

Solvable lattice gas models of random heteropolymers at finite density: II. Dynamics and transitions to compact states

H. Chakravorty^a, J. van Mourik^b, and A.C.C. Coolen

Department of Mathematics, King's College London, The Strand, London WC2R 2LS, UK

Received 15 March 2001 and Received in final form 24 June 2001

Abstract. In this paper we analyse both the dynamics and the high density physics of the infinite dimensional lattice gas model for random heteropolymers recently introduced in [1]. Restricting ourselves to site-disordered heteropolymers, we derive exact closed deterministic evolution equations for a suitable set of dynamic order parameters (in the thermodynamic limit), and use these to study the dynamics of the system for different choices of the monomer polarity parameters. We also study the equilibrium properties of the system in the high density limit, which leads to a phase diagram exhibiting transitions between swollen states, compact states, and regions with partial compactification. Our results find excellent verification in numerical simulations, and have a natural and appealing interpretation in terms of real heteropolymers.

PACS. 61.41.+e Polymers, elastomers and plastics – 75.10.Nr Spin-glass and other random models

1 Introduction

The practical advantage of infinite dimensional models for random heteropolymers over the more standard ones (which are either defined on finite dimensional lattices, or in real space) is the absence of the polymer chain constraint. In infinite dimensional (or ‘lattice gas’) models, where every site is defined to be a neighbouring site of every other site, the constituent monomers of a polymer are given unrestricted freedom to occupy positions. The movements of the latter are controlled only by energetic considerations, by the maximum number of monomers allowed to occupy any given site, and by trivial restrictions on the number of moves allowed to occur simultaneously. In a recent paper [1], which was devoted to the equilibrium solution of a particular infinite dimensional model, this property was shown to enable an exact solution, in spite of the presence of quenched disorder. (*viz.* the realization of the constituent monomers of the chain, which are assumed to come in different species. The quenched disorder is entirely specified by the proportions of the various monomeric species present. No further aspects of the system are quenched.) In those cases where one’s interest is mostly in issues which are not directly related to spatial properties of the heteropolymer, the compromise of infinite dimensionality might be a price worth paying for exact solvability.

The present paper is concerned with the further analysis and exploration of the model introduced in [1], to

which we refer for a more detailed and extensive discussion of the relevant literature and of the motivation underlying the study of random heteropolymers in general (*e.g.* their role as precursors of models for proteins), and lattice gas models of random heteropolymers in particular (see also *e.g.* [2–4] and references therein). We would like to emphasize that this model does not represent realistically the polymeric features of a heteropolymer, but it does capture some of the characteristics, and for this reason it can be considered as a precursor for a model of random heteropolymers. We complement the equilibrium study [1] in two distinct ways. Firstly, we solve the dynamics of the model (which we define as a stochastic process of the Glauber type), by deriving exact closed deterministic laws in the thermodynamics limit for a suitably chosen set of dynamic order parameters. This can be done, at least for site-disordered models, as a direct consequence of the simplification induced by the infinite dimensionality of the model, and allows for an exact quantitative analysis of the approach to equilibrium. Secondly, we study the equilibrium properties of the model of [1] in the limit of high local density. Since the original procedure followed in [1] for solving the model (which applies to both site-disordered and bond-disordered systems), does not allow for this limit to be taken, we first present an alternative (but equivalent) equilibrium solution (which applies only to site disorder) where taking the high local density limit is allowed. We then find, after an appropriate re-scaling of the system temperature (in order to prevent entropy dominance of the free energy) a non-trivial phase diagram with a distinct swollen phase, a compact phase, and a phase characterised by partial compactification of

^a e-mail: hirak@mth.kcl.ac.uk

^b Present address: The Neural Computing Research Group, Aston University, Birmingham B4 7ET, UK

the heteropolymer. Especially near the transition lines in this diagram, the system is found to exhibit a highly non-trivial and slow plateau-type dynamics.

We supplement our theoretical analysis by extensive numerical simulations, which support our predictions regarding the dynamics and regarding the high-density phase diagram very satisfactorily (within the experimental limitations imposed by the slow dynamics).

2 Microscopic rules of the model

The model of [1] describes a system of N monomers, labeled $i \in \{1, \dots, N\}$, and $R = \alpha N$ possible sites for these monomers, labeled $r \in \{1, \dots, R\}$. The degrees of freedom of the system are the locations $r_i \in \{1, \dots, R\}$ of the N monomers, and the overall system state is defined by the state vector $\mathbf{r} = (r_1, \dots, r_N)$. Each site, however, can occupy at most n_c monomers; this imposes R constraints on the state vector of the form $\sum_i \delta_{r_i, r} \leq n_c$ (one for each r), and limits the global monomer density according to $\alpha \geq n_c^{-1}$. Those positions at a given site which are not occupied by monomers, are assumed to be occupied by solvent molecules. The monomers are allowed to interact (in pairs) only when they find themselves at the same site. Within equilibrium statistical mechanics this is described by a Hamiltonian of the form

$$H(\mathbf{r}) = \sum_{i < j} \lambda_{ij} \delta_{r_i, r_j}. \quad (1)$$

The disorder in the system is in the realisation of the interaction energies λ_{ij} . Unlike [1], we will in this paper restrict ourselves to models with site disorder, where

$$\lambda_{ij} = \phi(\lambda_i, \lambda_j) \quad (2)$$

such as the Random Hydrophobic Model (RHM) [5–7], where $\phi(\lambda, \lambda') = \frac{1}{2}(\lambda + \lambda')$, or the Random Charge Model (RCM) [8, 9], where $\phi(\lambda, \lambda') = \epsilon \lambda_i \lambda_j$. Here the variable λ_i defines the monomer species of monomer i , with $\lambda_i > 0$ for hydrophilic/polar monomers, and $\lambda_i < 0$ for hydrophobic monomers. The RHM describes a system where hydrophobic monomers try to avoid contact with the solvent, and all hydrophilic monomers try to optimise contact with the solvent. The parameter $\epsilon \in \{-1, 1\}$ in the RCM controls whether monomers of identical polarity attract or repel each other. We restrict ourselves, for simplicity, to the case whether the disorder is discrete in nature and where there are K monomer species, *i.e.* $\lambda_i \in \{A_1, \dots, A_K\}$.

We next endow this equilibrium system with a Glauber-type dynamics. Time is discretised, and the probability to find system state \mathbf{r} at iteration step ℓ is written as $p_\ell(\mathbf{r})$. The elementary transitions we will allow for are those where a single (randomly drawn) monomer is moved to a (randomly drawn) new site; to describe such transitions we define an operator F_{ix} which moves monomer i to site x : $F_{ix}(\mathbf{r}) = (r_1, \dots, r_{i-1}, x, r_{i+1}, \dots, r_N)$. The evolution of the system is a Markovian stochastic process of the

form

$$p_{\ell+1}(\mathbf{r}) = \sum_{\mathbf{r}'} W[\mathbf{r}; \mathbf{r}'] p_\ell(\mathbf{r}') \quad (3)$$

with the transition probabilities

$$W[\mathbf{r}, \mathbf{r}'] = \frac{1}{NR} \sum_{i=1}^N \sum_{x=1}^R \left[\prod_{j \neq i} \delta_{r_j, r'_j} \right] \{ \delta_{n_x(\mathbf{r}'), n_c} \delta_{r_i, r'_i} + \bar{\delta}_{n_x(\mathbf{r}'), n_c} [p(F_{ix}\mathbf{r}'; \mathbf{r}') \delta_{r_i, x} + [1 - p(F_{ix}\mathbf{r}'; \mathbf{r}')] \delta_{r_i, r'_i}] \} \quad (4)$$

and

$$p[\mathbf{r}; \mathbf{r}'] = \frac{e^{-\frac{1}{2}\beta\Delta H(\mathbf{r}; \mathbf{r}')}}{2 \cosh[\frac{1}{2}\beta\Delta H(\mathbf{r}; \mathbf{r}')] } \quad (5)$$

in which $\Delta H(\mathbf{r}; \mathbf{r}') = H(\mathbf{r}) - H(\mathbf{r}')$, $\bar{\delta}_{ij} = 1 - \delta_{ij}$, and $n_x(\mathbf{r}) = \sum_i \delta_{r_i, x}$ (the number of monomers found at site x). The process (3) is one where we forbid all transitions which would lead to violation of the constraint $n_x(\mathbf{r}) \leq n_c$, and where we allow all other single particle moves (where both the candidate particle i to be moved and the candidate site x for it to move to are drawn uniformly at random) with Glauber-type probabilities, such that the invariant distribution will be the Boltzmann state (provided the system is ergodic, which in the present case is intuitively clear). One easily convinces oneself that the transition matrix (4) is properly normalised, *i.e.* $\sum_{\mathbf{r}'} W[\mathbf{r}, \mathbf{r}'] = 1$ for all \mathbf{r} , and that the Markov chain (3) indeed obeys detailed balance, *i.e.* $W[\mathbf{r}, \mathbf{r}'] p_\infty(\mathbf{r}') = W[\mathbf{r}', \mathbf{r}] p_\infty(\mathbf{r})$ for all $\{\mathbf{r}, \mathbf{r}'\}$, with $p_\infty(\mathbf{r}) \sim \exp[-\beta H(\mathbf{r})]$.

Although the Markov chain (3) is the representation which is most conveniently implemented in numerical simulations, for a theoretical analysis of the dynamics a continuous time description would be preferable. This is achieved (see [10]) upon choosing the duration of each iteration step of (3) to be a Poisson distributed random number (with average τ), which converts the Markov chain into a master equation of the form

$$\tau \frac{d}{dt} p_t(\mathbf{r}) = \sum_{\mathbf{r}'} W[\mathbf{r}; \mathbf{r}'] p_t(\mathbf{r}') - p_t(\mathbf{r}).$$

Upon choosing $\tau = N^{-1}$ (so that every monomer is targeted for update on average once per unit time), and upon inserting (4), we then arrive at the master equation

$$\frac{d}{dt} p_t(\mathbf{r}) = \frac{1}{R} \sum_{ix} \sum_{\mathbf{r}'} p_t(\mathbf{r}') p[F_{ix}\mathbf{r}'; \mathbf{r}] \times \bar{\delta}_{n_x(\mathbf{r}'), n_c} [\delta_{\mathbf{r}, F_{ix}\mathbf{r}'} - \delta_{\mathbf{r}, \mathbf{r}'}]. \quad (6)$$

Using the properties of the Poisson distribution, one can show that the relative uncertainty on the time axis, induced by having random durations of iteration steps, will vanish for $N \rightarrow \infty$.

We finally note that the only remaining non-trivial Glauber term (5) occurring in (6) is of the following form:

$$p[F_{ix}\mathbf{r}; \mathbf{r}] = \frac{1}{2} - \frac{1}{2} \tanh \left[\frac{1}{2} \beta \Delta_{ix} H(\mathbf{r}) \right] \quad (7)$$

with $\Delta_{ix} H(\mathbf{r}) = H(F_{ix}\mathbf{r}) - H(\mathbf{r})$, which is here given by

$$\Delta_{ix} H(\mathbf{r}) = \sum_j \bar{\delta}_{ij} \phi(\lambda_i, \lambda_j) [\delta_{x,r_j} - \delta_{r_i,r_j}].$$

Upon defining $n_{\lambda,x}(\mathbf{r}) = \sum_i \delta_{r_i,x} \delta_{\lambda_i,\lambda}$ (counting the number of type- λ monomers at site x), we can write this as

$$\Delta_{ix} H(\mathbf{r}) = \sum_{\lambda} \phi(\lambda_i, \lambda) [n_{\lambda,x}(\mathbf{r}) - n_{\lambda,r_i}(\mathbf{r})] + \phi(\lambda_i, \lambda_i) \bar{\delta}_{x,r_i}. \quad (8)$$

3 Dynamics: derivation of closed macroscopic laws

The appropriate dynamic order parameters of the system are the smallest set of macroscopic observables which obey closed and deterministic evolution equations in the thermodynamic limit. In the present model this set turns out to be, at least for finite n_c and for pair energies of the separable form (2), the fraction $c_{\mathbf{n}}(\mathbf{r})$ of sites with a given local population (or 'local state') $\mathbf{n} = (n_{A_1}, \dots, n_{A_K})$ of monomers. Here n_{λ} indicates the number of monomers of species λ . Thus

$$c_{\mathbf{n}}(\mathbf{r}) = \frac{1}{R} \sum_{x=1}^R \left[\prod_{\lambda=A_1}^{A_K} \delta_{n_{\lambda,x}(\mathbf{r}), n_{\lambda}} \right] \quad (9)$$

with $\mathbf{n} \in \{0, \dots, n_c\}^K$. Note that $\sum_{\mathbf{n}} c_{\mathbf{n}}(\mathbf{r}) = 1$ and that $c_{\mathbf{n}}(\mathbf{r}) = 0$ as soon as $\sum_{\lambda} n_{\lambda} > n_c$ (due to our constraint on the maximum number of monomers at any given site), for any \mathbf{r} . We will write the constraint as $g[\mathbf{n}] \in \{0, 1\}$, where $g[\mathbf{n}] = 1$ if and only if $\sum_{\lambda} n_{\lambda} \leq n_c$.

The probability of finding the set of observables $\{c_{\mathbf{n}}(\mathbf{r})\}$ at time t is given by

$$P_t[\{c_{\mathbf{n}}\}] = \sum_{\mathbf{r}} p_t(\mathbf{r}) \prod_{\mathbf{n}} \delta[c_{\mathbf{n}} - c_{\mathbf{n}}(\mathbf{r})]. \quad (10)$$

Its evolution in time follows upon taking a temporal derivative and inserting the master equation (6):

$$\begin{aligned} \frac{d}{dt} P_t[\{c_{\mathbf{n}}\}] &= \frac{1}{2R} \sum_{\mathbf{r}} p_t(\mathbf{r}) \sum_{ix} \left[1 - \tanh \left[\frac{1}{2} \beta \Delta_{ix} H(\mathbf{r}) \right] \right] \\ &\times \bar{\delta}_{n_x(\mathbf{r}), n_c} \left\{ \prod_{\mathbf{n}} \delta[c_{\mathbf{n}} - c_{\mathbf{n}}(F_{ix}\mathbf{r})] - \prod_{\mathbf{n}} \delta[c_{\mathbf{n}} - c_{\mathbf{n}}(\mathbf{r})] \right\}. \end{aligned}$$

We expand this expression, using

$$\begin{aligned} c_{\mathbf{n}}(F_{ix}\mathbf{r}) &= c_{\mathbf{n}}(\mathbf{r}) \\ &+ \frac{1}{R} \bar{\delta}_{x,r_i} [\delta_{n_{\lambda_i,r_i}(\mathbf{r}), n_{\lambda_i}+1} - \delta_{n_{\lambda_i,r_i}(\mathbf{r}), n_{\lambda_i}}] \prod_{\lambda \neq \lambda_i} \delta_{n_{\lambda,r_i}(\mathbf{r}), n_{\lambda}} \\ &+ \frac{1}{R} \bar{\delta}_{x,r_i} [\delta_{n_{\lambda_i,x}(\mathbf{r}), n_{\lambda_i}-1} - \delta_{n_{\lambda_i,x}(\mathbf{r}), n_{\lambda_i}}] \prod_{\lambda \neq \lambda_i} \delta_{n_{\lambda,x}(\mathbf{r}), n_{\lambda}}. \end{aligned}$$

We also insert the partitionings $1 = \sum_{\mathbf{n}''} \left[\prod_{\lambda} \delta_{n_{\lambda,x}(\mathbf{r}), n_{\lambda}''} \right]$ and $1 = \sum_{\mathbf{n}''' } \left[\prod_{\lambda} \delta_{n_{\lambda,r_i}(\mathbf{r}), n_{\lambda}''' } \right]$, and we use (8). This then, together with the identity $n_x(\mathbf{r}) = \sum_{\lambda} n_{\lambda,x}(\mathbf{r})$, leads to the following differential equation for $P_t[\{c_{\mathbf{n}}\}]$:

$$\begin{aligned} \frac{d}{dt} P_t[\{c_{\mathbf{n}}\}] &= - \sum_{\mathbf{n}'} \frac{\partial}{\partial c_{\mathbf{n}'}} \left\{ P_t[\{c_{\mathbf{n}}\}] \sum_{\lambda} \sum_{\mathbf{n}''} \sum_{\mathbf{n}'''} c_{\mathbf{n}''} c_{\mathbf{n}'''} \right. \\ &\times \frac{1}{2} n_{\lambda}'' \left[1 - \tanh \left[\frac{1}{2} \beta \left[\sum_{\lambda'} \phi(\lambda, \lambda') (n_{\lambda'}'' - n_{\lambda'}''') + \phi(\lambda, \lambda) \right] \right] \right] \\ &\times \bar{\delta}_{\sum_{\lambda'} n_{\lambda'}'', n_c} \left[(\delta_{n_{\lambda}''', n_{\lambda}+1} - \delta_{n_{\lambda}''', n_{\lambda}}) \prod_{\lambda' \neq \lambda} \delta_{n_{\lambda'}'', n_{\lambda'}} \right. \\ &\left. \left. + (\delta_{n_{\lambda}''', n_{\lambda}-1} - \delta_{n_{\lambda}''', n_{\lambda}}) \prod_{\lambda' \neq \lambda} \delta_{n_{\lambda'}'', n_{\lambda'}} \right] \right\} + \mathcal{O} \left(\frac{1}{N} \right). \end{aligned}$$

Taking the limit $N \rightarrow \infty$ for finite n_c and finite K converts this into a Liouville equation, describing deterministic flow, *i.e.* solutions of the form $P_t[\{c_{\mathbf{n}}\}] = \prod_{\mathbf{n}} \delta[c_{\mathbf{n}} - c_{\mathbf{n}}(t)]$ where the deterministic trajectories $\{c_{\mathbf{n}}(t)\}$ are simply given by the solution of the associated flow equations. With the short-hands $F_{\lambda}^{\pm} \mathbf{n} = (n_{A_1}, \dots, n_{\lambda-1}, n_{\lambda} \pm 1, n_{\lambda+1}, \dots, n_{A_K})$ these (coupled) flow equations can be simplified to

$$\begin{aligned} \frac{d}{dt} c_{\mathbf{n}} &= \sum_{\lambda} \sum_{\mathbf{n}'} c_{\mathbf{n}'} \left\{ c_{F_{\lambda}^+ \mathbf{n}} (n_{\lambda} + 1) \bar{\delta}_{\sum_{\mu} n_{\mu}', n_c} \varepsilon_{\lambda} [\mathbf{n}' - \mathbf{n}] \right. \\ &+ c_{F_{\lambda}^- \mathbf{n}} n_{\lambda}' \bar{\delta}_{\sum_{\mu} n_{\mu}', n_c + 1} \varepsilon_{\lambda} [\mathbf{n} - \mathbf{n}'] \\ &- c_{\mathbf{n}} n_{\lambda} \bar{\delta}_{\sum_{\mu} n_{\mu}', n_c} \varepsilon_{\lambda} [F_{\lambda}^+ \mathbf{n}' - \mathbf{n}] \\ &\left. - c_{\mathbf{n}} n_{\lambda}' \bar{\delta}_{\sum_{\mu} n_{\mu}', n_c} \varepsilon_{\lambda} [\mathbf{n} - F_{\lambda}^- \mathbf{n}'] \right\} \quad (11) \end{aligned}$$

where

$$\varepsilon_{\lambda}[\mathbf{m}] = \frac{1}{2} - \frac{1}{2} \tanh \left[\frac{1}{2} \beta \sum_{\lambda'} \phi(\lambda, \lambda') m_{\lambda'} \right]. \quad (12)$$

The imposed limit on the maximum number of monomers at a given site is seen to be built into (11), since for any local monomer arrangement \mathbf{n} with $\sum_{\lambda} n_{\lambda} = n_c + 1$ equation (11) indeed dictates that $\frac{d}{dt} c_{\mathbf{n}} = 0$ if $c_{\mathbf{n}} = 0$.

Since the dimensionality of our macroscopic laws increases with n_c and the number K of monomer types, further reductions can be helpful. One reduction is to consider the fraction $p[n_+, n_-]$ of sites with local monomer arrangements where n_+ monomers have $\lambda_i > 0$ and n_- monomers have $\lambda_i < 0$ (for the RHM this implies counting hydrophobic and hydrophilic monomers; for the RCM one counts monomers with a specific charge sign):

$$p[n_+, n_-] = \sum_{\mathbf{n}} c_{\mathbf{n}} \delta_{\sum_{\lambda > 0} n_{\lambda}, n_+} \delta_{\sum_{\lambda < 0} n_{\lambda}, n_-} \quad (13)$$

with $n_{\pm} \in \{0, n_c\}$. We will, more specifically, inspect the (co-)variances $\sigma_{\pm}^2 = \langle n_{\pm}^2 \rangle - \langle n_{\pm} \rangle^2$ and $\rho = \langle n_+ n_- \rangle$. They give information about the spatial properties of the system state: a large σ_{\pm} implies spatial concentration of

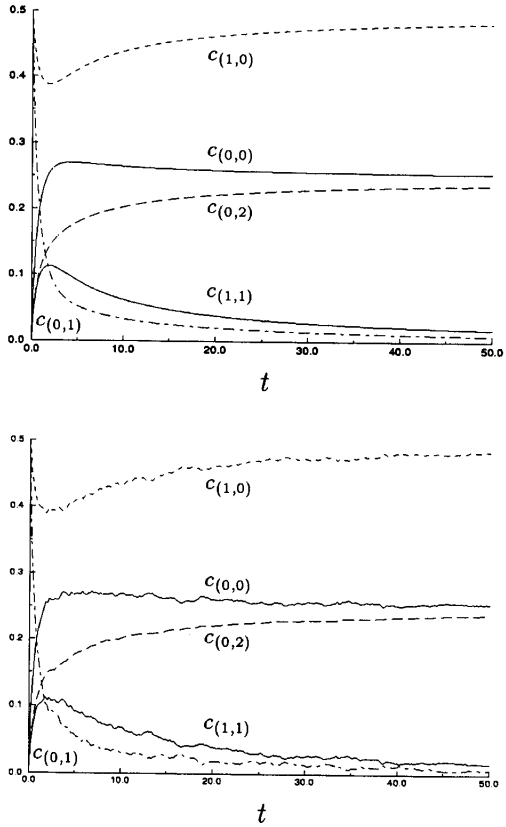


Fig. 1. Evolution of the order parameters $\{c_n\}$ for the Random Hydrophilic Model (RHM), with $K = 2$ (two monomer types, $\lambda \in \{-1, 1\}$, one hydrophobic and one hydrophilic), $n_c = 2$, $\alpha = 1$, and $\beta = 100\,000$. Upper graphs: theory, *i.e.* numerical solution of (11). Lower graphs: numerical simulations, with $N = 3000$. Note: in this example $c_{(2,0)} = 0$ for all t .

a monomer type in a relatively small number of sites, whereas ρ informs us of the degree of spatial separation of the monomer types. These quantities $p[n_+, n_-]$ will normally not obey closed equations, but are calculated *via* (13) from the (numerical) solution of (11).

4 Dynamics: comparison with numerical simulations

We first compare the numerical solution of the macroscopic laws (11) with the results of measuring the observables (9) directly in numerical simulations of the microscopic process (3, 4). All simulations are presented as single runs, as opposed to ensemble averages. A typical example is shown in Figure 1, for the RHM (*i.e.* $\phi(\lambda, \lambda') = \frac{1}{2}(\lambda + \lambda')$, with $K = 2$ (two monomer types, $\lambda \in \{-1, 1\}$, one hydrophobic and one hydrophilic), $n_c = 2$, $\alpha = 1$ and $\beta = 100\,000$. In this case

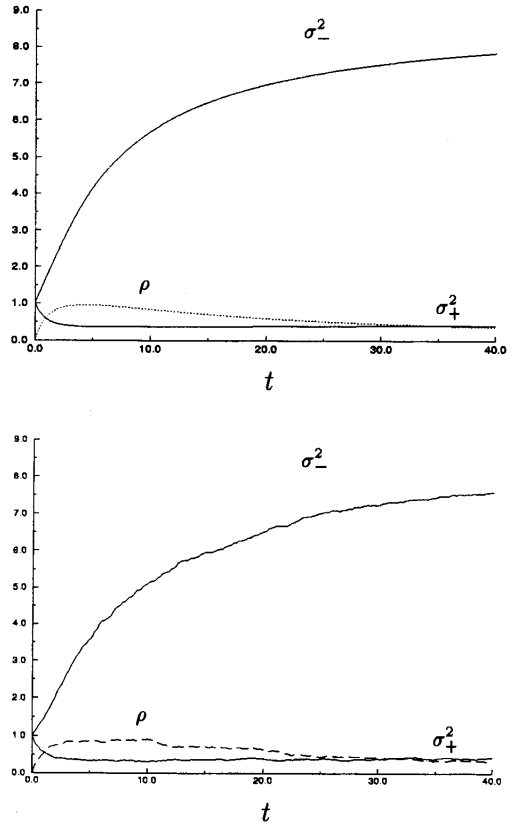


Fig. 2. Evolution of the (co-)variances $\sigma_{\pm}^2 = \langle n_{\pm}^2 \rangle$ and $\rho = \langle n_+ n_- \rangle - \langle n_+ \rangle \langle n_- \rangle$, for the Random Hydrophilic Model (RHM), with $K = 2$ ($\lambda \in \{-1, 1\}$), $n_c = 10$, $\alpha = \frac{1}{2}$, and $\beta = 1$. Initial conditions: $c_{(2,0)} = c_{(0,2)} = 0.5$ (perfectly separated monomer species). Upper graphs: theory, *i.e.* numerical solution of (11). Lower graphs: numerical simulations, with $N = 2000$. From top to bottom in both pictures: $\sigma_{-}^2(t)$ (with $\sigma_{-}^2(0) = 1$), $\rho(t)$ (dotted/dashed, with $\rho(0) = 0$) and $\sigma_{+}^2(t)$ (with $\sigma_{+}^2(0) = 1$).

$\mathbf{n} \in \{(0, 0), (1, 0), (0, 1), (1, 1), (2, 0), (0, 2)\}$. Note that $\sum_{\mathbf{n}} c_{\mathbf{n}} = 1$ at all times, by definition. We observe excellent agreement between theory and experiment. The behaviour of the system can be understood as follows. Hydrophilic monomers, with $\lambda = 1$, prefer local arrangements where contact with the solvent is maximised, *i.e.* where $n_+ \in \{0, 1\}$ (since the maximum number of molecules per site is $n_c = 2$). Hydrophobic monomers prefer the opposite, *i.e.* arrangement with $n_- \in \{0, 2\}$. Hence the observed evolution of the system towards a state where only $c_{(0,0)}$, $c_{(0,2)}$ and $c_{(1,0)}$ are nonzero.

Note that as soon as $K = 2$, with one hydrophobic and one hydrophilic monomer species (as in the example above), the observables in (13) simply reduce to $p[n_+, n_-] = c_{(n_+, n_-)}$, and the (co-)variances of $p[n_+, n_-]$ reduce to those of the c_n . In Figure 2 we show a first example of the evolution of $\sigma_{\pm}^2 = \langle n_{\pm}^2 \rangle$ and $\rho = \langle n_+ n_- \rangle$,

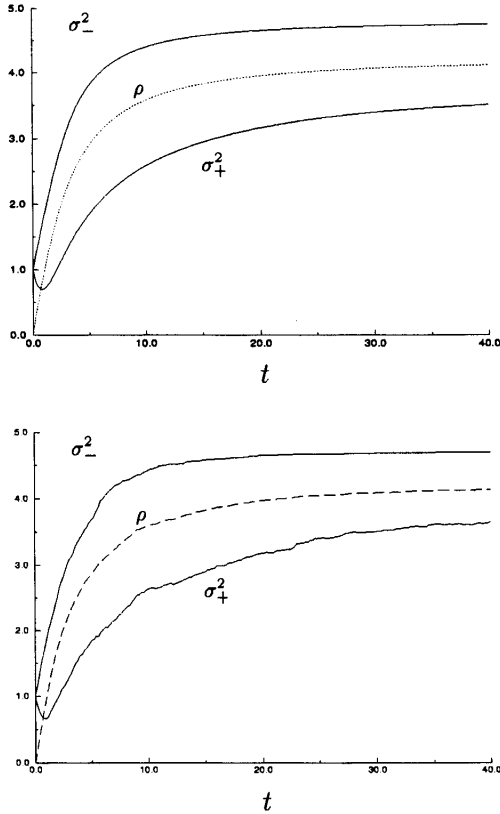


Fig. 3. Evolution of the (co-)variances $\sigma_{\pm}^2 = \langle n_{\pm}^2 \rangle$ and $\rho = \langle n_+ n_- \rangle - \langle n_+ \rangle \langle n_- \rangle$, for the Random Hydrophilic Model (RHM), with $K = 2$ ($\lambda \in \{-10, 1\}$, *i.e.* very strong hydrophobicity), $n_c = 10$, $\alpha = \frac{1}{2}$, and $\beta = 1$. Initial conditions: $c_{(2,0)} = c_{(0,2)} = 0.5$ (perfectly separated monomer species). Upper graphs: theory, *i.e.* numerical solution of (11). Lower graphs: numerical simulations, with $N = 2000$. From top to bottom in both pictures: $\sigma_-^2(t)$ (with $\sigma_-^2(0) = 1$), $\rho(t)$ (dotted/dashed with $\rho(0) = 0$) and $\sigma_+^2(t)$ (with $\sigma_+^2(0) = 1$).

for the RHM with $\beta = 1$, $n_c = 10$ and $\alpha = \frac{1}{2}$. Here $\lambda_+ = 1$ and $\lambda_- = -1$, and the initial conditions were $c_{(2,0)} = c_{(0,2)} = 0.5$ (*i.e.* perfectly separated monomer species). We observe again a perfect agreement between theory and (numerical) experiment. The development of a large variance σ_- for the hydrophobic monomers is explained by the observation that such monomers promote arrangements of the type $(n_+, n_-) = (0, n_c)$ or $(n_+, n_-) = (\star, 0)$ (*i.e.* $n_- \in \{0, n_c\}$, without intermediate values). Hydrophilic monomers, on the other hand, promote local arrangements of the type $(n_+, n_-) = (1, \star)$, *i.e.* with a single optimal value for n_+ , hence the low variance σ_+ . The low observed correlation ρ indicates that the various objectives can be achieved independently. The situation is quite different for the system illustrated in Figure 3, which only differs from the previous one in that the hydrophobicity

of the hydrophobic monomers has been strengthened from $\lambda = -1$ to $\lambda = -10$. Now the system will find it energetically favourable to ‘absorb’ hydrophilic monomers into sites where hydrophobic monomers are in contact with solvent (that is, in addition to the system aiming for its hydrophobic monomers to be in $(n_+, n_-) = (0, n_c)$ arrangements, wherever possible). The energetic gain of the hydrophobic monomers, which results upon having their hydrophilic cousins replacing solvent molecules, overrules the energetic loss of the hydrophilic ones, when the latter are taken away from solvent and put into contact with hydrophobic monomers. The primary aim of the hydrophobic monomers is to avoid water molecules, and it is acceptable for this to be achieved by sharing sites with hydrophilic monomers. Hence, virtually any proposed move is accepted which involves a monomer going to a site with an adequate number of hydrophobic monomers, regardless of the species of the mobile monomer. As a result we observe, in addition to the increase of the variance σ_-^2 of the hydrophobic monomers, an increase of both the variance σ_+^2 of the hydrophilic monomers and of the covariance ρ .

5 Statics: an alternative solution

In the remainder of this paper we will concentrate on transitions between low density (swollen) and high density (collapsed) states. We distinguish between the overall density $\alpha^{-1} = N/p$ (which will remain finite) and the local density as limited by n_c , which is the quantity which will be allowed to diverge. In [1] the (exact) equilibrium solution of site-disordered models (1, 2) was derived in a form which was specifically convenient to deal with many-valued or continuous disorder distributions; for large n_c , however, it is very cumbersome (taking the $n_c \rightarrow \infty$ limit would be ruled out). Since finding collapsed states requires $n_c \rightarrow \infty$, we will first give an alternative solution method, which is complementary to that of [1] in that it is convenient for large n_c , but inappropriate to deal with many-valued disorder distributions. In addition it will enable a natural and transparent connection with the dynamical solution (11). We write the partition function for the system (1, 2) as

$$\mathcal{Z} = \sum_{\mathbf{r}} \left[\prod_{x=1}^R \theta_{n_c, n_x(\mathbf{r})} \right] e^{-\beta \sum_{i < j} \phi(\lambda_i, \lambda_j) \delta_{r_i, r_j}} \quad (14)$$

with the short-hand

$$\theta_{x,y} = \begin{cases} 1 & \text{for } x > y \\ 0 & \text{for } x \leq y \end{cases}$$

(and its obvious vectorial generalisation $\theta_{\mathbf{x}, \mathbf{y}} = \prod_i \theta_{x_i, y_i}$). We introduce the vectors $\mathbf{m} = (m_1, \dots, m_R)$ with $m_x \in \{0, \dots, n_c\}$ to indicate the total number of monomers found at each of the R sites. Similarly we define the vector $\mathbf{m}_c = (n_c, \dots, n_c)$. This allows us to write the system’s free energy as $F = -\beta^{-1} \log \sum_{\mathbf{m}} \theta_{\mathbf{m}_c, \mathbf{m}} \mathcal{Z}[\mathbf{m}]$, where the

'conditioned' partition functions $\mathcal{Z}[\mathbf{m}]$ are given by

$$\mathcal{Z}[\mathbf{m}] = \sum_{\mathbf{r}} \left[\prod_{x=1}^R \delta_{m_x, n_x(\mathbf{r})} \right] e^{-\beta \sum_{i < j} \phi(\lambda_i, \lambda_j) \delta_{r_i, r_j}}. \quad (15)$$

We consider only the scaling regime where the limit $n_c \rightarrow \infty$ is taken *after* the thermodynamic limit $N \rightarrow \infty$. This allows us to solve our model in terms of the familiar sublattices $I_\lambda = \{i | \lambda_i = \lambda\}$, of size $|I_\lambda|$, and their associated local occupation numbers $n_{\lambda, x}(\mathbf{r}) = \sum_{i \in I_\lambda} \delta_{x, r_i}$. The latter are, in turn, also grouped into R -dimensional vectors according to $\mathbf{m}_\lambda(\mathbf{r}) = (n_{\lambda, 1}(\mathbf{r}), \dots, n_{\lambda, R}(\mathbf{r}))$, which allows us to write the conditioned partition function as

$$\mathcal{Z}[\mathbf{m}] = e^{\frac{1}{2}\beta \sum_i \phi(\lambda_i, \lambda_i)} \prod_{\lambda} \left[\sum_{\mathbf{m}_\lambda} \mathcal{N}_{|I_\lambda|}[\mathbf{m}_\lambda] \right] \times \delta_{\mathbf{m}, \sum_{\lambda} \mathbf{m}_\lambda} e^{-\frac{1}{2}\beta \sum_{\lambda, \lambda'} \phi(\lambda, \lambda') \mathbf{m}_\lambda \cdot \mathbf{m}_{\lambda'}} \quad (16)$$

where

$$\mathcal{N}_M[\mathbf{m}] = \frac{M!}{m_1! \dots m_R!} \delta_{\sum_x m_x, M}$$

is a combinatorial term, counting the number of ways in which M objects can be grouped into R subsets of prescribed sizes $\{m_1, \dots, m_R\}$. We note that $\lim_{N \rightarrow \infty} |I_\lambda|/N = p_\lambda$ (the a priori probability to draw a monomer of type λ randomly). In the thermodynamic limit $N \rightarrow \infty$ the free energy per monomer $f = \lim_{N \rightarrow \infty} F/N$ becomes

$$f = -\frac{1}{2} \sum_{\lambda} p_\lambda \phi(\lambda, \lambda) - \lim_{N \rightarrow \infty} \frac{1}{\beta N} \log \sum_{\mathbf{m}} \theta_{\mathbf{m}_c, \mathbf{m}} \sum_{\{\mathbf{m}_\lambda\}} \delta_{\mathbf{m}, \sum_{\lambda} \mathbf{m}_\lambda} \times \prod_{\lambda} \left[\mathcal{N}_{|I_\lambda|}[\mathbf{m}_\lambda] \right] e^{-\frac{1}{2}\beta \sum_{\lambda, \lambda'} \phi(\lambda, \lambda') \mathbf{m}_\lambda \cdot \mathbf{m}_{\lambda'}}. \quad (17)$$

Introducing integral representations for the δ -constraints in the terms $\mathcal{N}_{|I_\lambda|}[\mathbf{m}_\lambda]$, *i.e.* $\delta_{nm} = (2\pi)^{-1} \int_0^{2\pi} du e^{iu(n-m)}$, leads to a full factorisation over sites, and consequently to an expression to be evaluated by saddle-point integration (provided K remains finite as $N \rightarrow \infty$). This then allows us to write the non-constant part of the free energy per monomer as:

$$f = -\frac{1}{\beta} \text{extr}_{\mathbf{y}} \left\{ \alpha \log \left[\sum_{\{n_\lambda\}} \frac{\theta_{n_c, \sum_{\lambda} n_\lambda}}{\prod_{\lambda} n_\lambda!} \times e^{\sum_{\lambda} y_\lambda n_\lambda - \frac{1}{2}\beta \sum_{\lambda, \lambda'} \phi(\lambda, \lambda') n_\lambda n_{\lambda'}} \right] - \sum_{\lambda} p_\lambda y_\lambda \right\}. \quad (18)$$

Variation with respect to $\mathbf{y} = \{y_\lambda\}$ gives the K saddle-point equations

$$\frac{p_\mu}{\alpha} = \frac{1}{Z} \sum_{\{n_\lambda\}} \frac{n_\mu \theta_{n_c, \sum_{\lambda} n_\lambda}}{\prod_{\lambda} n_\lambda!} e^{\sum_{\lambda} y_\lambda n_\lambda - \frac{1}{2}\beta \sum_{\lambda, \lambda'} \phi(\lambda, \lambda') n_\lambda n_{\lambda'}} \quad (19)$$

with the normalisation factor

$$Z = \sum_{\{n_\lambda\}} \frac{\theta_{n_c, \sum_{\lambda} n_\lambda}}{\prod_{\lambda} n_\lambda!} e^{\sum_{\lambda} y_\lambda n_\lambda - \frac{1}{2}\beta \sum_{\lambda, \lambda'} \phi(\lambda, \lambda') n_\lambda n_{\lambda'}}. \quad (20)$$

The relation of this equilibrium solution (18, 19) to physical observables and to the previously obtained dynamical equations (11) becomes clear if one adds a suitable infinitesimal generating term to the Hamiltonian (1):

$$H(\mathbf{r}) \rightarrow H(\mathbf{r}) + \frac{\chi}{R} \sum_{x=1}^R \left[\prod_{\lambda=A_1}^{A_K} \delta_{n_{\lambda, x}(\mathbf{r}), n_\lambda} \right].$$

We repeat the above derivation for the extended Hamiltonian and use the identity $\lim_{\chi \rightarrow 0} \partial f / \partial \chi = \langle c_n(\mathbf{r}) \rangle$, where $\langle \dots \rangle$ denotes the thermal average in the Boltzmann state, and where the observables $c_n(\mathbf{r})$ are defined in (9). This reveals that the numbers $\mathbf{n} = \{n_\lambda\}$ in (18, 19) are identical to the local occupation numbers in (11), and that in equilibrium

$$c_n = \frac{1}{Z} \frac{\theta_{n_c, \sum_{\lambda} n_\lambda}}{\prod_{\lambda} n_\lambda!} e^{\sum_{\lambda} y_\lambda n_\lambda - \frac{1}{2}\beta \sum_{\lambda, \lambda'} \phi(\lambda, \lambda') n_\lambda n_{\lambda'}}. \quad (21)$$

Hence the saddle-point equations (19) simply state that $\langle n_\mu \rangle = p_\mu / \alpha$. Furthermore, one easily confirms that equation (21), together with (19), indeed gives the stationary solution of the macroscopic dynamical laws (11).

6 Two simple models in the limit $n_c \rightarrow \infty$

We now turn to the simplest case, where there is only one type of monomer in the system, such that a local arrangement at a given site is fully specified by the number of monomers present, *i.e.* $\mathbf{n} \rightarrow n$. Note also that just one interaction energy $\phi = \phi(\lambda, \lambda)$ will be left in our equations. Model types will thus only differ in terms of the interpretation given to the possible values of ϕ , *e.g.* for the RHM one has $\phi = \lambda$ ($\phi > 0$ for hydrophilic monomers, $\phi < 0$ for hydrophobic monomers), whereas for the RHM one has $\phi = \epsilon \lambda^2$ ($\phi > 0$ for mutually repelling monomers, $\phi < 0$ for mutually attracting monomers). In this section we will for simplicity use the RHM terminology, and refer to $\phi > 0$ and $\phi < 0$ as hydrophilic and hydrophobic systems, respectively. Equation (21) now becomes

$$c_n = \frac{\frac{1}{n!} e^{ny - \frac{1}{2}\beta \phi n^2}}{\sum_{m=0}^{n_c} \frac{1}{m!} e^{my - \frac{1}{2}\beta \phi m^2}} \quad (22)$$

where c_n gives the fraction of sites with n monomers, and where y is to be solved from the non-linear equation (19), which reduces to

$$\frac{1}{\alpha} = \sum_{n=0}^{n_c} n c_n. \quad (23)$$

This equation can also be written as

$$\frac{1}{\alpha} = \frac{\partial}{\partial y} \log Z(y) \quad Z(y) = \sum_{n=0}^{n_c} \frac{1}{n!} e^{ny - \frac{1}{2}\beta \phi n^2}. \quad (24)$$

The associated value of the free energy per monomer (18) is given by

$$-\beta f = \alpha \log Z(y) - y. \quad (25)$$

We will next study the solution y of (22, 23) and the monomer density variance $\sigma^2 = \langle n^2 \rangle - \langle n \rangle^2 = \partial^2 \log Z / \partial y^2$ in the different temperature regimes, for large n_c . Note that from $\partial^2 \log Z / \partial y^2 = \sigma^2 \geq 0$ it also follows that the solution of (24), and thus also of (22, 23), is unique.

In the high-temperature region, both cases (hydrophilic and hydrophobic) can be treated on the same footing. Expansion of (24) in powers of β yields

$$Z(y) = \left[1 - \frac{1}{2} \beta \phi \frac{\partial^2}{\partial y^2} + \mathcal{O}(\beta^2) \right] \sum_{n=0}^{n_c} \frac{e^{ny}}{n!}. \quad (26)$$

We take the limit $n_c \rightarrow \infty$ and insert the result into (24), which leads to the following quadratic equation for $x \equiv e^y$:

$$\alpha^{-1} = x - \frac{1}{2} \beta \phi (x + 2x^2) + \mathcal{O}(\beta^2). \quad (27)$$

The solution gives

$$y = -\log \alpha + \frac{1}{2} \beta \phi \left(1 + \frac{2}{\alpha} \right) + \mathcal{O}(\beta^2) \quad (28)$$

such that

$$\sigma^2 = \frac{1}{\alpha} \left[1 - \frac{\phi}{\alpha T} \right] + \mathcal{O}(T^{-2}) \quad (T \rightarrow \infty). \quad (29)$$

Hence, at finite overall density (*i.e.* for finite α) and for $T \rightarrow \infty$, the monomer density variance σ always remains finite (so the system is in a swollen state), due to entropy dominance.

In the low temperature regime, on the other hand, it will be clear that the hydrophilic ($\phi > 0$) and the hydrophobic ($\phi < 0$) models could exhibit qualitatively different behaviour. More specifically, for models with just one monomer type as studied in this section, we can write the Hamiltonian (1) as

$$\begin{aligned} H(\mathbf{r}) &= \frac{1}{2} \phi \sum_{i \neq j} \delta_{r_i, r_j} = \frac{1}{2} \alpha N \phi \left[\frac{1}{R} \sum_x n_x^2(\mathbf{r}) - \frac{1}{\alpha} \right] \\ &= \frac{1}{2} \alpha N \phi \left[\sigma^2 + \frac{1}{\alpha} \left(\frac{1}{\alpha} - 1 \right) \right]. \end{aligned} \quad (30)$$

Thus the ground state of the hydrophilic system is the one with minimal monomer density variance, and that of the hydrophobic system is the one with maximal variance. This allows us to determine the ground states without effort, which provides a welcome test of our general equations (22, 23). Extremisation of σ^2 by variation of the $\{c_n\}$, with two Lagrange parameters for the constraints $\sum_n c_n = 1$ and $\sum_n n c_n = \alpha^{-1}$, shows that an extremum will have at most two n with $c_n > 0$. However, since according to the theory we are not allowed to vary the $\{c_n\}$ at will (they must be of the form (22)), any extremum

thus obtained must be shown to correspond to a specific value of the order parameter y in (22).

We first analyse the ground state of the hydrophilic model, for $n_c \rightarrow \infty$. In order to minimize the energy the system must minimise the sum $\sum_{i < j} \delta_{r_i, r_j}$, so the monomers must spread out maximally over the available sites (*i.e.* the distribution $\{c_n\}$ must be as narrow as possible), given the constraints $\sum_n c_n = 1$ and $\sum_n n c_n = \alpha^{-1}$. We define $n_\alpha \in \{0, 1, 2, \dots\}$ such that $n_\alpha \leq \alpha^{-1} < n_\alpha + 1$. The minimum energy arrangement of the $\{c_n\}$ can now be written as

$$\ell = n_\alpha : \quad c_\ell = 1 - \alpha^{-1} + n_\alpha \quad (31)$$

$$\ell = n_\alpha + 1 : \quad c_\ell = \alpha^{-1} - n_\alpha \quad (32)$$

with $c_\ell = 0$ for all other ℓ . For the variance we find

$$\sigma^2 = (\alpha^{-1} - n_\alpha)(1 + n_\alpha - \alpha^{-1}). \quad (33)$$

The variance remains finite, and is zero when α^{-1} is an integer (in which case a state exists where all sites have equal monomer numbers). The system will be in a swollen state at any temperature; there are no phase transitions, not even for $n_c \rightarrow \infty$. We now show how this state can be extracted from (22, 23) for $\beta \rightarrow \infty$. It will be clear from (23) that a finite α solution requires $y^{-1} = \mathcal{O}(T)$ as $T \rightarrow 0$. Putting $y = \beta \phi x$ converts (23) into

$$\frac{1}{\alpha} = \frac{\sum_{n \geq 0} \frac{n}{n!} e^{-\frac{1}{2} \beta \phi (n-x)^2}}{\sum_{n \geq 0} \frac{1}{n!} e^{-\frac{1}{2} \beta \phi (n-x)^2}}. \quad (34)$$

For $\beta \rightarrow \infty$ the two terms $n = n_x$ and $n = n_x + 1$ will dominate the summations in (34), where $n_x \in \{0, 1, 2, \dots\}$ such that $n_x \leq x < n_x + 1$, and we obtain

$$\frac{1}{\alpha} = \lim_{\beta \rightarrow \infty} \frac{n_x + e^{-\beta \phi [n_x - x + \frac{1}{2}]}}{1 + \frac{1}{n_x + 1} e^{-\beta \phi [n_x - x + \frac{1}{2}]}}.$$

Upon making the ansatz $x = n_\alpha + \frac{1}{2} + \log(z) / \beta \phi$, where $n_\alpha \in \{0, 1, 2, \dots\}$ such that $n_\alpha \leq \alpha^{-1} < n_\alpha + 1$, and subsequently taking the $\beta \rightarrow \infty$ limit, we find:

$$\frac{1}{\alpha} = \frac{[n_\alpha + \lim_{\beta \rightarrow \infty} z] [n_\alpha + 1]}{n_\alpha + 1 + \lim_{\beta \rightarrow \infty} z}$$

with solution

$$\lim_{\beta \rightarrow \infty} z = \frac{[\alpha^{-1} - n_\alpha][1 + n_\alpha]}{1 + n_\alpha - \alpha^{-1}}. \quad (35)$$

This result is always positive and finite, so our ansatz indeed solves the saddle-point equation. We can also calculate the zero temperature limit of the fraction c_n (22) of sites with n particles. Inserting $y = \beta \phi [n_\alpha + \frac{1}{2}] + \log(z)$ into (22) gives

$$c_\ell = \frac{\frac{1}{\ell!} z^\ell e^{-\frac{1}{2} \beta \phi [\ell - \frac{1}{2} - n_\alpha]^2}}{\sum_{m \geq 0} \frac{1}{m!} z^m e^{-\frac{1}{2} \beta \phi [m - \frac{1}{2} - n_\alpha]^2}}. \quad (36)$$

Upon taking the $\beta \rightarrow \infty$ limit, using (35), this reduces to the expressions (31, 32) found earlier by direct variation of the $\{c_n\}$.

In contrast, the hydrophobic model must maximize the variance in order to minimize the energy (see (30)), and for $\phi < 0$ the system will consequently compactify (as we shall see) at any finite temperature, such that only c_{n_c} and c_0 are non-zero:

$$\ell = 0 : c_0 = 1 - (\alpha n_c)^{-1} \quad (37)$$

$$\ell = n_c : c_{n_c} = (\alpha n_c)^{-1} \quad (38)$$

with $c_\ell = 0$ for all other ℓ . The variance then equals, in leading order as $n_c \rightarrow \infty$,

$$\sigma^2 = n_c/\alpha - 1/\alpha^2 \quad (39)$$

and consequently diverges. Apparently, for $n_c \rightarrow \infty$ the equilibrium state of the hydrophobic system is one with just two types of site: those with n_c particles and those with no particles (*i.e.* it is a maximally compact state), for any temperature. We will now study the $n_c \rightarrow \infty$ behaviour of the saddle-point equation (23), for $\phi < 0$ and arbitrary β . We make the ansatz

$$y = -1 + \frac{1}{2}\beta\phi n_c - \frac{1}{n_c} \log(\alpha\sqrt{n_c}) + \log(n_c) + x \quad (40)$$

where $\lim_{n_c \rightarrow \infty} x = 0$. This converts (24) into $\alpha^{-1} = \partial \log \tilde{Z}(x)/\partial x$, with, upon application of Stirling's formula to the term $n = n_c$ in $Z(y)$, and modulo contributions to the exponents which vanish as $n_c \rightarrow \infty$:

$$\begin{aligned} \tilde{Z}(x) = & 1 + \frac{e^{xn_c}}{\alpha n_c \sqrt{2\pi}} e^{\mathcal{O}(\dots)} \\ & + \sum_{n=1}^{n_c-1} \frac{1}{n!} \left[e^{x-1 + \frac{1}{2}\beta\phi|(n-n_c) - \frac{1}{n_c} \log(\alpha\sqrt{n_c}) + \log(n_c)} \right]^n. \end{aligned}$$

One distinguishes two qualitatively different types of terms in the above summation. Firstly, those with $n \ll n_c$ or with $n = \mathcal{O}(n_c)$ (provided $n \ll n_c$) give a vanishing cumulative contribution to $\tilde{Z}(x)$ due to the appearance of $\frac{1}{2}\beta\phi|(n-n_c)$ in the exponent. Secondly, those terms with $n = n_c - k$, with finite $k > 0$, each give according to Stirling's formula and modulo contributions to the exponents which vanish as $n_c \rightarrow \infty$:

$$\begin{aligned} \frac{1}{n!} \left[e^{x-1 + \frac{1}{2}\beta\phi|(n-n_c) - \frac{1}{n_c} \log(\alpha\sqrt{n_c}) + \log(n_c)} \right]^n \\ = \frac{e^{xn_c}}{\alpha n_c \sqrt{2\pi}} e^{-\frac{1}{2}n_c\beta\phi|k - k(x-1 - \frac{1}{2}\beta\phi|k) + \mathcal{O}(\dots)}. \end{aligned}$$

Hence also the second type of terms in $\tilde{Z}(x)$ are exponentially damped as $n_c \rightarrow \infty$, relative to the two surviving terms $n = 0$ and $n = n_c$, and hence

$$\tilde{Z}(x) = 1 + \frac{e^{xn_c}}{\alpha n_c \sqrt{2\pi}} [1 + \mathcal{O}(\dots)]. \quad (41)$$

Thus for $n_c \rightarrow \infty$ the saddle-point equation reduces to

$$\frac{1}{\alpha} = \lim_{n_c \rightarrow \infty} \frac{n_c e^{xn_c}}{\alpha n_c \sqrt{2\pi} + e^{xn_c}}$$

and hence $x = n_c^{-1} \log \sqrt{2\pi} + \dots$ as $n_c \rightarrow \infty$, confirming the correctness of our ansatz. This gives for the fractions c_n (22) the state (37, 38) which resulted earlier from direct maximisation of σ^2 by variation of $\{c_n\}$.

7 The transitions to compact states

We have observed in the simple single-species models of the previous section that in the hydrophilic case there are no phase transitions for $n_c \rightarrow \infty$ (since here the entropic and energetic forces have the same objective: a swollen state), whereas in the hydrophobic case the system will for $n_c \rightarrow \infty$ be found in a maximally compact state at any temperature (*i.e.* there is a phase transition at $T = \infty$). Since the latter behaviour is strongly linked to the $n_c \rightarrow \infty$ limit, one should expect that proper phase transitions in the hydrophobic system can be identified in terms of a suitably re-scaled n_c -dependent temperature (or, equivalently, a suitably re-scaled interaction energy ϕ). It turns out the the proper scaling is obtained upon transforming according to

$$T = \tilde{T} |\phi| n_c^2 / \log(n_c!), \quad (42)$$

(where we have absorbed the interaction energy into the re-scaled temperature), together with $\tilde{\beta} = 1/\tilde{T}$. For $n_c \rightarrow \infty$ we observed that the fraction c_{n_c} of sites with n_c particles vanished, even in the maximally compact state, where $c_{n_c} = (\alpha n_c)^{-1}$. Hence we switch in this section from inspecting the fraction c_n (22) of sites with a given number n of particles, to an alternative representation of the macroscopic system state in terms of the fraction of particles $f_n = \alpha n c_n$ which find themselves at a site filled with n particles (in the maximally compact state one has $f_n = 0$ for all $n < n_c$ and $f_{n_c} = 1$). Note that $f_0 = 0$, by definition. We introduce the short-hand

$$h(n, y) = \frac{\tilde{\beta} n^2 \log(n_c!)}{2n_c^2} - \log(n!) + ny. \quad (43)$$

The saddle-point equation (23), from which to solve y , can now be written as

$$\sum_{n=1}^{n_c} f_n = 1, \quad f_n = \frac{\alpha n e^{h(n, y)}}{\sum_{m=0}^{n_c} e^{h(m, y)}}. \quad (44)$$

Equation (44) looks deceptively simple. However, it turns out that one cannot simply transform the previous solution y (40) of the $\beta = \mathcal{O}(n_c^0)$ regime by writing β in terms of the re-scaled $\tilde{\beta}$, due to the non-commuting of the various operations involved. Numerical solution of (44) indicates that for $n_c \rightarrow \infty$ we can neglect all those terms n in the summations with $1 \ll n < n_c$, relative to $n = n_c$, and that terms with $n \sim \mathcal{O}(1)$ are generally of the same

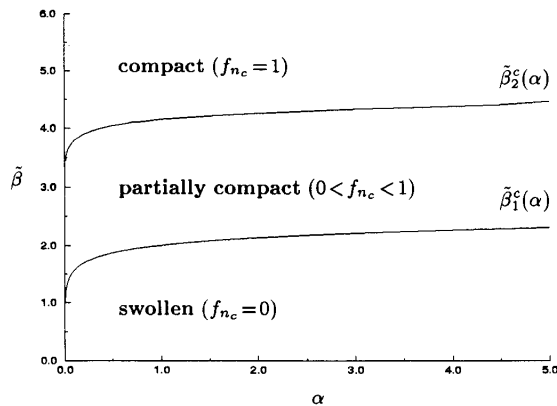


Fig. 4. Phase diagram of the hydrophobic model (with a single monomer species), as obtained by solving the saddle-point equation (44) numerically (with $n_c = 2 \times 10^5$). The re-scaled inverse temperature $\tilde{\beta}$ is defined in (42).

order of magnitude as that of $n = 1$. One can show that this property can be derived, from (44), as an immediate consequence of assuming that the leading orders of f_1 and f_{n_c} scale in the same way with n_c as $n_c \rightarrow \infty$. The latter, in turn, is equivalent to saying $y = (1 - \frac{1}{2}\tilde{\beta}) \log n_c + u$ with u finite for $n_c \rightarrow \infty$, exactly as in (40) (after carrying out the transformation from β to $\tilde{\beta}$). As a result, and in contrast to the scaling regime $\beta = \mathcal{O}(n_c^0)$, even for $n_c \rightarrow \infty$ equation (44) remains transcendental, and it generally resists analytical solution.

A systematic numerical analysis of (44) reveals the following picture. At high temperatures, $\tilde{\beta} = 0$, the system is in a swollen phase, which is defined by the condition $\lim_{n_c \rightarrow \infty} f_{n_c} = 0$ (the fraction of monomers found at a maximally filled site is zero). Upon increasing $\tilde{\beta}$ one then finds a phase transition at $\tilde{\beta}_1^c(\alpha)$ to a partially compact state, where $0 < \lim_{n_c \rightarrow \infty} f_{n_c} < 1$. A further increase of $\tilde{\beta}$ leads to a growth of the fraction of monomers in maximally filled sites, until at a second transition temperature $\tilde{\beta}_2^c(\alpha)$ the system enters the fully compact phase, where $\lim_{n_c \rightarrow \infty} f_{n_c} = 1$. For the special case $\tilde{\beta} = 2$ it turns out that (44) can be solved analytically: for $\alpha > 1$ one finds in leading order $y = -\log \alpha$ and $\lim_{n_c \rightarrow \infty} f_{n_c} = 0$, for $\alpha < 1$ one has $y = \mathcal{O}(n_c^{-1} \log n_c)$ and $0 < \lim_{n_c \rightarrow \infty} f_{n_c} < 1$, which confirms the picture obtained by numerical solution (see below).

The resulting phase diagram, obtained by extracting the two transition temperatures $\tilde{\beta}_{1,2}^c(\alpha)$ from a numerical solution of (44), is drawn in Figure 4, and is confirmed qualitatively by numerical simulations (see the next section). Note that even for the relatively high value $n_c = 2 \times 10^5$ as used in the numerical evaluation, the line $\tilde{\beta}_2^c(\alpha)$ can only be qualitatively correct, since in order to probe truly the $n_c \rightarrow \infty$ regime we need $\log(n_c) \rightarrow \infty$, ac-

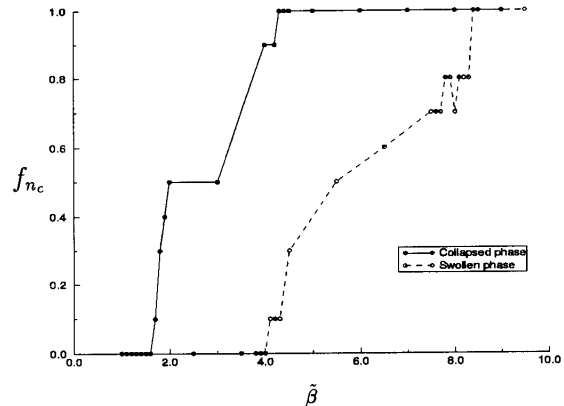


Fig. 5. Fraction f_{n_c} of monomers in a maximally filled site, measured in numerical simulations of the single-species hydrophobic system ($\phi < 0$), as a function of the re-scaled inverse temperature $\tilde{\beta}$. Simulation parameters: $N = R = 1000$ (so $\alpha = 1$), $n_c = 100$, and relaxation time $t = 10^6$ iterations/monomer. Initial conditions: either the compact state (solid), or the swollen state (dashed).

cording to the saddle-point equation. Note, furthermore, that in the limit $\alpha \rightarrow 0$ the system must be fully compact, so that both $\tilde{\beta}_1^c$ and $\tilde{\beta}_2^c$ must tend to 0. Finally, for $\alpha \gg 1$ our numerical solution shows that both transition temperatures grow as $\tilde{\beta}_{1,2}^c(\alpha) \sim \log(\alpha)$.

8 Numerical simulations for high n_c

We have carried out numerical simulations to verify the theoretical equilibrium results of the last section, and the kinetics of the approach to equilibrium in the $n_c \rightarrow \infty$ regime; since phase transitions only occur for $\phi < 0$ (the hydrophobic system) we have restricted our simulations accordingly. Due to often excessive equilibration times observed for large n_c (see below), it has been necessary to limit our experiments to $n_c = 100$, $N = 1000$. Hence, strong finite size effects could be anticipated, since the theory requires $N \rightarrow \infty$, $n_c \rightarrow \infty$ and $n_c/N \rightarrow 0$. Nevertheless, the numerical simulations were found to agree remarkably well with the theory.

Figure 5 shows the fraction f_{n_c} of monomers in a maximally filled site, as a function of the re-scaled inverse temperature $\tilde{\beta}$, after 10^6 updates per monomer. The solid curve shows the observed value following initialisation of the system in a maximally compact (collapsed) state; the dashed curve shows the observed f_{n_c} following initialisation in the fully swollen state. The hysteresis-type graph which results from these measurements (the dependence on initialisation) reflects extremely slow equilibration of the system towards more compact states, due to entropic barriers (similar to those in *e.g.* [12]). The full curve (which describes a process of de-compactification)

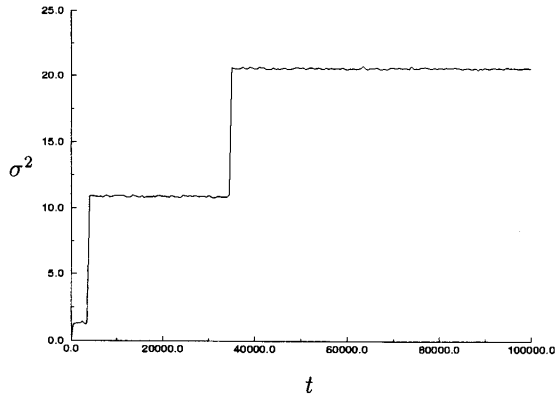


Fig. 6. Variance $\sigma^2 = \langle n^2 \rangle - \langle n \rangle^2$ (proportional to the energy) as a function of time, measured in numerical simulations of a hydrophobic system as in the previous figure, at $\beta = 3.8$, following fully swollen initial conditions.

does not suffer from this problem, and therefore gives the best experimental estimate for the true locations of the transitions. This latter curve predicts (for $\alpha = 1.0$) that $\tilde{\beta}_1^c(1) \simeq 1.7$ and $\tilde{\beta}_2^c(1) \simeq 4.1$, to be compared to the theoretical predictions $\beta_1^c(1) \simeq 2.0$ and $\beta_2^c \simeq 4.2$ (see Fig. 4). Given the anticipated finite size effects, these values are remarkably close. The non-monotonicity of some of the experimental curves are surely finite n_c effects; for finite n_c a maximally filled site can indeed temporarily lose one or more monomers, whereas the probability for this to happen vanishes for $n_c \rightarrow \infty$.

We now turn to the entropy barriers which cause the slow equilibration towards compact states. At sufficiently low temperatures, starting from a swollen (*i.e.* high temperature) state, monomers initially hop around until at some site a critical (temperature dependent) monomer number is obtained. After this has happened, this site fills up very rapidly, since its monomers can no longer escape (for energetic reasons). All monomers in fully filled sites are thus more or less immobile, and the remaining (mobile) monomers find themselves at a reduced effective density $\rho_{\text{eff}} = (1 - f_{n_c})/(\alpha - c_{n_c})$. Hence, as f_{n_c} increases, for the mobile monomers it will become increasingly harder to ‘find’ other monomers to form a new site with a critical monomer number. This entropic effect is illustrated by the simulation shown in in Figure 6. It is characterised by sudden jumps in the variance, corresponding to a site reaching the critical monomer number and filling up rapidly, with increasingly long plateaus in between, where the remainder of mobile monomers hop around at a reduced effective density. The actual process of the rapid filling of those sites which a some stage reach the critical monomer number, is illustrated in Figure 7. It shows, as snapshots of numerical simulations at three intermediate (relatively close) times, the distribution $\{f_n\}$ for a hydrophobic system as in the previous experiments which has been quenched

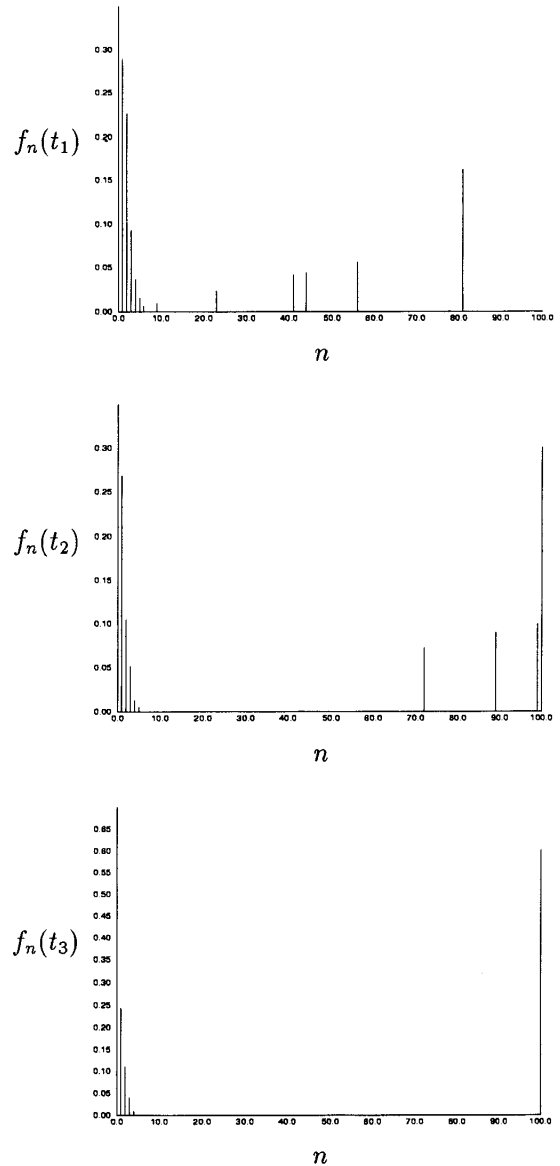


Fig. 7. The occupation fractions f_n as observed in a single numerical simulation at times $t_1 = 10^2$, $t_2 = 2 \times 10^2$, $t_3 = 10^4$, for a hydrophobic system with $N = R = 1000$, $n_c = 100$, and $\tilde{\beta} = 8.0$. It illustrates the formation, within a relatively small number of iteration steps, of a maximally filled site.

from a maximally swollen state to a lower temperature, well within the fully compact phase. Since the equilibrium state in that phase corresponds to $f_{n_c} = 1$, even at the largest observation time shown, $t_3 = 10^4$, the system clearly has not yet equilibrated.

Another interesting aspect of the equilibration dynamics is the fact that compactification is found to take

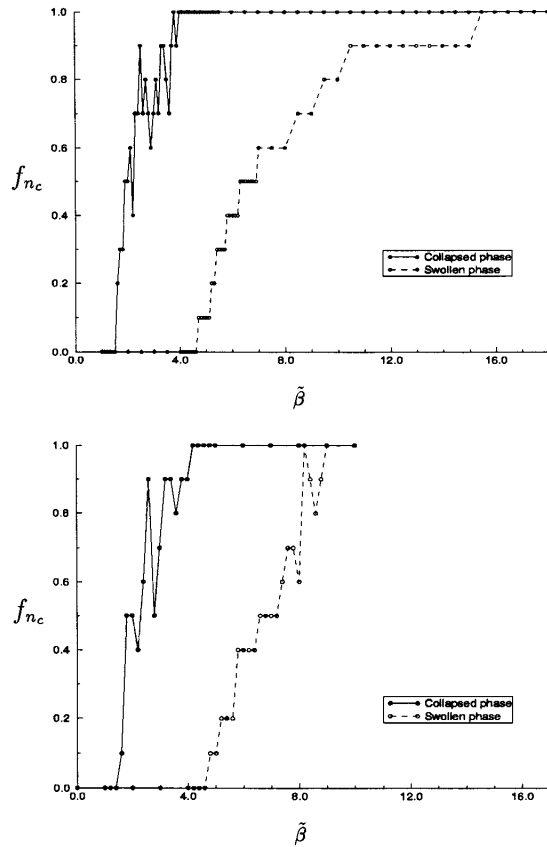


Fig. 8. Occupation fractions f_{n_c} versus the re-scaled inverse temperature $\tilde{\beta}$, as measured at $t = 1 \times 10^5$ in simulations of a hydrophobic system with $N = R = 1000$, $n_c = 100$. Upper graph: the system is not re-initialized when starting a new simulation at increased $\tilde{\beta}$. Lower graph: for each $\tilde{\beta}$ the system is re-initialized to a maximally compact or swollen state.

longer for a system which is not re-initialised at each new relaxation temperature, compared to a system that is. This is illustrated in Figure 8. Apparently, in a quench (re-initialisation to a swollen state), several sites reach the critical monomer density simultaneously, while the effective density is still reasonably high. Note that at short times one can indeed observe several peaks in Figure 7. On the other hand, in a system which is not re-initialised, the effective monomer density is already reduced when the temperature is changed and a new relaxation commences. As a result, the entropic barrier to be overcome before a critical monomer density at any site is achieved, is larger in the system which is not re-initialised than in the re-initialised system, and hysteresis effects are therefore stronger. Non-monotonicity is, again, a finite n_c effect.

9 Discussion

In this paper we have studied the dynamics and the compactification phase transitions of a class of simple lattice gas models for a random heteropolymer in infinite dimensions, as introduced in [1]. Within this class we have restricted ourselves to models with site disorder, in particular to the Random Hydrophobic Model (RHM). For finite n_c (the maximum number allowed at a given site), and a finite number of monomer species, we have been able to solve the dynamics exactly, by deriving exact closed macroscopic evolution equations for a suitable set of dynamic order parameters. The solution of the (non-linear) macroscopic equations, describing the approach to equilibrium, is shown to be in excellent agreement with numerical simulations.

In order to study the physics of the model also for $n_c \rightarrow \infty$ (a limit which could not easily be taken in the equilibrium solution proposed in [1]) we have presented an alternative (but equivalent) equilibrium solution. In the latter approach the limit $n_c \rightarrow \infty$ poses no problems (although the method would not apply to models with bond disorder, in contrast to that of [1]). Working out our equations for models which just a single monomer type shows that for $n_c \rightarrow \infty$ a hydrophilic model will not undergo phase transitions at any temperature (as expected), but that the hydrophobic system will be in a collapsed state, at any finite temperature. In the latter case the transition(s) apparently occur at $T = \infty$, as a consequence of having taken the $n_c \rightarrow \infty$ limit. In terms of a suitably re-scaled inverse temperature $\tilde{\beta}$, however, two transitions are found to occur, at $\tilde{\beta}_1^c$ (from a swollen to a partially compact state) and at $\tilde{\beta}_2^c$ (from a partially compact to a fully compact state). Both the existence and location of these two transitions are confirmed by numerical simulations, within the experimental limits imposed by having finite N and finite n_c (the theory requires $N \rightarrow \infty$, $n_c \rightarrow \infty$ with $n_c/N \rightarrow 0$). Especially when following a swollen (*i.e.* high temperature) initial state, equilibration is observed to take place *via* an extremely slow plateau-like dynamics, which is caused by entropic effects, particularly at the critical points and in the partially compact phase.

A natural extension of the work reported in this paper would be to solve the dynamics of the model for the case of bond disorder. In statics such models have been solved using replica theory, see [1, 3–6, 8, 9], and it is to be anticipated that the method of path integrals [13] would have to be used for solving the dynamics, as in *e.g.* [7].

Infinite dimensional models are obviously less realistic than finite dimensional ones in terms of the spatial properties of polymer folding processes. However, our present study shows that, due to the absence of any need for approximations (in contrast to their finite dimensional counterparts), they do allow for a fairly complete and fully exact quantitative understanding of the energetic and entropic interplay between hydrophobic and hydrophilic effects, including both statics and dynamics.

References

1. J. van Mourik, *Eur. Phys. J. E* **2**, 75 (2000).
2. M. Doi, S.F. Edwards, *The Theory of Polymer Dynamics* (Oxford University Press, Oxford, 1996).
3. T. Garel, H. Orland, D. Thirumalai, in *New Developments in Theoretical Studies of Proteins*, edited by R. Elber (World Scientific, Singapore, 1996).
4. T. Garel, H. Orland, E. Pitard, in *Spin Glasses and Random Fields*, edited by A.P. Young (World Scientific, Singapore, 1997).
5. S.P. Obukhov, *J. Phys. A* **19**, 3655 (1986).
6. T. Garel, L. Leibler, H. Orland, *J. Phys. II France* **4**, 2139 (1994).
7. D. Thirumalai, V. Ashwin, J.K. Bhattacharjee, *Phys. Rev. Lett.* **77**, 5385 (1996).
8. T. Garel, H. Orland, *Europhys. Lett.* **6**, 597 (1988).
9. C.D. Stafos, A.M. Gutin, E.I. Shakhnovich, *Phys. Rev. E* **48**, 465 (1993).
10. D. Bedeaux, K. Lakatos-Lindenberg, K. Shuler, *J. Math. Phys.* **12**, 2116 (1971).
11. N.G. Van Kampen, *Stochastic processes in Physics and Chemistry* (North-Holland, Amsterdam, 1997).
12. F. Ritort, *Phys. Rev. Lett.* **75**, 1190 (1995).
13. C. De Dominicis, *Phys. Rev. B* **18**, 4913 (1978).