Eur. Phys. J. B **6**, 63–73 (1998)

THE EUROPEAN
PHYSICAL JOURNAL B

© EDP Sciences
Springer-Verlag 1998

Swollen-collapsed transition in random hetero-polymers

A. Trovato[a], J. van Mourik, and A. Maritan

International School for Advanced Studies (SISSA), and Istituto Nazionale di Fisica della Materia, Via Beirut 2-4,
34014 Trieste, Italy
and
Abdus Salam International Center for Theoretical Physics, Strada Costiera 11, 34100 Trieste, Italy

Received: 10 April 1998 / Revised: 4 June 1998 / Accepted: 1st July 1998

Abstract. A lattice model of a hetero-polymer with random hydrophilic-hydrophobic charges interacting with the solvent is introduced, whose continuum counterpart has been proposed by Garel, Leibler and Orland [7]. The transfer matrix technique is used to study various constrained annealed systems which approximate at various degrees of accuracy the original quenched model. For highly hydrophobic chains an ordinary θ-point transition is found from a high temperature swollen phase to a low temperature compact phase. Depending on the type of constrained averages, at very low temperatures a swollen phase or a coexistence between compact and swollen phases are found. The results are carefully compared with the corresponding ones obtained in the continuum limit, and various improvements in the original calculations are discussed.

PACS. 05.70.Fh Phase transitions: general aspects – 61.41.+e Polymers, elastomers, and plastics – 64.75.+g Solubility, segregation, and mixing; phase separation – 75.10.Nr Spin-glass and other random models

1 Introduction

The main reason for the study of random hetero-polymers in solutions, is a possible connection of this problem with the protein folding problem [1–3]. Indeed, proteins are believed to be by nature selected special cases of random hetero-polymers. Before dealing with these special cases, it is of great importance to understand the typical behaviour of the various kinds of random hetero-polymer models that have been introduced, as it may give important insight in which types of interactions are indispensible for folding, and which types of interactions, on the other hand, are of secondary importance.

Several models of (quenched) randomness have been considered. Here, we study the role of the solvent (water) in the equilibrium [4] properties of the collapsed phase, as it is commonly believed that the hydrophobic effect [5] is the main driving force for the folding transition. Most proteins in nature consist of a strongly hydrophobic core, surrounded by hydrophilic (less hydrophobic) residues. We restrict ourselves to the simple coarse grained model, that was originally introduced by Obukhov [6], where the monomers of a single chain are randomly hydrophilic or hydrophobic (RHH), and interact with the solvent molecules through an effective two-body short range interaction. The statics of the continuum version of this model has been studied by Garel *et al.* [7], both in the case of annealed and quenched disorder, while the dynamics (with quenched disorder) has been studied by Thirumalai *et al.* [8]. The model has been studied also by Timoshenko *et al.* [9] and Moskalenko *et al.* [10] with the Gaussian self-consistent method.

We have choosen to study a (2d square) lattice version of the RHH model, and the method we used to assess the conformational entropy, is the transfer matrix (TM) method, which is most fit to study the case of annealed disorder. Furthermore, using the approximation scheme introduced by Morita [11], we are able to give lower bounds for the quenched free energy. It will turn out that the annealed case may exhibit a very rich phase diagram and re-entrant compact-swollen transitions, and we come to different conclusion than Garel *et al.* [7]. The case of the annealed average with fixed mean for the hydrophobic-hydrophilic charges, gives the same results one can get in the continuum limit for the quenched average using the method of reference [7] (see Sect. 7 for details). We go one step forward analyzing a better approximation to the quenched system which cures some problems present in the previous approximations and in the standard approach presented in [7].

---

[a] e-mail: `trovato@shannon.sissa.it`

This work is built up in the following way. In Section 2, we introduce the model in more detail. In Section 3, we introduce the concept of constrained annealing; in Section 4 we show that the effective models after averaging over the disorder involve 2- and 3-body interactions and the general phase diagram of such kind of models is discussed. In Section 5, the TM method is used to assess the conformational entropy of the polymer chain. The results are presented in Section 6, together with an outlook of the items that are still to be investigated. Finally, in Section 7 we give an interpretation of our results, and a detailed comparison with those obtained for the continuum model by [7].

2 Definition of the model

The polymer chain is represented by a self-avoiding walk (SAW) on a lattice where each site is either visited by the walk (*i.e.* is occupied by one monomer of the chain), or occupied by a solvent molecule. The interactions in the model are two-body short-range interactions. The only interactions we take into account, are those between solvent molecules and monomers if they occupy nearest-neighbor sites. Hydrophilicities λ_i are attached to each monomer i of the walk such that the Hamiltonian is given by

$$\mathcal{H} = -\sum_{i=0}^{N} \lambda_i z_i, \tag{1}$$

where the sum runs over the $N+1$ sites of the lattice occupied by the N-step walk, and z_i is the number of nearest-neighbor contacts of monomer i with solvent molecules, *i.e.* the number of nearest-neighbor sites of site i not occupied by the walk.

The hydrophilicities λ_i are supposed to be independent identically distributed random variables with a Gaussian distribution with mean λ_0 and variance λ:

$$P(\lambda_i) = \frac{1}{\sqrt{2\pi\lambda^2}} \exp\left(-\frac{(\lambda_i - \lambda_0)^2}{2\lambda^2}\right), \tag{2}$$

and the average over this (*a priori*) distribution is indicated by $\langle\!\langle \cdot \rangle\!\rangle$. If $\lambda_i > 0$, the corresponding monomer is hydrophilic and attracts solvent molecules, whereas if $\lambda_i < 0$, the monomer is hydrophobic and repels solvent molecules.

The canonical partition function for SAW of N steps with a fixed disorder configuration $\{\lambda_i\}$ is then

$$\mathcal{Z}_N(\{\lambda_i\}) = \sum_{W_N} \exp\left(\beta \sum_{i=0}^{N} \lambda_i z_i\right), \tag{3}$$

where the sum has to be taken over all N-step SAW starting from the origin.

If monomers can rearrange themselves along the chain and change their hydrophilicities, *e.g.* with chemical reactions, these have to be considered as thermal annealed variables, which approach equilibrium in the same time scale as the configurational degrees of freedom. The physics of such hetero-polymers is given by the average of the partition function over the disorder distribution (annealed average)

$$\langle\!\langle \mathcal{Z}_N(\{\lambda_i\}) \rangle\!\rangle = \sum_{W_N} \exp\left(\beta\lambda_0 \sum_{i=0}^{N} z_i + \frac{\beta^2\lambda^2}{2} \sum_{i=0}^{N} z_i^2\right). \tag{4}$$

Instead, if the monomer sequence of the chain is fixed, as it is the case for proteins, the hydrophilicities are frozen while the polymer is approaching thermal equilibrium; the average over the disorder distribution has then to be taken over the logarithm of the partition sum (quenched average) [12], to yield the quenched free energy

$$f_q = -\frac{1}{\beta} \langle\!\langle \ln[\mathcal{Z}_N(\{\lambda_i\})] \rangle\!\rangle, \tag{5}$$

which is a much harder task to accomplish.

3 Constrained annealing

In order to avoid the difficult direct computation of the quenched average (5), we have applied an idea first introduced by Morita [11]. This is the so-called Equilibrium Ensemble Approach (EEA) (see *e.g.* [13] for a recent review and discussion). The EEA consists of a systematic approximation procedure for the quenched free energy by annealed averages. It can be shown [13] that each successive approximation gives a better or equally good lower bound for the quenched free energy.

Each approximation consists in performing an annealed average over a new Hamiltonian $\mathcal{H}^* \equiv \mathcal{H} + \mathcal{H}_d$, where $\mathcal{H}$ is the original Hamiltonian (1), and $\mathcal{H}_d$ is a fictitious disorder potential, which contains a number of parameters. These parameters (Lagrange multipliers) have to be tuned in such a way that some moments of the *a posteriori* (annealed) distribution of the disorder are equal to the *a priori* (quenched) ones. In annealed averages, the *a posteriori* distribution $P^*(\{\lambda_i\},\{z_i\})$ is defined as

$$P^*(\{\lambda_i\},\{z_i\}) \equiv \frac{P(\lambda_i)\exp(-\beta\mathcal{H}^*(\{\lambda_i\},\{z_i\}))}{\langle\!\langle \mathcal{Z}_N^* \rangle\!\rangle} \cdot \tag{6}$$

The average over this distribution will be denoted by $\langle\cdot\rangle \equiv \int d\{\lambda_i\} \sum_{W_N} P^*(\{\lambda_i\},\{z_i\})\cdot$. In principle, one has to fix all the moments of $P^*_{ann}(\{\lambda_i\}) \equiv \sum_{W_N} P^*(\{\lambda_i\},\{z_i\})$ to obtain the quenched result, which is as difficult as the direct computation of (5). Nevertheless, one can hope to obtain a reasonable approximation of the quenched case by fixing a few suitably chosen moments. Moreover, the method is variational, and fixing more and more moments yields thighter lower bounds for the quenched free energy.

In this work, we have considered three different cases of annealing: without constraints (a_0), constraining the first moment of overall hydrophilicity (a_1), constraining the first and the second moment of overall hydrophilicity

(a_2). For all these cases, we obtain the same formal expression for the effective homo-polymer partition function

$$\mathcal{Z}_N^{eff} = \sum_{W_N} \exp\left[N\beta_0 + \beta_1 \sum_i z_i + \beta_2 \sum_i z_i^2\right], \qquad (7)$$

and any further complexity is hidden in the computation of β_0, β_1 and β_2 for the different cases. The strategy we will follow, is to study the general homo-polymer model defined by (7), in the (β_1, β_2)-plane (the β_0 dependence being trivial). Then, we investigate to which temperature dependent trajectories in the (β_1, β_2)-plane, the three annealed averages give rise.

In case (a_0), the simple annealed average (4) has already been computed in the preceding section, and equation (7) is recovered with the definitions

$$\beta_0 = 1, \qquad \beta_1 = \beta\lambda_0, \qquad \beta_2 = \frac{1}{2}\beta^2\lambda^2. \qquad (8)$$

We can immediately argue from equations (7) and (8) that even hydrophobic chains ($\lambda_0 < 0$) will be swollen at low enough temperature. Indeed, since $\beta_2 \gg |\beta_1|$ as $\beta \to \infty$, the repulsive $\beta_2 \sum_i z_i^2$ term causes the number of contacts with solvent to be maximized, independently of the sign and the value of λ_0.

In case (a_1), we fix the *a posteriori* overall hydrophilicity $\sum_i \lambda_i/N$ to its *a priori* value λ_0:

$$\frac{\langle \sum_i \lambda_i \rangle}{N} = \lambda_0. \qquad (9)$$

We impose this constraint by defining a generalized partition function $\mathcal{Z}_N^{a_1}(\{\lambda_i\}, \hat{h})$ which depends on the Lagrange multiplier $\hat{h}$, and by finding the effective value $\hat{h}^*$ for which (9) holds:

$$\mathcal{Z}_N^{a_1}(\{\lambda_i\}, \hat{h}) = \sum_{W_N} \exp\left[\beta \sum_i \lambda_i z_i - \beta\hat{h}\left(\sum_i \lambda_i - N\lambda_0\right)\right]. \qquad (10)$$

We recover (7) with the following definitions

$$\beta_0 = \frac{\beta^2\lambda^2\hat{h}^2}{2}, \quad \beta_1 = \beta\lambda_0 - \beta^2\lambda^2\hat{h}, \quad \beta_2 = \frac{\beta^2\lambda^2}{2}\cdot \qquad (11)$$

In terms of $\hat{h}$, condition (9) becomes

$$\hat{h}^* = \frac{\langle \sum_i z_i \rangle (\hat{h}^*)}{N}, \qquad (12)$$

and the free energy f_{a_1} is thus given by

$$f_{a_1} = -\frac{1}{\beta N} \ln \left\langle\!\left\langle \mathcal{Z}_N^{a_1}(\hat{h}^*) \right\rangle\!\right\rangle. \qquad (13)$$

Note that the quenched free energy computed in the continuum model by Garel *et al.* [7], is exactly the free energy f_{a_1} (13), if one performs the annealed average with constraint (9) within the analytic calculation scheme of reference [7]. We will comment on this in the final discussion of Section 7.

In case (a_2), in addition to (9), we also put a constraint on the overall variance $\sum_i (\lambda_i - \lambda_0)^2/N$:

$$\frac{\langle \sum_i \lambda_i^2 \rangle}{N} = \lambda^2 + \lambda_0^2. \qquad (14)$$

In the same way as before, we introduce a second Lagrange multiplier $\hat{s}$, and we define a generalized partition function

$$Z_N^{a_2}(\{\lambda_i\}, \hat{h}, \hat{s}) = \sum_{W_N} \exp\left[\beta \sum_i \lambda_i z_i - \beta\hat{h}\left(\sum_i \lambda_i - N\lambda_0\right) - \frac{\beta\hat{s}}{2}\left(\sum_i \lambda_i^2 - N(\lambda^2 + \lambda_0^2)\right)\right]. \qquad (15)$$

After performing the average, we recover (7) with the following definitions

$$\beta_0 = \frac{1}{2}\left(\frac{\beta^2\lambda^2}{s'}(\hat{h} + \lambda_0\hat{s})^2 + \beta\lambda^2\hat{s} - \ln s'\right),$$
$$\beta_1 = \frac{\beta\lambda_0 - \beta^2\lambda^2\hat{h}}{s'}, \quad \beta_2 = \frac{\beta^2\lambda^2}{2s'} \quad s' = 1 + \beta\lambda^2\hat{s}. \qquad (16)$$

Conditions (9) and (14) yield two coupled equations for $\hat{h}^*$ and $\hat{s}^*$ with solution:

$$\hat{h}^* + \lambda_0\hat{s}^* = \Delta_2(\hat{h}^*, \hat{s}^*), \qquad (17)$$
$$\hat{s}^* = \frac{\sqrt{1 + 4\beta^2\lambda^2(\Delta_3(\hat{h}^*, \hat{s}^*) - \Delta_2^2(\hat{h}^*, \hat{s}^*))} - 1}{2\beta\lambda^2}, \qquad (18)$$

where $\Delta_2 \equiv \langle \sum_i z_i \rangle / N$ and $\Delta_3 \equiv \langle \sum_i z_i^2 \rangle / N$. In the next section, we will show that Δ_2 and Δ_3 are closely connected to effective 2-, respectively 3-body interactions between the monomers. The free energy f_{a_2} is now given by

$$f_{a_2} = -\frac{1}{\beta N} \ln \left\langle\!\left\langle \mathcal{Z}_N^{a_2}(\hat{h}^*, \hat{s}^*) \right\rangle\!\right\rangle. \qquad (19)$$

As mentioned before, it can be shown [13] that at any temperature $f_{a_0} \leq f_{a_1} \leq f_{a_2} \leq ... \leq f_q$. Hence, fixing more and more moments, we get a better approximation for the quenched free energy.

4 Phase diagram

We now discuss the phase diagram of the model defined in (7), in the (β_1, β_2)-plane, with β_0 constant:

$$\mathcal{Z}_N^{eff} = C \sum_{W_N} \exp\left[\beta_1 \sum_i z_i + \beta_2 \sum_i z_i^2\right]. \qquad (20)$$

The number z_i of nearest-neighbor contacts of monomer i (not at the chain ends) with solvent molecules can be expressed in terms of the number n_i of nearest-neighbor monomers of monomer i not along the chain using

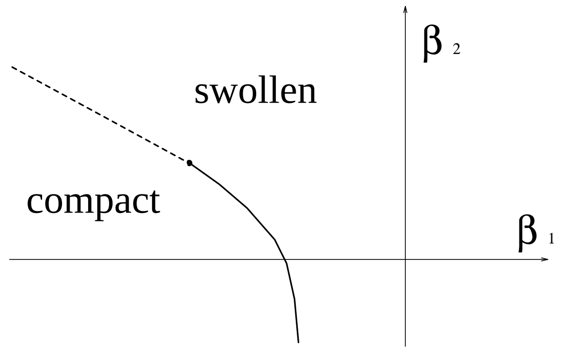

Fig. 1. Qualitative phase diagram in the (β_1,β_2) plane: the solid line is the tri-critical θ-line which ends in the multi-critical point; the dashed line is the coexistence line.

$n_i = 2(d-1) - z_i$, where we have exploited the incompressibility condition. In this way the quantities $\sum_i z_i$ and $\sum_i z_i^2$ can be related to effective 2-body and 3-body interactions between the monomers. We define N_2 to be the number of nearest-neighbor 2-monomer contacts not along the chain, while N_3 is the number of *nearest* 3-monomer contacts not along the chain. On a hypercubic lattice in $I\!R^d$, we have a nearest 3-monomer contact when two monomers are both nearest neighbors (not along the chain) to the third monomer and we have that

$$N_2 = \frac{\sum_i n_i}{2}, \qquad N_3 = \frac{(\sum_i n_i^2 - \sum_i n_i)}{2}\cdot \tag{21}$$

In terms of N_2 and N_3, the reduced Hamiltonian can be rewritten

$$\begin{aligned} -\beta\mathcal{H}^{eff} &= 2(d-1)[\beta_1 + 2(d-1)\beta_2]N \\ &\quad - [2\beta_1 + 2(4d-5)\beta_2]N_2 + 2\beta_2 N_3. \end{aligned} \tag{22}$$

Since $\beta_2 \geq 0$ for the models that we have considered, the 3-body term is either attractive or absent. Note that β_2 enters also in the 2-body term with a *repulsive* effect; as already noted, the total contribution of the β_2 term in equation (20) has to be repulsive. We recall that the self-avoidance constraint automatically introduces effective n-body repulsive terms, with $n = 2, 3, 4$ and so on.

For $\beta_2 = 0$, we have a self-avoiding walk with only 2-body interactions, each of energy $2\beta_1/\beta$. For these models, the existence of a critical value $\beta_1^* < 0$ at which the chain undergoes a second order θ-transition, is well-known [15–17]. The transition takes place when the 2-body attractive interaction exactly balances the 2-body steric repulsion. For $\beta_1 > \beta_1^*$, the chain is in the swollen phase, while for $\beta_1 < \beta_1^*$, the chain is in the collapsed phase. The θ-point is a tri-critical point [18] corresponding to a ϕ^6 field theory with a Landau-Ginzburg functional; the necessary stabilizing 3-body term is provided by the self-avoidance constraint.

When $\beta_2 > 0$, the attractive 3-body term $2\beta_2 N_3$ competes with the corresponding steric repulsion. At the mean field level [14], with increasing β_2, a tri-critical line departs from the θ-point at $\beta_2 = 0$. The tri-critical line ends at a multi-(tetra-)critical point (β_1^m, β_2^m), when the 3-body attractive interaction exactly balances the steric repulsion. This corresponds to a ϕ^8 Landau-Ginzburg theory, since the necessary stabilizing 4-body term is provided by the self-avoidance constraint. Increasing β_2 further, the transition line becomes a coexistence line between the swollen and the compact phase. This phase diagram is qualitatively sketched in Figure 1.

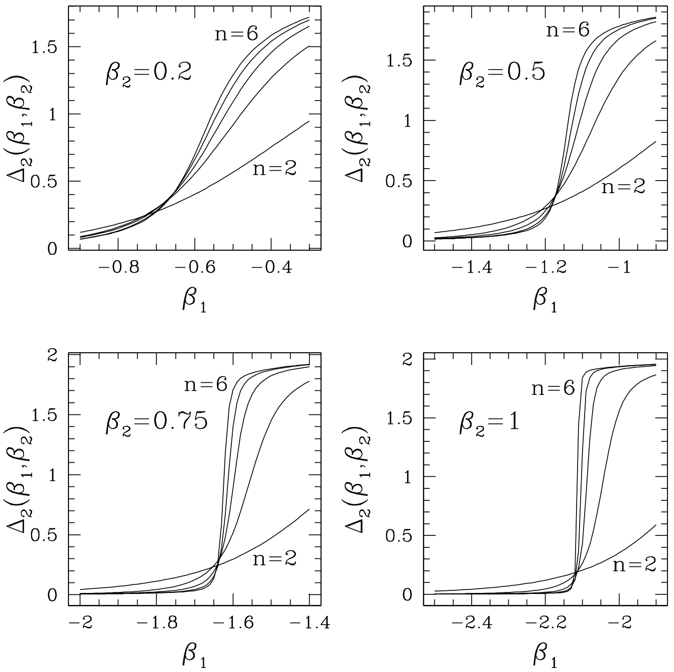

Fig. 2. Mean number of monomer-solvent close contacts $\Delta_2\,(\beta_1, \beta_2)$ at varying β_1 for different fixed values of β_2, with strip width from 2 to 6. The compact-to-swollen transition is continuous for $\beta_2 < \beta_2^m$ and first order for $\beta_2 > \beta_2^m$, with the (very raw) estimate $\beta_2^m \simeq 0.75$.

At zero temperature, when the entropy is negligible with respect to the energy, we can give rigorous results for the asymptotic behavior of the coexistence line. If $a = \beta_1/\beta_2$ is fixed and $\beta_2 \to \infty$, we can rewrite the Hamiltonian

$$-\beta\mathcal{H}^{eff} = \beta_2\sum_i(az_i + z_i^2) = -N\beta_2\frac{a}{4} + \beta_2\sum_i\left(\frac{a}{2} + z_i\right)^2. \tag{23}$$

Since $0 < z_i < 2(d-1)$, for $a > -2(d-1)$, the ground state is at $z_i = 2(d-1)$ for any i, and the walk is swollen. On the other hand, for $a < -2(d-1)$, the ground state is at $z_i = 0$ for any i, and the walk is maximally compact.

For $a = -2(d-1)$ (*i.e.* $\beta_1 \sim -2(d-1)\beta_2$), the energy of the two competing ground states is the same, and there is phase coexistence.

The presence of a multi-critical point, if not rigorously proved, is numerically established, as one can see from Figure 2, where the order parameter $\Delta_2(\beta_1, \beta_2)$ is plotted as a function of β_1 for different values of β_2.

Camacho and Schanke [19], using exact enumerations, have obtained a phase diagram which exhibits similar features as our. A transition line between the swollen and the collapsed phase is present, which is first order at low temperatures, and becomes second order at higher temperatures through a multi-critical point. However, they treat the quenched case and describe a slightly different model (*i.e.* the HP-model). A translation of this model in terms of hydrophobic charges would introduce an extra interaction term which depends on the product of the charges, and is absent in our model.

5 Transfer matrix

We have addressed the numerical study of the lattice model defined by (7), by the transfer matrix technique on a two-dimensional square lattice. With this method it is possible to consider infinite polymers on a lattice of finite width (strip) [20–22]. The price to pay is the uncertain extrapolation of the thermodynamic limit, caused by the limited width of the strip that we can achieve.

In a grand-canonical context, the generalized two point correlation function is defined as

$$\mathcal{G}(x,r,\beta_1,\beta_2)=\sum_{N=1}^{\infty}\sum_{W_N}x^N\exp\left[\beta_1\sum_i z_i+\beta_2\sum_i z_i^2\right], \tag{24}$$

where x is the step fugacity and the second sum runs over the SAW of N steps which connect the origin with an arbitrary point at distance r. We have neglected the dependence on β_0 because it only affects a simple rescaling of fugacity. The two-point correlation function decreases exponentially in r at long distances, if x is less than the critical fugacity $x_c(\beta_1,\beta_2)$. This defines the correlation length $\xi(x,\beta_1,\beta_2)$:

$$\mathcal{G}(x,r,\beta_1,\beta_2)\sim\exp\left(-\frac{r}{\xi(x,\beta_1,\beta_2)}\right)\cdot \tag{25}$$

The correlation length $\xi_n(x,\beta_1,\beta_2)$ can be calculated exactly on a lattice strip of infinite length and finite width n, with the TM method. The idea is to write recursion relations between a strip of length r and a strip of length $r+1$. We consider a walk on the strip, which goes from the left to the right, and we cut the strip at column r. The local configuration at r is then given by the set of occupied sites of column r and how these are connected to each other by the part of the walk at the left of r. Since the interaction $\beta_2\sum_i z_i^2$ gives rise to effective 3-body interactions, it is necessary to define the local configurations at stage r taking the lattice bonds occupied by the walk between the columns r–2 and r into account. We combine all the possible local configurations i at column r, with all the possible local configurations j at column $r+1$. They yield a non-zero TM element T_{ij} if it is possible to connect them, without producing disconnected pieces, and T_{ij} is given by

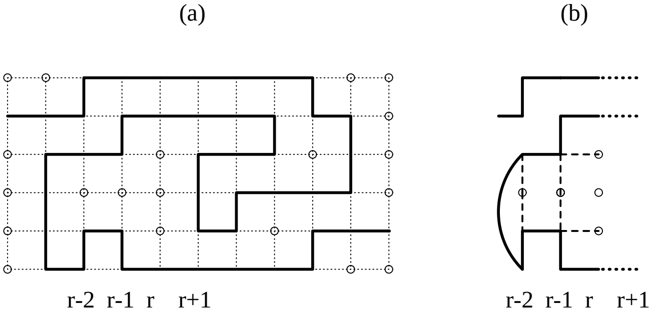

Fig. 3. Example of a transfer matrix element. Empty circles are solvent molecules and dashed lines show the nearest-neighbor monomer-solvent contacts. Configuration i is defined at column r and takes into account how the walk steps back to column $r-2$ (solid line in (b)). Configuration j is defined at column $r+1$ and takes into account how the walk steps back to column $r-1$, and thus, partially overlaps with configuration i. The dotted line in (b) shows the non overlapping part of configuration j. In this example the quantities needed for the computation of the matrix elements (26) are $t_{ij}=5$ and $z^{ij}=(0,0,2,0,2,0)$ (the sites of column $r-1$ are ordered from the top to the bottom of the strip).

$$T_{ij}=x^{t_{ij}}\exp\left[\beta_1\sum_{\alpha=1}^{n}z_\alpha^{ij}+\beta_2\sum_{\alpha=1}^{n}(z_\alpha^{ij})^2\right]; \tag{26}$$

where t_{ij} is the number of occupied bonds between columns r–1 and r, and z_α^{ij} is the number of non occupied nearest-neighbor sites of the site at row α and column r–1, if this is occupied by the walk (see Fig. 3).

The number of possible configurations, and therefore the computational effort, can be strongly reduced by considering periodic boundary conditions (the strip becomes a cylinder) and then by exploiting all the symmetry properties of the strip. Furthermore, periodic boundary conditions reduce the finite size effects. In this way, within reasonable time, we are able to study strip widths up to $n=6$, corresponding to 5387 configurations and 154 149 non-zero matrix elements.

The correlation function can be expressed in terms of the trace of the rth power of the TM T:

$$\mathcal{G}(x,r,\beta_1,\beta_2)\sim\mathrm{Tr}\,T^r, \tag{27}$$

and the correlation length (25) is related to the largest eigenvalue $\lambda_n^{max}(x,\beta_1,\beta_2)$ of T, for a strip of width n:

$$\xi_n(x,\beta_1,\beta_2)=-\frac{1}{\ln\lambda_n^{max}(x,\beta_1,\beta_2)}\cdot \tag{28}$$

The critical fugacity x_c^n is determined by the value at which the correlation length diverges, *i.e.*

$$\lambda_n^{max}(x_c^n,\beta_1,\beta_2)=1. \tag{29}$$

The computation of the free energy per monomer $f=-\ln\mathcal{Z}_N^{eff}/(\beta N)$ and of any other quantity of physical interest (*e.g.* the mean number of monomer-solvent contacts

Δ_2), is now straightforward in terms of the critical fugacity:

$$f = \frac{1}{\beta}\ln x_c, \qquad \Delta_2 = -\frac{\partial \ln x_c}{\partial \beta_1}, \qquad \Delta_3 = -\frac{\partial \ln x_c}{\partial \beta_2}\,\cdot \tag{30}$$

The thermal exponent ν, which characterizes the divergence of the correlation length at the critical fugacity: $\xi \sim (x_c - x)^{-\nu}$, is a good indicator of a collapse transition. A SAW in two dimensions has the value $\nu = 3/4$ in the swollen phase [23], $\nu = 1/2$ in the collapsed phase and $\nu = 4/7$ on the tri-critical line [24,25].

In order to compute the thermal exponent, we first calculate the density $\rho_n(\beta_1, \beta_2)$ of monomers in the strip

$$\rho_n(\beta_1,\beta_2) = -\frac{x_c^n(\beta_1,\beta_2)}{n}\frac{\partial \xi_n^{-1}(x_c^n,\beta_1,\beta_2)}{\partial x}\,\cdot \tag{31}$$

Then, we use a phenomenological renormalization (PR) group procedure [26,27] to obtain finite size estimates for the thermal exponent; the underlying hypothesis is the finite size scaling behavior [28] of the correlation length for $n \gg 1$ and $(x_c - x) \ll 1$:

$$\xi_n\left(x,\beta_1,\beta_2\right) = n\; g\left[n^{1/\nu}\left(x_c - x\right),\beta_1,\beta_2\right], \tag{32}$$

where g is a scaling function. Using the single strip critical fugacity estimate (29) leads to

$$\nu_{n,n-1} = \left(\frac{\ln(\rho_n/\rho_{n-1})}{\ln(n/(n-1))} + 2\right)^{-1}\cdot \tag{33}$$

Note that we compare the derivative of the correlation length at criticality for two consecutive strip widths, but criticality is determined in a different way for different widths.

This is not the most accurate way of applying the ideas of the PR. In fact, the critical fugacity can be determined for two consecutive widths at once by

$$\frac{\xi_n\left(x_c^{n,n-1},\beta_1,\beta_2\right)}{n} = \frac{\xi_{n-1}\left(x_c^{n,n-1},\beta_1,\beta_2\right)}{n-1}\,\cdot \tag{34}$$

This estimate is better than (29) and the thermal exponent can easily be obtained

$$\nu_{n,n-1} = \left(\frac{\ln\left(\frac{\partial \xi_n(x_c^{n,n-1},\beta_1,\beta_2)}{\partial x}\Big/\frac{\partial \xi_{n-1}(x_c^{n,n-1},\beta_1,\beta_2)}{\partial x}\right)}{\ln(n/(n-1))} - 1\right)^{-1}\cdot \tag{35}$$

Nevertheless, we have used the rougher formulae (29) and (33), because solving (34) numerically, is a much harder task, especially in proximity of coexistence.

6 Results

We will show that the trajectory that the system follows in the (β_1,β_2)-plane, for the different models with decreasing temperature, only depends on the fraction $\lambda_0/\lambda \equiv \lambda_{\mathit{eff}}$. The position on the trajectory at a given temperature, however, does depend on λ_0 and λ separately. Although the trajectories can not be calculated analytically, we give a general qualitative picture, which will be confirmed by the numerical data.

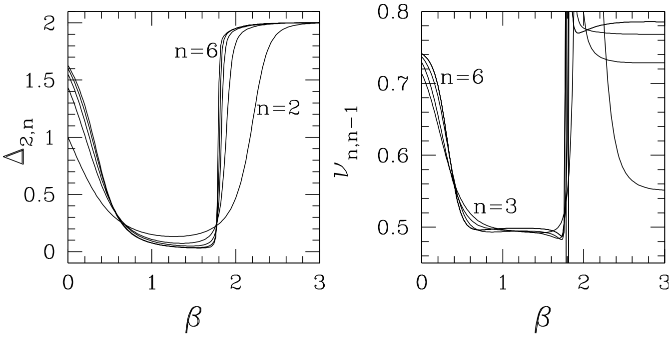

Fig. 4. Mean number of monomer-solvent close contacts $\Delta_{2,n}$ and thermal exponent $\nu_{n,n-1}$ at varying temperature, with strip width n from 2 to 6, in the case of unconstrained annealing with $\lambda_0 = -1$ and $\lambda = 0.7$. Evidence is provided for a second order swollen-to-compact θ-transition (see the crossings of different n-estimates of the thermal exponent around the θ-value $\nu_\theta \simeq 0.57$), and for a first order compact-to-swollen transition (see the abrupt jump of the order parameter $\Delta_{2,n}$). The thermal exponent strongly fluctuates at the first order transition due to the phenomenological renormalization method employed for its calculation.

The maximum strip width avalaible, $n = 6$, is rather small. Nevertheless, the data in the various plots already show a nice convergence, and we think that the obtained results are reliable. Moreover, we are mainly interested in qualitative features of the phase diagram and not in precise quantitative values of critical exponents or transition temperatures.

6.1 Simple annealing (a_0)

After elimination of the temperature in (8), the locus of the trajectory in the (β_1,β_2)-plane is given by the equation

$$g_{a_0}(\beta_1,\beta_2) \equiv \beta_2 - \frac{\lambda_{\mathit{eff}}^2\beta_1^2}{2} = 0, \tag{36}$$

which describes a parabola. As the temperature is positive, the $(\beta_1 > 0)$-branch of this parabola has to be considered, for $\lambda_{\mathit{eff}} > 0$, while the $(\beta_1 < 0)$-branch is relevant for $\lambda_{\mathit{eff}} < 0$. Hence, at sufficiently low temperature, the chain will be always swollen, no matter how strongly hydrophobic λ_{eff} is (*i.e.* $\lambda_{\mathit{eff}} < 0$). We note here that the opposite result, *i.e.* even highly hydrophilic chains are compact at sufficiently low temperature, has been found by Garel *et al.* [7] in the corresponding continuum model, due to improper consideration of the incompressibility condition. This condition is automatically accounted for in the definition of our lattice model.

The typical behaviour for a strongly hydrophobic chain ($\lambda_{\mathit{eff}} \ll -1$) is as follows. At high temperatures, it will be swollen for entropic reasons. Then with decreasing temperature, it will undergo a 2nd order θ-transition from swollen to collapsed. Finally, at even lower temperature, it will undergo a 1st order collapsed to swollen transition. We present a numerical evidence of this remarkable re-entrant behaviour in Figure 4: the crossings of the various n-estimates of the thermal exponent are typical of θ-point [16] and are just around the value $\nu_\theta \simeq 0.57$, and the jump of the order parameter $\Delta_2 = \langle \sum_i z_i/N \rangle$ provides strong evidence for the first order transition from the compact to the swollen phase; the value of the compact phase being $\Delta_2 \simeq 0$, and in the swollen phase $\Delta_2 \simeq 2$. We have thus shown that considering 3-body interaction does not change the universality class of the θ-transition, as long as one is referring to the tri-critical line. Although not surprising, this result is not trivial in two dimensions. A more interesting theoretical question concerns the value of the thermal exponent in $d = 2$ at the multi-critical point, but the TM approach employed here, is uneffective because of the limited strip width we are able to study.

We note here that we have considered the overall hydrophilicity λ_0 to be constant. The chemical reactions that give rise to annealed hydrophilicities, however, may be temperature dependent and may cause λ_0 to vary with temperature [29]. Nevertheless, the re-entrant behavior is a quite robust feature: even when λ_0 diverges exponentially to $-\infty$ ($\lambda_0^* < 0$) with a rate $\alpha > 0$

$$\lambda_0(\beta) = \lambda_0^* \exp(\alpha\beta), \tag{37}$$

re-entrant behavior is still observed for α not too big.

6.2 Fixing the mean (a$_1$)

After elimination of the temperature in (11), the locus of the trajectory in the (β_1, β_2)-plane is given by the equation

$$g_{a_1}(\beta_1,\beta_2) \equiv \beta_1 - \sqrt{2\beta_2}\lambda_{\mathit{eff}} + 2\beta_2\Delta_2(\beta_1,\beta_2) = 0, \tag{38}$$

which has to be combined with condition (12) (fixing the mean)

$$\Delta_2(\beta) \equiv \Delta_2(\beta_1,\beta_2) = \frac{1}{N}\frac{\partial \ln \mathcal{Z}_N^{\mathit{eff}}(\beta_1,\beta_2)}{\partial \beta_1} = \frac{1}{N}\left\langle \sum_i z_i \right\rangle (\beta_1,\beta_2). \tag{39}$$

First, we show that equation (38) defines a unique trajectory $\beta_1^{tr}(\beta_2)$. It is easy to see that $\partial g_{a_1}/\partial\beta_1 > 0$, $\forall \beta_1$, β_2. Hence, we can apply the implicit function theorem, but only if $\beta_2 < \beta_2^m$, when $g_{a_1}(\beta_1,\beta_2)$ is a continuous function of its arguments. For $\beta_2 > \beta_2^m$, $g_{a_1}(\beta_1,\beta_2)$ (in particular Δ_2), is discontinuous on the coexistence line $\beta_1^{co}(\beta_2)$. Since the discontinuity is developed in the thermodynamic limit, and since $\partial g_{a_1}/\partial\beta_1 > 0$ for any finite N, the only possibility at coexistence is $g_{sw} \equiv g_{a_1}(\beta_1 \downarrow \beta_1^{co}(\beta_2),\beta_2) > g_c \equiv g_{a_1}(\beta_1 \uparrow \beta_1^{co}(\beta_2),\beta_2)$ (the chain is collapsed for $\beta_1 < \beta_1^{co}(\beta_2)$ and swollen otherwise). If g_{sw} and g_c have the same sign, (38) is still uniquely satisfied far away from coexistence. Instead, if $g_{sw} > 0$ and $g_c < 0$, (38) can only be satisfied on the coexistence line. In this case, a fraction f_c of the chain is collapsed and the remaining fraction $1 - f_c$ is swollen, so that physical mean values are mixtures of the pure phase values:

$$\Delta_2(\beta) = f_c\ \Delta_{2,c}(\beta) + (1-f_c)\ \Delta_{2,sw}(\beta). \tag{40}$$

Thus, equation (38) becomes a condition on f_c:

$$f_c\ g_c + (1-f_c)\ g_{sw} = 0. \tag{41}$$

We now prove that at zero temperature ($\beta_2 \to \infty$) the only way to satisfy (38) for the chain, is to be at the coexistence line with $f_c = \frac{1}{2}$, *i.e.* half collapsed and half swollen, independently of the value of λ_{eff}, as far it is kept fixed. For $\beta_1/a \simeq \beta_2 \to \infty$, (38) becomes

$$a + 2\Delta_2(\infty) = 0. \tag{42}$$

As we have seen at the end of Section 4, for $a > -2$ the chain is swollen, and $\Delta_{2,sw}(\infty) = 2$, but (42) implies $a = -4$ which is a contradiction. Similarly for $a < -2$, the chain is compact and $\Delta_{2,c}(\infty) = 0$, implying $a = 0$. Hence, the only remaining possibility is $a = -2$, *i.e.* coexistence of the swollen and the compact phase:

$$\Delta_2(\infty) = f_c\Delta_{2,c} + (1-f_c)\Delta_{2,sw}, \tag{43}$$

where f_c is the collapsed fraction of the chain at coexistence. Plugging this in (42), yields $f_c = \frac{1}{2}$, such that $\Delta_2(\infty) = 1$.

The phase separation already occurs at finite temperature, since condition (38) implies that $\Delta_2(\beta_1,\beta_2)$ is a continuous function along the trajectory, and since the only way to reach the value $\Delta_2(\infty) = 1$ continuously, is to move along the coexistence line.

All this, in combination with the numerical data, gives the following qualitative scenario of what happens lowering the temperature:

- There exists a particular value $\lambda_m < 0$ such that for $\lambda_{\mathit{eff}} = \lambda_m$ the trajectory passes through the multi-critical point, and then follows the coexistence line.
- For $\lambda_{\mathit{eff}} > \lambda_m$, the trajectory hits the coexistence line coming from the swollen phase, and this will happen further away from the multi-critical point the larger $\lambda_{\mathit{eff}} - \lambda_m$ is. Then, it follows the coexistence line, and the collapsed fraction (f_c) of the chain steadily increases to become $\frac{1}{2}$ at zero temperature.
- For $\lambda_{\mathit{eff}} < \lambda_m$, the trajectory hits the coexistence line coming from the compact phase. This means that the chain first collapses with a 2nd order θ-transition, before the trajectory hits the coexistence line. Then, it follows the coexistence line, and the collapsed fraction (f_c)

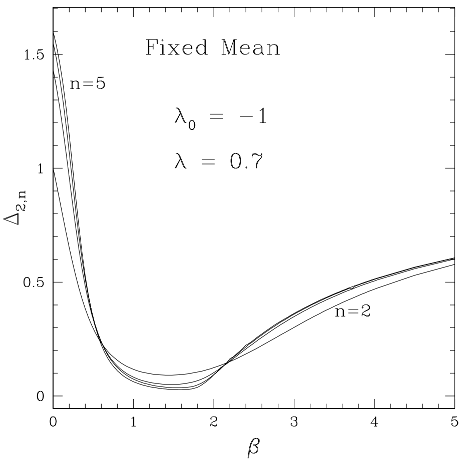

Fig. 5. Mean number of monomer-solvent close contacts $\Delta_{2,n}$ at varying temperature, with strip width n from 2 to 5, in the fixed mean case. The behavior of the order parameter is the same (swollen at high temperature and then θ-transition to the collapsed phase) as in the annealed case (see Fig. 3) until coexistence starts at $\beta \simeq 1.7$ and $\Delta_{2,n}$ starts increasing slowly; the asymptotic ($\beta \to \infty$) value is $\Delta_2 = 1$. The behavior of the thermal exponent is the same as in the annealed case until coexistence starts. At coexistence the thermal exponent should be the same as in the swollen phase, as far as a finite fraction of the chain is swollen, but due to limited numerical precision we get higly fluctuating values.

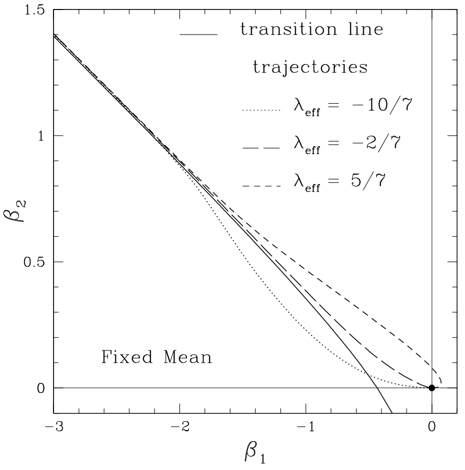

Fig. 6. Trajectories in the (β_1,β_2) plane for different values of λ_{eff} in the fixed mean case, with strip width $n = 5$. The transition line has been located at the crossing of two consecutive n-estimates of Δ_2 (see Fig. 2).

of the chain steadily decreases to become $\frac{1}{2}$ at zero temperature.

This qualitative scenario is confirmed by the numerical evidence shown in Figures 5 and 6. They respectively show the numerical results for the trajectory in the (β_1, β_2)-plane, and the variation of the order parameter $\Delta_2(\infty)$ with temperature.

6.3 Fixing the mean and the variance (a$_2$)

Using (17) and (16), we obtain $\lambda\beta = (\beta_1 + 2\Delta_2\beta_2)/\lambda_{\mathit{eff}} \equiv \chi > 0$, and the locus of the trajectory in the (β_1,β_2) plane is given by the following equation

$$g_{a_2}(\beta_1, \beta_2) \equiv \frac{\chi^2}{2\beta_2} - \frac{1}{2}\sqrt{\frac{1}{4} + \chi^2\left(\Delta_3 - \Delta_2^2\right)} = 0, \quad (44)$$

where $\Delta_2(\beta)$ is defined as in (39), and

$$\begin{aligned} \Delta_3(\beta) \equiv \Delta_3(\beta_1, \beta_2) &= \frac{1}{N}\frac{\partial \ln \mathcal{Z}_N^{\mathit{eff}}(\beta_1, \beta_2)}{\partial \beta_2} \\ &= \frac{1}{N}\left\langle \sum_i z_i^2 \right\rangle (\beta_1, \beta_2). \quad (45) \end{aligned}$$

The qualitative behaviour of the trajectories in the phase plane is very similar to that of the previous subsection. The collapsed fraction f_c of the chain, however, does depend on λ_{eff} at zero temperature. In order to show this, we repeat the same argument as before. For $\beta_1/a \equiv \beta_2 \to \infty$, condition (44) simplifies to

$$a + 2\left(\Delta_2(\infty) - \lambda_{\mathit{eff}}\sqrt{\Delta_3(\infty) - \Delta_2^2(\infty)}\right) = 0, \quad (46)$$

which is the analogous of equation (42). For $a > -2$ the chain is swollen and $\Delta_{2,sw}(\infty) = 2$, $\Delta_{3,sw}(\infty) = 4$, but (46) implies $a = -4$ which is a contradiction. For $a < -2$ the chain is collapsed, and $\Delta_{2,c}(\infty) = \Delta_{3,c}(\infty) = 0$ leads to $a = 0$. Again, we conclude that the chain is at coexistence at zero temperature, but in this case using (43) and the analogous formula for $\Delta_3(\infty)$, we get the following condition for the collapsed fraction f_c of the chain:

$$f_c = \frac{1}{2}\left[1 - \frac{\lambda_{\mathit{eff}}}{\sqrt{1 + \lambda_{\mathit{eff}}^2}}\right]. \quad (47)$$

In Figure 7 the numerical results are shown for the variation of the order parameter Δ_2 with temperature. The analogous of Figure 6 with the trajectories in the (β_1, β_2)-plane turns out to be indistinguishable from Figure 6 itself, and is therefore not shown.

The behaviour of the chain seems qualitatively unchanged adding the constraint on the variance with respect to the fixed mean case. It can easily be verified that for $\lambda_0 = 0$ equations (38) and (44), defining the trajectories in the (β_1, β_2)-plane, are equal, and they are qualitatively very similar for $\lambda_0 \neq 0$. If we compare the free energies (13) and (19), however, taking the proper values of β_0 into account, we find that the constraint on the variance is crucial for the low temperature behaviour of the free energy. In the fixed mean case (like in the simple annealed case), the free energy diverges linearly to $-\infty$ with β, whereas fixing also the variance yields a finite free energy. The divergence can easily be understood. A fraction (all, in the simple annealed case) of the monomers want to be as hydrophilic as possible and to maximize their solvent

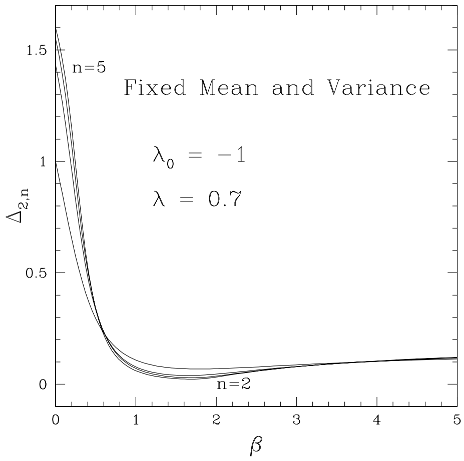

Fig. 7. As in Figure 5, in the fixed mean and variance case. The asymptotic ($\beta \to \infty$) value is $\varDelta_2 \simeq 0.18$ (see Eq. (47)).

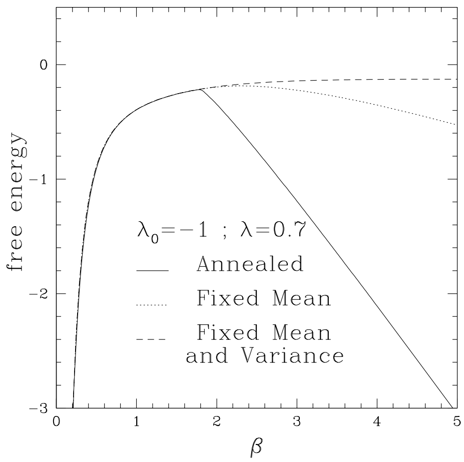

Fig. 8. Free energies f_{a_0}, f_{a_1} and f_{a_2}, of the three considered annealed cases, at varying temperature, with the same λ_{eff}, and strip width $n = 5$. While f_{a_0} and f_{a_1} diverge linearly to $-\infty$ as $\beta \to \infty$, f_{a_2} is constant in the same limit.

contacts, in order to minimize the energy, while the other fraction has to be very hydrophobic to keep the mean finite. For entropical reasons the fractions are exactly $\frac{1}{2}$. This is illustrated in Figure 8, where the free energies are compared for the various cases for the same values of λ_0 and λ.

6.4 Towards the quenched average

So far, we have only fixed overall moments of the type $\sum_i \lambda_i^l$, $l \in \mathbb{N}$. In this way, we do not impose the λ_i to be independent variables. Or equivalently, even if we fix all the overall moments, the λ_i still have the complete freedom to rearrange themselves along the chain. Hence, we can assume that fixing both mean and variance may be a good approximation for a hetero-polymer, whose hydrophobicities are fixed, but are allowed to migrate. In a protein, however, not only the hydrophilicities, but also the positions along the chain are fixed. This corresponds to the quenched case.

In order to get a reasonable approximation by means of annealed averages, one should also ensure the independence of the λ_i. A first try might be to impose *e.g.* $\left\langle \sum_i \lambda_i \lambda_{i+1} \right\rangle / N = \lambda_0^2$, but after performing the average, one discovers immediately that this would involve a coupling between all the z_i, which, obviously, can not be done by the TM method. Instead as a first approach, one could start with

$$\left\langle\!\!\left\langle \exp\Big(-\frac{1}{2} \sum_{j,k} (\lambda_j - \lambda_0) M_{jk} (\lambda_k - \lambda_0) + \sum_j \lambda_j \beta (z_j - \hat{h}) - \log(P(\{\lambda_i\})) \Big) \right\rangle\!\!\right\rangle,$$

$$M_{jk} \equiv \int_0^{2\pi} \frac{dq}{2\pi} \frac{\exp(iq(j-k))}{(\hat{s} + \hat{d}\cos(q))} \,. \tag{48}$$

After performing the average, this results in the following expression (up to constants)

$$\exp\Big(\frac{\beta^2}{2} \sum_{j,k} (z_j - \hat{h}) M_{jk}^{-1} (z_k - \hat{h}) + \lambda_0 \beta \sum_j (z_j - \hat{h}) + \frac{N}{2} \log \left[\hat{s} + \sqrt{\hat{s}^2 - \hat{d}^2}\, \right] \Big),$$

$$M_{jk}^{-1} \equiv \hat{s}\ \delta_{j,k} + \frac{\hat{d}}{2} (\delta_{j,k-1} + \delta_{j,k+1}). \tag{49}$$

We have introduced the Lagrange multipliers $\hat{h}$, $\hat{s}$ and $\hat{d}$, which combined fix $\left\langle \sum_j \lambda_j \right\rangle = N\lambda_0$, $\left\langle \sum_j \lambda_j^2 \right\rangle = N(\lambda_0^2 + \lambda^2)$, and $\left\langle \sum_{j>k} \lambda_j M_{jk} \lambda_k \right\rangle = \lambda_0^2 \sum_{j>k} M_{jk}$, which ensures the independence of a linear combination of the λ_j. We expect that fixing the latter, may already qualitatively describe the quenched case very well. The only reason we did not do the numerics of this case, is of a purely practical nature. In order to calculate $\left\langle \sum_i z_i z_{i+1} \right\rangle$, we have to consider configurations on 3 colums instead of on 2. This increases the size of the transfer matrix so drastically that we would have to limit ourselves to very narrow strip widths. Furthermore, we have an extra self consistency equation (*i.e.* for $\hat{d}$) to solve numerically. All this makes it unfeasible (in terms of CPU time) for us at the moment, to perform this calculation for a reasonable strip width (*i.e.* ≥ 4).

Nevertheless, one may anticipate that some of the typical behaviour found for annealed averages, should not be present for the quenched case. The re-entrant behaviour at intermediate temperatures is due to the competition between the configurational entropy and the energy on the one hand, and the entropy of the λ_i distribution on the other hand. In the quenched case, the entropy of the λ_i distribution is absent, and hence re-entrant behaviour, if present at all, can not have its origin there.

The phase separation (in low dimensions) is due to the possibility for the monomers to rearrange and to form a hydrophobic compact core. Since this is not possible for the quenched polymer, we do not expect macroscopic phase separation in that case. Instead, microscopic phase separation seems to play an important role for quenched sequences [9,10]. However, one might expect the quenched polymer to behave as an effective homo-polymer, where the groundstate is either swollen or compact, depending on the value λ_{eff}.

7 Discussion of results

We have studied a simple lattice model for a random hydrophobic-hydrophilic chain in a solvent, with a Gaussian distribution for the hydrophilicities. We have considered the case of annealed disorder, without constraints, and with constraints on the first and second moments of the overall hydrophobicity.

We have obtained both exact analytical results (mainly at $T = 0$), and numerical ones, employing the transfer matrix technique on a $2d$ square lattice. One may ask whether the $2d$ results are relevant in $3d$ too. For example, in the random sequence model with charge product interaction, a simple mean field argument shows that $d = 2$ is a very peculiar case [30]. For the considered model, analytical results at the mean field level [7] do not show any qualitative difference between different spatial dimensions, and our exact results at $T = 0$ exhibit the same qualitative behavior for any $d > 1$. Hence, we believe the TM results in $2d$ to be meaningful in any dimension $d > 1$.

We now discuss our results and compare them with the ones obtained by Garel *et al.* [7] in the corresponding continuum model. The main result of [7] is the fact that the annealed and quenched cases are very similar. They find that, at sufficiently low temperature, the polymer is always collapsed, even for hydrophilic chains $\lambda_0 > 0$. Depending on the average degree of hydrophobicity, the transition to the collapsed phase is either first or second order.

- For the simple annealed case (a_0), we have shown that for any hydrophobicity $\lambda_0 > 0$ the chain is always swollen, while for any $\lambda_0 < 0$, the chain is swollen at sufficiently low temperature. Using transfer matrix techniques, we have found that, for sufficiently negative λ_0, a temperature interval (T_1, T_2) exists, where the chain is collapsed. Coming from the high temperature region the chain undergoes a standard 2nd order θ-point transition at T_2 (in the same universality class as homo-polymers [25]). Lowering the temperature further, the chain undergoes a 1st order (re-entrant) transition at T_1 towards the swollen phase.

Hence, in the annealed case, we come to the opposite conclusion for the low temperature behavior as predicted by Garel *et al.* [7]. Nevertheless, if one takes the incompressibility of the monomer-solvent system properly into account (*i.e.* putting an upper bound on the monomer density), it is possible to recover the same qualitative picture in the continuum model of [7], too.

- In the annealed case with fixing the mean (a_1), we have found that, for any λ_{eff}, there is coexistence of the swollen and the collapsed phase (phase separation) at sufficiently low temperature, and the chain is exactly half collapsed and half swollen at $T = 0$. For $\lambda_{eff} < \lambda_m$, a temperature interval (T_1, T_2) exists where the chain is collapsed. At T_2 the chain undergoes a 2nd order θ-transition from the swollen phase, while at T_1 a fraction of the chain swells, and lowering the temperature, the swollen fraction steadily increases to the $T = 0$ value $\frac{1}{2}$. For $\lambda_{eff} > \lambda_m$ a temperature T_1 exists above which the chain is swollen. At T_1 a fraction of the chain collapses, and lowering the temperature, the collapsed fraction of the chain increases steadily to the $T = 0$ value $\frac{1}{2}$.

As already noted in Section 3, the expressions for the quenched case of [7] are exactly the same as the ones we obtain for the continuum model in the case (a_1). Using a one-parameter Gaussian trial wave function for the monomer density, Garel *et al.* [7] find a collapsed phase for any λ_{eff} at low temperature. Inspired by the observation that they in fact describe the case (a_1), and by the phase separation observed in our lattice model, we tried a one-parameter trial function with a hydrophobic compact core (fraction f_c) and hydrophilic swollen tails (phase separation). In the low temperature limit, we recover the result $f_c = \frac{1}{2}$, and obtain a free energy that is considerably lower than the one obtained using the trial function of [7].

Note that the free energy diverges linearly to $-\infty$ for $T \to 0$, as in the annealed case (see Fig. 8). The same kind of divergence appears in the quenched free energy computed in [7], but the quenched free energy should not diverge at zero temperature.

- In the annealed case with fixed mean and variance (a_2), we have found very similar results as for the case (a_1). The main differences are that the collapsed fraction of the chain at $T = 0$ depends on λ_{eff} (47), and that the free energy does not diverge at $T = 0$. We repeated these calculations for the continuum model, and again we find that at $T = 0$ the phase separation trial function yields a finite groundstate energy lower than the one obtained with a Gaussian trial function, and the collapsed fraction of the chain (f_c) is found to be exactly (47).

We conclude that, on the one hand we have good evidence that lattice and continuum models exhibit the same qualitative behavior, if constrained annealing is considered. On the other hand, a good equilibrium description in the case of quenched disorder for this model seems to be lacking at present, and carrying on the constrained annealing approximation procedure (*e.g.* by fixing correlations between different hydrophilicities, as explained in the previous section) may be one way to address the problem of quenched disorder.

We would like to thank H. Orland for stimulating discussions.

References

1. J. Bryngelson, P.G. Wolynes, Proc. Natl. Acad. Sci. USA **84**, 7524 (1987).
2. T. Garel, H. Orland, Europhys. Lett. **6**, 307 (1988).
3. E.I. Shakhnovich, A.M. Gutin, Biophys. Chem. **34**, 187 (1989).
4. Following C.B. Anfinsen, Science **181**, 223 (1973), we assume that proteins in solution are in thermodynamic equilibrium.
5. K.A. Dill, Biochem. **29**, 7133 (1990).
6. S.P. Obukhov, J. Phys. A **19**, 3655 (1986).
7. T. Garel, L. Leibler, H. Orland, J. Phys. II France **4**, 2139 (1994).
8. D. Thirumalai, V. Ashwin, J.K. Bhattacharjee, Phys. Rev. Lett. **77**, 5385 (1996).
9. E.G. Timoshenko, Yu.A. Kuznetsov, K.A. Dawson, Phys. Rev. E **53** (**4**), 3886 (1996); cond-mat/9802230 (1998).
10. A. Moskalenko, Yu.A. Kuznetsov, K.A. Dawson, cond-mat/9702175 (1997).
11. T.J. Morita, Math. Phys. **5**, 1401 (1964).
12. R. Brout, Phys. Rev. **115**, 824 (1959).
13. R. Kühn, Z. Phys. B **100**, 231 (1996).
14. C. Itzykson, J.M. Drouffe, *Statistical Field Theory*, (Cambridge University Press, Cambridge, 1989)
15. P.G. de Gennes, *Scaling Concept in Polymer Physics*, (Cornell University Press, Ithaca, New York, 1979).
16. H. Saleur, J. Stat. Phys. **45**, 419 (1986).
17. F. Seno, A.L. Stella, J. Phys. France **49**, 739 (1988); Europhys. Lett. **7**, 605 (1988).
18. P.G. de Gennes, J. Phys. Lett. **39**, L55 (1975).
19. C.J. Camacho, T. Schanke, Europhys. Lett. **37**, 603 (1997).
20. O.J. Klein, J. Stat. Phys. **23**, 561 (1980).
21. B. Derrida, J. Phys. A: Math. Gen. **14**, L5 (1981).
22. H. Saleur, B. Derrida, J. Stat. Phys. **44**, 225 (1986).
23. B. Nienhuis, Phys. Rev. Lett. **49**, 1062 (1982).
24. A. Coniglio, N. Jan, I. Majid, E. Stanley, Phys. Rev. B **35**, 3617 (1987).
25. B. Duplantier, H. Saleur, Phys. Rev. Lett. **59**, 539 (1987).
26. M.P. Nightingale, Physica A **83**, 561 (1976).
27. B. Derrida, L. de Seze, J. Phys. France **43**, 475 (1982).
28. M.E. Fisher, N.B. Barber, Phys. Rev. Lett. **28**, 1516 (1972).
29. H. Orland, private communications.
30. C.D. Sfatos, A.M. Gutin, E.I. Shakhnovich, Phys. Rev. E **48**, 465 (1993).