

# The Statistics of Energy States of a Hookean Model for Proteins\*

**Ayşe ERZAN**

*Department of Physics, Faculty of Sciences and Letters  
Istanbul Technical University, Maslak 80626, Istanbul-TURKEY  
Gürsey Institute, P. O. Box 6, Çengelköy 81220, Istanbul-TURKEY*

**Erkan TÜZEL**

*Department of Physics, Faculty of Sciences and Letters  
Işık University, Maslak 80670, Istanbul-TURKEY*

Received 23.04.2000

## Abstract

A generic model of a random polypeptide chain, with discrete torsional degrees of freedom and Hookean springs connecting pairs of hydrophobic residues, reproduces the energy probability distribution of real proteins over a very large range of energies. We show that this system with harmonic interactions, under dissipative dynamics driven by random noise, leads to a distribution of energy states obeying a modified one-dimensional Ornstein-Uhlenbeck process with reflecting boundary conditions, and giving rise to distributions of the Wigner or inverse Gaussian form. A continuum approximation leads to a path integral formulation of the problem. PACS 5.65+b,5.70Ln,87.17.Aa

## 1. Introduction

It has recently begun to be appreciated [1, 2, 3, 4, 5], that such features of real proteins as the density of vibrational energy states [6] may be reproduced by coarse-grained model hamiltonians which capture the essential mechanism driving the folding process, namely hydrophobic interactions. [3, 4, 5] At least for relatively short polypeptide chains, it may not be unreasonable to assume that a kind of self-similarity holds over the entire energy landscape, such that not only vibrational but conformational energy states obey the same overall statistics. This self-similarity of the energy spectrum is indeed encountered in other complex systems such as large nuclei. [7, 8]

---

\*Talk presented in Regional Conference on Mathematical Physics IX held at Feza Gürsey Institute, Istanbul, August 1999.

We take the view here that the protein in its native state must essentially correspond to a self-organized system, i.e., the “native state” should be conceived of as the attractor of a dynamics. This typically corresponds not to a unique conformation but to a set of conformations to which the trajectory of the phase point representing the molecule is confined after asymptotically long times. Therefore, the aim should eventually be to model the dynamics of the molecule which puts it on the correct attractor.

Our model involves  $N$  coupled, discrete, over-damped torsional degrees of freedom coupled by Hookean forces and driven by random noise. We find that the distributions of energy states may be very well represented by a Wigner distribution [7], or “Wigner surmise” for the statistics of level spacings  $S$  encountered [8] in the study of large nuclei,

$$P(S) = \frac{\pi}{2} S \exp\left(-\frac{\pi}{4} S^2\right) . \quad (1)$$

The coarse grained energy level distributions are comparable with the statistics of the  $n$ 'th ( $n = 1, 2, 3 \dots$ ) neighbor energy level spacing. The energy histograms can also be very well fitted with an “inverse gaussian” (IG) distribution

$$P(E) = \sqrt{\frac{A}{2\pi E^3}} \exp\left[-\frac{A(E-B)^2}{2B^2 E}\right] . \quad (2)$$

We find [9] that preferentially relaxing the maximal torques drives the system to less stable, high energy states, whereas choosing the dihedral angles either with uniform probability or preferring minimal torques give rise to more successful strategies for reaching low lying energy states [10].

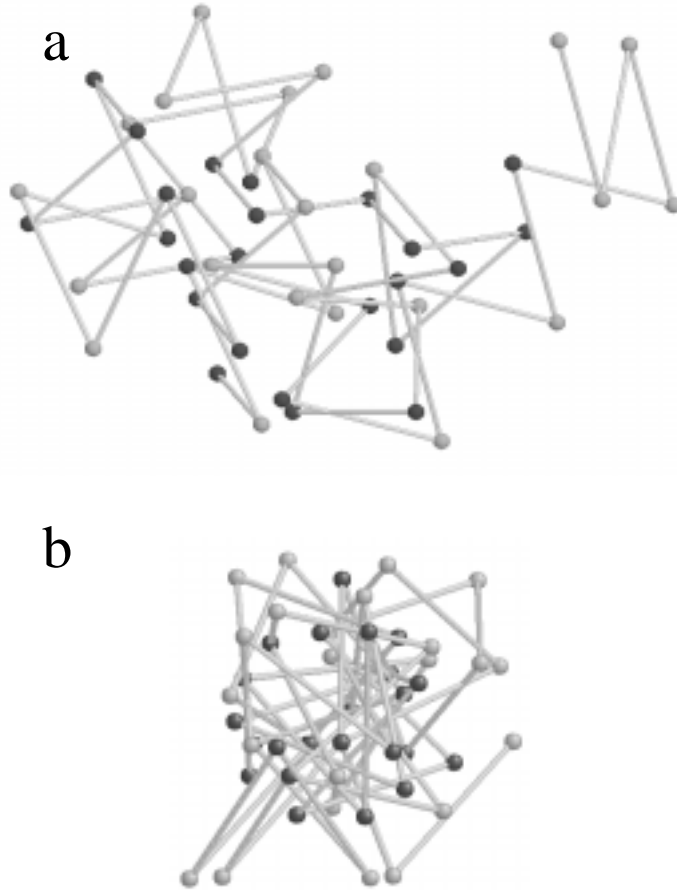
We are able to show that for harmonic potentials, quite independently of the nature of the sequence of hydrophobic and polar residues, or the dimensionality of the space, the energy of the system obeys an Ornstein-Uhlenbeck (OU) process [11]. The steady state distribution for this process with reflecting boundary conditions introduced due to constraints may then be related to the energy distribution. As a bonus, we are thus also able to understand quantitatively the distribution of relaxation times found for global optimization problems [12] by Li and coworkers.

The paper is organized as follows. In section 2 we define our model, in section 3 we present our simulation results for the “ $\eta$ -dynamics”, in section 4 we show that the energy obeys a modified OU process. In section 5 we present a discussion of the path integral formulation of the problem.

## 2. The Model

We consider a model [9, 10] consisting of  $N$  residues, treated as point vertices, interacting via Hookean potentials. We have been motivated by the model proposed by Haliloğlu, Bahar, Erman [3] where all interactions between different residues are governed by confining square-law potentials [3, 4, 5]. In our model, however, the covalent bonds between residues are treated as fixed rods of equal length. The residues located at the vertices may be polar  $P$  or hydrophobic  $H$ . All the hydrophobic vertices are to be connected to each other with springs of equal stiffness. This feature mimicks the effective

pressure that is exerted on the hydrophobic residues by the ambient water molecules, and results in their being driven to the relatively less exposed center of the molecule in the low lying energy states, whereas the polar residues are closer to the surface (see Fig.1). It is important to note that we treat all  $H - H$  pairs on an equal footing, regardless of whether they end up close to each other in the “native” configuration.



**Figure 1.** A chain of  $N = 48$  residues, half of which are randomly chosen to be hydrophobic, (darker beads) shown a) in a random initial configuration and b) in a folded state reached under Metropolis dynamics. The chain has folded in such a way as to leave the polar residues on the outside. (Generated using RasMol V2.6)

It is known that real proteins are distinguished by  $H - P$  sequences that lead to unique ground states while a randomly chosen  $H - P$  sequence will typically give rise to a highly degenerate ground state. In the absence of detailed knowledge regarding the rules singling out the realistic  $H - P$  sequences we considered a generic  $H - P$  sequence obtained by choosing fifty percent of the residues to be hydrophobic and distributing them randomly along the chain. We have checked that our results were quite robust with respect to changing the sequence of hydrophobic or hydrophilic residues, or even taking all of them to be hydrophobic.

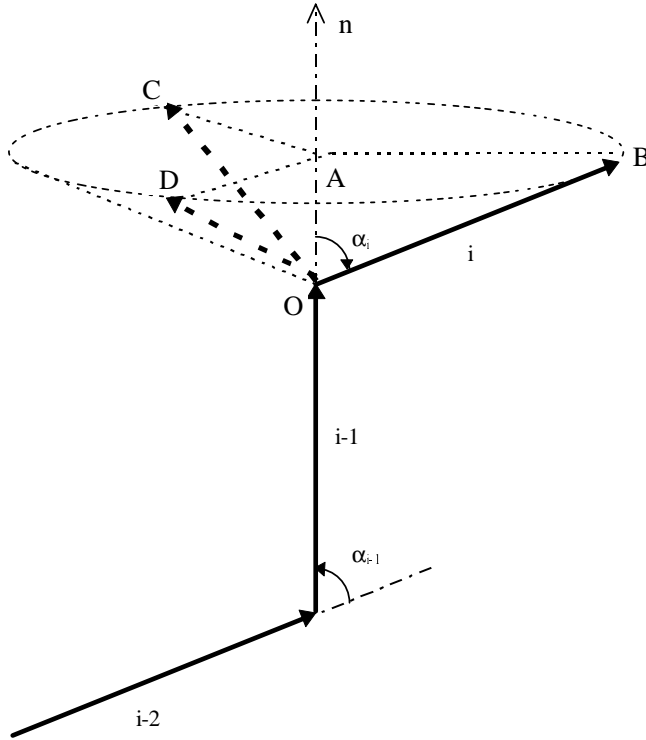
The energy of the molecule is

$$E = \frac{K}{2} \sum_{i,j} c_{i,j} |\mathbf{r}_i - \mathbf{r}_j|^2 = K \sum_{i,j} \mathbf{r}_i^\dagger V_{ij} \mathbf{r}_j \quad (3)$$

If we define  $Q_i = 1$  for the  $i$ 'th vertex being occupied by a hydrophobic residue, and  $Q_i = 0$  otherwise, we may write  $c_{i,j} = Q_i Q_j$  and

$$V_{ij} = [(N_H - 1)c_{i,i} - c_{i,j-1} - c_{i,j+1}] \delta_{i,j} - (1 - \delta_{i,j})(1 - \delta_{i,j-1} - \delta_{i,j+1})c_{i,j} \quad (4)$$

The chemical bonds can be considered as directors  $\mathbf{R}_i$  of unit length (see Fig.2). The conformation of the chain is specified by the ‘‘bond’’ (or ‘‘chemical’’) angles and the dihedral angles. The bond angles between the directors of successive bonds are measured from the director of the previous bond; thus the  $i$ th bond angle  $\alpha_i$  is measured from the director of the  $i - 1$ st bond to the director of the  $i$ th bond. We take the bond angles  $\alpha_i, i = 1 \dots, N - 1$ , to have the alternating values of  $(-1)^i \alpha$ , with  $\alpha = 68^\circ$ . The dihedral angles are related to the angle the  $i$ th bond makes with the plane of the  $i - 1$ st and  $i - 2$ nd bonds as shown in Fig.2. The angle is defined on the base of the cone described by the  $i$ th bond around the direction fixed by the  $i - 1$ st bond. It is measured as the angle between the projection of the  $i$ th bond and the projection of the director of the  $i - 2$ nd bond on the base of this cone. We allow the dihedral angles  $\phi_i$  to take on the values of 0 and  $\pm 2\pi/3$ . The state (conformation) of the system is uniquely specified once the numbers  $\{\phi_i\}$  are given. The constraints placed on the conformations due to the rigid chemical bond lengths and by restricting the chemical and dihedral angles to discrete values prevent the molecule from trivially collapsing to a point. The residues effectively reside on the vertices of a tetrahedral lattice. The position vectors  $\mathbf{r}_i$  of each of the vertices in the chain can be expressed in terms of a sum over the directors  $\mathbf{R}_i$ , which may be obtained from  $\mathbf{R}_1$  by successive rotations  $\mathbf{M}_k(\alpha_k)$  and  $\mathbf{T}_k(\phi_k)$  through the bond and the dihedral angles [15], viz.,



**Figure 2.** The bond angles  $\alpha_i$ , defined to be the angles between successive bonds, are shown in the figure. Each  $\alpha_i$  is measured from the director of the  $i - 1$ st bond to the director of the  $i$ th bond. The dihedral angles are related to the angle the  $i$ th bond ( $\overline{OB}$ ,  $\overline{OC}$  or  $\overline{OD}$  in the figure) makes with the plane of the  $i - 1$ st and  $i - 2$ nd bonds. If we denote the direction fixed by the  $i - 1$ st bond as  $\hat{n}$ , then the dihedral angles are defined between line segments projected on to the base of the cone described by the  $i$ th bond around  $\hat{n}$ . If the  $i$ th bond takes the position  $\overline{OD}$ , then the dihedral angle is the angle between the projection ( $\overline{AD}$  in the figure) of the director of the  $i - 2$ nd bond and the projection ( $\overline{AB}$  in the figure) of the director of  $\overline{OD}$ , namely  $\widehat{BAD}$ . If the  $i$ th bond is indicated by  $\overline{OC}$ , then the dihedral angle  $\widehat{BAC}$  is measured from  $\overline{AB}$  to the projection of  $\overline{OC}$  on the base of the cone, namely  $\overline{AC}$ . Finally, if the  $i$ th bond is the line segment  $\overline{OB}$ , as shown in the figure, then the dihedral angle takes the value 0 since the projection of the  $i$ th bond and  $\overline{AB}$  coincide. In the figure, the angles  $\widehat{BAD}$ ,  $\widehat{BAC}$  are  $\pm 2\pi/3$  respectively.

$$\mathbf{r}_i = \sum_{j=1}^{i-1} \prod_{k=j}^2 \mathbf{T}_k(\phi_k) \mathbf{M}_k(\alpha_k) \mathbf{R}_1 \quad . \quad (5)$$

If we choose, without loss of generality,  $\mathbf{R}_1 = \hat{y}$  in our Cartesian laboratory frame, this

is accomplished by use of the matrices

$$M_k = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha_k & -\sin \alpha_k \\ 0 & \sin \alpha_k & \cos \alpha_k \end{pmatrix} \quad (6)$$

and

$$T_k = \begin{pmatrix} \cos \phi_k & 0 & -\sin \phi_k \\ 0 & 1 & 0 \\ \sin \phi_k & 0 & \cos \phi_k \end{pmatrix}. \quad (7)$$

We obtain the torques that act at each of the vertices  $i$  by substituting this in equation (3) and taking the partial derivative with respect to  $\phi_i$ , viz.,

$$\tau_i = -\partial E / \partial \phi_i. \quad (8)$$

Clearly the torque on the  $\ell$ 'th vertex will be

$$\tau_\ell = -2K \sum_{i,j} \mathbf{r}_i^\dagger V_{ij} \frac{\partial \mathbf{r}_j}{\partial \phi_\ell}. \quad (9)$$

If we define the derivative of matrix  $\mathbf{T}$  as follows

$$\tilde{T}_\ell \equiv \frac{\partial T_\ell}{\partial \phi_\ell}, \quad (10)$$

then we can write

$$\frac{\partial \mathbf{r}_i}{\partial \phi_\ell} = \sum_{k=\ell}^{i-1} (T_1 M_1 \dots \tilde{T}_\ell M_\ell \dots T_k M_k) \mathbf{R}_1. \quad (11)$$

The system is assumed to evolve within a viscous environment, with friction coefficient  $\zeta_r$ , subject to random kicks from the ambient molecules. The dynamics is overdamped, so that there is no acceleration, and the impulse received with each kick dies immediately. We may write the Langevin equation for the positions of the vertices as,

$$\frac{d\mathbf{r}_i(t)}{dt} = \frac{1}{\zeta_r} \mathbf{F}_i + \boldsymbol{\xi}_r(i, t) \quad (12)$$

where  $\boldsymbol{\xi}_r(i, t)$  is a Gaussian distributed noise term, delta correlated in  $i$  and in time. Equivalently, for the state vector  $\boldsymbol{\phi} = (\phi_1, \dots, \phi_N)$ , we have the Langevin equation

$$\frac{d\phi_i(t)}{dt} = \frac{1}{\zeta_\tau} \tau_i + \xi_\tau(i, t) \quad (13)$$

where the torque  $\tau_i$  is a function of all the angles  $\{\boldsymbol{\phi}\}$ ,  $\zeta_\tau$  is the appropriate friction coefficient and  $\xi_\tau$  is again a Gaussian random “force” delta correlated in  $i$  and time. Viewed in this way the dynamics is similar to a pinned interface [16, 17] or a charge

density wave system [18, 19, 20, 21] in 1 + 1 dimensions, where the phase angles of the charge density wave are also cyclic variables. On the other hand, in the present system the interactions are long ranged.

For the discrete, sequential numerical simulation of the evolution of this system, we postulate the following set of rules:

1. Form the self-similar probability distribution

$$P(i) = |\tau_i|^\eta / \sum_i |\tau_i|^\eta \quad , \quad (14)$$

2. Choose a pair of vertices  $(i, i')$  according to this probability distribution over  $\{\tau_i > 0\}$  and  $\{\tau_i < 0\}$ ,
3. Set  $\phi_i(t+1) = \phi_i(t) + \text{sign}(\tau_i)(2\pi/3)$ .

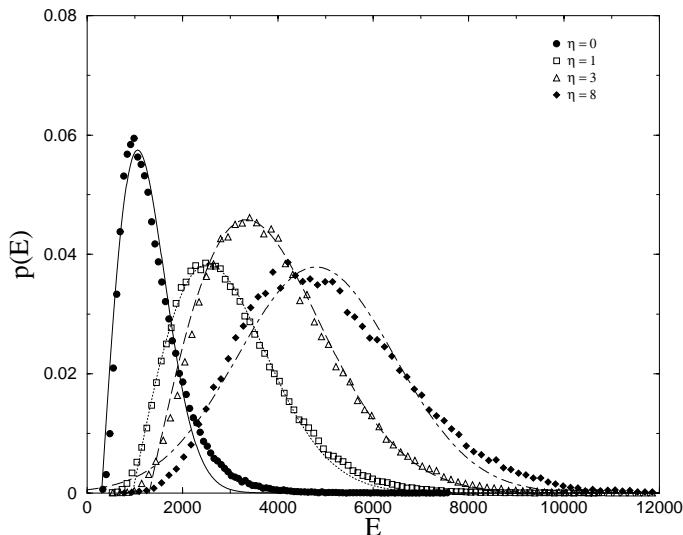
Here  $\eta$  is a tunable parameter defining the dynamics. For large positive values of  $\eta$ , those angles  $\phi_i$  with the maximal conjugate torques are incremented; for negative values of  $\eta$  the small values of the torque are preferred. For  $\eta = 0$  the angles to be incremented are picked randomly. If one choses  $\eta$  to be very large, then we find that there is a large probability that the most recently updated  $\phi_i$  still carries a very large torque, resulting in a jamming of the dynamics. Incrementing the dihedral angles with the large conjugate torques resulted not in the relaxation of these torques but in pumping energy into the system, as when pushing a swing at the top of its arc. After applying the search strategy based on changing the torques according to a distribution, we found that updating the maximal torques ( $\eta > 0$ ) drives the system to a state with relatively high energies, whereas a random search ( $\eta = 0$ ) or preferentially choosing the minimal torques ( $\eta < 0$ ) gives rise to more successful strategies for reaching low lying energy states.

It can be said that  $\eta$  here plays the role of a coarse-graining parameter in the exploration of the energy landscape. We would like to recall a recent paper [22] where an effective inertial effect was introduced into the evolution of a pinned interface, by giving an advantage to that point along the interface which moved last. This led to the coarse graining of the interface, by introducing a persistence time and an associated length scale into the system, whose long time and large scale scaling behaviour, however, was not altered. (In contrast, discouraging the same point from moving at the next time step led to no appreciable change, since this occurred rarely, to start with.) In the present case, changing  $\eta$  has a similar effect; for large  $\eta$  we get very big persistence effects, while values of  $\eta \leq 0$  seem to be qualitatively similar to each other.

### Distribution of energy states and level spacings

The distribution of the energies of the discrete configurational states explored by the chain of  $N = 48$  residues shown in Fig.1, as it evolves under the above dynamics, is shown in Figs.3-6, for both positive and negative  $\eta$ . After the first 5000 steps were discarded, the statistics were taken over 5000 steps of the trajectory. It can be seen that

the shape of the curve does not essentially change with  $\eta$ , while for positive  $\eta$  the peak shifts to successively higher values of the energy, and the distribution is distorted towards a Gaussian, indicating that the states explored are less correlated. These figures should be compared with those reported by ben-Avraham [6] for the density of vibrational states and by Mach et al. [23] for the ultraviolet absorption spectra, and also with the energy histograms obtained by Socci and Onuchic [24] for a Monte Carlo simulation on a lattice model of a proteinlike heteropolymer. Our model seems to be very successful in producing realistic distributions of energy states over the whole range of relevant energies.



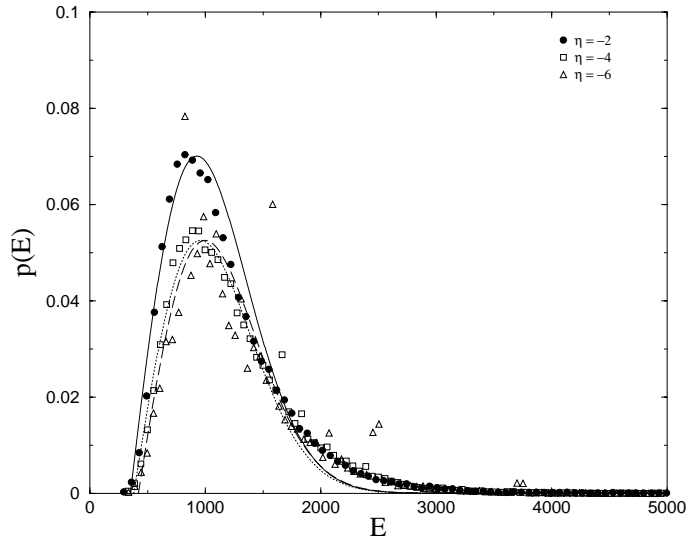
**Figure 3.** The normalized energy histograms, averaged over 10 random initial states for chains of  $N = 48$ , for different  $\eta \geq 0$ , along paths of  $10^4$  steps, with the first 5000 steps discarded. The fits are to the Wigner distribution for  $\eta = 0, 1, 3$  and Gaussian distribution for  $\eta = 8$ .

We have been able to fit the simulation results very successfully with a distribution of the Wigner form (Figs.3,4)

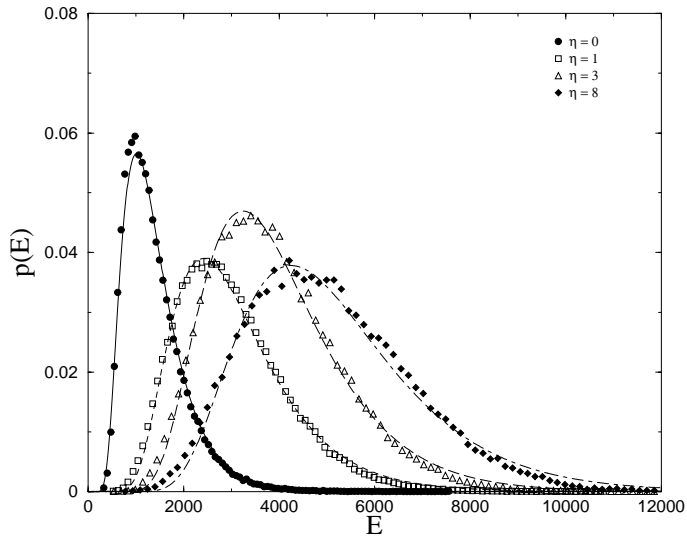
$$f_W(E) = a(E - E_0)e^{-b(E-E_0)^2} , \quad (15)$$

for  $\eta = -6$  to  $\eta = 3$ . Here  $E_0$  corresponds to the offset due to the lowest energy state attained for the different  $\eta$ , and it can be seen that the distribution is shifted to higher values of the energy for higher values of  $\eta$ . The curves become Gaussian for  $\eta = 6$  and  $\eta = 8$ . (See Ref. [9] for the fit parameters.)

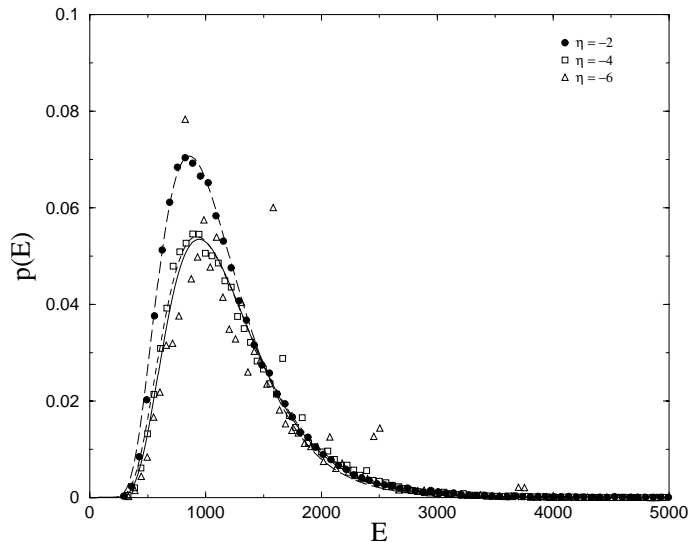




**Figure 4.** The normalized energy histograms, for chains of  $N = 48$ , for different  $\eta < 0$  (see Fig.3). The fits are to the Wigner distribution.



**Figure 5.** The normalized energy histograms along trajectories in phase space for the  $N = 48$  chain, for  $\eta \geq 0$  as in Fig.3, fitted with the "inverse gaussian" distribution given in Eq. (2).



**Figure 6.** Energy histograms for  $\eta < 0$  as in Fig.5, fitted with the “inverse gaussian” distribution given in Eq. (2), for the  $N = 48$  chain.

It should be mentioned that the same energy distributions may be fitted equally well (see Figs.5,6) by the “inverse Gaussian” [12],

$$f_{\text{IG}}(E) = \sqrt{\frac{A}{2\pi E^3}} \exp\left[-\frac{A(E-B)^2}{2B^2 E}\right]. \quad (16)$$

It will be noted that this has the same functional form as the distribution of first passage times over a distance  $d$  for an Ornstein Uhlenbeck process [11] with diffusion coefficient  $D = \sigma^2/2$  and initial drift velocity  $v$ , in the regime of small times, if one makes the further identifications  $A = d^2/(2D)$  and  $B = d/v$ . We postpone until section 4 a discussion of this result. The fit parameters  $A$  and  $B$  are plotted against  $\eta$  in Fig.7. We find that both the “diffusion constant (mobility)” and the “drift velocity” of the phase point along its trajectory in phase space depend on  $\eta$ , being maximum for  $\eta = 0$  and decreasing for positive values of  $\eta$ . For  $\eta < 0$  they essentially stay the same.

#### 4. Ornstein-Uhlenbeck Process and the Wigner Distribution

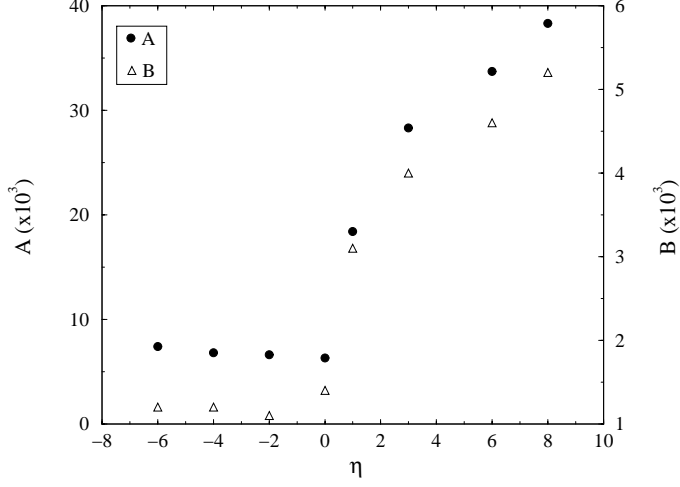
We would now like to show that the energy obeys a stochastic process which can be modelled by Fokker-Planck equations with the Wigner (15) or the inverse gaussian (2) forms as stationary solutions.

We remind the reader that an OU process describes the diffusive motion of a particle subject to a drift velocity proportional to the distance from the origin [11]. It can easily

be seen that such a process for a single particle in one dimension would be described by the Langevin equation,

$$\frac{dx}{dt} = -\frac{1}{\zeta}gx + \xi(t) \quad (17)$$

with a Hookean force  $F(x) = -gx$  and a delta correlated random force  $\xi(t)$ ,  $\langle(\xi(t))^2\rangle = \sigma^2$ .



**Figure 7.** The fit parameters  $A$  and  $B$  are plotted against  $\eta$ . Both  $A$  and  $B$  increase with increasing  $\eta$ , therefore the diffusion coefficient and the drift velocity decrease as  $\eta$  gets bigger.

We would like to show that the energy  $E$  given in Eq.(3) obeys an OU process, under the dynamics given by (12). Since there is no explicit time dependence of  $E$ , we have

$$\frac{dE}{dt} = \sum_i \frac{\partial E}{\partial \mathbf{r}_i} \cdot \frac{\partial \mathbf{r}_i}{\partial t} . \quad (18)$$

Substituting from (12) we get,

$$\frac{dE}{dt} = -\frac{1}{\zeta_r} \sum_i \left( \frac{\partial E}{\partial \mathbf{r}_i} \right)^2 + \sum_i \frac{\partial E}{\partial \mathbf{r}_i} \cdot \xi_i(t) . \quad (19)$$

From (3) we may compute that

$$\sum_i \left( \frac{\partial E}{\partial \mathbf{r}_i} \right)^2 = \frac{NE}{\zeta_r} + \sum_{\substack{i,j,k \\ i \neq j}} c_{ik}c_{jk}(\mathbf{r}_i - \mathbf{r}_k) \cdot (\mathbf{r}_j - \mathbf{r}_k) . \quad (20)$$

We see that the second term is like an average of the products  $(\mathbf{r}_i - \mathbf{r}_k) \cdot (\mathbf{r}_j - \mathbf{r}_k)$  over  $(i, j)$  pairs ( $i \neq j$ ), and for a reasonably isotropic configuration, it vanishes. To the

same approximation, we may assume that the second term in Eq.(19) is itself equal to a Gaussian stochastic noise, i.e., set  $\xi_E(t) = K \sum_{ij} c_{ij}(\mathbf{r}_i - \mathbf{r}_j) \cdot \xi_i(t)$  . This yields the required result, namely,

$$\frac{dE}{dt} = -\frac{NE}{\zeta_r} + \xi_E(t) \quad , \quad (21)$$

which should be compared with (17).

This stochastic equation is equivalent [25] to the Fokker-Planck equation

$$\frac{\partial P(E, t)}{\partial t} = -\frac{\partial}{\partial E} \left[ -\frac{\partial \tilde{\Phi}(E)}{\partial E} P(E, t) - D \frac{\partial P(E, t)}{\partial E} \right] \quad , \quad (22)$$

for the probability distribution of  $E$ , where  $D = 2\langle \xi_E^2 \rangle$  and  $\tilde{\Phi}(E) = \int_0^E (N/\zeta_r) x dx = bE^2/2$ , with  $b = N/\zeta_r$ . The constraints we have placed on our configurational degrees of freedom (see Eq.(22)ff.) require that there be some minimum value of the energy where the probability current vanishes, implying reflecting boundary conditions there, as well as at some  $E_{\max}$ , which we may take to  $\infty$  for all practical purposes. To mimic these constraints we introduce an infinitely high potential barrier at  $E_0$ , while at the same time shifting the point of equilibrium of the Hookean “force” to this point. A convenient choice for a singular potential to add to  $\tilde{\Phi}$ , is  $-\ln(E - E_0)$ . These reflecting boundary conditions at  $E_0$  and at  $\infty$  then lead to a stationary solution  $P(E)$ ,

$$P(E) = a e^{-\Phi(E)} \quad , \quad (23)$$

where  $\Phi(E) = \tilde{\Phi} - \ln(E - E_0)$ , or,

$$\Phi(E) = \frac{1}{2} b(E - E_0)^2 - \ln(E - E_0) \quad (24)$$

and  $a$  is a normalization constant. Substituting (24) in (23) leads to the Wigner formula (15).

A stationary distribution of the inverse gaussian form may be obtained if we modify the quadratic potential  $\tilde{\Phi}$  in a different way to model the constraints in the system, viz.,

$$\Phi(E) = \frac{A}{2B^2E} (E - B)^2 + \frac{3}{2} \ln E \quad . \quad (25)$$

This also leads to reflecting boundary conditions, at  $E = 0$  and  $E \rightarrow \infty$ , and a point of equilibrium at  $E = B$ . As the stationary solution we obtain the inverse gaussian distribution (2), as can be seen from direct substitution into (23).

The distribution of first passage times for the attainment of the optimum solution in such diverse high dimensional optimization problems as fits to X-ray patterns, travelling salesman problems and determination of the lowest energy state for lattice models of protein configurations, have been reported by Li and coworkers [12], to obey the Ornstein-Uhlenbeck form. The plots of these distributions all display a striking similarity to each other, and to the distribution of energy states which we have found in the present problem.

Now we see that if an optimization problem has a quadratic cost function  $C$  which, in terms of the large number of variational parameters in a reasonably isotropic phase space, has the same form as our energy Eq.(3), then the optimization algorithm defines a *dynamics* for  $C$  which may be described by means of an OU process as in Eq.(17), with a repulsive barrier at  $C_{\min}$  and at  $\infty$ . This may be modelled by the same Fokker Planck equation (22), and potentials  $\Phi$ , as we have discussed above.

Recall that for an OU process, with an initial displacement  $x(0) = d$ , the solution for the distribution of first passage times  $t$  through the origin is given by [11],

$$f(t) = \frac{2yd}{\pi^{1/2}\sigma} \left( \frac{\rho}{1-y^2} \right)^{3/2} e^{-\frac{\rho y^2 d^2}{\sigma^2(1-y^2)}} , \quad (26)$$

where  $\rho = g/\zeta$  and  $y = \exp(-\rho t)$ . We see that (26) goes over, in the limit of large times, i.e.  $y \ll 1$ , to

$$f_W(y) = \frac{2d\rho^{3/2}}{\pi^{1/2}\sigma} y e^{-\frac{\rho d^2 y^2}{\sigma^2}} . \quad (27)$$

On the other hand, for very small times, (26) becomes, to leading order,

$$f_{IG}(t) = \frac{2\pi d\sigma^2}{(2\pi\sigma^2 t)^{3/2}} e^{-\frac{(d-vt)^2}{2\sigma^2 t}} \quad (28)$$

where we have defined  $\rho d = v$ . It should be noted that these functions (27) and (28) have the same form, as functions of  $y$  and  $t$ , respectively, as the Wigner and inverse gaussian distributions which we have found above.

## 5. Path Integral Formulation

In this section we discuss a path integral approach to the computation of the statistics of the energy levels. This work is not yet complete, and we will only indicate the general direction.

Let us postulate the following weight for a given configuration of a continuously deformable chain, parametrized by  $s$ , the distance along the chain,

$$P[\mathbf{r}(s)] = \exp \left[ - \int_0^L ds [\lambda |\mathbf{r}(s)|^2 + a_0 |\partial_s \vec{\Gamma}(s)|^2 + a_1 |\partial_s^2 \mathbf{r}(s)|^2 + \dots] \right] . \quad (29)$$

The first term in the exponential is proportional to a radius of gyration squared, and tends to prevent the chain from folding out completely; the value of  $\lambda$  is immaterial for the moment. The gradient squared term is the same as for a path integral representation of a random walk. Its derivation may be found in standard text books on path integrals. [28] The next term makes sure that the chain does not fold too abruptly - i.e., mimicks a sort of self-avoiding potential. In the Fourier transform representation, the integral in the exponential becomes  $\int \frac{dq}{2\pi} [\lambda - a_0 q^2 + a_1 q^4] \tilde{\mathbf{r}}(q) \tilde{\mathbf{r}}(-q)$  . To simplify life, we will take *all*

the residues to be hydrophobic, so that the energy functional, in this continuum approach becomes,

$$\begin{aligned}
 E[\mathbf{r}(s)] &= K/2 \int ds ds' |\mathbf{r}(s) - \mathbf{r}(s')|^2 \\
 &= K/2 \int ds \int d\ell |\mathbf{r}(s) - \mathbf{r}(s + \ell)|^2 \\
 &= LK \left\{ \int \frac{dq}{2\pi} \tilde{\mathbf{r}}(q) \tilde{\mathbf{r}}(-q) - \int \frac{dq}{2\pi} \tilde{\mathbf{r}}(q) \tilde{\mathbf{r}}(-q) \delta(q) \right\} , \quad (30)
 \end{aligned}$$

where  $\delta(q)$  is the Dirac delta-function. Since, strictly speaking, the chain consists of rigid segments, (which, to be able to keep track of the dimensional quantities we will henceforth take to be of length  $a$ ), and is of finite length  $L = Na$ , the  $q$  values are restricted to  $q_n = n\pi/L$ , ranging from  $\pi/L$  to  $N\pi/L \equiv \pi a$ . This tells us that the second term in the above expression for the energy actually vanishes.

Now we may write the probability distribution for the energy states  $E$ ,

$$P(E) = \int \mathcal{D}(\mathbf{r}(s)) P[\mathbf{r}(s)] \delta(E - K/2 \int ds \int d\ell |\mathbf{r}(s) - \mathbf{r}(s + \ell)|^2) . \quad (31)$$

Going over to the Fourier representation, we get,

$$P(E) = \int \frac{d\omega}{2\pi} e^{i\omega E} \int \mathcal{J} \mathcal{D}(\tilde{\mathbf{r}}(q)) \exp\left[- \int \frac{dq}{2\pi} [i\omega k + \lambda - a_0 q^2 + a_1 q^4] \tilde{\mathbf{r}}(q) \tilde{\mathbf{r}}(-q)\right] . \quad (32)$$

where we have defined  $k \equiv LK$ , and  $\mathcal{J}$  is the (as yet to be calculated) Jacobean of the transformation of variables from the  $\mathbf{r}(s)$  to the  $\tilde{\mathbf{r}}(q)$ . The meaning of the functional integral is

$$\mathcal{D}(\tilde{\mathbf{r}}(q)) = \prod_{n=1}^N \frac{d\tilde{\mathbf{r}}(q_n)}{(2\pi)^{d/2}} [\epsilon(\lambda - a_0 q_n^2 + a_1 q_n^4)]^{d/2} \quad (33)$$

where  $d$  is the Euclidean dimensionality of space and  $\epsilon = 1/L$ . The square brackets is the normalization factor for the Gaussian measure. Performing the Gaussian integrals one gets,

$$P(E) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} e^{i\omega E} \prod_{n=1}^N \left[ \frac{1}{i\omega k + \lambda - a_0 q_n^2 + a_1 q_n^4} \right]^{d/2} . \quad (34)$$

For  $d = 2$  and  $d = 4$  the integral over  $\omega$  may actually be done via the calculus of residues, and with  $\lambda$  large enough, all the poles may be arranged to lie in the upper half plane. Nevertheless, the resulting integrals over  $q$  are very rapidly varying and have so far defied computation.

On the basis of the spectral representation of the Wiener integral, however, we believe that we will be able to elucidate the connection between the form we have found for the distribution of energy states and the Wigner distribution, which arises as the distribution of eigenvalue spacings for Gaussian orthogonal matrices [7, 8, 26, 27].

### Acknowledgements

We express our gratitude to the organizers of the IX. Regional Conference on Mathematical Physics. It is a pleasure to thank Ersin Yurtsever for pointing out to us the similarity between our numerical results and the Wigner distribution. One of us (A.E.) acknowledges partial support by the Turkish Academy of Sciences.

### References

- [1] K.A. Dill, S. Bromberg, K. Yue, K.M. Feibig, D.P. Yee, P.D. Thomas and H.S. Chan, *Protein Science* **4**, 561 (1995).
- [2] M.M. Tirion, *Phys. Rev. Lett.* **77**, 1905 (1996).
- [3] T. Haliloglu, I. Bahar, and B. Erman, *Phys. Rev. Lett.* **79**, 3090 (1997).
- [4] B. Erman and K. Dill, *J.Chem. Phys.*, in press.
- [5] B. Erman, "Hydrophobic Collapse of Proteins into their Near-Native Configurations," unpublished.
- [6] D. ben-Avraham, *Phys. Rev. B* **47**, 14 559 (1993).
- [7] E. Wigner, *Proc. Cambridge Phil Soc.* **47**, 790 (1951); *Ann. Math.* **62**, 548 (1955).
- [8] C.E. Porter, *Statistical Theories of Spectra: Fluctuations* (Academic Press, New York, 1965).
- [9] E. Tüzel, A. Erzan, cond-mat/9909350, in print. *J. Stat. Phys.*
- [10] E. Tüzel, A. Erzan, cond-mat/9907397, to appear in *Phys. Rev. E*, **61**, R1041 (2000).
- [11] W. Feller, *An Introduction to Probability Theory and its Applications* (Wiley, N.Y. 1957), Vol. I and II.
- [12] C.N. Chen, C.I. Chou, C.R. Hwang, J. Kang, T.K. Lee, and S.P. Li, *Phys. Rev. E* **60**, 2388 (1999).
- [13] J.L. Green, J. Fan, and C.A. Angell, *J. Phys.Chem.* **98** 13780 (1994).
- [14] B. Erman, I. Bahar, *Macromol.Symp.* **133**, 33 (1998).
- [15] P.J. Flory, *Statistical Mechanics of Chain Molecules*, (Interscience, N.Y., 1969).
- [16] For an insightful review see T. Halpin-Healy and Y.-C. Zhang, *Phys. Rep.* **254**, 215 (1995).
- [17] J. Krug, H. Spohn, "Kinetic roughening of growing surfaces", in C. Godreche ed., *Solids Far from Equilibrium* (Cambridge University Press, Cambridge, 1992).
- [18] A. Erzan, E. Veermans, R. Heijungs and L. Pietronero, *Phys. Rev. B* **41**, 11522 (1990).
- [19] E. Veermans, A. Erzan, R. Heijungs, L. Pietronero, *Physica A* **166**, 447 (1990).

- [20] G. Parisi and L. Pietronero, *Physica A* **179**, 16 (1991).
- [21] O. Narayan and D. Fisher, *Phys. Rev. B* **46**, 11520 (1992).
- [22] Ö. Kayalar, A. Erzan, *Phys. Rev. E.*, to appear.
- [23] H. Mach, D.B. Volkin, C.J. Burke, C.R. Middaugh, "Ultraviolet Absorption Spectroscopy," in B.A. Shirley ed., *Methods in Molecular Biology, Vol. 40: Protein Stability and Folding* (Humana Press, Totowa, N.J. 1995)p. 91-114.
- [24] N. D. Socci, J.N. Onuchic, *J. Chem. Phys.* **103**, 4732 (1995).
- [25] H. Risken, *The Fokker-Planck Equation* (Springer, Berlin, 1984).
- [26] T. Brody, J. Flores, J.B. French, P.A. Mello, A. Pandey, and S.S.S. Wong, *Rev. Mod. Phys.* **53**, 385 (1981).
- [27] M.L. Mehta, *Random Matricies and the Statistical Theory of Energy Levels* (Academic Press, N.Y. 1967).
- [28] F.W. Weigel, *Introduction to Path-Integral Methods in Physics and Polymer Science* (World Scientific, Singapore, 1986); H. Kleinert, *Path Integrals in Quantum Mechanics, Statistics and Polymer Physics* (World Scientific, Singapore, 1990).