Lattice model for cold and warm swelling of polymers in water

Pierpaolo Bruscolini* and Lapo Casetti†

Istituto Nazionale per la Fisica della Materia (INFM) and Dipartimento di Fisica, Politecnico di Torino,
Corso Duca degli Abruzzi 24, I-10129 Torino, Italy
(Received 11 October 1999)

We define a lattice model for the interaction of a polymer with water. We solve the model in a suitable approximation. In the case of a nonpolar homopolymer, for reasonable values of the parameters, the polymer is found in a noncompact conformation at low temperature; as the temperature grows, there is a sharp transition towards a compact state, then, at higher temperatures, the polymer swells again. This behavior closely reminds one of that of proteins that are unfolded at both low and high temperatures.

PACS number(s): 61.25.Hq, 05.20.−y, 05.40.Fb, 87.10.+e

The theoretical modeling of the behavior of polymers in aqueous solution is a long-standing problem which has received considerable attention also in connection with the problem of protein folding. It is commonly believed [1] that the organization of water molecules in quasiordered structures around nonpolar monomers plays a fundamental role in stabilizing the folded ''native'' state of a protein, namely, the biologically active state. In simple models of protein folding, however, water is usually taken into account within the definition of the interactions between amino acids, rather than explicitly considering water-monomer interactions. The rationale for this approach [2] is that water behaves as a bad solvent for hydrophobic monomers, and this can be modeled by an effective Hamiltonian with attractive monomer-monomer interactions. However, an ''exact'' effective Hamiltonian, obtained from a partial trace over water degrees of freedom, should in principle depend also on the temperature and on the conformation of the polymer. Hence, it is interesting to study simple models where the role of water can be considered explicitly.

Here we introduce a lattice model where each site is occupied either by a monomer or by water molecules. Water-monomer interactions depend on the state of water; water-water interactions are dealt with in a coarse-grained way. To distinguish between the effect of the presence of water and any other, monomer-monomer interactions are neglected. The model can be thought of as a generalization of the one proposed by De Los Rios and Caldarelli (DLRC) [3]. Our generalization is an attempt to model more carefully the water degrees of freedom, yet keeping the description as simple as to obtain a model that can be approached by analytical techniques. Specializing to the case of a nonpolar homopolymer, we introduce a suitable approximation scheme and evaluate analytically the partition function, the specific heat and the average number of water-polymer contacts, which is a measure of the compactness of the polymer conformation. We find that, for reasonable values of the parameters, the polymer behaves as follows: at low temperatures, it is found in a noncompact conformation. As the temperature grows, there is a sharp transition towards a compact state, associated

with a sharp peak in the specific heat of the water-polymer system; then, at higher temperatures, the polymer swells again, smoothly. This behavior reminds one of proteins, that are unfolded at both low and high temperatures [4].

Let us now define our model. We consider a d-dimensional lattice with N_s sites and coordination number z. Each lattice site is occupied either by a monomer of the polymer or by a water ''cluster.'' The polymer is made of N monomers. Each monomer can be either hydrophobic (H) or polar (P), to which cases we associate $\sigma = 0,1$, respectively. The sequence of H's and P's is fixed, so that the σ's are not dynamical variables. The number of water clusters is $N_w = N_s - N$. Each water cluster is a system of m water molecules, to account for the fact that several water molecules are affected by the presence of a monomer. For each cluster, the state-space has an ordinary part, characterized by $\chi = 0$ and made up of an infinity of states $\vartheta = 0,1,2,\dots,\infty$, and a special part, with $\chi = 1$, containing only the state ''*.'' This state is special in the sense that it more favorably interacts with hydrophobic monomers, as we will see in the following. The choices of a discrete spectrum and of a unique special state are made for the sake of simplicity.

The model Hamiltonian is written as a sum of a noninteracting water term plus a water-polymer interaction; no energy is associated to the polymer alone:

$$\mathcal{H} = \mathcal{H}_{\text{water}} + \mathcal{H}_{\text{int}}. \tag{1}$$

The first term has the form

$$\mathcal{H}_{\text{water}} = \sum_{j=1}^{N_w} [\vartheta_j(1-\chi_j) + E^* \chi_j], \tag{2}$$

where $E^* > 0$. For each label j the state $\vartheta_j = 0$ represents ice, i.e., a completely ordered state, with two hydrogen bonds per molecule. Each water-filled site, representing a cell of m molecules, in the $\vartheta = 0$ state has $2m$ hydrogen-bonds, a fraction of which is buried in the bulk of the cell, while the rest connects neighboring cells; we set the energy of this state equal to zero. The states ($\chi_j = 0$; $\vartheta_j = 1,2,\dots$) are excited states obtained by adding kinetic energy to the water cluster: the hydrogen bonds are weakened and, eventually, broken. According to this coarse-grained description, which disregards the details of water-water interactions, the degeneracy

_____
*Electronic address: pbr@athena.polito.it
†Electronic address: lapo@polito.it

$g(\vartheta)$ of the energy level ϑ is given by the number of ways in which one can assign ϑ ''quanta'' of energy to the m molecules, i.e., the number of ordered partitions of ϑ objects in m classes: $g(\vartheta)=\binom{m+\vartheta-1}{\vartheta}$.

The special state ($\chi=1$; *) models the partial ordering that water molecules are believed to assume in presence of a non-polar residue [1]. It is indeed an excited states of energy E^*, with a peculiar character: the m water molecules are thought of as geometrically ordered in the cell, so that hydrogen-bond weakening and breaking affects only the surface molecules in the water-filled site. An accurate evaluation of the degeneracy g^* of this state would require a detailed specification of the microscopic geometry of water, which is beyond the scope of the present analysis; nevertheless, a crude estimate of it can be given by noticing that, in the presence of a water-monomer contact, $m_{eff}\approx zm^{(d-1/d)}$ water molecules shall rearrange in an ordered conformation to maximize hydrogen-bonding. Then, reasoning as before, g^* will be given by the number of ways to assign $\vartheta^*=E^*$ quanta to the m_{eff} molecules, i.e., $g^*=\binom{[[m_{eff}]]+\vartheta^*-1}{\vartheta*}$, where $[[\bullet]]$ is the integer part of $\bullet$.

The second term in Hamiltonian (1) is

$$\mathcal{H}_{int}=\sum_{i=1}^{N}\sum_{j=1}^{N_w}\Delta_{ij}\,\sigma_i[J\,\chi_j+K(1-\chi_j)],\qquad (3)$$

where $\Delta_{ij}=1$ if i and j are nearest-neighbors and $\Delta_{ij}=0$ otherwise, $K>0$ (respectively $J\leq0$) is the energy cost (respectively gain) of a contact between a H monomer and a water site in an ordinary state (respectively special * state). The form of Eq. (3) can be understood if we consider a ''droplet'' of hydrophobic monomers in a water-filled lattice, and analyze the energy balance in exchanging a water site with a monomer (both taken from the respective bulks). If the water is in an ordinary state, each contact with a monomer will break up to $\sim m^{(d-1/d)}$ hydrogen bonds between neighboring water cells. If $a>0$ is the energy involved in a bond (in units of the spacing between the levels), this process yields an energy cost of $K\simeq m^{(d-1/d)}a$, while the cost will be zero if the water is in the special * state. At the same time, nonpolar residues interact more favorably with water than with each other, due to the permanent dipole moment of water [5]. This is taken into account by assuming an energy gain $J(-a\leq J\leq0)$ for a contact between water in the * state and a hydrophobic monomer. It should be clear that this theoretical framework is not meant to provide a detailed microscopic description of the physics of water, but rather to account for the basic ingredients of nonpolar solvation: the existence of an icelike ground state of zero entropy for water, and of a set of excited states, some of which may be particularly suited for interaction with nonpolar solutes, but involve a substantial entropy loss.

Let us now study the equilibrium thermodynamical properties of the model. To compute the partition function it is useful to introduce some notations. The maximum number of contacts between water and H monomers is obtained when the polymer is in an extended configuration, so that $M=(z-2)N_H+\sigma_1+\sigma_N$, where N_H is the number of non-polar residues. Given a configuration C of the polymer, the energy of the system can be written as

$$\mathcal{H}(C)=\sum_{j=1}^{N_w}[(1-\chi_j)(\vartheta_j+K\ell_j)+\chi_j(E^*+J\ell_j)],\quad (4)$$

where $\ell_j\equiv\ell_j(C)$ is the number of contacts between the jth water cluster and H monomers. Denoting by $n_\ell(C)$ the number of water clusters that have ℓ contacts with H monomers, the partition function of the model can be written as

$$\mathcal{Z}=\sum_C\mathcal{Z}(C)=\sum_{\{n_\ell\}}\zeta(\{n_\ell(C)\};N)\,Z(\{n_\ell(C)\}),\quad (5a)$$

$$Z(\{n_\ell(C)\})=\sum_{\{\vartheta\}}e^{-\beta\mathcal{H}(C)},\qquad (5b)$$

where $\beta=1/T$ and $\zeta(\{n_\ell(C)\};N)$ is the number of equivalent conformations for a polymer of length N, i.e., of the conformations with the same set of n_ℓ's. Now, to ease the notation let us put $Z(C)=Z(\{n_\ell(C)\})$. Using Eq. (4), and the fact that water cells do not interact with each other, so that it is possible to factor $Z(C)$ according to the labels $j=1,\ldots,N_w$, $Z(C)$ can be written as

$$Z(C)=P_0^{N_w}\prod_{\ell=1}^{z}x_\ell^{n_\ell(C)},\qquad (6a)$$

where $x_\ell=P_\ell/P_0$, and

$$P_\ell=g^*e^{-\beta(E^*+J\ell)}+e^{-\beta K\ell}\frac{1}{(1-e^{-\beta})^m}.\qquad (6b)$$

The mean energy and the specific heat are evaluated in the standard way as $U=\langle\mathcal{H}\rangle=-\partial\log\mathcal{Z}/\partial\beta$ and $C_V=\partial U/\partial T$. The average number of water sites presenting ℓ contacts with the polymer can be computed as $\langle n_\ell\rangle=x_\ell(\partial\log\mathcal{Z}/\partial x_\ell)$, from which one can obtain the average (total) number of H-water contacts, $\langle n_c\rangle=\Sigma_\ell\ell\langle n_\ell\rangle$, which is a measure of the compactness of the polymer. Another interesting observable is the average number of contacts with water sites in the * state, given by $\langle n_c^*\rangle=-T\partial\log\mathcal{Z}/\partial J$.

Notice that, even in the simple hydrophobic homopolymer case, our model cannot be mapped onto a polymer model with monomer-monomer contact interactions, because the partition function (5) cannot be written in the form

$$\mathcal{Z}=\sum_{n_{HH}}\xi(n_{HH})\,e^{-\beta hn_{HH}},\qquad (7)$$

where $n_{HH}=(M-\Sigma\ell n_\ell)/2$ is the number of monomer-monomer contacts and $\xi(n_{HH})$ is the number of polymer configurations with n_{HH} internal contacts, with h a true coupling constant, independent of T and $\{n_\ell\}$.

Before going further with the discussion of the thermodynamical properties of our model, let us note that the partition function of the DLRC model [3] can be obtained from the present one with the substitutions $g^*\to1$, $E^*\to0$, $(1-e^{-\beta})^{-m}\to q-1$, due to the fact that in the DLRC model all the states are equivalent (with zero energy) for pure water, and when a water-monomer interaction takes place, the special state has zero entropy, while the degeneracy of the excited states is independent of the temperature. In both

models one has to perform a sum over the conformations of the polymer to calculate the thermodynamical functions. This is a hard numerical task when the length of the polymer grows: for this reason we introduce an approximation that allows us to evaluate them analytically. First, we test our approximation against the exact numerical results obtained by DLRC in the case of a homopolymer (with $N \leq 25$) on a two-dimensional (2D) Manhattan lattice; then we apply it to our model.

From now on we specialize to the case of a nonpolar homopolymer of length N on a 2D square lattice ($z=4$). Let us now come to the main point of our approach, i.e., let us introduce an approximation for the unknown quantity $\zeta(\{n_l(C)\};N)$ in Eq. (5), representing the number of polymer conformations characterized by the same set $\{n_\angle\}$. We observe that this number must be strongly dependent on n_c, the total number of water-monomer contacts, which in turn is related to the number of monomer-monomer n_{HH} contacts by $n_c = M - 2n_{HH}$, and hence is a rough measure of how compact the polymer is. Then, our approximation is built in the following three steps: (i) We assume that $\zeta(\{n_l(C)\};N)$ depends essentially on n_c and not on each of the n_l. More precisely, we assume that, at fixed number of contacts n_c, the number of conformations in which there are water sites with three or more contacts with the polymer is negligible (single and double contacts are much more likely to appear, due to geometrical reasons), so that $\zeta(\{n_\angle(C)\};N)$ does not depend on n_3 and n_4. Moreover, we assume also that ζ is the same for each n_1, n_2 such that $n_1 + 2n_2 = n_c$, i.e.,

$$\zeta(\{n_\angle(C)\};N) = \zeta(n_c;N) \equiv C_N(n_{HH}), \qquad (8)$$

where $C_N(n_{HH})$ is the number of walks of length N with $n_{HH} = (M - n_c)/2$ internal contacts. (ii) We consider $\zeta(n_c;N)$ as characterized by two regimes, a globular and an extended one, referred to as ζ_c and ζ_g, respectively, which come into play depending on the value of n_c,

$$\zeta(\{n_\angle(C)\}) = \begin{cases} \zeta_c(n_c;N) & \text{if } n_c > \tilde{n}_c \\ \zeta_g(n_c;N) & \text{if } n_c \leq \tilde{n}_c. \end{cases} \qquad (9)$$

This is justified by the fact that the polymer shows, at different temperatures, either a globular, compact phase or an extended one. The latter involves a high number of water-monomer contacts, n_c, while the former is characterized by small values of n_c, namely, $n_c \propto N^{d-1/d}$, as in the case of a maximally compact state. We assume that a step function in the number of contacts separates these regimes; the position of the step, $\tilde{n}_c$, is at present unknown. (iii) We consider the extended conformations as self-avoiding walks (SAWs) and the compact ones as Hamiltonian walks (HWs): their numbers are both exponential in N, and are related by

$$\zeta_c(N) \simeq \exp(\alpha N) \zeta_g(N), \qquad (10a)$$

where α is given by

$$\alpha = \log \mu_{SAW} - \log \mu_{HW} \qquad (10b)$$

and μ_{SAW}, μ_{HW} are the respective connectivity constants [6].

Resorting to the above assumptions we can evaluate the partition function. In the homopolymer case n_c is always even, so that we set $k = n_c/2$, $k^{min} = n_c^{min}/2$, where $n_c^{min} \simeq z\sqrt{N}$ is the minimum number of contacts, $\tilde{k} = \tilde{n}_c/2$ and sum over k, obtaining

$$\mathcal{Z} = P_0^{N_w} \zeta_g(N)[\psi(x_1, x_2; k^{min}, \tilde{k}) + e^{\alpha N} \psi(x_1, x_2; \tilde{k} + 1, M/2)], \qquad (11a)$$

where ψ is given by

$$\begin{aligned} \psi(x,y;p,q) = &\frac{x^{2p}}{(x^2-1)(x^2-y)(y-1)} \\ &\times \{x^2 + x^{2(q-p+2)}(y-1) \\ &- y[1+(x^2-1)y^{q-p+1}]\}. \end{aligned} \qquad (11b)$$

The thermodynamic observables, U, C_V, $\langle n_\angle \rangle$, and $\langle n_c^* \rangle$ are then computed according to their definitions.

The approximation we introduced contains the threshold parameter, $\tilde{n}_c$, which is not a free parameter, but is unknown; its evaluation would require the knowledge of the number of walks of arbitrary length and number of internal contacts on a lattice. In order to circumvent this difficulty, we fix $\tilde{n}_c$ applying our approximation to the DLRC model, for which we know the results from exact enumeration on a 2D Manhattan lattice [3]. Using our method, the best approximation of the DLRC model specific heat is obtained using $\tilde{n}_c = n_c^{min} 23/20$; we choose this value for $\tilde{n}_c$. For the parameter α, Eq. (10b) holds, where, for a Manhattan lattice, $\mu_{SAW} = 1.7335$ [7] and $\mu_{HW} \approx \exp(G/\pi) = 1.3385$ (G is the Catalan constant) [8].

Let us now move to our model. Since the discretization of the water energy levels is artificial (and we set it as our arbitrary energy unit), the ratio a of the hydrogen bond energy to the level spacing is a free parameter. We choose $a = 100$; yet, we verified that even remarkable changes in a do not imply major modifications of the behavior of the thermodynamic quantities. The values of the other parameters have been fixed according to their physical meaning, i.e., we have chosen $m = 20$, $E^* = 2a$, $K = a\sqrt{m}$. Given the length N of the
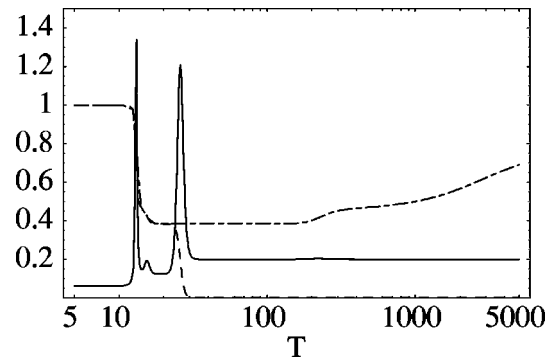

FIG. 1. Specific heat $c = C_V/(mM)$ (solid line, rescaled by a factor of 0.2 for graphical reasons), average number of water-monomer contacts $\langle n_c/M \rangle$ (dash-dotted line), and average number of contacts with water in the * state $\langle n_c^*/M \rangle$ (dashed line) as a function of T. Here $N=25$ and $J=-E^*/2$.
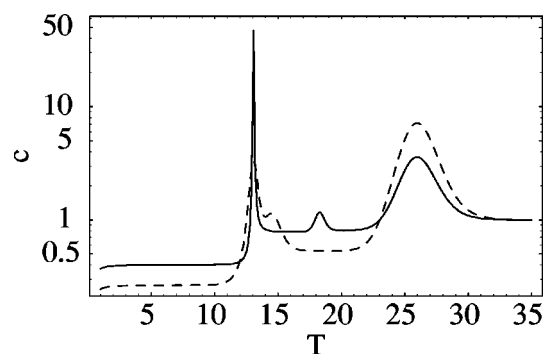
FIG. 2. $c = C_V/(mM)$ for $N = 16$ (dashed line) and $N = 100$ (solid line). $J = -E^*/2$. Here and in Fig. 1 T is in (arbitrary) energy units, so that c is dimensionless.

polymer, we set $N_w = M = 2N + 2$ (which is the maximum number of sites which can wet the chain). If $J \lesssim -E^*/2$, at low T the polymer is extended and completely wet by the solvent, as should be a neighbor-avoiding walk. Then, raising the temperature above a certain $T = T_f \simeq 13$, the polymer becomes compact ($\langle n_c \rangle$ drops to that of a maximally compact conformation), and finally, as T grows further, it swells again smoothly (see Fig. 1). The specific heat has a peak at T_f and another one at a higher temperature, when the polymer is compact. The latter is related to the excitation of the water sites around the polymer in compact conformation out of the * state, as witnessed by the drop of $\langle n_c^* \rangle$ to zero in correspondence of the peak; the former peak at T_f, whose height grows with the length N (see Fig. 2), could be related to a true phase transition, reminding cold unfolding in proteins. The thermal swelling of the polymer is present here, but does not have the characteristic of a phase transition, as it happens, for instance, in [9]. Yet this is not a surprise, due to the absence of any kind of imposed cooperativity in our model. It is interesting to notice that, at low temperatures, the polymer is expanded even for moderate values of J, i.e., $J \approx -E^*/2$: this means that, in the presence of a monomer, water prefers to stay in the excited special state, even when the energetic gain J does not apparently compensate the energy loss E^*. For values of J closer to zero the cold collapse is lost: at low temperatures the polymer is compact, then it swells smoothly at high temperatures.

To summarize, we have defined and discussed a lattice model for water-polymer interaction, where water is explicitly considered, and introduced an approximation scheme that allowed us to analytically evaluate the relevant thermodynamical averages. We remark that our model cannot be mapped onto a model with monomer-monomer contact interactions alone. We addressed the homopolymer case and observed that, for reasonable values of the parameters, the polymer is in a compact conformation at intermediate temperatures, and swells when the temperature is lowered as well as when it is raised. This recalls similar protein behavior. While the ''cold unfolding'' process is sharp, and possibly represents a phase transition as $N \to \infty$, thermal swelling is smooth. The absence of a sharp transition at high temperature could possibly be attributed to the lack of specificity in interactions: we expect that the heteropolymer case (to be studied next) will show more similarities to real proteins. The fact that our results are indeed qualitatively similar to those of DLRC [3], in spite of the differences in the model studied (many states versus two, temperature-dependent degeneracy, etc.), suggests that the polymer behavior we both find out is not a peculiarity of a particular model, but a property of a class of them: it is likely that the crucial thing is to take into account explicitly the degrees of freedom of the solvent, though in a simplified way. Indeed it appears (see, e.g., Ref. [10], where a model of random heteropolymers is introduced and studied) that if the solvent's degrees are neglected from the beginning, and hydrophobicity is attributed to monomers like a ''charge,'' cold unfolding will not be found for non-polar homopolymers, and probably even for quenched random heteropolymers.

[1] T. E. Creighton, *Proteins, Structure and Molecular Properties* (Freeman, New York, 1993); K. A. T. Silverstein, A. D. J. Haymet, and K. A. Dill, J. Am. Chem. Soc. **120**, 3166 (1998), and references therein.

[2] T. Garel, H. Orland, and D. Thirumalai, in *New Developments in Theoretical Studies of Proteins*, edited by R. Elber (World Scientific, Singapore, 1996).

[3] P. De Los Rios and G. Caldarelli, e-print cond-mat/9903394.

[4] See, e.g., P. L. Privalov, CRC Crit. Rev. Biochem. Mol. Bio. **25**, 181 (1990).

[5] G. E. Schulz and R. H. Schirmer, *Principles of Protein Structure* (Springer, Heidelberg, 1979).

[6] Equation (10) follows from the expressions for the number of SAWs and HWs on a lattice: see, e.g., C. Vanderzande, *Lattice Models of Polymers* (Cambridge University Press, Cambridge, England, 1998).

[7] D. Bennett-Wood, J. L. Cardy, I. G. Enting, A. J. Guttmann, and A. L. Owczarek, Nucl. Phys. B **528**, 533 (1998).

[8] B. Duplantier and F. David, J. Stat. Phys. **51**, 327 (1988).

[9] A. Hansen, M. H. Jensen, K. Sneppen, and G. Zocchi, Physica A **250**, 355 (1998); Eur. Phys. J. B. **6**, 157 (1998); **10**, 193 (1999).

[10] A. Trovato, J. van Mourik, and A. Maritan, Eur. Phys. J. B **6**, 63 (1998).