Thermodynamics of a Brownian bridge polymer model in a random environment

1996 J. Phys. A: Math. Gen. 29 1267

(http://iopscience.iop.org/0305-4470/29/6/013)

J. Phys. A: Math. Gen. **29** (1996) 1267–1279. Printed in the UK

Thermodynamics of a Brownian bridge polymer model in a random environment

Servet Martínez†§ and Dimitri Petritis‡‖

† Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas, Departamento de Ingeniería Matemática, Casilla 170-3 Correo 3, Santiago, Chile

‡ Institut de Recherche Mathématique, Université de Rennes I and CNRS URA 305, Campus de Beaulieu, 35042 Rennes Cedex, France

Received 25 September 1995

Abstract. We consider a directed random walk making either 0 or +1 moves and a Brownian bridge, independent of the walk, conditioned to arrive at point b at time T. The Hamiltonian is defined as the sum of the square of the increments of the bridge between the moments of jump of the random walk and is interpreted as an energy function over the bridge configuration; the random walk acts as the random environment. This system provides a soluble model inspired by a closely related but distinct model of proteins. The thermodynamic limit of the quenched specific free energy is shown to exist and to be self-averaging, i.e. it is equal to a trivial—explicitly computed—random variable. An estimate of the asymptotic behaviour of the ground-state energy is also obtained.

1. Definition of the model and main results

Disordered systems are nowadays thoroughly studied as condensed matter models in the presence of impurities. Interesting questions, both mathematically and physically, concern the sample dependence or independence of intensive quantities: the specific free energy in statistical mechanics of spin glasses or the integrated density of states in spectral theory. Closely related models are random walks in random environments; in this context, interesting questions concern the possible modifications of the critical exponents governing the asymptotic behaviour or the need for unusual normalization for the classical limit theorems of probability theory to hold [17].

More recently, the methods developed in the statistical mechanical study of disordered systems have been applied to other systems like long protein or RNA molecules. It is argued that a protein molecule is very much like a random walk with random charges attached at the vertices of the walk; these charges are interacting through local interactions mimicking Lennard–Jones or hydrogen-bond potentials.

The purpose of this paper is to rigorously study a soluble model arising in protein conformation. The model studied here is inspired by a recent article [6] and is very reminiscent, although distinct, of this latter model. Protein conformation has a long history (see [8] for a recent update of the physical theories). On the basis of replica trick heuristics, it is claimed in [7] that the protein undergoes a phase transition between an unfolded state

§ E-mail: smartine@dim.uchile.cl

‖ E-mail: petritis@levy.univ-rennes1.fr

0305-4470/96/061267+13$19.50

and a folded biologically active state. Kantor and Kardar [11] have introduced a model defined on the product space, $\mathcal{W}_N \times \mathcal{Q}_N$, of N-step one-dimensional nearest-neighbour random walks

$$\mathcal{W}_N = \{r : \{1, \dots, N\} \to \mathbb{Z} \text{ with } r_{i+1} - r_i \in \{-1, 1\} \text{ for } 1 \leqslant i \leqslant N-1\}$$

equipped with the uniform probability, with the probability space $(\mathcal{Q}_N, \mathcal{F}, \mathbb{P})$ where a collection $q = (q_i)_{1 \leqslant i \leqslant N}$ of symmetric, independent, and identically distributed random variables are defined. For given (r, q), the Hamiltonian used in [11] is given by

$$H_N^{KK}(r, q) = \sum_{1 \leqslant i < j \leqslant N} q_i q_j \delta_{r_i, r_j} \tag{1}$$

and has the interpretation of the conformational energy of a protein molecule with N units whose geometrical shape is given by r and whose charges over the constituting units are given by q; these charges interact through an ultra-local interaction. In [11], the random variables q are symmetric Bernoulli variables with values in the range $\{-1, 1\}$. This model is quite degenerate due to the discrete character of the q variables. For that reason, in [5], the model having the same Hamiltonian as that of formula (1) but with the variables q being independent Gaussian was studied. Both models are, however, quite intractable and, therefore, in [6] a further simplification was proposed: namely, the variables q are independent Gaussian, the Hamiltonian is again given by the same formula (1), but it is defined on the space $\mathcal{D}_N$ of *directed* nearest-neighbour random walks

$$\mathcal{D}_N = \{r : \{1, \dots, N\} \to \mathbb{N} \text{ with } r_{i+1} - r_i \in \{0, 1\} \text{ for } 1 \leqslant i \leqslant N\}$$

instead of being defined on $\mathcal{W}_N$. This latter simplification allowed the use of transfer matrix techniques which, combined with numerical experiments, gave the asymptotic (for large N) behaviour of the ground-state energy, $\inf_{r \in \mathcal{D}_N} H^{KK}(r, q)$, of this simplified model under the constraint that the total charge, $Q = \sum_{i=1}^N q_i$, is conditioned to remain fixed to a constant b.

All the models introduced so far have been studied in the literature by using severe approximation methods. In most of the previous studies the self-averaging property of the free energy has been taken for granted [14]. Notice, however, that the Hamiltonian (1) is of short range in the image space of the random walk but it is a 'complete graph' Hamiltonian on the internal clock of the random walk. *Therefore classical subadditivity arguments do not apply in order to guarantee the existence of the specific free energy*.

Most of the previously published works use the replica trick to study the aforementioned models [11, 7, 14]. However, the replica trick is a drastic approximation of the original model *believed* to hold for mean-field models and is known to predict wrong results when applied to short-range models [13] (see [16] for a recent review). The Hamiltonian (1) is neither a short range nor mean-field one, as we explained above. It is therefore questionable as to whether it is legitimate even to hope that the replica trick should be applicable for the study of these models.

To go beyond all these unjustified approximations, Derrida and Higgs [6] introduced a simplification of the model that, supported by numerical evidence, allowed the study of the ground-state energy. Our aim is to go beyond the ground-state energy and to study the thermodynamics of this model. The main difficulty, however, is to satisfy the constraint $Q = b$, implying a conditioning on this event, known to be a delicate procedure [12]. To avoid such subtleties, we have decided to work on the space $\{Q = b\}$. Our starting point is the Hamiltonian (1) considered as a function on the space $\mathcal{Q}_N$ in a random environment $r \in \mathcal{D}_N$. Therefore our model, although closely related to that of [6], is distinct from it as it considers charges as the thermodynamic variables and geometries as the environment.

Thus, in [6], for a given charge distribution, how the geometry of the polymer is adjusted to minimize the energy has been studied. Here, a complementary point of view is adopted; namely, for a given geometry and given total charge, we study how the charge distributes over the polymer to minimize the energy. Of course, since the geometry plays the role of random environment, awkward geometries can arise as is explained in section 5 later. It proved computationally easier to consider the continuous version of the previous model; it is expected, however, that both discrete and continuous models have the same asymptotic behaviour. We now finish speculative considerations about the physical interpretation of our model and define it mathematically.

Let T be a fixed strictly positive integer and b an arbitrary real number. We denote by $\mathcal{C}^{0,b}_{[0,T]}$ the space of continuous real functions X defined on $[0,T]$ such that $X(0)=0$ and $X(T)=b$. We denote by $(X_t)_{t\in[0,T]}$ the standard Brownian bridge process on the space $\mathcal{C}^{0,b}_{[0,T]}$. Let $(Y_k)_{k\geqslant 0}$ be a sequence of independent Bernoulli variables, of parameter $p=\mathbb{P}\{Y_1=1\}=1-\mathbb{P}\{Y_1=0\}$, also independent of the Brownian bridge and defined on a probability space Ω. Define, on the probability space Ω, the renewal process, i.e. the strictly increasing sequence of random positive integers $(U_i)_{i\geqslant 0}$, with $U_0=0$ and recursively $U_k=\inf\{i>U_{k-1}:Y_i=1\}$. By $N_T=\sup\{i:U_i<T\}$ we denote the last index i such that $U_i<T$. Thus

$$0=U_0<U_1<\cdots<U_{N_T}<T\leqslant U_{N_T+1}<\cdots.$$

For $U=(U_i)_{i\in\mathbb{N}}$ and $X=(X_t)_{t\in[0,T]}$, the energy of our model is given by the Hamiltonian $H_T:\Omega\times\mathcal{C}\to\mathbb{R}^+$, defined by

$$H_T(U,X)=\sum_{i=1}^{N_T}(X_{U_i}-X_{U_{i-1}})^2+(b-X_{U_N})^2. \tag{2}$$

This Hamiltonian is the continuous analogue of the Hamiltonian (1) considered in [11]. The length of the protein is identified with the parameter T and the Gaussian charges are modelled by the Brownian bridge process. The constraint of fixed total charge b is automatically satisfied by all configurations in $\mathcal{C}^{0,b}_{[0,T]}$. The directed polymer acts as a random environment for the Brownian bridge. Notice that the number N_T of terms in the first sum if the Hamiltonian is a random variable, depending on the environment.

We now define the thermodynamics of the model. Denote by $\mathbb{E}\{\cdot\}$ the mean expected value, by $\mathbb{E}\{\cdot|\mathcal{R}\}$ the mean expected value conditioned to the process or variable or event $\mathcal{R}$, and by $\mathbb{E}_{\mathcal{R}}\{\cdot\}$ the mean expected value with respect to the distribution of $\mathcal{R}$. With these notations, the partition function is

$$Z_T(\beta)=\mathbb{E}\{\exp(-\beta H_T)|(N_T,U)\}. \tag{3}$$

The parameter T obviously plays the *role* of volume. The finite-volume, specific, quenched free energy is defined, as usual, by

$$f_T(\beta)=-\frac{1}{\beta T}\log Z_T(\beta) \tag{4}$$

and the finite-volume, specific, annealed free energy is given by

$$\overline{f}_T(\beta)=-\frac{1}{\beta T}\log\mathbb{E}_{(N_T,U)}Z_T(\beta). \tag{5}$$

This model, while reminiscent of the system studied in [15], has, however, a much richer structure since the random environment given by the renewal process is independent of the bridge. A model related to the one studied here was also treated in [3] where the unconstrained asymptotic behaviour of the process was given.

With the formulation that we use, we impose the total charge to be fixed, thus avoiding the usual subtleties [12, 1] of conditioning with respect to the final point. Moreover, with the total charge being left as a free parameter from the very beginning, several scaling laws can be tested.

We study the full thermodynamics of this model and we prove the following.

Theorem 1.1. Denote by $V_i = U_i - U_{i-1}$ the increment of the renewal process. Let $\lim_{T\to\infty} b^2/T = z \in [0, \infty[$. For every inverse temperature $\beta \in \mathbb{R}^+$ and for every asymptotic behaviour of the total charge $z \in [0, \infty[$, the quenched specific free energy, f_T, converges almost surely, for $T \to \infty$, to a trivial random variable f_∞, where

$$f_\infty(\beta) = \frac{p}{2\beta}\mathbb{E}\log(1 + 2\beta V_1)$$

and $p = \mathbb{P}(Y_1 = 1)$.

Remark. The previous theorem is established for a finite limit z. It is not difficult to see, however, that this result remains true even in the case $z = \infty$ provided that the divergence of the ratio b^2/T is not very violent (sublinear).

The specific free energy can easily be shown to be a smooth convex function of β. In fact, the series $\sum_{k=1}^{\infty} g_k(\beta)\mathbb{P}(V_1 = k)$, with $g_k(\beta) = [\log(1 + 2\beta k)]/\beta$ that defines $\mathbb{E}[\log(1 + 2\beta V_1)]/\beta$, converges uniformly in β. Since the function g_k is C^r, with arbitrary order r, and convex (i.e. $\partial^2 g_k/\partial\beta^2 \geqslant 0$), these properties also hold for the expected value by the uniform convergence of the series.

Thus, the theorem indicates that there is no phase transition at the thermodynamic level. Comparing the model that we studied with standard statistical mechanics models, we see that, expanding the interaction term $(X_{U_i} - X_{U_{i-1}})^2$ of the Hamiltonian, we get a contribution $-2\sum_i^{N_T} X_{U_i} X_{U_{i-1}}$ reminiscent of the term $-2\sum_i^N \sigma_i\sigma_{i-1}$ for the standard Ising model. Notice that here the range of pair interactions $V_i = U_i - U_{i-1}$ is an unbounded random variable introducing arbitrarily long-range correlations. However, as far as $\mathbb{E}V < \infty$, the model behaves, on average, as a finite-range one-dimensional model. It is expected that this model undergoes a phase transition only if V becomes non-integrable; however, in this regime the model is not soluble anymore.

Besides the thermodynamic behaviour, we also obtain results on the scaling of the ground-state energy. For the model studied in [11] it is argued that when the total charge b scales asymptotically like $b \sim T^x$ (where $f \sim g$ means $\lim_{T\to\infty} f(T)/g(T) = 1$) then the minimum of the energy scales like $\min H \sim T^{\theta(x)}$, where $\theta(x)$ is a critical exponent continuously depending on x. These results are based on heuristic arguments and numerical simulations. Here we prove the following.

Theorem 1.2. Let $0 < p = \mathbb{P}(Y_1 = 1) < 1$, $q = 1 - p$, $b \in \mathbb{R}$, and $T \in \mathbb{R}^+$. Then, we have

$$\mathcal{I}_T \equiv -\lim_{\beta\to\infty}\frac{1}{\beta}\log[\mathbb{E}(\exp(-\beta H_T)|(N, U))] = \frac{b^2}{(N_T + 1)}.$$

Moreover,

$$\mathbb{E}(\mathcal{I}_T) = \frac{b^2}{pT}(1 - q^T).$$

The previous result suggests that for fixed p and $q = 1 - p$, if $b \sim T^x$, the dominant behaviour of the minimum of the Hamiltonian will be $\mathbb{E}[\min_X H] \sim T^{2x-1}$. For p

and q depending on T, however, the precise asymptotic behaviour can be changed as is demonstrated in section 5.

This paper is organised as follows. The next section deals with technical explicit computations and intermediate results on finite-volume systems. These results are used in section 3 to obtain the thermodynamic limit of the specific free energy. The zero temperature limit and ground-state energy are studied in section 4. In the final section some open problems are presented. The appendices contain some well known results so that the paper is self-contained.

2. The conditional expectation of the Boltzmann factor

Let $\boldsymbol{u} = (u_1, \ldots, u_n) \in \mathbb{R}^n$ be a vector with $0 < u_1 < \cdots < u_n < T$ and $R(n, \boldsymbol{u})$ denote the event

$$R(n, \boldsymbol{u}) = \{N_T = n; U_1 = u_1, \ldots, U_{N_T} = u_n\}.$$

On this event, the Hamiltonian becomes a function solely of the Brownian bridge and reads

$$\sum_{i=1}^{n} (X_{u_i} - X_{u_{i-1}})^2 + (b - X_{u_n})^2.$$

Introduce the random vector $\Xi = (\Xi_1, \ldots, \Xi_n)$ with $\Xi = (X_{u_i} - X_{u_{i-1}})$ together with its expectation vector $\mu = (\mu_1, \ldots, \mu_n)$ with (see appendix A) $\mu_i = \mathbb{E}\Xi_i = bv_i/T \equiv b(u_i - u_{i-1})/T$ and its convariance matrix $\mathbf{Q} = (Q_{ij})_{i,j=1,\ldots,n}$ with $Q_{ij} = \mathbb{E}(\Xi_i \Xi_j) - \mu_i \mu_j$, yielding

$$\mathbf{Q} = \begin{pmatrix} v_1 & & \\ & \ddots & \\ & & v_n \end{pmatrix} - \frac{1}{T} \begin{pmatrix} v_1 v_1 & \cdots & v_1 v_n \\ v_2 v_1 & \cdots & v_2 v_n \\ \vdots & & \vdots \\ v_n v_1 & \cdots & v_n v_n \end{pmatrix}.$$

Notice that since $(1/T)\sum_{i=1}^{n} v_i = u_n/T < 1$, the matrix $\mathbf{Q}$ is invertible and its inverse reads

$$\mathbf{Q}^{-1} = \begin{pmatrix} v_1^{-1} & & \\ & \ddots & \\ & & v_n^{-1} \end{pmatrix} + \frac{1}{T - u_n} \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{pmatrix}.$$

Moreover, its determinant is explicitly computed (see appendix B) and gives

$$\det \mathbf{Q}^{-1} = \frac{T}{(T - u_n)\prod_{j=1}^{n} v_j}.$$

Therefore, the conditional expectation of the Boltzmann factor is expressed as

$$\mathbb{E}[\exp(-\beta H_T(X, U))|R(n, u)] = (2\pi)^{-n/2} |\det \mathbf{Q}^{-1}|^{1/2} \times \int_{\mathbb{R}^n} \exp[-\{\beta H_T(\xi, u) + \tfrac{1}{2}(\xi - \mu)^{\mathrm{t}} \mathbf{Q}^{-1}(\xi - \mu)\}] \, \mathrm{d}\xi_1 \ldots \mathrm{d}\xi_n$$

where $(\cdot)^{\mathrm{t}}$ denotes the transpose. The next step is to evaluate this integral by reducing the exponent of the integrand into a quadratic form of the integration variable. For that purpose, introduce the auxiliary matrix

$$\mathbf{Q}' = \begin{pmatrix} v_1^{-1} + 2\beta & & \\ & \ddots & \\ & & v_n^{-1} + 2\beta \end{pmatrix} + \left(2\beta + \frac{1}{T - u_n}\right) \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots 1 & \end{pmatrix}.$$

Using the results proved in appendix B, we deduce that this matrix is positive definite and admits a regular square root $\mathbf{\Gamma}$ (i.e. $\mathbf{Q}' = \mathbf{\Gamma}^2$) whose non-vanishing determinant yields

$$\det \mathbf{\Gamma} = \left(\prod_{j=1}^{n} \lambda_j \right)^{1/2} \left(1 + \left(\frac{1}{T - u_n} + 2\beta \right) \sum_{j=1}^{n} \lambda_j^{-1} \right)^{1/2}$$

with $\lambda_j = v_j^{-1} + 2\beta$. Introducing the auxiliary vector $\boldsymbol{r} = \mathbf{\Gamma}^{-1}(2\beta b\mathbf{1} + \mathbf{Q}^{-1}\mu)$ with $\mathbf{1} = (1, \ldots, 1)$, we have

$$\mathbb{E}(\exp(-\beta H_T)|R(n, \boldsymbol{u})) = (2\pi)^{-n/2} |\det \mathbf{Q}^{-1}|^{1/2} \int \exp(-M(\boldsymbol{u}, \xi))\, \mathrm{d}\xi_1 \ldots \mathrm{d}\xi_n$$

with

$$M(\boldsymbol{u}, \xi) = \tfrac{1}{2}(\mathbf{\Gamma}\xi - \boldsymbol{r})^{\mathrm{t}}(\mathbf{\Gamma}\xi - \boldsymbol{r}) + c(\beta)$$

and

$$c(\beta) = \beta b^2 + \tfrac{1}{2}\boldsymbol{\mu}^{\mathrm{t}}\mathbf{Q}^{-1}\boldsymbol{\mu} - \tfrac{1}{2}\boldsymbol{r}^{\mathrm{t}}\boldsymbol{r}.$$

Now the ξ integration can be explicitly performed by the change of variables

$$\xi' = \mathbf{\Gamma}\xi - \boldsymbol{r}$$

that introduces the Jacobian determinant in the volume element $\mathrm{d}\xi_1 \ldots \mathrm{d}\xi_n = |\det \mathbf{\Gamma}^{-1}|\, \mathrm{d}\xi_1' \ldots \mathrm{d}\xi_n'$ to yield

$$\mathbb{E}(\exp(-\beta H_T)|R(n, \boldsymbol{u})) = |\det \mathbf{Q}^{-1}|^{1/2} |\det \mathbf{\Gamma}^{-1}| \exp(-c(\beta)).$$

The only thing remaining is the explicit computation of the constant $c(\beta)$. Straightforward algebraic manipulations, although extremely cumbersome, are needed. Introducing the auxiliary quantities

$$A_1(\beta) = \frac{2b^2\beta^2}{[\gamma(\beta) + (\sum_{i=1}^{n}(v_i^{-1} + 2\beta)^{-1})^{-1}]}$$

$$A_2(\beta) = \frac{2b^2\beta}{T}\left(\sum_{i=1}^{n}(v_i^{-1} + 2\beta)^{-1} \right)$$

$$A_3(\beta) = \frac{2b^2\beta}{T}\left(\frac{1}{\gamma(\beta)(\sum_{i=1}^{n}(v_i^{-1} + 2\beta)^{-1})} + 1 \right)^{-1} \times \left\{ (T - u_n)^{-1}\gamma(\beta)^{-1}u_n - \sum_{i=1}^{n}(v_i^{-1} + 2\beta)^{-1} \right\}$$

$$A_4(\beta) = \frac{b^2}{2(T - u_n)^2}\left(\gamma(\beta) + \left(\sum_{i=1}^{n}(v_i^{-1} + 2\beta)^{-1} \right)^{-1} \right)^{-1}$$

with

$$\gamma(\beta) = (T - u_n)^{-1} + 2\beta$$

we can prove that

$$c(\beta) = \beta b^2 + \frac{b^2}{2T}\frac{u_n}{T - u_n} - A_1(\beta) - A_2(\beta) - A_3(\beta) - A_4(\beta).$$

Remark. Since explicit formulae for the conditional expectation with respect to the event

$$R(n, \boldsymbol{u}) = \{(N_T, U) = (n, \boldsymbol{u})\}$$

are obtained, the conditional expectation $\mathbb{E}(\cdot|N_T, U)$ is also explicitly known. In fact, if for an integrable function F,

$$\mathbb{E}(F|(R(n, \boldsymbol{u})) = \phi(n, \boldsymbol{u})$$

then

$$\mathbb{E}(F|N_T, U) = \phi(N_T, U).$$

So that, if we replace in the above formulae n by N_T and u_i or v_i by U_i or V_i, respectively, we obtain the corresponding expressions for the conditional expectations $\mathbb{E}(\cdot|N_T, U)$.

We have thus established the following.

Lemma 2.1. Assuming $\beta \geqslant 0$ and using the notation introduced so far, we have

$$\mathbb{E}\left(\exp(-\beta H_T)|N_T, U\right) = \psi(N_T, U)$$

where

$$\psi(n, \boldsymbol{u}) \equiv \mathbb{E}[\exp(-\beta H_T)|R(n, \boldsymbol{u})]$$

$$= \left\{\prod_{j=1}^{n}\left(\frac{1}{1+2\beta v_j}\right)\right\}^{\frac{1}{2}}\left\{\frac{T(T-u_n)^{-1}}{(1+\gamma(\beta)(\sum_{j=1}^{n}(v_j^{-1}+2\beta)^{-1}))}\right\}^{\frac{1}{2}}\exp(-c(\beta)).$$

The annealed partition function is also immediately computed. We have

$$\mathbb{E}Z_T(\beta) = \sum_{n=0}^{T-1} p^n q^{T-1-n}\sum_{u_1=1}^{T-1-n}\cdots\sum_{u_n=u_{n-1}+1}^{T-1}\mathbb{E}[\exp(-\beta H_T)|R(n, \boldsymbol{u})].$$

3. Thermodynamic limit of the specific free energy

In this section the $T \to \infty$ limit of the specific free energy will be evaluated. Some notation and well known results on the renewal process are necessary. Since $U_0 = 0$ and $U_n = \inf\{k > U_{n-1} : Y_k = 1\}$, the variables Y_i being independent and identically distributed, the renewal process (U_i) can be written as a sum of independent geometric variables $U_n = \sum_{i=1}^{n} V_i$, where $V_i = U_i - U_{i-1}$. For every integer $x \geqslant 1$, we have that $\mathbb{P}(V_1 = x) = pq^{x-1}$, where $p = \mathbb{P}(Y_1 = 1) = 1 - q$. Therefore, $m \equiv \mathbb{E}V_i = 1/p$. Also of special interest is the random variable $N_T = \sup\{k : U_k < T\}$. Obviously, $\lim_{T\to\infty} N_T = \infty$ and, using the strong law of large numbers, almost certainly (see [2] for instance)

$$\lim_{T\to\infty}\frac{N_T}{T} = \frac{1}{m} = p.$$

Notice also that although N_T is not a stopping time with respect to the σ-algebra $\sigma(Y_1, \ldots, Y_T)$, the random variable $N_T + 1$ is, in contrast, a stopping time.

Lemma 3.1. The finite-volume specific free energy, f_T, is a random variable reading

$$f_T(\beta) = \frac{1}{2\beta T}\sum_{i=1}^{N_T}\log(1+2\beta V_i) + \frac{1}{2\beta T}\log T - \frac{1}{2\beta T}\log(T-U_{N_T})$$

$$-\frac{1}{2\beta T}\log\left[1+((T-U_{N_T})^{-1}+2\beta)\sum_{i=1}^{N_T}(V_i^{-1}+2\beta)^{-1}\right]$$

$$+\frac{b^2}{T}+\frac{1}{2\beta}\frac{b^2}{T}\frac{U_{N_T}}{T(T-U_{N_T})}-\frac{A(\beta)}{\beta T}$$

where $A(\beta)$ is a random variable such that

$$\mathbb{E}(A(\beta)|n, \boldsymbol{u}) = A_1(\beta) + A_2(\beta) + A_3(\beta) + A_4(\beta)$$

and $A_1, \ldots, A_4$ are the expressions given in the previous section.

Proof. Recalling the remark on conditional expectations given at the end of the previous section, lemma 2.1, and the expression for $c(\beta)$, the present result follows immediately. □

Studying the thermodynamic limit of f_T is equivalent in studying the $T \to \infty$ limit in the above formula. Notice, however, that besides random terms whose limit must be computed, there are also deterministic terms of the form b^2/T where b is the total charge of the protein, supposed to scale with the size T of the molecule like $b \sim T^\alpha$. It follows that

$$z \equiv \lim_{T\to\infty} \frac{b^2}{T} = \begin{cases} 1 & \text{if } \alpha = 1/2 \\ 0 & \text{if } \alpha < 1/2 \\ \infty & \text{if } \alpha > 1/2. \end{cases}$$

We do not fix the exact behaviour here. We only assume that $z \in [0, \infty]$ in the following.

Proof of theorem 1.1. It is enough to study the $T \to \infty$ limit of all individual terms appearing in the expression for f_T given in lemma 3.1.

First remark that $\lim_{T\to\infty} N_T = \infty$ almost surely. Let $X_j \equiv \log(1 + 2\beta V_j) \geqslant 0$. On the event $\{\lim_{T\to\infty} N_T = \infty\}$, for the independent identically distributed random variables X_j the strong law of large numbers holds. Hence,

$$\frac{1}{2\beta T}\sum_{j=1}^{N_T} \log(1 + 2\beta V_j) = \frac{N_T}{2\beta T}\frac{1}{N_T}\sum_{j=1}^{N_T} \log(1 + 2\beta V_j) \overset{\text{a.s.}}{\to} \frac{p}{2\beta}\mathbb{E}[\log(1 + 2\beta V_1)]$$

because $N_T/T \overset{\text{a.s.}}{\to} p$.

The second term trivially vanishes in the limit. As for the third term, remark that $1 \leqslant T - U_N < T$, therefore it also vanishes in the limit. For the fourth term, remark that $(T - U_N)^{-1} + 2\beta \leqslant 1 + 2\beta$ and $\sum_{j=1}^{N_T}(V_j^{-1} + 2\beta)^{-1} \leqslant \sum_{j=1}^{N_T} V_j = U_{N_T} < T$, hence this term gives also a vanishing limiting contribution. The fifth term yields, by definition, z.

For the study of subsequent terms we introduce the random variable $\Delta_T = T - U_{N_T}$ and the numerical sequence $z_T = b^2/T$ converging to z. With this notation the sixth term reads

$$z_T \frac{U_{N_T}}{T}\frac{1}{2\beta\Delta_T}.$$

Because $(V_i^{-1} + 2\beta)^{-1} \geqslant (1 + 2\beta)^{-1}$ and $N_T \overset{\text{a.s.}}{\to} \infty$ it follows that the inverse sum, appearing in the expression for the seventh term, vanishes almost certainly in the limit. Hence, the seventh term is almost certainly asymptotically equivalent to

$$-z_T \frac{2\beta\Delta_T}{1 + 2\beta\Delta_T}.$$

The arguments used for the study of the first term guarantee that the eighth term

$$-\frac{2b^2\beta}{T}\frac{1}{T}\left(\sum_{i=1}^{N_T}(V_i^{-1} + 2\beta)^{-1}\right) \overset{\text{a.s.}}{\to} -2pz\mathbb{E}[(V_1^{-1} + 2\beta)^{-1}].$$

For the ninth term we obtain

$$-2\frac{b^2}{T}\frac{[(U_{N_T}/T)(1 + 2\beta\Delta_T)^{-1} - (1/T)\sum_{i=1}^{N_T}(V_i^{-1} + 2\beta)^{-1}]}{1 + [(2\beta + \Delta_T^{-1})\sum_{i=1}^{N_T}(V_i^{-1} + 2\beta)^{-1}]^{-1}}.$$

Again, using the observations that $2\beta + \Delta_T^{-1} \geqslant 2\beta + 1$ and $\sum_{i=1}^{N_T}(V_i^{-1} + 2\beta)^{-1} \geqslant N_T(1+2\beta)^{-1}$, the large denominator $1 + [\cdots]$ reduces to 1 in the limit $T \to \infty$. This fraction, therefore, gives rise to two contributions, asymptotically equivalent to

$$-2z_T \frac{1}{1+2\beta\Delta_T} + 2zp\mathbb{E}[(V_1^{-1}+2\beta)^{-1}].$$

The same arguments can be used to study the last term

$$-\frac{b^2}{2\beta T(T-U_{N_T})^2}\left[2\beta + \Delta_T^{-1} + \left(\sum_{i=1}^{N_T}(V_i^{-1}+2\beta)^{-1}\right)^{-1}\right]^{-1}.$$

Obviously, again, the term $(\sum_{i=1}^{N_T}(V_i^{-1}+2\beta)^{-1})^{-1}$ vanishes in the limit. It remains a contribution asymptotically equivalent to

$$-\frac{z_T}{2\beta\Delta_T} + \frac{z_T}{1+2\beta\Delta_T}.$$

Adding together all but the first terms, we get a net contribution

$$z_T\left(\frac{U_{N_T}}{T} - 1\right)\frac{1}{2\beta\Delta_T} + 2z_T\left(1 - \frac{U_{N_T}}{T}\right)\frac{1}{1+2\beta\Delta_T}.$$

Now the proof is completed by remarking that $\Delta_T \geqslant 1$—so that both $(2\beta\Delta_T)^{-1}$ and $(1+2\beta\Delta_T)^{-1}$ are bounded from above, uniformly in T—that $z_T \to z \in [0,\infty[$, and that the ratio $U_{N_T}/T \overset{\text{a.s.}}{\to} 1$. □

4. Finite-volume ground-state energy

We can now consider the behaviour of the ground-state energy. First notice that a standard Laplace argument should imply that $\mathcal{I}_T$ coincides with $\inf_{X\in\mathcal{C}^{0,b}_{[0,T]}} H_T(U,X)$ and, therefore, results about $\mathcal{I}_T$ must be interpreted as results concerning the finite-volume ground-state energy of the model.

Proof of the theorem 1.2. The proof proceeds by examining how the individual terms in the expression of $F_T(\beta) = -(1/\beta)\log Z_T(\beta) = Tf_T(\beta)$, where $f_T(\beta)$ is given in lemma 3.1, behave in the limit $\beta \to \infty$. Namely, we examine the behaviour when $\beta \to \infty$ of the individual terms for the following expression:

$$\begin{aligned}F_T(\beta) = Tf_T(\beta) = {} & \frac{1}{2\beta}\sum_{i=1}^{N_T}\log(1+2\beta V_i) + \frac{1}{2\beta}\log T - \frac{1}{2\beta}\log(T-U_N)\\ & -\frac{1}{2\beta}\log\left[1 + ((T-U_N)^{-1}+2\beta)\sum_{i=1}^{N_T}(V_i^{-1}+2\beta)^{-1}\right]\\ & +b^2 + \frac{1}{2\beta}\frac{U_{N_T}}{(T-U_{N_T})} - \frac{A_1(\beta)}{\beta} - \frac{A_2(\beta)}{\beta} - \frac{A_3(\beta)}{\beta} - \frac{A_4(\beta)}{\beta}.\end{aligned}$$

It is easy to see that the only terms remaining in the limit $\beta \to \infty$ are the fifth one—giving trivially a contribution equal to b^2—and the seventh one—namely,

$$-\lim_{\beta\to\infty}\frac{A_1(\beta)}{\beta} = -\lim_{\beta\to\infty}\frac{b^2}{1+(2\beta\Delta_T)^{-1}+[\sum_{i=1}^{N_T}(1+(2\beta V_i)^{-1})^{-1}]^{-1}} = -\frac{b^2}{1+N_T^{-1}}.$$

Adding the two non-vanishing contributions, we get

$$\lim_{\beta\to\infty} F_T(\beta) = b^2 - \frac{b^2}{1+N_T^{-1}} = \frac{b^2}{1+N_T}$$

and the first assertion of the theorem follows.

The second assertion is obtained by an elementary summation:

$$\mathbb{E}\left(\frac{1}{N_T+1}\right)=\frac{1}{pT}\sum_{n=0}^{T-1}\binom{T}{n+1}p^{n+1}q^{T-(n+1)}=\frac{1}{pT}(1-q^T).$$

Therefore, collecting all terms,

$$\mathbb{E}(\mathcal{I}_T)=\mathbb{E}(\lim_{\beta\to\infty}F_T(\beta))=\frac{b^2}{pT}(1-q^T).$$

□

5. Conclusion and open problems

It is established that the model studied here has well-defined thermodynamics in the sense that the infinite volume limit of the quenched specific free energy exists; moreover, this limit is a trivial random variable—i.e. almost certainly a constant. This result shows the sample independence of the thermodynamic properties of the model at every temperature; a feature common to all short-range disordered systems.

The large scale behaviour of the average ground-state energy is also very instructive. Suppose first that the parameters p and q are fixed—independent of T—and non-trivial i.e. $0<p,q<1$. Then

$$\mathbb{E}[\inf_X H_T(U,X)]\sim z/p \qquad \text{for large } T$$

and it has a finite value only when $z<+\infty$. Now, a finite and strictly positive z means that the total charge scales as $T^{1/2}$ for large T, while the number of visited sites from the directed random walks behaves like pT. Allowing for $z=\infty$ by choosing $b\sim T^x$ with $x>1/2$, for instance, leads to an asymptotic behaviour of the total energy of the form T^{2x-1}. On the other hand, nothing prevents us from choosing p and q depending on T. To illustrate what may happen, suppose that we choose p to scale asymptotically as α/T, with $0<\alpha<+\infty$. Then

$$\mathbb{E}[\inf_X H_T(U,X)]\sim\frac{b^2}{\alpha}(1-\exp(-\alpha))$$

and this expectation is finite provided that the total charge remains finite. For a scaling of the form $b\sim T^x$, the total energy behaves like T^{2x}. A change of the critical exponent is observed. At the same time, the total number of visited sites does not diverge. Of course various other asymptotic behaviours can be obtained by choosing different scaling laws for the parameters p and q (such as $p\sim\alpha T/\log T$, for instance). While constant p gives a more or less regular distribution of the charge over the protein, volume-dependent environments give rise to highly irregular charge adjustments. The result we have obtained allows an exhaustive study of all these scaling behaviours.

What remains an open problem, for the moment, is the statistical mechanics study of the system. Namely, for every realization U of the random environment, we can define a random finite-volume Gibbs measure, μ_T, on the space $\mathcal{C}^{0,b}_{[0,T]}$ as a measure absolutely continuous with respect to the standard Gaussian measure, γ_T, of the Brownian bridge. This Gibbs measure can be defined through its Radon–Nikodým derivative

$$\frac{\mathrm{d}\mu_T}{\mathrm{d}\gamma_T}(U,X)=\frac{\exp(-\beta H_T(U,X))}{Z_T(\beta)}$$

where $Z_T(\beta)$ is the partition function that depends also on U. It is not yet clear whether a Dobrushin–Lanford–Ruelle construction is possible for this model, and neither is it clear whether there is a phase transition or not, whether it be in the DLR or the weak sense. These fundamental questions for the statistical mechanics formulation of the model are presently under investigation.

A second problem that is still under investigation concerns the thermodynamics of the model defined for genuine random walks and not directed random walks. This destroys the explicit renewal process structure of the model and makes the computations much more complicated, but not untractable.

Summarizing, inspired by the protein folding statistical mechanics problems, we are confronted with a new class of systems, reminiscent of both spin glasses and random walks in random environments.

Acknowledgments

SM thanks the Insitut de Recherche Mathématique de l'Université de Rennes I and DP the Departamento de Ingeniería Matemática of the Universidad de Chile, where this paper was written. This work was partially supported by FONDECYT grant 1940405 and EU grants CI1-CT92-0046 and CHRX-CT93-0411.

Appendix A. Identities for the Brownian bridge

Let $(W_s : s \geqslant 0)$ be a Brownian motion starting from zero and introduce the process $(Y_s : s \geqslant 0)$ by

$$Y_t = (T-t)\int_0^t \frac{\mathrm{d}W_s}{T-s}.$$

Then, the Brownian bridge between $X_0 = 0$ and $X_T = b$ is expressed by

$$X_t = b\frac{t}{T} + Y_t \qquad \text{for } t \in [0, T].$$

It is immediate to verify the following identities that are reported here only for completeness:

$$\mathbb{E}(Y_t) = 0 \qquad \mathbb{E}(X_t) = \frac{bt}{T} \qquad \mathbb{E}(X_t - X_s) = \frac{b}{T}(t-s)$$

and

$$\mathbb{E}(Y_t^2) = \frac{T-t}{T}t \qquad \mathbb{E}(X_t^2) = \frac{b^2t^2}{T^2} + \frac{(T-t)}{T}t.$$

For $s \leqslant t$, we also have

$$\mathbb{E}(Y_sY_t) = \frac{(T-t)}{T}s \qquad \mathbb{E}(X_sX_t) = \frac{b^2st}{T^2} + \frac{(T-t)}{T}s$$

$$\mathbb{E}((Y_t - Y_s)^2) = \frac{(t-s)}{T}(T-(t-s))$$

$$\mathbb{E}((X_t - X_s)^2) = \left(\frac{b^2}{T^2} - \frac{1}{T}\right)(t-s)^2 + (t-s) \leqslant t$$

and

$$\mathrm{Var}((X_t - X_s)) = (t-s) - \frac{1}{T}(t-s)^2.$$

Finally, for $s_1 < s_2 \leqslant t_1 < t_2$, we have

$$\mathbb{E}((Y_{t_2} - Y_{t_1})(Y_{s_2} - Y_{s_1})) = -\frac{(s_2 - s_1)(t_2 - t_1)}{T}$$

$$\mathbb{E}((X_{t_2} - X_{t_1})(X_{s_2} - Y_{s_1})) = \left(\frac{b^2}{T^2} - \frac{1}{T}\right)(s_2 - s_1)(t_2 - t_1)$$

and

$$\mathrm{Cov}((X_{t_2} - X_{t_1}), (X_{s_2} - X_{s_1})) = -\frac{1}{T}(s_2 - s_1)(t_2 - t_1).$$

Appendix B. Determinant and positive definiteness for a class of matrices

For $\ell = (\ell_1, \ldots, \ell_n) \in \mathbb{R}^n$ denote by $\mathbf{D}(\ell)$ the diagonal matrix, $\mathbf{D}(\ell)_{ij} = \ell_i \delta_{ij}$ for $i, j = 1, \ldots, n$. For $a \in \mathbb{R}$ denote by $\tilde{\mathbf{a}}$ the constant matrix whose elements are $\widetilde{a}_{ij} = a$ for $i, j = 1, \ldots, n$. From [18, p 92], we get the following property:

$$\det(\mathbf{D}(\ell) + \tilde{\mathbf{a}}) = \prod_{j=1}^{n} \ell_j + a \sum_{i=1}^{n} \prod_{\substack{j=1 \\ j \neq i}}^{n} \ell_j.$$

Consequently, if $\ell_j \neq 0$ for every j, then

$$\det(\mathbf{D}(\ell) + \tilde{\mathbf{a}}) = \left(\prod_{j=1}^{n} \ell_j\right)\left(1 + a \sum_{i=1}^{n} \ell_i^{-1}\right). \tag{6}$$

Remark, moreover, that the matrix $(\mathbf{D}(\ell) + \tilde{\mathbf{a}})$ is symmetric and that, in the case $\ell_i > 0$, for all $i = 1, \ldots, n$ and $a \geqslant 0$ the matrix $(\mathbf{D}(\ell) + \tilde{\mathbf{a}})$ is positive definite. Therefore there exists a positive definite symmetric matrix $\mathbf{\Gamma}(\ell, a)$ that satisfies

$$(\mathbf{D}(\ell) + \tilde{\mathbf{a}}) = (\mathbf{\Gamma}(\ell, a))^2$$

and

$$\det \mathbf{\Gamma}(\ell, a) = \left(\prod_{j=1}^{n} \ell_j\right)^{\frac{1}{2}} \left(1 + a \sum_{j=1}^{n} \ell_j^{-1}\right)^{\frac{1}{2}}.$$

Appendix C. Inversion for a class of matrices

We use the notation already introduced in appendix B. For $\ell = (\ell_1, \ldots, \ell_n) \in \mathbb{R}^n$ we define the matrix $\mathbf{S}(\ell)$ by $\mathbf{S}(\ell)_{ij} = \ell_i \ell_j$ for $i, j = 1, \ldots, n$. Also we denote $\ell^{-1} = (\ell_1^{-1}, \ldots, \ell_n^{-1})$, so that $\mathbf{D}(\ell^{-1}) = (\mathbf{D}(\ell))^{-1}$.

Lemma C.1. Assume $\ell_i \neq 0$ for every $i = 1, \ldots, n$.

- If $\gamma \in \mathbb{R}$ verifies $\gamma \sum_{i=1}^{n} \ell_i \neq 1$, then the matrix $(\mathbf{D}(\ell) - \gamma \mathbf{S}(\ell))^{-1}$ exists and verifies

$$(\mathbf{D}(\ell) - \gamma \mathbf{S}(\ell))^{-1} = \mathbf{D}(\ell)^{-1} + \gamma \left(1 - \gamma \sum_{i=1}^{n} \ell_i\right)^{-1} \tilde{\mathbf{1}}.$$

- Reciprocally, if $\gamma \sum_{i=1}^{n} \ell_i^{-1} \neq -1$, then the matrix $(\mathbf{D}(\ell) + \gamma \tilde{\mathbf{1}})^{-1}$ exists and verifies

$$(\mathbf{D}(\ell) + \gamma \tilde{\mathbf{1}})^{-1} = \mathbf{D}(\ell^{-1}) - \gamma \left(1 + \gamma \sum_{i=1}^{n} \ell_i^{-1}\right)^{-1} \mathbf{S}(\ell^{-1}).$$

Proof. Denote by $\mathbf{R}(\ell)$ the matrix $\mathbf{R}(\ell)_{ij} = \ell_j$, for $i, j = 1, \ldots, n$. It is immediate to verify the identities $\mathbf{D}(\ell)\mathbf{R}(\ell) = \mathbf{S}(\ell)$, $\mathbf{R}(\ell)^2 = (\sum_{i=1}^n \ell_i)\mathbf{R}(\ell)$, and $\mathbf{R}(\ell)\mathbf{D}(\ell)^{-1} = \tilde{\mathbf{1}}$. Hence, $\mathbf{D}(\ell) - \gamma\mathbf{S}(\ell) = \mathbf{D}(\ell)(\mathbf{I} - \gamma\mathbf{R}(\ell))$ and, therefore, it follows by direct computation that $(\mathbf{I} - \gamma\mathbf{R}(\ell))^{-1} = \mathbf{I} + \gamma(1 - \gamma\sum_{i=1}^n \ell_i)^{-1}\mathbf{R}(\ell)$. Now

$$\mathbf{D}(\ell) - \gamma\mathbf{S}(\ell))^{-1} = \mathbf{D}(\ell)^{-1} + \gamma\left(1 - \gamma\sum_{i=1}^n \ell_i\right)\mathbf{R}(\ell)\mathbf{D}(\ell)^{-1}$$

which is enough to conclude. □

Hence, under the condition of this lemma and if $\gamma \neq 0$,

$$(\mathbf{D}(\ell) - \gamma\mathbf{S}(\ell))^{-1} = \mathbf{D}(\ell^{-1}) + \left(\gamma^{-1} - \sum_{i=1}^n \ell_i\right)^{-1}\tilde{\mathbf{1}}$$

and

$$(\mathbf{D}(\ell) + \gamma\tilde{\mathbf{1}})^{-1} = \mathbf{D}(\ell^{-1}) - \left(\gamma^{-1} + \sum_{i=1}^n \ell_i^{-1}\right)^{-1}\mathbf{S}(\ell^{-1}).$$

References

[1] Bertoin J and Doney R A 1994 On conditioning random walk in an exponential family to stay nonnegative *Séminaire de Probabilités XXVIII (Lecture Notes in Mathematics 1583)* pp 116–21

[2] Bhattacharya R N and Waymire E C 1990 *Stochastic Processes with Applications* (New York: Wiley)

[3] Buffet E and Hannigan P 1991 Directed random walk in random environments *J. Stat. Phys.* **65** 645–72

[4] Chow Y S and Teicher H 1988 *Probability Theory, Independence, Interchangeability, Martingales* (Berlin: Springer)

[5] Derrida B, Griffiths R B and Higgs R G 1992 A model of directed walks with random self interactions *Europhys. Lett.* **18** 361–6

[6] Derrida B and Higgs P G 1994 Low temperature properties of directed walks with random self interactions *J. Phys. A: Math. Gen.* **27** 5485–93

[7] Garel T and Orland H 1988 Mean-field model for protein folding *Europhys. Lett.* **6** 307–10; 1988 Chemical sequence and spatial structure in simple models of biopolymers *Europhys. Lett.* **6** 597–601

[8] M Kardar 1994 Lectures on directed paths in random media *Lectures given at Les Houches Summer School on 'Fluctuating Geometries in Statistical Mechanics and Field Theory' (August 1994)* available at cond-mat@babbage.sissa.it under reference 9411022

[9] Karlin S and Taylor H M 1975 *A First Course in Stochastic Processes* (New York: Academic)

[10] Kantor Y and Ertaş D 1994 Extremal segments in random sequences *J. Phys. A: Math. Gen.* **27** L907–11

[11] Kantor Y and Kardar M 1991 Polymers with random self-interactions *Europhys. Lett.* **14** 421

[12] Martínez S and San Martín J 1994 Quasi-stationary distributions for a Brownian motion with drift and associated limit laws *J. Appl. Prob.* **31** 911–20

[13] Newman C M and Stein D L Non-mean-field behaviour or realistic spin glasses *Preprint* New York University, available at adap-org@xyz.lanl.gov under reference 9508006

[14] Pande V S, Grosberg A Yu and Tanaka T 1995 Freezing transition of random heteropolymers consisting of an arbitrary set of monomers *Preprint* available at cond-mat@babbage.sissa.it under reference 9412006

[15] Petritis D 1990 The zero-level set of the Brownian bridge as a model of disordered physical system *Proc. Prob. Th. and Math. Stat. (Vilnius 1989)* vol 2 (Vilnius: Mosklas)

[16] Petritis D 1994 Equilibrium statistical mechanics of frustrated spin glasses: a survey of mathematical results *Talk Presented at the Satellite Conf. to the XIth Int. Congress of Mathematical Physics Devoted to the 'Mathematical Physics of Disordered Systems' (Paris, 25–27 July 1994)* to appear in the proceedings, available at cond-mat@babbage.sissa.it under reference 9411098

[17] Révész P 1990 *Random Walks in Random and Non-random Environments* (Singapore: World Scientific)

[18] Satake I 1975 A series on monographs and textbooks *Linear Algebra, Pure and Applied Mathematics* (New York: Marcel Dekker)