains grow, i.e., based on the actual (incomplete) weight factors. In some cases it is advantageous to use not the present weights but (implicit) estimates of future weights to control the algorithm [34], but this is not done in the present paper. Instead, we use some of the special tricks used for strongly collapsed systems in Ref. [27].

When the weight is too low, configurations are not simply killed (this would imply systematic errors), but instead they are killed with probability 1/2, and those which are not killed get their weights doubled. Similarly, in the case of cloning the weight is spread uniformly among all clones. Technically, cloning is performed by means of recursive subroutine calls. A pseudocode of the basic algorithm is given in the appendix of Ref. [25].

The most important shortcoming of the Rosenbluth method is that the distribution of weights can become extremely wide for large systems and at low temperatures. The only exception is for interacting homopolymers on the simple cubic lattice at the theta point, where Rosenbluth and Boltzmann factors nearly cancel. In less favorable cases, even a very large statistical sample can be dominated by just a handful of high weight events, and statistical errors grow out of bounds. Even worse, in extreme cases the events that would carry (in average!) most of the weight are so rare that they are missed completely with high probability, and the free energy is underestimated systematically.

Pruning and enrichment guarantee that the weights of individual configurations stay within narrow bounds, and the above cannot happen. But in very difficult situations (large N, low T, large disorder) it may happen that due to cloning the configurations are strongly correlated, and the weights of clusters of such correlated configurations play essentially a similar role as the weights of individual configurations in the above discussion. In the following, we will call such a cluster, which is composed of all configurations having a common root, a ''tour.'' In order to check that the weights of tours do not become too uneven, we have measured their distribution. Let us call the weights W, and the distribution $P(W)$. We are on safe grounds if this distribution is so narrow that $P(W)$ and $WP(W)$ have basically the same support. In particular, we should require that the maximum of $WP(W)$ occurs at such values where $P(W)$ is still appreciable, and the distribution is well sampled. We have verified that this is the case for all data shown in the present paper. Notice, however, that this is a very stringent requirement. If it is not satisfied, this would not necessarily mean that the data are wrong, since configurations within one tour are only partially correlated.

The generalization of this algorithm to pairs of chains growing simultaneously is straightforward [32]. One just has to add monomers alternately. When the total number of monomers in both chains is even, the addition of the next monomer is attempted at chain 1; when this number is odd, the next monomer is added to chain 2.

[1] T.E. Creighton, *Protein Folding* (W.H. Freeman, New York, 1992).
[2] J.D. Bryngelson and P.G. Wolynes, Proc. Natl. Acad. Sci. U.S.A. **84**, 7524 (1987).
[3] T. Garel and H. Orland, Europhys. Lett. **6**, 307 (1988).
[4] E.I. Shakhnovich and A.M. Gutin, Biophys. Chem. **34**, 187 (1989).
[5] E. I. Shakhnovich and M. Karplus, in *Protein Folding: Theoretical Studies of Thermodynamics and Dynamics*, edited by T.E. Creighton (W.H. Freeman, New York, 1992).
[6] E.I. Shakhnovich and A.M. Gutin, J. Chem. Phys. **93**, 5967 (1990).
[7] E.I. Shakhnovich and A.M. Gutin, Nature (London) **346**, 773 (1990).
[8] E.I. Shakhnovich and A.M. Gutin, J. Mol. Biol. **3**, 1614 (1990).
[9] C. Camacho and D. Thirumalai, Proc. Natl. Acad. Sci. U.S.A. **90**, 6369 (1996).
[10] E.I. Shakhnovich and A.M. Gutin, Proc. Natl. Acad. Sci. U.S.A. **90**, 7195 (1993); E.I. Shakhnovich, Phys. Rev. Lett. **24**, 3907 (1994).
[11] G. Iori, E. Marinari, and G. Parisi, J. Phys. A **24**, 5349 (1991); Europhys. Lett. **25**, 491 (1994).
[12] Y. Kantor and M. Kardar, Europhys. Lett. **28**, 169 (1994).
[13] S.S. Plotkin, J. Wang, and P.G. Wolynes, Phys. Rev. E **53**, 6271 (1996); S.S. Plotkin, J. Wang, and P.G. Wolynes, J. Chem. Phys. **53**, 2932 (1997).
[14] V.S. Pande, A.Y. Grosberg, C. Joerg, M. Kardar, and T. Tanaka, Phys. Rev. Lett. **76**, 3565 (1996).
[15] V.S. Pande, A.Y. Grosberg, C. Joerg, and T. Tanaka, Phys. Rev. Lett. **76**, 3987 (1996).
[16] C.J. Camacho and T. Shanke, Europhys. Lett. **37**, 603 (1997).
[17] D.K. Klimov and D. Thirumalai, Phys. Rev. Lett. **76**, 4070 (1997).
[18] P. Monari, A.L. Stella, C. Vanderzande, and E. Orlandini, Phys. Rev. Lett. **83**, 112 (1999).
[19] T. Garel, H. Orland, and E. Pitard cond-mat/9706125 1997 (unpublished).
[20] U. Bastolla, H. Frauenkron, and P. Grassberger, J. Mol. Liq. **84**, 111 (2000).
[21] P.G. de Gennes, *Scaling Concepts in Polymer Physics* (Cornell University Press, Ithaca, 1979).
[22] A. Trovato, J. van Mourik, and A. Maritan, Eur. Phys. J. B **6**, 63 (1998).
[23] T. Garel, H. Orland, and Leibler, J. Phys. II **4**, 2139 (1994).
[24] U. Bastolla, E.W. Knapp, and M. Vendruscolo, Proteins (to be published).
[25] P. Grassberger, Phys. Rev. E **56**, 3682 (1997).
[26] U. Bastolla and P. Grassberger, J. Stat. Phys. **89**, 1061 (1997).
[27] H. Frauenkron, U. Bastolla, E. Gerstner, P. Grassberger, and W. Nadler, Phys. Rev. Lett. **80**, 3149 (1998).
[28] H. Frauenkron and P. Grassberger, J. Chem. Phys. **107**, 9599 (1997).
[29] G.T. Barkema, U. Bastolla, and P. Grassberger, J. Stat. Phys. **90**, 1311 (1998).
[30] B. Duplantier, J. Chem. Phys. **86**, 4233 (1987).
[31] J. Hager and L. Schäfer, Phys. Rev. E **60**, 2071 (1999).
[32] B. Coluzzi, M.S. Causo, and P. Grassberger, cond-mat/9910188 1999 (unpublished).
[33] M.N. Rosenbluth and A.W. Rosenbluth, J. Chem. Phys. **23**, 356 (1955).
[34] H. Frauenkron, M.S. Causo, and P. Grassberger, Phys. Rev. E **59**, R16 (1999).
[35] U. Bastolla, Eur. Phys. J. E (to be published).