

Adsorption of a random heteropolymer at a potential well revisited: location of transition point and design of sequences

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

2001 J. Phys. A: Math. Gen. 34 5625

(<http://iopscience.iop.org/0305-4470/34/28/301>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 171.66.16.97

The article was downloaded on 02/06/2010 at 09:08

Please note that [terms and conditions apply](#).

Adsorption of a random heteropolymer at a potential well revisited: location of transition point and design of sequences

Alexei Naidenov¹ and Sergei Nechaev^{1,2}

¹ L D Landau Institute for Theoretical Physics, 117940, Moscow, Russia

² UMR 8626, CNRS–Université Paris XI, LPTMS, Bat.100, Université Paris Sud, 91405 Orsay Cedex, France

Received 11 January 2001, in final form 26 April 2001

Published 6 July 2001

Online at stacks.iop.org/JPhysA/34/5625

Abstract

The adsorption of an ideal heteropolymer loop at a potential point well is investigated within the framework of standard random matrix theory. On the basis of a semi-analytical/semi-numerical approach a histogram of transition points for the ensemble of quenched heteropolymer structures with bimodal symmetric distribution of types of chain links is constructed. It is shown that the sequences having transition points in the tail of the histogram display correlations between nearest-neighbour monomers.

PACS numbers: 0550, 0510C, 6460

1. Introduction

The problem of adsorption of an ideal heteropolymer chain with random quenched chemical structure (i.e. sequence of links) is far from being new and is well represented in the literature.

The simple diffusive approach [1–4] provides us with complete understanding of the homopolymer and block-copolymers adsorption in various geometries. More advanced renormalization group (RG) methods [5, 6] and power series analysis [7] applied to random chains with a disordered sequence of links delivers important information about the thermodynamic properties of ideal polymers near the transition point from delocalized (Gaussian) to localized (adsorbed) regimes. A mathematical formalism based on perturbation theory gives reasonable results for a phase transition in solid-on-solid (SOS) models with quenched impurities [8]. Many conclusions of [8] correlate with those obtained by RG analysis in [6]. Problems dealing with localization of disordered polymers at selective interfaces [9–11] should be also mentioned in the context of discussed problems.

However, despite almost 20 years of intent attention, some important questions of heteropolymer adsorption still remain open. One of them, the most intriguing, in our view,

concerns the location of a point of a phase transition from coil to localized (adsorbed at a potential well) state.

In this paper we consider the model of an ideal, self-intersecting *ring* polymer chain consisting of two types of links, ‘black’ and ‘white’, organized in different chemical structures having different energies in the single point-like potential well. We distinguish between the following models of chemical structures:

- (1) ‘*Canonically quenched*’ sequence positions of ‘black’ and ‘white’ links are uniquely fixed for a given chain structure and cannot exchange in the course of chain fluctuations; the chemical structure is prepared from ‘black’ and ‘white’ links at random with prescribed probabilities.
- (2) ‘*Microcanonically quenched*’ sequence positions of ‘black’ and ‘white’ links are again uniquely fixed for a given chain structure and satisfy an extra constraint: the total number of ‘black’ and ‘white’ links per chain is conserved for all realizations of chain structures.

Our consideration is semi-analytical/semi-numerical. Namely, the closure of an ideal N -link chain in a ring enables us to integrate from the very beginning over all space degrees of freedom for quenched sequences of links and explicitly rewrite the Green function of a chain as a quotient of determinants of two $N \times N$ -matrices with random coefficients on the main diagonals. The transition point corresponds to the situation when the denominator of the quotient becomes zero. This last question for large N and different models of chemical structures is analysed numerically.

The results are displayed in the form of histograms showing how many sequences have the transition points at a given energy. Choosing the sequences which have the transition points in the tail of the histogram, we found that these sequences have effective correlations between nearest-neighbour monomers for model (2) and do not exhibit any correlations for model (1).

Despite the model under consideration seeming rather abstract at first sight, it, as well as some of its modifications, have been successfully exploited for description of various types of coil-to-globule phase transitions in synthetic and biological random heteropolymers. Among these are polymer adsorption at impenetrable substrates and interfaces between solvents of different qualities. For a review of biophysical aspects of related statistical problems see [12, 13].

2. The model

Consider an N -link ideal (i.e. self-intersecting) Gaussian polymer chain in d -dimensional space. Suppose that the chain forms a ring and attach the first and the last chain’s segments to the origin. All segments are labelled ‘black’ or ‘white’, designating different interaction energies with the potential well located at the origin. There is no interaction of chain links between themselves or between a link and any point in d -dimensional space distinct from the origin. Increasing the interaction energy, we can provoke the transition from extended (coil) to localized (adsorbed) state of the macromolecule. The transition is governed by the interplay between entropic forces tending to expel the chain and contact potential interactions which gain the energy in a compact polymer configuration.

The above can be easily rewritten in a formal way. Let us introduce the joint distribution function $P_N(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$ having probability that the links with numbers $1, 2, \dots, N$ of the N -step chain are located at the points $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ in d -dimensional space. The Markov structure of our polymer chain ensures the presence of correlation only between the neighbouring segments of length a and allows us to set

$$P_N(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) = g(\mathbf{x}_1, \mathbf{x}_2)g(\mathbf{x}_2, \mathbf{x}_3) \dots g(\mathbf{x}_{N-1}, \mathbf{x}_N)$$

where the transition probability $g(\mathbf{x}_1, \mathbf{x}_2)$ satisfies the normalization condition $\int g(\mathbf{x}, \mathbf{x}') d\mathbf{x} = 1$ and for simplicity is supposed to be Gaussian:

$$g(\mathbf{x}, \mathbf{x}') \equiv g(\mathbf{x} - \mathbf{x}') = \left(\frac{d}{2\pi a^2} \right)^{d/2} e^{-\frac{d|\mathbf{x}-\mathbf{x}'|^2}{2a^2}}. \quad (1)$$

The Green function $G(\mathbf{x}_N, N) \equiv \int G_N(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) d\mathbf{x}_1 d\mathbf{x}_2 \dots d\mathbf{x}_{N-1}$ is defined as follows:

$$G(\mathbf{x}_N, N) = \int \prod_{i=1}^N [g(\mathbf{x}_{i-1}, \mathbf{x}_i) e^{-\phi(\mathbf{x}_i)/T}] d^d x_1 \dots d^d x_{N-1} \quad (2)$$

where $e^{-\phi(\mathbf{x}_i)/T}$ is the Boltzmann weight of the i th chain segment in the potential well at the origin (compare to [6]):

$$e^{-\phi(\mathbf{x}_i)/T} = 1 + \beta_i \delta(\mathbf{x}_i) \quad \beta_i = \begin{cases} e^{-u_b} - 1 & \text{if the segment } i \text{ is 'black'} \\ e^{-u_w} - 1 & \text{if the segment } i \text{ is 'white'} \end{cases} \quad (3)$$

and $u_{b,w}$ is the dimensionless energy (everywhere below we set $T = 1$).

Now we can write the recursion equation for the Green function (2), expressing $G(\mathbf{x}_{i+1}, i+1)$ in terms of $G(\mathbf{x}_i, i)$ and corresponding Boltzmann weights for $0 \leq i \leq N$. Performing Fourier transform and taking into account (3) we arrive at the following integral equation in the momentum space:

$$G(\mathbf{k}, i+1) = g(\mathbf{k})G(\mathbf{k}, i) + (2\pi)^{-d} \beta_{i+1} \int g(\mathbf{k}') G(\mathbf{k}', i) d^d k' \quad (4)$$

where

$$g(\mathbf{k}) = e^{-|\mathbf{k}|^2 a^2 / 2d}. \quad (5)$$

In order to make the system of equations (4) closed and nonuniform we should take into account the initial and boundary conditions. Let us point out that for open chains the precise form of initial and boundary conditions does not affect in the thermodynamic limit the phase transition point (see, e.g., [14]). We expect the same also for closed chains.

Suppose that the last segment (with $i = N$) is linked to the origin by some sufficiently weak dimensionless potential f . Without the loss of generality we may choose f in the simplest form: $f(\mathbf{k}) \equiv f = \text{const}$. The presence of the potential f increases the probability $\tilde{G}(\mathbf{k}, N)$ to find the N chain link near the origin. If $|f| \ll 1$ we may set in the linear approximation

$$\tilde{G}(\mathbf{k}, i = N) = G(\mathbf{k}, i = N) + f. \quad (6)$$

Using the fact that the chain is closed, we complete equation (4) with the following one:

$$G(\mathbf{k}, i = 1) = g(\mathbf{k})\tilde{G}(\mathbf{k}, i = N) + (2\pi)^{-d} \beta_1 \int g(\mathbf{k}') \tilde{G}(\mathbf{k}', i = N) d^d k' \quad (7)$$

(as we shall see later, f does not enter the final expression for the transition point). Roughly speaking, the closure condition is considered as an extra constraint (the Dirac δ -function in \mathbf{x} -space) which produces effective potential keeping the last N segment at the origin.

The system of N equations (4)–(7) is closed. This enables us to write it as a single *matrix-integral* equation for the N -component ‘spinor’ Green function $\mathbf{G}(\mathbf{k}) = \{G(\mathbf{k}, 1), G(\mathbf{k}, 2), \dots, G(\mathbf{k}, N-1), G(\mathbf{k}, N)\}$. Hence, we get

$$\mathbf{G}(\mathbf{k}) = \hat{M} g(\mathbf{k}) \mathbf{G}(\mathbf{k}) + (2\pi)^{-d} \hat{\beta} \hat{M} \int g(\mathbf{k}') \mathbf{G}(\mathbf{k}') d^d k' + \mathbf{F}(\mathbf{k}) \quad (8)$$

with

$$\hat{M} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \end{pmatrix} \quad \hat{\beta} = \begin{pmatrix} \beta_1 & 0 & 0 & \cdots & 0 \\ 0 & \beta_2 & 0 & \cdots & 0 \\ 0 & 0 & \beta_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \beta_N \end{pmatrix} \quad (9)$$

and

$$\mathbf{F}(\mathbf{k}) = \left((g(\mathbf{k}) + (d/2\pi a^2)^{\frac{d}{2}})f, 0, \dots, 0 \right)^\top \quad (10)$$

where ‘ \top ’ denotes transposition.

Let us rearrange the terms in (8) as follows:

$$\mathbf{G}(\mathbf{k}) = (\hat{I} - g(\mathbf{k})\hat{M})^{-1} \left[(2\pi)^{-d} \hat{\beta} \hat{M} \int g(\mathbf{k}') \mathbf{G}(\mathbf{k}') d^d k' + \mathbf{F}(\mathbf{k}) \right] \quad (11)$$

where \hat{I} is $N \times N$ identity matrix. Integrating the lhs and rhs of (11) with the weight $g(\mathbf{k})$ over all \mathbf{k} , we arrive at the *algebraic matrix* equation for the function $\mathbf{A} = \{A_1, A_2, \dots, A_N\}$, where $A_i = \int g(\mathbf{k}) G(\mathbf{k}, i) d^d k$:

$$\mathbf{A} = (2\pi)^{-d} \left[\int g(\mathbf{k}) (\hat{I} - g(\mathbf{k})\hat{M})^{-1} \hat{\beta} \hat{M} d^d k \right] \mathbf{A} + \Phi. \quad (12)$$

Straightforward computations show that the $N \times N$ matrix $\hat{C} = (\hat{I} - g(\mathbf{k})\hat{M})^{-1} \hat{\beta} \hat{M}$ has the form

$$\hat{C} = \begin{pmatrix} \beta_2 c_1 & \beta_3 c_2 & \beta_4 c_3 & \cdots & \beta_1 c_N \\ \beta_2 c_N & \beta_3 c_1 & \beta_4 c_2 & \cdots & \beta_1 c_{N-1} \\ \beta_2 c_{N-1} & \beta_3 c_N & \beta_4 c_1 & \cdots & \beta_1 c_{N-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \beta_2 c_2 & \beta_3 c_3 & \beta_4 c_4 & \cdots & \beta_1 c_1 \end{pmatrix} \quad (13)$$

with the coefficients

$$c_i \equiv c_i(\mathbf{k}) = -\frac{g^{N-i}(\mathbf{k})}{g^N(\mathbf{k}) - 1} \quad i \in [1, N] \quad (14)$$

and the vector $\Phi = (\Phi_1, \Phi_2, \dots, \Phi_N)^\top$ has components

$$\Phi_i = \frac{f}{(2\pi)^d} \int (g(\mathbf{k}) + (d/2\pi a^2)^{\frac{d}{2}}) c_{N-i+1}(\mathbf{k}) d^d k. \quad (15)$$

Hence the vector \mathbf{A} obeys the system of equations

$$\begin{aligned} A_1 &= \beta_2 a_1 A_1 + \beta_3 a_2 A_2 + \cdots + \beta_N a_{N-1} A_{N-1} + \beta_1 a_N A_N + F_1 \\ A_2 &= \beta_2 a_N A_1 + \beta_3 a_1 A_2 + \cdots + \beta_N a_{N-2} A_{N-1} + \beta_1 a_{N-1} A_N + F_2 \\ A_3 &= \beta_2 a_{N-1} A_1 + \beta_3 a_N A_2 + \cdots + \beta_N a_{N-3} A_{N-1} + \beta_1 a_{N-2} A_N + F_3 \\ &\dots \\ A_{N-1} &= \beta_2 a_3 A_1 + \beta_3 a_4 A_2 + \cdots + \beta_N a_1 A_{N-1} + \beta_1 a_2 A_N + F_{N-1} \\ A_N &= \beta_2 a_2 A_1 + \beta_3 a_3 A_2 + \cdots + \beta_N a_N A_{N-1} + \beta_1 a_1 A_N + F_N \end{aligned} \quad (16)$$

where the coefficients a_i for $1 \leq i \leq N$ are defined as follows:

$$a_i = (2\pi)^{-d} \int g(\mathbf{k}) c_i(\mathbf{k}) d^d k = \left(\frac{d}{2\pi a^2 N} \right)^{d/2} \zeta \left(\frac{d}{2}, \frac{N+1-i}{N} \right). \quad (17)$$

Solving the corresponding system of linear equations via the standard Kramers method, we arrive at the following expression for the values $\beta_{i+1}A_i$:

$$\beta_{i+1}A_i = \frac{\det \hat{A}'(i)}{\det \hat{A}} \quad i \in [1, N] \quad (\text{by definition } \beta_{N+1} = \beta_1) \quad (18)$$

where the $N \times N$ matrix \hat{A} reads as

$$\hat{A} = \begin{pmatrix} a_1 - \beta_2^{-1} & a_2 & a_3 & \cdots & a_{N-1} & a_N \\ a_N & a_1 - \beta_3^{-1} & a_2 & \cdots & a_{N-2} & a_{N-1} \\ a_{N-1} & a_N & a_1 - \beta_4^{-1} & \cdots & a_{N-3} & a_{N-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ a_3 & a_4 & a_5 & \cdots & a_1 - \beta_N^{-1} & a_2 \\ a_2 & a_3 & a_4 & \cdots & a_N & a_1 - \beta_1^{-1} \end{pmatrix} \quad (19)$$

and the $N \times N$ matrix $\hat{A}'(i)$ is obtained from the matrix \hat{A} by replacing the i th column by the vector $-\Phi$.

Equations (18), (19) enable us to rewrite N linearly independent components of the ‘spinor’ $\mathbf{G}(\mathbf{k}) = \{G_1(\mathbf{k}), G_2(\mathbf{k}), \dots, G_N(\mathbf{k})\}$ (see (11)) in a compact form:

$$G_i(\mathbf{k}) = (2\pi)^{-d} [\det \hat{A}]^{-1} \sum_{j=1}^N c_{j-i+1}(\mathbf{k}) \det \hat{A}'(j) + (\hat{I} - g(\mathbf{k})\hat{M})^{-1} \mathbf{F}(\mathbf{k}) \quad (20)$$

where $c_{j-i+1}(\mathbf{k}) = c_{N-(j-i+1)}(\mathbf{k})$ if $j - i + 1 \leq 0$.

The transition point from delocalized (coil) to the adsorbed (globule) state of a heteropolymer ring is manifested in the divergence of the Green function $G_i(\mathbf{k})$ (20) for any $i \in [1, N]$. Thus, the transition point is determined by the equation

$$\det \hat{A} = 0 \quad (21)$$

where the matrix \hat{A} is given by (19). Note that equation (21) is very general: it is valid for any type of disorder and number of species (i.e. sorts of the links).

For any finite chain lengths we can define only a transition region which, due to the supposed self-averaging, becomes sharper and sharper as the chain length increases, tending to a single point in the thermodynamic limit $N \rightarrow \infty$.

Before passing to numerical solution of (21) for different randomly generated sequences belonging to canonically and microcanonically quenched chemical structures, let us derive analytically the transition point for the effective homopolymer ring chain. In this case one can pre-average the partition function of the chain over the distribution of ‘black’ and ‘white’ links. Let us take $u_b = u$, $u_w = -u$ and suppose that the number of ‘black’ links is equal to the number of ‘white’ ones. Then

$$\beta_i \equiv \beta_{av} = \frac{e^{-u} + e^u}{2} - 1 = \cosh u - 1 \quad (22)$$

for all $1 \leq i \leq N$.

Recall that for an N -link *open* homopolymer chain attached by one end at the origin in a potential well, the critical value β_{open} of a transition point in d -dimensional space in the limit $N \rightarrow \infty$ reads [4, 6] as

$$\beta_{open}^{-1} = \left(\frac{d}{2\pi a^2} \right)^{d/2} \zeta \left(\frac{d}{2} \right) \Big|_{d=3}^{d=1} \approx 0.86188. \quad (23)$$

The transition point of a *closed* chain can be analytically evaluated in the effective homopolymer case (22). If all β_i are equal, then the matrix \hat{A} (see equation (19)) is so-called *circulant* and its eigenvalues are [15]

$$\lambda_m = \sum_{j=1}^N a_j e^{i2\pi m(j-1)/N} \quad m \in [0, N-1]. \quad (24)$$

Equation (21) is true when β^{-1} equals one of the eigenvalues λ_m . The transition point corresponds to the largest real λ_m (that is the smallest β). Thus, $\beta_{\text{av}}^{-1} = \lambda_0$:

$$\begin{aligned} \beta_{\text{av}}^{-1} &= \sum_{j=1}^N a_j = \left(\frac{d}{2\pi a^2 N} \right)^{d/2} \sum_{i=1}^N \sum_{k=0}^{\infty} \frac{1}{(k+1+(1-i)/N)^{d/2}} \\ &= \left(\frac{d}{2\pi a^2} \right)^{d/2} \sum_{i=0}^{N-1} \sum_{k=1}^{\infty} \frac{1}{(Nk-i)^{d/2}} = \left(\frac{d}{2\pi a^2} \right)^{d/2} \sum_{j=1}^{\infty} \frac{1}{j^{d/2}} \\ &= \left(\frac{d}{2\pi a^2} \right)^{d/2} \zeta \left(\frac{d}{2} \right) = \beta_{\text{open}}^{-1}. \end{aligned} \quad (25)$$

It should be noted that the above equation is true for any given N . Hence, in 3D space for $a = 1$ we have

$$u_{\text{tr}}^{\text{av}} = \text{arccosh}[\beta_{\text{av}} + 1] = \text{arccosh} \left[\left(\frac{3}{2\pi} \right)^{3/2} \zeta \left(\frac{3}{2} \right) + 1 \right] \approx 1.405.$$

Let us pay attention to the fact that the adsorption points of ring and open chains coincide in the thermodynamic limit. This is consistent with the statement on independence of the point of second-order phase transition on the boundary conditions [16]. Moreover, the interesting feature of (25) consists of the fact that the value β_{av} does not depend on N for ring chains.

3. Numerical results for quenched sequences of links

In this section we solve numerically equation (21) for two different types of bimodal distribution of links with $u_b = u$, $u_w = -u$: (1) canonically quenched and (2) microcanonically quenched. Recall that in case (1) ‘white’ and ‘black’ links appear in the sequence independently with equal probability $\frac{1}{2}$, while in case (2) the total number of ‘black’ and ‘white’ links per chain is exactly $\frac{N}{2}$. We have generated ensembles of 10 000 chains of N segments each for both of models (1) and (2). The distribution of transition points $W(u)$ in cases (a) and (b) is depicted in figure 1 for $N = 100$.

The shape of histograms in figures 1(a) and (b) are correspondingly fitted by Gaussian curves

$$W^c(u_{\text{tr}}) = A^c e^{-\frac{(u_{\text{tr}} - \langle u_{\text{tr}}^c \rangle)^2}{2(\sigma^c)^2}} \quad W^m(u_{\text{tr}}) = A^m e^{-\frac{(u_{\text{tr}} - \langle u_{\text{tr}}^m \rangle)^2}{2(\sigma^m)^2}}.$$

For $N = 100$ the parameters $A^{c,m}$, $\langle u_{\text{tr}}^{c,m} \rangle$, $\sigma_{c,m}$ are as follows:

$$\begin{aligned} A^c &= 644.78 & A^m &= 908.49 \\ \langle u_{\text{tr}}^c \rangle &= 1.756 & \langle u_{\text{tr}}^m \rangle &= 1.778 \\ \sigma_c &= 0.116 & \sigma_m &= 0.039. \end{aligned}$$

To predict the numerical values of $\langle u_{\text{tr}}^c(N) \rangle$ and $\langle u_{\text{tr}}^m(N) \rangle$ at $N \rightarrow \infty$ we have studied numerically these dependencies for different N . The results of our simulations are displayed

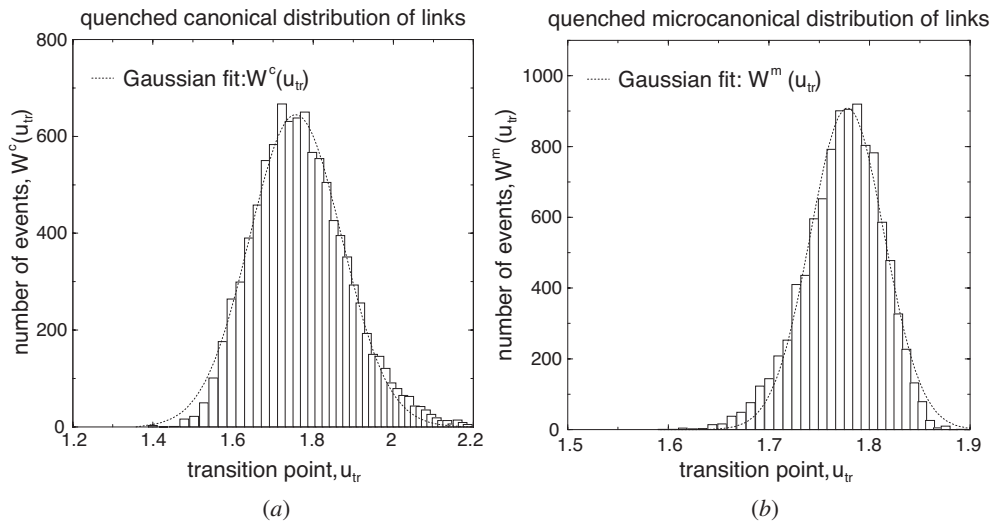


Figure 1. Distribution of transition points for (a) canonical and (b) microcanonical sequences of links.

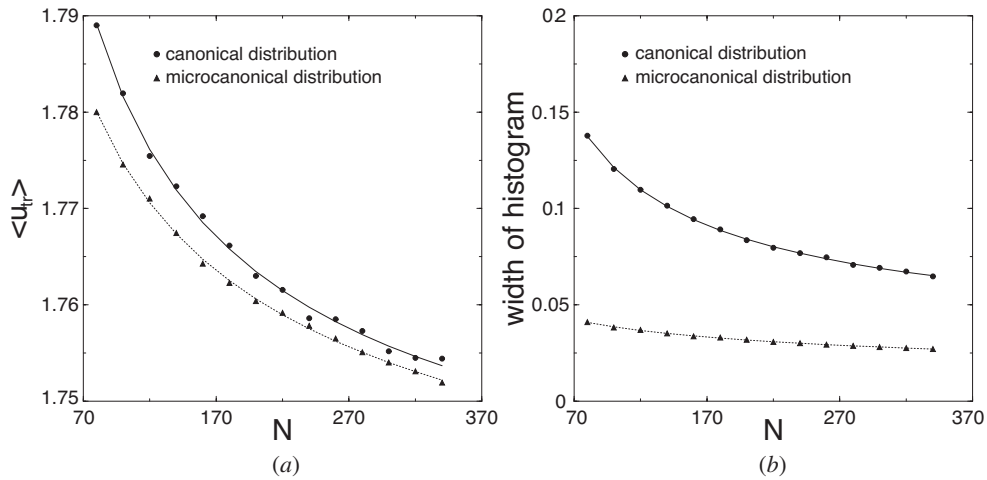


Figure 2. Dependence of (a) mean value of transition point and (b) width of the histogram on chain length.

in figure 2(a), where the chain lengths N have varied in the interval $80 \leq N \leq 400$. The fitting curves are as follows:

$$\langle u_{tr}^c(N) \rangle = u_{\infty}^c + u_1^c N^{-u_2^c} \quad \langle u_{tr}^m(N) \rangle = u_{\infty}^m + u_1^m N^{-u_2^m}$$

where

$$\begin{aligned} u_{\infty}^c &= 1.73 & u_{\infty}^m &= 1.72 \\ u_1^c &= 0.76 & u_1^m &= 0.36 \\ u_2^c &= 0.57 & u_2^m &= 0.39. \end{aligned}$$

The crucial question concerns the behaviour of the width of histograms (see figure 1) in

the thermodynamic limit $N \rightarrow \infty$. We have analysed numerically the behaviour $\sigma^{c,m}(N)$ for chains of length N in the interval $80 \leq N \leq 360$ and found no evident saturation of $\sigma^{c,m}(N)$ at $N \gg 1$. The numerical data and corresponding power-law fits are shown in figure 2(b), where

$$\sigma^c(N) = \sigma_0^c N^{-\alpha^c} \quad \sigma^m(N) = \sigma_0^m N^{-\alpha^m}$$

and the numerical values are as follows:

$$\begin{aligned} \sigma_0^c &= 1.331 & \alpha^c &= 0.52 & \text{for canonical distribution} \\ \sigma_0^m &= 0.143 & \alpha^m &= 0.285 & \text{for microcanonical distribution.} \end{aligned}$$

4. Discussion

4.1. Location of transition points in thermodynamic limit

The results of our numerical simulations of equation (21) permit us to conclude that within the standard error the mean values of phase transition energies $\langle u_{\text{tr}}^c \rangle$ and $\langle u_{\text{tr}}^m \rangle$ in chains with symmetric bimodal canonically and microcanonically quenched sequences tend to the same value $\langle u_{\text{tr}}^c \rangle = \langle u_{\text{tr}}^m \rangle \equiv u_\infty \approx 1.75 \pm 0.02$ in the limit $N \rightarrow \infty$. At the same time the transition point in an effective homopolymer chain defined by equation (22) is $u_{\text{tr}}^{\text{av}} = 1.405$. These numerical results seem to be sufficiently strong to conjecture the difference between the mean value of transition point for ensembles of heteropolymers with quenched sequences of links and the transition point in an effective homopolymer chain. However, one should honestly say that the expressed conjecture deserves more investigation and still cannot be considered as a rigorous statement. The reason we cannot regard it yet as an exact one is as follows.

- (1) We have analysed the distribution of transition points in sufficiently long but finite chains ($N \sim 80\text{--}360$). Denote by $\Delta u_{\text{fin}}(N)$ the transition region for finite N . As we know from the theory of second-order phase transitions [14, 16],

$$\Delta u_{\text{fin}}(N) \sim N^{-1/2} \quad (26)$$

as $N \rightarrow \infty$ due to the *finiteness* of N for homopolymer chains.

- (2) The quenched randomness in types of links leads to an extra uncertainty in the location of a transition point for heteropolymer chains. The transition region $\Delta u_{\text{ran}}(N)$ due to the *randomness* in sequences of links can be estimated by the width of the histograms—see figure 1. So, we may set

$$\Delta u_{\text{ran}}(N) \sim \begin{cases} N^{-\alpha^c} & \text{for canonically quenched sequences} \\ N^{-\alpha^m} & \text{for microcanonically quenched sequences.} \end{cases} \quad (27)$$

The behaviour (27) is consistent with a general self-averaging hypothesis for disordered systems [17] claiming the existence of a unique transition point for *every* disordered sequence in the thermodynamic limit $N \rightarrow \infty$, i.e.

$$\Delta u_{\text{ran}}(N)|_{N \rightarrow \infty} \rightarrow 0.$$

Considering (26) and (27) as independent contributions we arrive at the statement on the difference between u_∞ and $u_{\text{tr}}^{\text{av}}$ in the limit $N \rightarrow \infty$. However one cannot exclude the possible ‘interference’ between (26) and (27) which may lead to the smearing of the transition region. This last possibility has not yet been analysed. Moreover, the asymptotics (27) are based on the interpolation of numerical results for finite N to $N \rightarrow \infty$.

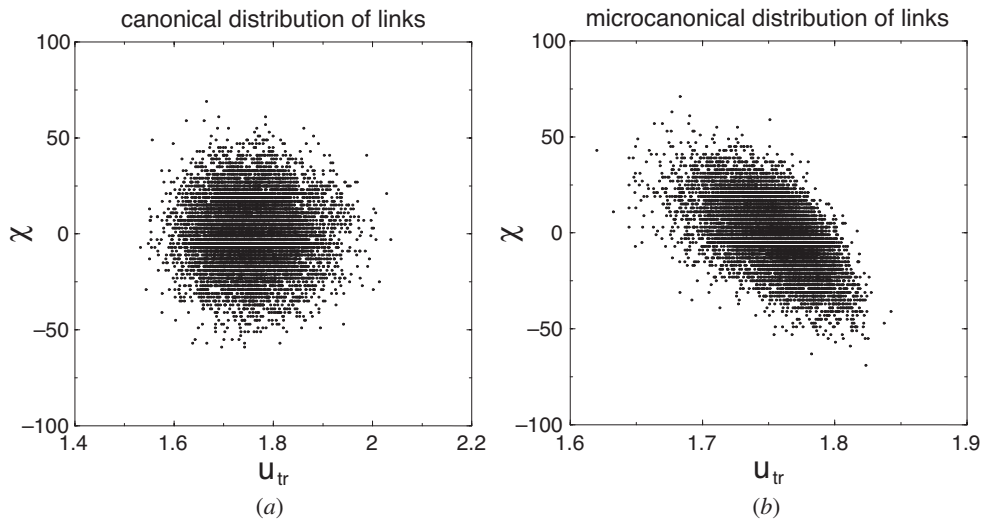


Figure 3. The two-dimensional set $(u, \chi(u))$ for (a) canonically and (b) microcanonically quenched sequences; $N = 300$.

4.2. Design of sequences

The data of our numerical simulations stored in the course of construction of the histograms (see figure 1) allow us to address the following problem. Let us define the correlation function χ for each ring chain

$$\chi = \frac{1}{N} \sum_{j=1}^{N-1} \sigma_j \sigma_{j+1} \quad (28)$$

where the sum runs along all the chain's segments $j = [1, \dots, N]$ and

$$\sigma_j = \begin{cases} +1 & \text{for 'black' segment} \\ -1 & \text{for 'white' segment.} \end{cases}$$

Knowing the transition point (the solution of equation (21)) for every one of 10 000 sequences (for both models (1) and (2)), let us assign to each sequence the point in the coordinates $(u, \chi(u))$ on the plane. The corresponding 'spots' for the models (1) and (2) are shown in figures 3(a) and (b) for $N = 300$. These figures lead to the following important conclusions.

- (1) There are no essential visible correlations between nearest-neighbour links in the ensemble of canonically quenched sequences, which is manifested in the fact that the corresponding 'spot' is circular—see figure 3(a).
- (2) There are correlations between nearest-neighbour links in the ensemble of microcanonically quenched sequences, which is manifested in the fact that the corresponding 'spot' is asymmetric. Moreover, figure 3(b) allows us to conclude how the primary structure of the heteropolymer chain is organized: the positive correlations $\langle \chi(u) \rangle > 0$ for $u < \langle u^m \rangle$ show preferable clustering in subsequences of the same colour, while the negative correlations $\langle \chi(u) \rangle < 0$ for $u > \langle u^m \rangle$ show preferable mixing of opposite colours.

Thus, knowing the value of the transition point for the chain of ‘microcanonically quenched ensemble’ we can arrive at an approximate conclusion about the primary sequence of monomers in a given chain. The fact that the width of the histogram shrinks to zero at $N \rightarrow \infty$ means that apparently at $N \rightarrow \infty$ the corresponding ‘spot’ in figure 3(b) becomes more and more symmetric and the nontrivial correlations disappear in the thermodynamic limit. However, every real physical polymer system consists of a finite number of monomers, which means that the above-mentioned effect should, in principle, exist.

The approach proposed in this paper moves the initial problem into the well posed subject of statistical theory of random matrices, allowing us to apply the standard methods of random matrix theory for heteropolymer adsorption.

Acknowledgments

We are grateful to A Grosberg for critical remarks and fruitful discussions and to M Mezard for useful suggestions.

References

- [1] Binder K 1983 *Phase Transitions and Critical Phenomena* vol 10, ed C Domb and J Lebowitz (New York: Academic)
- [2] Birshtein T M 1982 *Vysokomol. Soedin. A* **24** 1828
- [3] Birshtein T M and Borisov O V 1991 *Polymer* **32** 916
- [4] Grosberg A Yu, Izrailev S F and Nechaev S K 1994 *Phys. Rev. E* **50** 1912
- [5] Grosberg A Yu and Shakhnovich E I 1986 *Sov. Phys.–JETP* **64** 493
- [6] Grosberg A Yu and Shakhnovich E I 1986 *Sov. Phys.–JETP* **64** 1284
- [7] Obukhov S P 1987 *Sov. Phys.–JETP* **66** 1125
- [8] Forgas G *et al* 1991 *Phase Transitions and Critical Phenomena* eds C Domb and J Lebowitz (New York: Academic) p 14
- [9] Garel T and Orland H 1988 *Europhys. Lett.* **6** 307
- [10] Nechaev S K and Zhang Y C 1995 *Phys. Rev. Lett.* **74** 1815
- [11] Monthus C, Garel T and Orland H 2000 *Preprint* cond-mat/0004141
- [12] Grosberg A Yu and Shakhnovich E I 1985 *Biophysica (USSR)* **85** 2341
- [13] Garel T, Orland H and Pitard E 1997 Protein folding and heteropolymers *Spin Glasses and Random Fields* (Singapore: World Scientific)
- [14] Grosberg A Yu and Khokhlov A R *Statistical Physics of Macromolecules*
- [15] Mehta M 2000 *Matrix Theory* (EDP Sciences)
- [16] Landau L D and Lifshitz E M *Statistical Physics: Part 1* (Oxford: Pergamon) 3rd edn
- [17] Mézard M, Parisi G and Virasoro M 1987 *Spin Glass Theory and Beyond* (Singapore: World Scientific)