Polymer Collapse, Protein Folding, and the Percolation Threshold

HAGAI MEIROVITCH
University of Pittsburgh, School of Medicine, Center for Computational Biology and Bioinformatics (CCBB), Suite 601 Kaufmann Building, 3471 Fifth Avenue, Pittsburgh, Pennsylvania 15213

Received 21 March 2001; Accepted 22 May 2001

Abstract: We study the transition of polymers in the dilute regime from a swollen shape at high temperatures to their low-temperature structures. The polymers are modeled by a single self-avoiding walk (SAW) on a lattice for which l of the monomers (the H monomers) are self-attracting, i.e., if two nonbonded H monomers become nearest neighbors on the lattice they gain energy of interaction ($\epsilon = -|\epsilon|$); the second type of monomers, denoted P, are neutral. This HP model was suggested by Lau and Dill (Macromolecules 1989, 22, 3986–3997) to study protein folding, where H and P are the hydrophobic and polar amino acid residues, respectively. The model is simulated on the square and simple cubic (SC) lattices using the scanning method. We show that the ground state and the sharpness of the transition depend on the lattice, the fraction g of the H monomers, as well as on their arrangement along the chain. In particular, if the H monomers are distributed at random and g is larger than the site percolation threshold of the lattice, a collapsed transition is very likely to occur. This conclusion, drawn for the lattice models, is also applicable to proteins where an *effective* lattice with coordination number between that of the SC lattice and the body centered cubic lattice is defined. Thus, the average fraction of hydrophobic amino acid residues in globular proteins is found to be close to the percolation threshold of the effective lattice.

 J Comput Chem 23: 166–171, 2002

Key words: percolation threshold; polymer chains; computer simulation; collapse transition; protein folding

Introduction

The behavior of dilute polymer systems under various solvent conditions has been the subject of an extensive research for many years[1–3] (see also discussions in refs. 4 and 5, and references cited therein). Such systems have been commonly modeled by self-avoiding walks (SAWs) of N steps (bonds), i.e., $N+1$ monomers with self-attractions on a regular lattice; thus, two nonbonded monamers that are nearest neighbors on the lattice gain energy of interaction ϵ, ($\epsilon = -|\epsilon|$). With this model, a high absolute temperature T corresponds to a good solvent conditions, where the excluded volume interactions dominate the attractions and the chain is swollen. Thus, the radius of gyration R_g scales as $R_g \sim N^{\nu}$, where $\nu = 3/4 = 0.75$ in two dimensions[6] ($2d$) and $\nu \simeq 0.59$ in $3d$ (e.g., see ref. 7). However, as T is decreased (i.e., the solvent conditions worsen) the attractions become more effective, and at a temperature θ (the Flory θ temperature[1–3]) they cancel to a large extent the excluded volume repulsions, and in $3d$ the chain behaves in many respects like a Gaussian chain (e.g., $\nu_\theta = 1/2$); in $2d$, on the other hand,[8] $\nu = 4/7 = 0.571$. At $T < \theta$ the attractions prevail and the chain collapses, i.e., $\nu_c = 1/d$. The *degenerate* ground state is of maximal density and minimal surface to volume ratio (see Figs. 1a and 1b). The collapse transition at θ has been identified by de Gennes[3] as a tricritical point with an upper critical dimension three. It should be pointed out that at T slightly smaller than θ only very long chains will collapse, whereas a short chain will show a θ behavior, collapsing only at T, which is significantly smaller that θ. The above values of ν have been approximately obtained also in the experiment.[9–12]

One may consider another SAW model in which only l, out of the $N+1$ monomers are attracting, i.e., their fraction is $g = l/(N+1)$; it is of interest to study how the value of g and the particular arrangement of the attracting monomers along the chain affect the character of the transition and the ground state of the chain. Indeed, such models have been employed to describe the behavior of self-associating polymers in a good solvent, i.e., polymers that attracting groups are attached to a small fraction of their monomers (ionomers, for example). Joanny[13] and Gates and Witten[14] studied analytically SAWs with a small fraction g of attracting monomers in the θ and the strong excluded volume regimes, respectively. They derived expressions that describe the polymer shape in terms of g and other parameters and defined the conditions for which a col-

Correspondence to: H. Meirovitch; e-mail: hagaim@pitt.edu

Contract/grant sponsor: U.S. Department of Energy; contract/grant number: DE-FG05-95ER62070

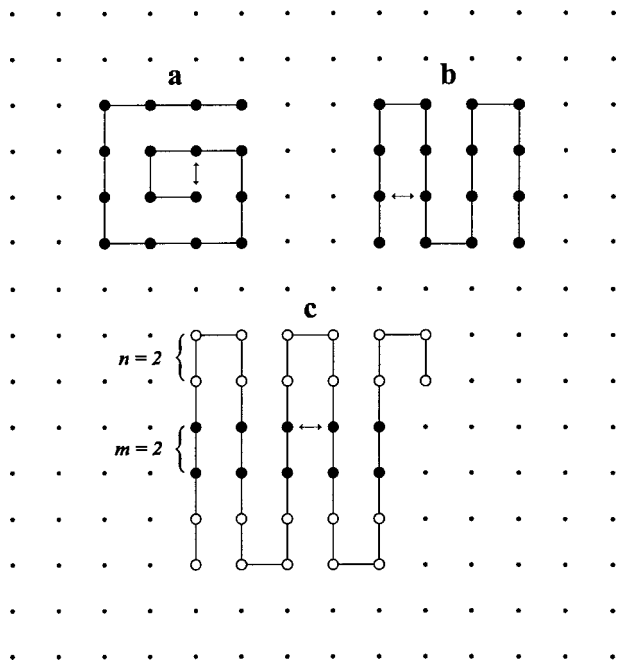

Figure 1. Ground states of SAWs consisting of attracting H monomers (full circle; their fraction is denoted by g) and neutral P monomers (empty circles). Nonbonded nearest-neighbor H monomers interact with negative energy $\epsilon = -|\epsilon|$ (e.g., see arrows). (a) and (b): Different ground states for $g = 1$ with energy 8ϵ. (c) The ground state for an arrangement of a repeated group of succesive $m = 2$ H monomers followed by $2n = 4$ P monomers.

lapse occurs; however, they did not consider the arrangement of the attraction monomers along the chain.

SAWs on a lattice with l attracting monomers have also been studied extensively in the context of protein folding, starting with the pioneering work of Gō and collaborators,[15] which was followed by others[16–19] (for reviews on more recent work see refs. 20–24). In particular, the present model is the HP model of Lau and Dill,[16, 25] where H denotes an attracting hydrophobic amino acid residue and P is a polar residue, which is considered to be neutral; for simplicity we shall use this notation as well.

In the present article we study the HP model on the square and the simple cubic (SC) lattices using the scanning simulation method.[26, 27] We discuss the effect of g and the arrangement of the H monomers along the chain on the ground state(s), the type of transition, and the transition temperature. In particular, we are interested in the case where the H monomers are distributed at random. We provide theoretical arguments supported by simulation results that a collapse transition will occur with high probability as long as $g > p_c$, where p_c is the site percolation threshold for the lattice studied. This conclusion is in accord with the fraction of hydrophobic amino acids found in proteins.

Methods

Simulation of self-attracting SAWs at low temperatures with the dynamical Metropolis method is inefficient because of the difficulty to induce global conformational changes. In this respect the scanning method is expected to perform better because the chains are constructed step by step with the help of transition probabilities that depend on scanning f future steps. However, the chain may get trapped in a dead end during construction; in this case it is discarded, and the construction of a new SAW is started. Therefore, the number of surviving chains is smaller than the number of chains attempted. Also, the generated chains are not distributed according to the Boltzmann probability; however, this bias can be removed by *importance sampling* or equivalently by selecting an effectively smaller set of *accepted* chains.[27] The larger is f the higher the efficiency, i.e., the larger is the number of the surviving chains as well as the accepted ones. The scanning method has the advantage that it provides the free energy as a by-product of the simulation. In this study we apply $f = 4$ for SAWs on the square and the SC lattices, which, however, requires scanning five steps ahead to take into account also the attracting interactions. Because the excluded volume interaction is much stronger and the attracting interaction is much weaker for a SAW on the square than on the SC lattice (maximum energy of 2ϵ vs. 4ϵ per H monomer, respectively), a simulation with $f = 4$ is much more efficient on the SC than the square lattice. Each simulation run is based on 3×10^5 attempted SAWs, where results based on less than 10^4 accepted SAWs are not considered.

The transition temperature for $g < 1$ is denoted by T_t to be distinguished from θ defined for $g = 1$. We also use the notations $K = -\epsilon/k_B T$ and K_t for the reciprocal temperature and its transition value, respectively, where k_B is the Boltzmann constant; the shape exponent of the collapsed state $(1/d)$ is denoted by ν_c. As discussed later, for $g < 1$ the ground state might be rod-like or layer-like in 2 and $3d$, respectively, and even when a collapse occurs it is not clear whether a θ-point scaling behavior is guaranteed. Also, even in the case of a θ-point transition, a highly accurate determination of T_t (for the infinite chain) from simulation data of short chains would require large samples and a relatively complex scaling analysis.[28] Therefore, for the relatively short chains studied, T_t is defined as the temperature at which the radius of gyration, $R_g(N)$ scales as $N^{1/d}$; this K_t is expected to overestimate significantly the corresponding K_θ temperature if it exists.

Results and Discussion

In Table 1 we present the analytical results[1, 2, 8] for ν_θ and the most accurate values of K_θ obtained for the square and the SC lattices by computer simulation;[4, 28, 29] as expected, K_θ decreases as d is increased. These values of K_θ should be distinguished from those for K_t in Table 2, obtained from the scaling of R_g as $N^{1/d}$. Thus, for the SC lattice $K_\theta \simeq 0.27$ in Table 1 is significantly smaller than the value $K_t = 0.45$ obtained in Table 2 for $g = 1$.

Ordered Arrangements of the H Monomers

Let us first examine a SAW on a square lattice where m successive attracting monomers H are followed by an even number $2n$ of nonattracting P monomers repeatedly ($P_1P_2 \ldots P_{2n}H_1H_2 \ldots H_m P_1P_2 \ldots P_{2n} \ldots$), i.e., $g = m/(m + 2n)$. This system has a unique ground state depicted in Figure 1c for $m = 2$ and $n = 2$, where

Table 1. Critical Exponents for $d = 2$ and 3, Transition Temperatures, and Site Percolation Thresholds for the Square and the SC Lattices.[a]

d	ν_θ	ν_c	$K_\theta = -\epsilon/k_B T_\theta$	p_c
2	4/7[b]	1/2	0.658(4)[c]	0.5927...
3	1/2	1/3	0.274(6)[d]	0.3116
3	1/2	1/3	0.2690(3)[e]	0.3116

[a] The percolation thresholds, p_c are taken from ref. 30.
[b] The exact results of Duplantier and Saleur.[8]
[c] From ref. 28.
[d] From ref. 4.
[e] From ref. 29. The statistical error is given in parenthesis, e.g., $0.658(4) = 0.658 \pm 0.004$.

every H monomer has the maximal number (2) of interactions, except for the $2m$ H monomers on the surface, which have only one interaction. Thus, at low temperature a long enough chain will always become rod-like (with a width of $m + 2n$) for any value of m and n, meaning that $\nu_c = 1$ rather than 1/2, the value for a collapsed chain. However, a chain of length $(m + 2n)^2$ has a collapsed ground state. It should also be pointed out that a rod-like structure will be obtained even for an extremely dilute concentration of the H monomers ($g \to 0$), where $m = 1$ and n is increased at will. However, as n is increased (for constant number l of H monomers) the entropy as well as the average energy increase (at a given temperature) meaning that the ground state becomes dominant only at a very low temperature. Also, such a rod-like structure will be difficult to generate by any computer simulation method because its stabilization is based on large loops of size $m + 2n$; in particular, an efficient simulation with the scanning method would require $f \sim m + 2n$, which allows "sensing" at each step the existence of the next H monomer. The other extreme case is the high concentration limit of H monomers, $g \to 1$, where $n = 2$ and m is large. Here, unlike the dilute limit, the energy dominates the entropy and

Table 2. The Reciprocal Transition Temperature K_t for Different Fractions g of Randomly Selected H Monomers.[a]

d	ν_c	g	$K_t = -\epsilon/k_B T_t$	$N_{\max}$
2	1/2	0.8	1.2	90
2	1/2	0.7	1.4	70
2	1/2	0.6	1.5	70
3	1/3	1.0	0.45	210
3	1/3	0.8	0.6	180
3	1/3	0.6	0.9	160
3	1/3	0.5	1.4	100
3	1/3	0.4	2.0	100
3	1/3	0.32	2.0	90

[a] The dimension $d = 2$ and 3 refer to the square and the simple cubic lattices, respectively. ν_c is the shape exponent for the collapsed chain, and $N_{\max}$ is the longest chain considered in the scaling analysis. For each g, different arrangements based on different random number sequences lead to different values of K_t; therefore, the results for K_t are representative values defined up to ± 0.3.

$f = 5$ or 6 would suffice for an efficient simulation with the scanning method. The ground state of these SAW models on a cubic lattice is a two-dimensional layer of width $m + 2n$, i.e., $\nu_c = 1/2$.

Simulations of SAWs of $N \leq 160$ ($m = 1$ and $n = 1$, i.e., $g = 1/3$) on the square lattice have shown that at high T (low K) the chain, as expected, is swollen, i.e., $\nu \simeq 0.75$. However, when K was increased, no sign of collapse was observed, i.e., ν was not decreased but rather slightly increased. Thus, at the relatively cold temperature, $K = 3$ we found $\nu \simeq 0.81$, where at higher K the scanning method (with $f = 4$) becomes inefficient; this means that a rod-like structure was not obtained either. We, therefore, checked the simulated configurations by computer graphics and found that they consisted basically of small "rods" (of the type depicted in Fig. 1c) linked together by flexible segments, where the length of these rods increases with increasing K. This suggests that a transition occurs between a swollen and a rod-like structure but the transition is not sharp due to strong entropic effects, which is expected for an one-dimensional model with short-range interactions.[31] We simulated this model ($N \leq 300$) also on a simple cubic lattice, where again a collapsed structure was not observed: As K was increased the corresponding values of ν were decreased from ~0.59 for small K to ~0.5 for $K = 1.9$, 2, and 2.1 (at higher K the simulation becomes inefficient). Thus, these results are in accord with a ground state that is a $2d$ layer, as discussed earlier.

We also simulated on the square lattice the model $m = 3$ and $2n = 2$ ($g = 0.6$) at various temperatures and again obtained that even at the lowest temperature studied ($K = 1.6$), $\nu \simeq 0.78$, i.e., a rod-like structure was not obtained as yet. Indeed, from free energy considerations the ground state is found to be unstable at $K = 1.6$ due to entropic effects: We obtain that the free energy F for $N = 100$ is $F/(Nk_BT) = -1.207$, which is significantly lower than $F_{gs}/(Nk_BT) = -0.944$, the free energy of the ground state alone at the same K (i.e., the entropy is zero and $F_{gs}/(NkT) = KE_{gs}/N$, where E_{gs} is the ground state energy). In other words, the contribution of the entropy (~0.757) to the free energy at $K = 1.6$ is significant. The ground state of the cubic lattice is a two-dimensional layer of width of five monomers. However, the structure of short chains ($\leq$125) is expected to be compact. Indeed, at $K = 1$ a collapse has been detected, that is, $\nu_c = 1/3$ rather than 1/2. As in $2d$, a transition to the ground state is also expected here at a higher K.

One may argue that the structural properties discussed thus far are not realistic because they reflect the geometrical restrictions imposed by the square and the SC lattices (in which, e.g., two H monomers separated by an odd number n of P monomers along the chain cannot become nearest neighbors; thus, if $n = 1$, $g = 0.5$ with zero energy). However, the structure of Figure 1c (for $m = 1$, $n \geq 1$) is reminiscent of the α-helical rod-like structure occurring frequently in polypeptides, which is mainly stabilized by short-range hydrogen bonds. In fact, the helix-coil transition is not sharp, and it can be described within the framework of the $1d$ Ising model.[31] Also, the loop created, for example, for $n = 1$ and $m > 1$ is similar to the a β-sheet structure prevalent in proteins, which is based on short- and medium-range hydrogen bonds. Thus, even though the ground states of our models on the cubic lattice are layers rather than rods or sheets, these models reflect the experimental reality that the arrangement of the attracting monomers along the

chain, their density g, and the type of interaction determine the ground state that is not necessarily a collapsed structure.

Random Distribution of the H Monomers

In the examples studied thus far the attracting monomers were arranged in a well-defined order along the chain; we now discuss the case where they are distributed at random, i.e., a H (P) monomer is determined with a probability g $(1 - g)$ by a random number. The probability of an arrangement with l monomers of type H is $g^l(1-g)^{(N+1-l)}$, and the probability to obtain l of them is defined by the binomial distribution where the average is $(N+1)g$ and the standard deviation $\sigma = [(N+1)g(1-g)]^{1/2}$. Thus, the *typical* number of H monomers obtained in a random selection will pertain to the range $(N+1)g \pm \sigma$ and the chance that the corresponding arrangement is an ordered one is low. For large enough g the ground state(s) is expected to be a collapsed (compact) structure, and the question is whether a critical value g_c exists, where for $g > g_c$ a collapse is guaranteed (in the probabilistic sense discussed above). In what follows we argue that $g_c \geq p_c$, where p_c is the site percolation threshold of the lattice studied. It should be pointed out that in a typical percolation experiment each site of a large empty lattice is visited and a monomer is placed there with probability p. p_c is the lowest probability at which percolation occurs, i.e., a cluster of monomers connecting the opposite sides of the lattice is generated.[30]

Assume that a SAW on a square lattice is arranged in a perfectly compact configuration (e.g., the configurations depicted in Figs. 1a and 1b); thus, if the H monomers are distributed at random, the process becomes a percolation experiment, where for $g \geq p_c$ ($p_c \simeq 0.590$ for the square lattice[30]) a percolating cluster defined by the H monomers will be created. This cluster will connect the opposite sides of the square structure, which most probably will be "held" together by the attracting H monomers. The energy of this cluster is low, because on average a nonsurface H monomer will have close to $2d_{\text{square}} - 2$ nearest-neighbor nonbonded H monomers, where $d_{\text{square}} = 1.896$ is the fractal dimension of the percolation cluster on the square lattice. Therefore, it is plausible to assume that for $g > p_c$ randomly distributed H monomers will lead to a collapsed degenerate ground state (notice that the energy of the cluster will be different for different compact structures). This argument can be extended to the simple cubic lattice where a H monomer of the percolation cluster will have on average close to $2d_{\text{SC}} - 2$ nearest-neighbor nonbonded H monomers; $d_{\text{SC}} = 2.5$ is the fractal dimension of a percolation cluster on the SC lattice. Because a collapsed structure is stabilized by short-, medium-, as well as long-range interactions, the corresponding transition is expected to be sharp[20] (notice that throughout this article the term "sharp transition" means a transition that is significantly sharper than a transition depending on short-range interactions, such as the helix-coil transition discussed earlier).

Again, the probabilistic nature of this statement should be emphasized, meaning that specific distributions of H monomers with $g < p_c$ can still lead to compact ground states, and as discussed earlier, for $g > p_c$, noncompact ground states are possible. Also, compact structures that are not "perfect" (i.e., squares with the maximal density in $2d$, as those depicted in Figure 1a and b) can also lead to $\nu_c = 1/2$. Therefore, the percolation threshold should be considered only as an approximate guiding value for collapse; notice also that at $g = p_c$ the P monomers on the square lattice will not percolate $(1 - g = 0.4 < p_c)$, but they create a percolating cluster on the SC lattice $(1 - p_c > p_c)$.

These theoretical considerations can be checked (even though not proven) by computer simulation. We carried out a large number of simulations as specified in the Methods section on the square and SC lattices. For each value of g the arrangement of the H monomers was determined at random, and simulations were performed at different temperatures to determine K_t at which $\nu_c \simeq 1/2$ and 1/3 for these lattices, respectively. Also, for each g, several arrangements of the H monomers were tested based on different random number sequences. Table 2 demonstrates the expected increase in K_t (i.e., the decrease in the temperature) as g is decreased (because stronger "glue" between the interacting H monomers is needed to hold the compact structure together, as their number is reduced and entropic effects strengthen). It should be pointed out, however, that each $K_t(g)$ shown should be considered only as a representative value obtained for a certain arrangement of the H monomers. We have found that other random arrangements for a given g can lead to K_t results that deviate by ± 0.3 from the presented value; this stems from both, the fluctuations in the number of monomers l discussed earlier, and changes in the specific ordering of the H monomers along the chain, an effect that increases with the decrease of g. Such fluctuations are demonstrated in Figure 2, where results for $\log(R_g^2/N)$ vs. $\log(N)$ obtained for $g = 0.4$ for the SC lattice are plotted for different sequences of the H monomers together with the corresponding best-fit lines. It is evident that the chain length leading to efficient simulations (measured by the number of ac-

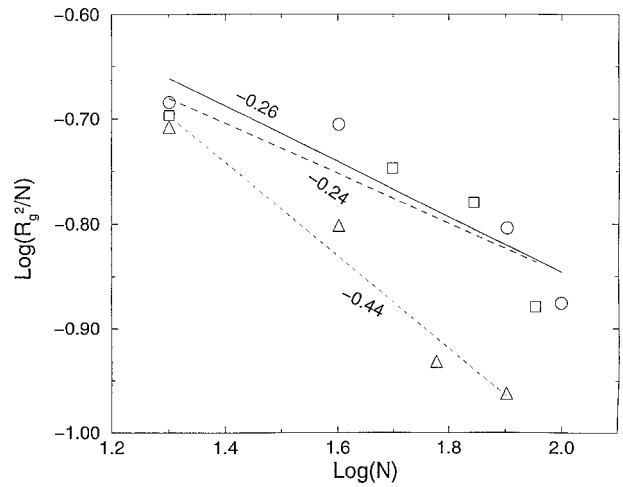

Figure 2. Log–log plots of the average radius of gyration, R_g^2/N vs. the chain length N for three SAWs with different arrangements of the H monomers on the SC lattice at the reciprocal temperature $K_t = -\epsilon/K_BT = 2.0$. The H monomers of each chain were selected at random with probability $g = 0.4$ but with different random number sequences. The best-fit lines denoted by full, dotted, and dashed lines correspond to the circles, squares, and triangles, respectively; the slopes of these lines are also provided where the collapsed value is $-\nu_c = -1/3$.

cepted chains), while relatively short, is sufficient to demonstrate the existence of a collapse.

The most interesting result is that collapse transitions were detected very close to the percolation thresholds for $g = 0.6$ and 0.32 for the square and SC lattices, respectively, which supports the main claim of this article. It should be emphasized, however, that for the corresponding lattices and g values, out of four H arrangements only 3/4 and 2/4 arrangements have led to collapse; on the other hand, collapse was obtained for specific arrangements based on $g = 0.5$ and 0.24, respectively.

When g is much smaller than p_c a collapsed ground state is not expected. The most likely scenario is that several successive H monomers along the chain that are separated by an even number of monomers create local "blobs" separated (on the square lattice) by flexible linear segments, and therefore, for large N, $\nu = 3/4$. Indeed, our simulations for $g = 1/3$ (random distribution) on the square lattice led to $\nu \sim 0.7$ at both $K = 2.5$ and 2.8; however, the corresponding values of G^2/N for $N = 100$, 0.77, and 0.70, respectively, are significantly smaller than 1.09, obtained for SAW without attractions at $K = 0$ (where ν is 0.75), which is because of the effectively shorter self-attracting chain due to the blobs.

A continuum chain is expected to behave in many respect similar to the lattice models discussed above; however, for a continuum chain the flexibility is not defined by the lattice but by geometrical restrictions induced by the intrachain interactions. Thus, a large persistence length will prevent the formation of short- and medium-range contacts between the attracting monomers but will have little effect on the creation of long-range loops; in this respect the lattice is more restrictive, where H monomers separated by an odd number of monomers along the chain can never become nearest neighbors. When the attracting monomers appear in a specific order along the chain, ordered structures can be formed, such as α-helices and β-sheets, as has already been discussed. On the other hand, if g is small and the attracting monomers are distributed at random, it is difficult to predict the ground state(s), which, however, is expected to become compact for a large enough g.

Because a continuum chain does not occupy a definite lattice, identification of g with a lattice percolation threshold is not straightforward. However, for globular proteins a typical coordination number for an internal residue is[32] 7–8, i.e., between the coordination numbers of the SC lattice (6) and the body-centered cubic (BCC) lattice (8). Therefore, it is plausible to assume that the percolation threshold for an *effective* protein lattice is between the corresponding p_c values, 0.3116 and 0.246, probably closer to the latter. In refs. 33 and 34, based on various criteria, we have classified as hydrophilic the nine residues, Asp, Glu, Gln, Asn, Lys, Pro, Arg, Ser, and Thr, as neutral the amino acid His and as ambivalent, Ala, Gly, and Tyr. The seven remaining amino acids, Cys, Phe, He, Leu, Met, Val, and Trp are the hydrophobic ones, which constitute $7/20 = 0.35$ of the 20 occurring amino acids. However, their fraction in the 19 smaller and larger proteins studied was found to be $\simeq$0.29, slightly above the effective percolation threshold and thus in accord with our picture.

It should be pointed out, however, that the distribution of hydrophobic residues along a protein chain is not random but very specific, as required for the stabilization of the native structure (however, no clear correlation has been found between arrangements of hydrophobic residues in proteins from different families; in this respect the distribution of the hydrophobic residues can be considered as random).

The hydrophobic residues avoid the contact with the surrounding water by concentrating in the interior of the protein structure. If their fraction was much smaller than p_c they would not be able to percolate through the compact structure (assuming a random distribution), but they would cluster in smaller groups "wrapped" by hydrophilic, neutral, or ambivalent residues. This, however, would not be the most stable structure because the same degree of "coverage" (from water) of the hydrophobic residues also occurs in linear structures consisting of blobs (see earlier discussion). In these structures part of the interactions between the polar residues are replaced by comparable polar–water interactions, while extra stability is gained from the increase in the chain entropy. Such local coverage, however, is unlikely to occur for a percolating cluster of hydrophobic residues, where the most stable structure is thus a compact one. Therefore, a fraction p_c of the hydrophobic residues is necessary to guarantee compactness.

Analysis of protein structures has shown[34] that the hydrophobic residues are not distributed homogeneously over the structure, but their concentration in spherical layers around the center of mass decreases significantly in going from the core towards the surface of the protein, whereas an opposite trend is observed for the hydrophilic residues. These changes in the distributions, however, stem partially from the fact that the structures of proteins are not spherical but have ramified surfaces, and even internal spherical layers might contain surface residues that preferably are hydrophilic. To optimize the electrostatic energy the hydrophilic residues in each layer are arranged in localized clusters, where the fraction of these residues is larger than that in the entire layer; this induces similar clustering of the hydrophobic residues as well.[35] Therefore, it is very likely that both types of residues percolate through the protein structure. In this context it should be pointed out that while the clustering of the H (P) monomers in the simplified HP model is induced by the HH attractions, the model still reflects the global features of the structural organization of proteins, and the present study, therefore, sheds new light on the hydrophobicity-driven mechanism of protein collapse in terms of the percolation theory.

Summary

We have studied the structural transition of polymers with attracting monomers in the dilute regime from a swollen shape at high temperatures to the ground state (which might be degenerate) at low temperatures. Extensive simulations of the HP model on the square and SC lattices were carried out with the scanning method, where the criterion for a structural change is the change in the shape exponent ν, describing the scaling of the radius of gyration with the chain length. As expected, the ground state depends on the lattice, the fraction of the H monomers, and their specific arrangement along the chain, which can lead to rod-like ($\nu = 1$) and $2d$ layer-like ($\nu = 1/2$) ground states on the square and the SC cubic lattices, respectively. Because these ground states are "held" by short- and medium-range interactions the corresponding transitions are not expected to be sharp but to occur through a range of temperatures. Whereas these ground

states are lattice dependent, they reflect the experimental reality that ordered structures, such as α helices and β sheets appear frequently in proteins and polypeptides due to orderly placed donor and acceptor groups of hydrogen bonds along the chain. When the H monomers are distributed at random and their fraction g is larger than the site percolation threshold of the lattice, a collapsed ground state (which can be degenerate) with a sharp transition is expected. This conclusion, drawn for lattice models, also applies to globular proteins where the residues can approximately be described as occupying an effective lattice with coordination number between the coordination numbers of the SC and the BCC lattices; hence, with an intermediate percolation threshold. This threshold, indeed, is very close the average fraction of hydrophobic residues in proteins. Thus, the percolation theory applied to the HP model sheds new light on the hydrophobicity-driven protein collapse.

Acknowledgments

I thank Professor I. Rabin for valuable discussions, and Dr. B. Das for her help.

References

1. Flory, P. J. Principles of Polymer Chemistry; Cornell University Press: Ithaca, NY, 1953.
2. Flory, P. J. Statistical Mechanics of Chain Molecules; Hanser Publishers: New York, 1988.
3. de Gennes, P. G. Scaling Concepts in Polymer Physics; Cornell University Press: Ithaca, NY, 1985.
4. Meirovitch, H.; Lim, H. A. J Chem Phys 1990, 92, 5144.
5. Chang, I.; Meirovitch, H. Phys Rev E 1993, 48, 3656.
6. Nienhuis, B. Phys Rev Lett 1982, 49, 1062.
7. Sokal, A. D. In Monte Carlo and Molecular Dynamics Simulations in Polymer Science; Binder, K., Ed.; Oxford University Press: New York, 1995, p. 47.
8. Duplantier, B.; Saleur, H. Phys Rev Lett 1987, 59, 539.
9. Mazur, J.; McIntyre, D. Macromolecules 1975, 8, 464.
10. Nierlich, M.; Cotton, J. P.; Farnoux, B. J Chem Phys 1978, 69, 1379.
11. Swislow, G.; Sun, S.-T.; Nishio, I.; Tanaka, T. Phys Rev Lett 1980, 44, 796.
12. Perzynski, R.; Adam, M.; Delsanti, M. J Phys (Paris) 1982, 43, 129.
13. Joanny, J. F. Polymer 1980, 21, 71.
14. Gates, M. E.; Witten, T. A. Macromolecules 1986, 19, 732.
15. Taketomi, H.; Ueda, Y.; Gō, N. Int J Pept Protein Res 1975, 7, 449.
16. Dill, K. A. Biochemistry 1985, 24, 1501.
17. Kolinski, A.; Skolnick, J.; Yaris, R. Proc Natl Acad Sci USA 1986, 83, 7267.
18. Covell, D. G.; Jernigan, R. L. Biochemistry 1990, 29, 3287.
19. Hinds, D. A.; Levitt, M. Proc Natl Acad Sci USA 1992, 89, 2536.
20. Dill, K. A.; Bromberg, S.; Yue, K.; Fiebig, K. M.; Yee, D. P.; Thomas, P. D.; Chan, H. S. Protein Sci 1995, 4, 541.
21. Dill, K. A. Protein Sci 1999, 8, 1166.
22. Shakhnovich, E. I. Fold Des 1998, 3, R45.
23. Gutin, A. M.; Abkevich, V. I.; Shakhnovich, E. I. Fold Des 1998, 3, 183.
24. Berriz, G. F.; Shakhnovich, E. I. Curr Opin Colloid Interface Sci 1999, 4, 72.
25. Lau, K. F.; Dill, K. A. Macromolecules 1989, 22, 3986.
26. Meirovitch, H. J Phys A 1982, 15, L735.
27. Meirovitch, H. J Chem Phys 1988, 89, 2514.
28. Chang, I.; Meirovitch, H. Phys Rev E 1993, 48, 3656.
29. Grassberger, P. Phys Rev E 1997, 56, 3682.
30. Stauffer, D.; Aharony, A. Introduction to Percolation Theory; Taylor & Francis: London, 1992.
31. Poland, D.; Scheraga, H. A. Theory of Helix-Coil Transitions in Biopolymers; Academic Press: New York, 1970.
32. Raghunathan, G.; Jernigan, R. L. Protein Sci 1997, 6, 2072.
33. Meirovitch, H.; Rackovsky, S.; Scheraga, H. A. Macromolecules 1980, 13, 1398.
34. Meirovitch, H.; Scheraga, H. A. Macromolecules 1980, 13, 1406.
35. Meirovitch, H.; Scheraga, H. A. Macromolecules 1981, 14, 340.