Phase Diagram of Random Heteropolymers

A. Montanari,[1] M. Müller,[2] and M. Mézard[2]

[1]*Laboratoire de Physique Théorique de l'Ecole Normale Supérieure, Paris, France**
[2]*Laboratoire de Physique Théorique et Modèles Statistiques, Université Paris-Sud, bâtiment 100, F-91405 Orsay, France*
(Received 2 July 2003; published 6 May 2004)

We propose a new analytic approach to study the phase diagram of random heteropolymers, based on the cavity method. For copolymers we analyze the nature and phenomenology of the glass transition as a function of sequence correlations. Depending on these correlations, we find that two different scenarios for the glass transition can occur. We show that, beside the much studied possibility of an abrupt freezing transition at low temperature, the system can exhibit, upon cooling, a first transition to a soft glass phase with fully broken replica symmetry and a continuously growing degree of freezing as the temperature is lowered.

The statistical mechanics of random heteropolymers has attracted a lot of interest in the past two decades, mainly for its close relation to the long standing problem of protein folding [1], but also because of its applications to material sciences [2]. It has soon been recognized that the energy landscape and low temperature physics of heteropolymers bear a strong resemblance with that of spin glasses, and many theoretical concepts have been fruitfully adapted to that particular class of disordered systems [3,4].

A popular guiding idea in this field was the conjecture that typical low energy configurations of a random heteropolymer behave as independent energy levels [5] as it happens in the random energy model (REM) [6]. Mean-field replica calculations corroborate such a discontinuous glass transition with one-step of replica symmetry breaking, both for random bond interactions [7] and for copolymers [8]. However, the applicability of the simple REM-like picture to heteropolymers and proteins has been questioned [9]. The numerical study of self-consistent dynamical equations for short chains [10] hints at a more refined scenario which we shall compare with our own results below.

The polymer freezing manifests itself in persisting correlations between two times t_w and $t_w + t$, after the molecule has been prepared in a random conformation at time 0 [11]. The following family of correlation functions for a chain of length N proves very useful [12]: $C_l(t, t_w) = N^{-1}\sum_i \langle \Delta \mathbf{r}_{i,l}(t) \cdot \Delta \mathbf{r}_{i,l}(t_w) \rangle$. Here $\Delta \mathbf{r}_{i,l} = \mathbf{r}_{i+l} - \mathbf{r}_i$ and $\mathbf{r}_i$ is the spatial position of the ith monomer of the chain. Within the glassy phase, these correlation functions should exhibit a *plateau* at $q_l = \lim_{t \to \infty} \lim_{t_w \to \infty} C_l(t, t_w)$ (t and t_w are much smaller than the equilibration time τ_{eq}).

The REM-like scenario predicts an abrupt freezing in a metastable conformation, implying a discontinuous jump of the plateau height q_l from zero to a finite value at the glass transition. However, many other disordered systems, like spin glasses, exhibit a continuous glass transition (in the "overlap" parameter, q_l). This has recently been predicted for the random amphiphilic chain [12], as does our analysis for a certain class of correlated sequences. The folding properties of a given heteropolymer will be crucially affected by the type of glass transition it undergoes, since the dynamics (e.g., the aging behavior) are rather different in these two scenarios [11].

Apart from numerical simulations, most of our fundamental knowledge about heteropolymers relies on two main analytical methods: the replica approach, with some kind of Gaussian variational ansatz [7], and the use of dynamical equations, with some self-consistent closure approximation [10,13]. Considering that these methods use some rather severe approximations which are not really independent from each other, it is clearly useful to have a new analytic method emerging. We propose here such a new approach, based on the cavity method [14], which is an extension of the classical Bethe approximation to frustrated systems. It is the simplest mean-field approximation that handles exactly the short distance correlations [15], and in this respect it is very different from the usual replica approach. Moreover, as we will see below, the existence of a nontrivial statistical mechanics model on a "Bethe lattice" for which the method is exact is very helpful. We expect the cavity approximation to be faithful as long as crystalline, or more complex, spatially ordered structures are suppressed by the randomness of the chain sequence.

We show that for fixed monomer-monomer interactions, both types of glass transition can be found. Depending on the correlations in the chain sequence, one finds either a transition to a "frozen glass" phase, with a jump in q_l (we shall call $T_d = 1/\beta_d$ the corresponding transition temperature), or a continuous transition to a "soft glass" phase, where q_l develops continuously from zero (we denote by β_i the inverse critical temperature in this case); see Figs. 1 and 2. The
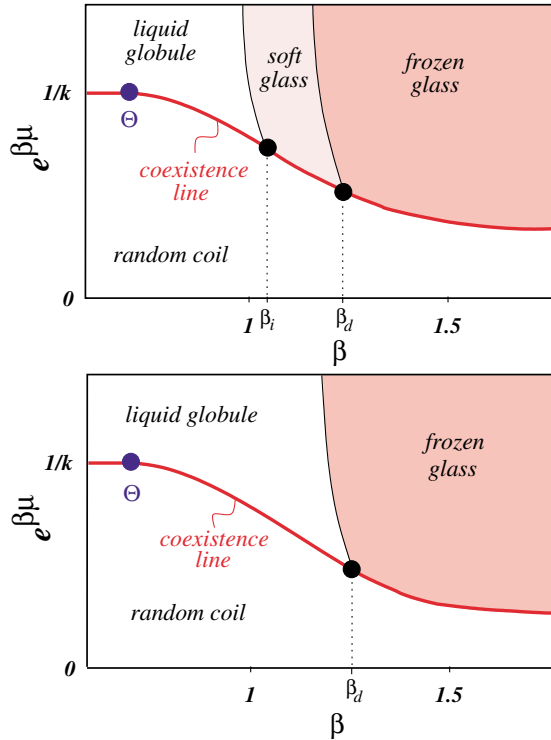
FIG. 1 (color online). The phase diagram in the grand canonical ensemble. An unconstrained polymer lives on the coexistence line [$\mu = \mu_c(T)$]. In correlated HP polymers and anticorrelated ampholytes (top) a continuous glass transition occurs at the liquid instability. It is followed by a discontinuous freezing transition. Oppositely correlated sequences (bottom) exhibit a direct freezing transition from the liquid globule.

predicted soft glass phase bears some resemblance to the intermediate glass phase found in the dynamic approaches [10].

Consider a lattice polymer whose conformation is given by a self-avoiding walk. Two monomers interact with an energy $e_{\sigma,\sigma'}$ if they occupy two neighboring sites i and j. Here $\sigma, \sigma' \in \{1, -1\} \leftrightarrow \{A, B\}$ denote the monomer types. We will restrict to the symmetric case $e_{AA} = e_{BB} = -e_{AB}$ and only consider neutral polymers (equal number of A's and B's). The case $e_{AA} = -1$ is close to the popular hydrophobic-polar (HP) model for proteins where hydrophobic and polar constituents tend to cluster, while $e_{AA} = 1$ corresponds to charged polymers (ampholytes) with short-range interactions.

On a lattice with V sites, we consider the polymer in the grand-canonical ensemble with a chemical potential μ per monomer. One can identify two phases depending on the scaling of the average length $\langle N \rangle$ of the polymer [15]. For small μ [$\mu < \mu_c(T)$] there exists an infinitely diluted phase where $\langle N \rangle / V \to 0$ as $V \to \infty$. For $\mu > \mu_c$, there is a dense phase with a finite $\langle N \rangle / V$. To describe a polymer in equilibrium with the solvent the chemical potential has to be adjusted to the critical value μ_c. The Θ point where the polymer collapses to the liquid globule occurs as a tricritical point on this line [16]; see Fig. 1.
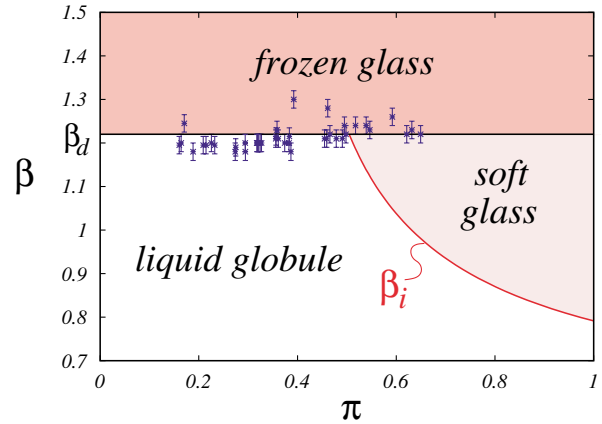
FIG. 2 (color online). The phase diagram as a function of sequence correlations π for HP polymers. At high temperature (small β) the polymer is found in a liquid globule phase. For correlated sequences ($\pi \gtrsim 0.5$), the molecule undergoes a continuous phase transition at inverse temperature β_i given by (5), followed by a discontinuous freezing at β_d. Anticorrelated sequences ($\pi \lesssim 0.5$) freeze discontinuously at β_d without intermediate phase. The inverse temperature of abrupt freezing, β_d, turns out to be nearly independent of the sequence, as indicated by the horizontal line at $\beta_d \approx 1.23$. The data points show the β_d's computed for a few particular sequences with $L = 20$.

Before explaining the cavity approach, we notice that its approximations become exact on a "Bethe lattice" whose appropriate definition for a frustrated system is a random graph with N vertices and fixed connectivity (taken equal to $k + 1 \equiv 6$ here to mimic cubic lattices) [14]. Since the size of typical loops in such a graph diverges as $O[\log(N)]$, the graph is locally treelike, allowing for iterative solutions of the statistical physics problem.

For technical reasons we work with a periodic sequence of monomers with period L and take the limit $L \to \infty$ in the end. Let us now consider an oriented edge of the graph, going from a "root" site R to another site S. It can be in one of the following states: (0): neither root nor edge are occupied; $(+, a)$ or $(-, a)$: the root is occupied by monomer $a \in [1, L]$ and the backbone continues on the edge, the sign $\pm$ indicating that S is occupied by monomer $a \pm 1$; (2a): the root is occupied by monomer a, but the edge is free. We now exploit the local treelike structure to recursively express the (Gibbs) probability $p_\alpha^{(0)}$ of finding the oriented edge "0" in the conformation $\alpha \in \{0, (+, a), (-, a), 2a\}$ in terms of the corresponding probabilities $p_{\alpha'}^{(i)}$ on the remaining edges ($i = 1, \ldots, k$) that are connected to the root of the edge "0":

$$p_0^{(0)} = C^{-1} \prod_{i=1}^{k} \left[p_0^{(i)} + \sum_{a'=1}^{L} p_{2a'}^{(i)} \right], \quad (1)$$

$$p_{\pm,a}^{(0)} = C^{-1} e^{\beta\mu} \sum_{i=1}^{k} p_{\pm,a\mp1}^{(i)} \prod_{j\neq i}^{k} \psi_a^{(j)}, \quad (2)$$

$$p_{2a}^{(0)} = C^{-1} e^{\beta\mu} \sum_{i_1 \neq i_2}^{k} p_{+,a-1}^{(i_1)} p_{-,a+1}^{(i_2)} \prod_{j \neq i_1, i_2}^{k} \psi_a^{(j)}, \qquad (3)$$

where C is a normalization and we have introduced $\psi_a^{(j)} = p_0^{(i)} + \sum_{a'=1}^{L} p_{2a'}^{(j)} e^{-\beta e_{\sigma_a, \sigma_{a'}}}$.

The recursion is exact on a tree, but it also holds asymptotically on a large N Bethe lattice, in the liquid phase or inside a pure state where the probabilities on the k parent links are uncorrelated.

The liquid phase is described by a fixed point $p_\alpha^{(i)} \equiv p_\alpha^*$ of Eqs. (1)–(3). The liquid free energy ϕ_{liq} is obtained as a sum of site and edge contributions which are expressed in terms of $\{p_\alpha^*\}$ following the general prescription of Ref. [14]. It turns out that the liquid globule of neutral copolymers bears no trace of disorder since ϕ_{liq} is independent of the sequence.

For any heterogeneous sequence the entropy of the liquid solution turns negative at sufficiently low temperatures, indicating the existence of a glass phase which breaks the translational invariance. The latter is also clearly observed in Monte Carlo simulations both on the Bethe and the cubic 3D lattice [17]. To treat the glass phase, the above framework has to be extended. We assume that below the glass transition the phase space splits into a large number of pure states. Restricting the Gibbs measure to a state γ gives rise to probabilities $p_\alpha^{(i;\gamma)}$, and we again obtain a closed recursion relation if we look at the distribution of local probabilities over all pure states [14], $\rho(\mathbf{p}) = \mathcal{N}^{-1} \sum_\gamma w_\gamma \delta[\mathbf{p} - \mathbf{p}^{(i;\gamma)}]$. It will not depend on the site i since the graph is regular. The cavity recursion then reads

$$\rho(\mathbf{p}) = \frac{1}{Z} \int \prod_{i=1}^{k} \rho(\mathbf{p}^{(i)}) d\mathbf{p}^{(i)} \; \delta(\mathbf{p} - \mathbf{p}^{(0)}[\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(k)}])$$
$$\times e^{-m\beta\Delta f[\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(k)}]} \qquad (4)$$

where $\mathbf{p}^{(0)}[\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(k)}]$ is the pure state recursion (1)–(3).

We have introduced in (4) a reweighting with respect to the free energy change Δf corresponding to the addition of the site 0, in order to take into account the exponential increase of the number of pure states as a function of the total free energy. Δf is related to the normalization factor in (1)–(3) by $\exp(-\beta\Delta f) = C$. The reweighting depends on $m \in [0, 1]$ which is the cavity analog of the breakpoint parameter appearing in Parisi's one-step replica symmetry breaking scheme. As explained in [14,17], one can write down an m-dependent free energy $\phi_1(m)$, and the dominant metastable states correspond to m^* which maximizes $\phi_1(m)$.

The local stability of the liquid phase with respect to the glass transition can be studied by computing appropriately defined glass susceptibilities. The simplest one can be written by associating to each site i of the lattice a "spin" s_i. We define $s_i = 0$ if i is empty and $s_i = \sigma$, if i is occupied by a monomer of type σ. The susceptibility is given by $\chi_g = V^{-1} \sum_{ij} \overline{\langle s_i s_j \rangle^2}$. A divergence of χ_g signals the appearance of a glass phase.

The instability of the liquid can be analyzed for arbitrary sequences of monomers. It occurs at the temperature β_i determined by

$$\frac{1 + z_2(\beta_i) \cosh(\beta_i)}{z_2(\beta_i) \sinh(\beta_i)} = -k^{1/2} + \frac{2(k-1)Q_{\text{seq}}}{k^{1/2}(1 - k^{-L/2})}, \quad (5)$$

where $z_2(\beta_i) = L p_2^*/p_0^*$ is independent of the sequence [17]. The latter only enters through the term $Q_{\text{seq}} = \sum_{b=0}^{L-1} q_b(\pm\sqrt{k})^{-b}$, where $q_b = 1/L \sum_{a=1}^{L} \sigma_a \sigma_{a+b}$ is the sequence autocorrelation function and the upper/lower sign corresponds to the HP model and the ampholyte, respectively. Equation (5) can be regarded as the generalization of a classic result by de Gennes [18] to strongly frustrated systems.

In Fig. 2 we plot the local instability β_i obtained from Eq. (5) as a function of sequence correlations in the simple case of Markovian chains described by the probability π for two consecutive monomers to be equal. If the instability occurs too close to the entropy crisis of the liquid, it will be prevented by a discontinuous transition at β_d where a nontrivial solution of the cavity Eq. (4) appears. This effect can be studied by a numerical solution of this equation using a population dynamics method [14,17]. Upon increasing the period L, we found the transition points to approach a correlation-independent constant $\beta_d \approx 1.23 \pm 0.03$ that we believe to be a good approximation in the large L limit; see Fig. 2.

For ampholytic sequences with a bias to alternation, $\pi \lesssim 0.5$, and for HP-like polymers with a preference for blocks, $\pi \gtrsim 0.5$, there is a continuous glass transition at $\beta_i < \beta_d$ as predicted by (5). It is followed by a discontinuous freezing transition at lower temperature β_d. In copolymers with opposite correlations, the liquid instability is irrelevant and the freezing transition occurs directly from the liquid globule phase. The grand-canonical phase diagrams for the two scenarios are presented in Fig. 1 for two fixed values of π.

At the continuous phase transition β_i, the distribution $\rho(\mathbf{p})$ continuously departs from the liquid fixed point $\delta(\mathbf{p} - \mathbf{p}^*)$. The transition is of third order in the free energy, and as usual, the presence of a continuous transition indicates that the replica symmetry is fully broken. The states will organize in an ultrametric hierarchy of clusters with small "preferences" towards some particular conformations. On lowering the temperature, these preferences grow and result in increasing correlations at larger distances. This effect might be related to the scale-dependent freezing predicted in [12,13] and the intermediate glass phase found in the numerics of [10].

The frozen phase at low temperature is of a rather different character. The site probabilities $p_\alpha^{(i)}$ are strongly biased towards one particular site-dependent conformation. The pure states correspond to almost frozen conformations and have very small internal entropy, as in the REM.
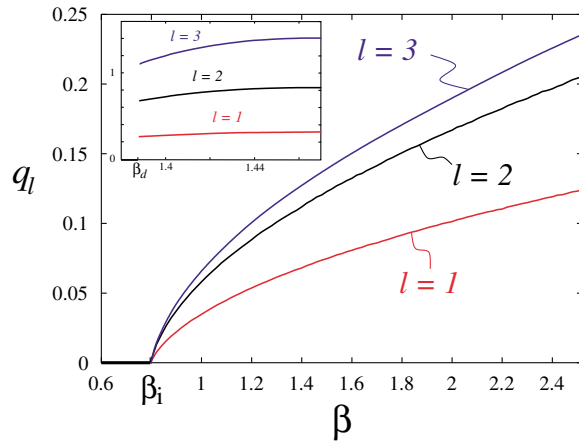
FIG. 3 (color online).   Temperature dependence of the corre-
lation plateaus q_l, for three types of correlations with $l = 1, 2, 3$, computed using the cavity method. The main graph
shows the continuous transition of strongly correlated HP
sequences. The inset displays the discontinuous transition of
anticorrelated HP sequences.

The freezing phenomenon can be quantitatively de-
scribed by the correlation plateaux q_l defined above. In
Fig. 3 we present our analytical results for $q_l(T)$ for two
sequences with opposite correlations.

What are the modifications of the above mean-field
picture for realistic (finite-dimensional) models? At least
two features must be seriously reconsidered. First, within
our approximation, the formation of ordered structures is
generally inhibited. Let us consider, for instance, the HP
copolymer with highly correlated sequences $\pi \approx 1$. In
dimension D, at low temperature, the system can separate
into A-rich and B-rich regions of linear size $\ell \sim \log(1 - \pi)^{1/D}$, with a frustration energy of order ℓ^{-1} per mono-
mer. The phase diagram in Fig. 2 must therefore be
modified in the $\pi \approx 1$ (for the HP model) and $\pi \approx 0$
(for the ampholyte) regions due to the emergence of
spatially ordered phases. A proper treatment of these
phenomena (which are crucial for block copolymers [2])
is beyond the scope of this Letter.

Second, the discontinuous dynamical phase transition
to the frozen glass phase cannot survive in finite dimen-
sions because of nucleation phenomena. According to
the usual scenario of discontinuous glass transitions,
we expect it to become a sharp crossover associated
with the emergence of very large ("activated") relaxation
times [11].

In conclusion, we have described a new approach to
heteropolymeric systems. The mean-field approximation
is introduced in a well-defined and consistent way, and
any further calculation can be checked through simula-
tions on the Bethe lattice. The flexibility of the approach
should allow for many applications (e.g., designed se-
quences) and further refinements (e.g., clusters of sites

in the cavity). We considered the case of copolymers with
correlated sequences and found that the frozen glass
phase can be preceded by a soft glass phase with very
different dynamical properties. Since the sequence corre-
lations turned out to be the decisive parameter, it would
be interesting to analyze protein sequences in the light of
our findings [19].

This work has been supported in part by the EC net-
work HPRN-CT-2002-00319, "Stipco."

*UMR 8549, Unité Mixte de Recherche du Centre
National de la Recherche Scientifique et de l'Ecole
Normale Supérieure.

[1] V. S. Pande, A. Y. Grosberg, and T. Tanaka, Rev. Mod.
Phys. **72**, 259 (2000).

[2] C. D. Sfatos and E. I. Shakhnovich, Phys. Rep. **288**, 77
(1997).

[3] P. G. Wolynes, in *Spin Glass Ideas in Biology*, edited by
D. Stein (World Scientific, Singapore, 1991).

[4] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin-Glass
Theory and Beyond* (World Scientific, Singapore, 1987).

[5] J. D. Bryngelson and P. G. Wolynes, Proc. Natl. Acad. Sci.
U.S.A **84**, 7524 (1987).

[6] B. Derrida, Phys. Rev. B **24**, 2613 (1981).

[7] E. I. Shakhnovich and A. M. Gutin, Biophys. Chem. **34**,
187 (1989); Europhys. Lett. **8**, 327 (1989); T. Garel and
H. Orland, Europhys. Lett. **6**, 307 (1988).

[8] T. Garel, L. Leibler, and H. Orland, J. Phys. II (France) **4**,
2139 (1994); C. D. Sfatos, A. M. Gutin, and E. I.
Shakhnovich, Phys. Rev. E **48**, 465 (1993).

[9] V. S. Pande, A. Y. Grosberg, C. Joerg, and T. Tanaka, Phys.
Rev. Lett. **76**, 3987 (1996).

[10] E. G. Timoshenko, Y. A. Kuznetsov, and K. A. Dawson,
Phys. Rev. E **54**, 4071 (1996); Phys. Rev. E **55**, 5750
(1997)**57**, 6801 (1998).

[11] J.-P. Bouchaud, L. F. Cugliandolo, J. Kurchan, and
M. Mézard, in *Spin Glasses and Random Fields*, edited
by A. P. Young (World Scientific, Singapore, 1997).

[12] A. Moskalenko, Y. A. Kuznetsov, and K. A. Dawson,
Europhys. Lett. **40**, 135 (1997).

[13] D. Thirumalai, V. Ashwin, and J. K. Bhattacharjee, Phys.
Rev. Lett. **77**, 5385 (1996).

[14] M. Mézard and G. Parisi, Eur. Phys. J. B **20**, 217 (2001).

[15] J. F. Nagle, Proc. R. Soc. London A **337**, 569 (1974);
S. Lise, A. Maritan, and A. Pelizzola, Phys. Rev. E **58**,
R5241 (1998).

[16] P. G. de Gennes, J. Phys. (Paris) Lett. **36**, L55 (1975).

[17] A. Montanari, M. Müller, and M. Mézard , J. Chem. Phys.
(to be published).

[18] P. G. de Gennes, Faraday Discuss. Chem. Soc. **68**, 96
(1979).

[19] V. S. Pande, A. Y. Grosberg, and T. Tanaka, Proc. Natl.
Acad. Sci. U.S.A **91**, 12972 (1994); A. Irbäck,
C. Peterson, and F. Potthast, Proc. Natl. Acad. Sci.
U.S.A **93**, 9533 (1996).