Dissipative Dynamics and the Statistics of Energy States of a Hookean Model for Protein Folding

Erkan Tüzel[1, †] **and Ayşe Erzan**[1, 2]

A generic model of a random polypeptide chain, with discrete torsional degrees of freedom and Hookean spring 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 and giving rise to the so-called Wigner distribution. A tunably fine- or coarse-grained sampling of the energy landscape yields a family of distributions for the energies and energy spacings.

I. INTRODUCTION

We present a generic model hamiltonian for polypeptide chains which we believe captures the essential mechanism driving the folding process, namely hydrophobic interactions.[1–7]

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 concieved 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.

Our model involves N coupled, discrete, over-damped torsional degrees of freedom coupled by Hookean forces and driven by random

---

[1] Department of Physics, Faculty of Sciences and Letters, Istanbul Technical University, Maslak 80626, Istanbul, Turkey.

[2] Gürsey Institute, P.O. Box 6, Çengelköy 81220, Istanbul, Turkey.

[†] Present address: Department of Physics, Işik University, Maslak 80670, Istanbul, Turkey.

noise. As a numerical realization of this dissipative system, we explore the phase space under a dynamics based on relaxing pairs of rotational degrees of freedom, namely the dihedral angles, sampled with a probability which is a function of the conjugate torques,

$$P(i) = \frac{|\tau_i|^\eta}{\sum_i |\tau_i|^\eta}. \tag{1}$$

The energy landscape is effectively coarse- or fine-grained by tuning the parameter η. For $\eta = 0$, the dynamics is identical to very high temperature Monte Carlo simulations.

We find that the energy probability distribution obtained from our simulations may be very well represented by a Wigner distribution[8-10] which, for a random quantity S, is given by

$$P(S) \sim S \exp\left(-\frac{\pi}{4} S^2\right) \tag{2}$$

while the coarse grained energy level distributions are comparable with the nth ($n = 1, 2, 3,...$) neighbor energy level spacing statistics encountered[11] in the study of large nuclei. The energy histograms can also be very well fitted with an "inverse Gaussian" (IG) distribution.

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 a modified Ornstein–Uhlenbeck (OU) process.[12] The stationary 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 also able to understand the distribution of relaxation times found for global optimization problems[13] by Li and coworkers.

We have already reported in a separate publication[7] that under Metropolis Monte Carlo dynamics, with random initial conditions, the model exhibits power law relaxation for the initial stages of decay, and at the later stages the relaxation obeys a stretched exponential with the exponent $\beta \simeq 1/4$. This Kohlrausch–Williams–Watts type relaxation behaviour is observed experimentally for real proteins.[14-17] At zero temperature the probability distribution function of the energy steps encountered along a relaxation path in phase space also obeys a stretched exponential form, with another exponent $\alpha \simeq 0.39$. In ref. 7 we show that $\beta = \alpha/(\alpha + 1)$, which yields a value for β in very good agreement with our simulation results.

The paper is organized as follows. In Section II we define our model, in Section III we present our simulation results, in Section IV we show that

the energy obeys an OU process. In Section V we discuss the relationship between this model and other complex systems and outline work in progress.

II. THE MODEL

We consider a model[7] consisting of N residues, treated as point vertices, interacting via Hookean potentials. We have been motivated by the model proposed by Haliloğlu, Bahar, Erman[4] where all interactions between the different residues are governed by confining square-law potentials.[4–6, 18] In our model, however, the covalent bonds between residues are treated as fixed rods of equal length (see Fig. 1). 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 mimics the effective pressure that is exerted on the hydrophobic
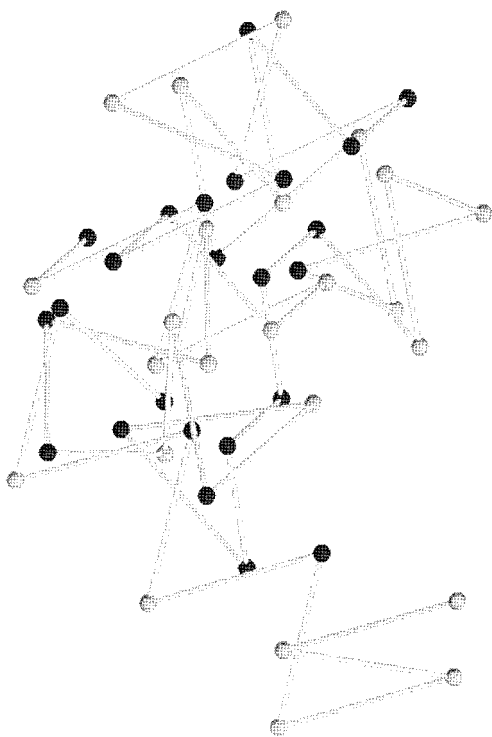

Fig. 1. A chain of $N = 48$ residues, half of which are randomly chosen to be hydrophobic, (darker beads) shown in a random initial configuration. (Generated using RasMol V2.6).

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. It is important to note that we treat all $H-H$ pairs on an equal footing, i.e., there is no "teleological" information that is fed into the system by connecting only those $H-H$ pairs which are close to each other in the native configuration for a particular sequence. 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

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

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

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

We take the bond angles α_i, $i = 1,..., N-1$, to have the alternating values of $(-1)^i \alpha$, with $\alpha = 68°$. The dihedral angles ϕ_i can 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. This has a similar effect to placing the chain on a tetrahedral lattice; however, since the chemical angles are slightly different from $\pi/3$, this is not exactly true, and the configurations are off lattice when compared to a tetrahedral structure. Since the chain has a certain rigidity and persistence, the volume of the folded structures grows with N, the number of residues, for relatively small N. However, for asymptotically large N, we do not expect this to continue to hold; the density of the collapsed state is apt to increase

without bound, and therefore the collapsed configurations cannot be taken seriously in this limit. This is a weak point of the Hookean model.[19]

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 $\mathbf{R}_i$ of unit length representing the chemical bonds, 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,[20] viz.,

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

where we may choose $\mathbf{R}_1$ to lie along any of the Cartesian directions in our laboratory frame without loss of generality. Although the allowed values of the dihedral angles are discrete, the energy is a perfectly well defined, differentiable function of these angles. We obtain the torques that act at each of the vertices i by substituting Eq. (5) in Eq. (3), taking the partial derivative with respect to ϕ_i, viz.,

$$\tau_i = -\partial E/\partial \phi_i \tag{6}$$

and evaluating these derivatives at the given set of values $\{\phi_i\}$.

The system is assumed to evolve within a viscous environment, subject to random kicks from the ambient molecules. We may write the Langevin equation,

$$\frac{d\mathbf{r}_i(t)}{dt} = \frac{1}{\zeta_r} \mathbf{F}_i + \xi_r(i, t) \tag{7}$$

where ζ_r is a friction coefficient and $\xi_r(i, t)$ is a Gaussian distributed noise terms, delta correlated in i and in time. Equivalently, in terms of the state vector $\boldsymbol{\phi} = (\phi_1, ..., \phi_N)$, we have the Langevin equation

$$\frac{d\phi_i(t)}{dt} = \frac{1}{\zeta_\tau} \tau_i + \xi_\tau(i, t) \tag{8}$$

where the torque τ_i is a function of all the angles $\{\boldsymbol{\phi}\}$, ζ_τ is the appropriate friction coefficient and ξ_τ is again a Gaussian random force delta correlated in space and time. Viewed in this way the dynamics is similar to a pinned interface[21, 22] or a charge density wave system[23–25] in $1+1$ dimensions.

The dynamical rules we will employ for the sequential updating of the system loosely correspond to a discrete version of the dissipative system

envisaged in Eq. (8) above. In order to mimic the conservation of angular momentum, we will choose pairs of vertices at a time, turning the ϕ_i in opposite directions.[7] (This does not strictly conserve angular momentum, due to the fact that the axes of rotation are not necessarily parallel; however since the motion is highly dissipative, we do not think this is a big problem.) The choice of vertices for each updating operation is done according to the distribution of torques over the vertices of the chain.

1. The most natural probability distribution we can form out of the torques, without introducing any special scale into the problem, is $P(i) = |\tau_i|^\eta / \sum_j |\tau_j|^\eta$. At each step, for that given configuration of the chain, we form two such independent distributions, one for $\{\tau_i > 0\}$ and another for $\{\tau_i < 0\}$. Note that η is a parameter of our dynamics, and we will be discussing later how changing η affects our results.

2. We choose a pair of vertices (one with positive, the other with a negative torque on them), according to the above distribution.

3. We then update the dihedral angles at the selected vertices, by incrementing them according to $\phi_i(t+1) = \phi_i(t) + \text{sign}(\tau_i) \times (2\pi/3)$.

The fact that the increments are of a fixed size leads in many cases to an "overshooting" of the nearby minimum of the respective torques, thereby resulting in a randomization of the impulses received.

For large positive values of η, those angles ϕ_i with the maximal conjugate torques are incremented; for negative values of η the small values of the torque are preferred. For $\eta = 0$ the angles to be incremented are picked randomly. If one choses η to be vary large, then we find that there is a large probability that the most recently updated ϕ_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. Thus it can be said that η here plays the role of a coarse- or fine-graining parameter in the exploration of the energy landscape.

It should be noted that incrementing preferentially those vertices with high torques on them corresponds, in the language of the Langevin Eq. (8) to relatively small friction coefficients ζ_τ; when $\eta = 0$, one simply has thermal noise, and no force term.

III. DISTRIBUTION OF ENERGY STATES AND LEVEL SPACINGS

In Figs. 2–5 we display the energy distribution of the discrete configurational states explored by our chain of $N = 48$ residues (Fig. 1) under the dynamics outlined in Section II. 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 change too much as the peak shifts to successively higher values of the energy for positive η. For relatively large values of η, however, the distribution is distorted towards a Gaussian, indicating that the states explored are less correlated.

These figures compare very favorably with the energy histograms obtained by Socci and Onuchic[26] for a Monte Carlo simulation on a lattice model of proteinlike heteropolymers.

We have been able to fit the simulation results very successfully with a distribution of the Wigner form (Figs. 2 and 3)

$$f_{\mathrm{W}}(E) = a(E - E_0)\, e^{-b(E - E_0)^2} \tag{9}$$

for $\eta = -6$ to $\eta = 3$, and the parameters for the fit are given in Table Ia. Here E_0 corresponds to the offset due to the lowest energy state attained for the different η, and it can be seen that it shifts the distribution to higher values of the energy for higher values of η. The distributions become
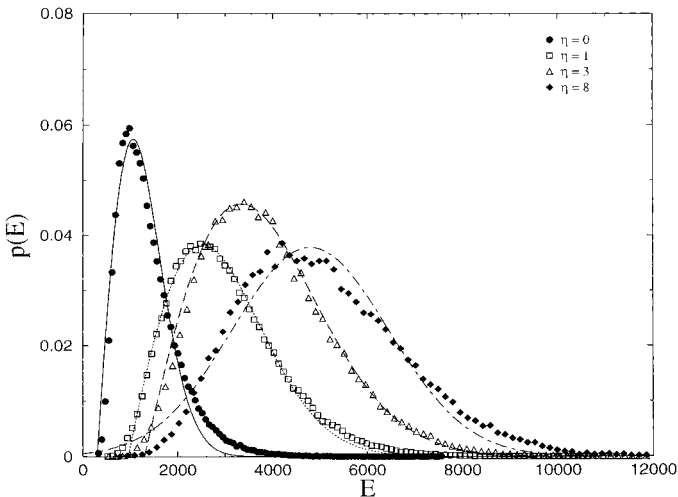

Fig. 2. The normalized energy histograms, averaged over 10 random initial states for chains of $N = 48$, for different $\eta \geqslant 0$, along paths of 10^4 steps, with the first 5000 steps discarded. The fits are to the Wigner distribution for $\eta = 0, 1, 3$ and Gaussian distribution for $\eta = 8$.
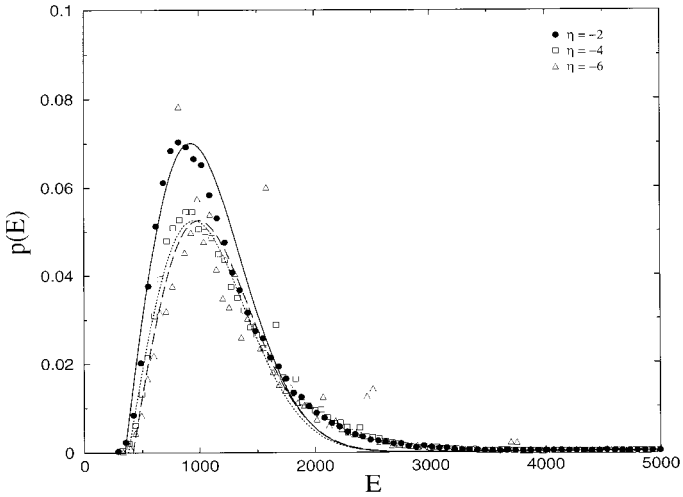
Fig. 3.   The normalized energy histograms, for chains of $N = 48$, for different $\eta < 0$ (see Fig. 2). The fits are to the Wigner distribution.
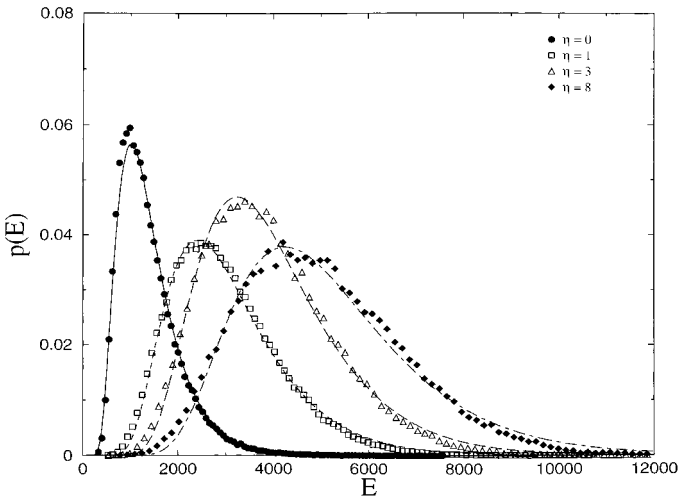

Fig. 4.   The normalized energy histograms along trajectories in phase space for the $N = 48$ chain, for $\eta \geqslant 0$ as in Fig. 2, fitted with the "inverse Gaussian" distribution given in Eq. (10).
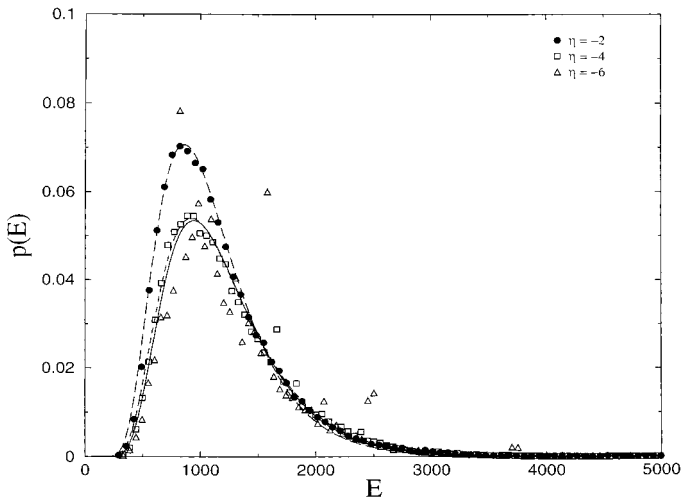
Fig. 5. Energy histograms for $\eta < 0$ as in Fig. 4, fitted with the "inverse Gaussian" distribution given in Eq. (10), for the $N = 48$ chain.

Table Ia. The Parameters a, b and E_0 Used for Fitting the Energy Histograms to the Wigner Distribution $f_W(E) = a(E - E_0) e^{-b(E - E_0)^2}$

η	a (10^{-4})	b (10^{-7})	E_0
-6	1.50	15.0	420
-4	1.50	15.0	380
-2	2.00	15.0	350
0	1.25	8.7	300
1	0.40	2.0	950
3	0.37	1.2	1300

Table Ib. The Mean $\tilde{a}$ and the Variance $\tilde{b}$ Used for Fitting the Energy Histograms to the Gaussian Distribution

η	$\tilde{a}$	$\tilde{b}$ (10^6)
6	4300	2.2
8	4800	2.7

Gaussian for $\eta = 6$ and $\eta = 8$; the results of the Gaussian fits are presented in Table Ib.

It should be mentioned that the same energy distributions may be fitted equally well or better by the "inverse Gaussian,"[13] where the probability density is given by (see Figs. 4 and 5),

$$f_{\mathbf{IG}}(E) = \sqrt{\frac{A}{2\pi E^3}} \exp\left[-\frac{A(E-B)^2}{2B^2E} \right]. \qquad (10)$$

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[12] with diffusion coefficient D 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 parameters for the fits to the parameters A and B are given in Table II. The estimated errors for each fit are also given in the table. We find that both the "diffusion constant (mobility)" and the "drift velocity" of the phase point along its trajectory in phase space depend on η, being maximum for $\eta = 0$ and decreasing for positive values of η. For $\eta < 0$ they essentially stay the same.

We have also considered the statistics of energy differences between successive energy states visited along a trajectory obeying the above dynamics. This does not necessarily mean that the energy differences considered here are nearest neighbors on the energy spectrum; rather these statistics may be considered as a classical analogue of an absorption (or emission) spectrum. We found that the distributions were symmetric for ΔE

Table II. The Parameters A, B Used for Fitting the Energy Histograms to the Inverse Gaussian Distribution [see Eq. (10)]^a

η	A ($\times 10^3$)	ΔA ($\times 10$)	B ($\times 10^3$)	ΔB ($\times 10$)
-6	7.4	48.0	1.2	2.3
-4	6.8	10.0	1.2	0.5
-2	6.6	6.4	1.1	0.3
0	6.3	4.9	1.4	0.4
1	18.4	7.9	3.1	0.8
3	28.3	32.2	4.0	1.2
6	33.7	39.5	4.6	1.4
8	38.3	59.2	5.2	2.0

^{a} The estimated errors are also provided. (Calculated using Levenberg–Marquart algorithm).
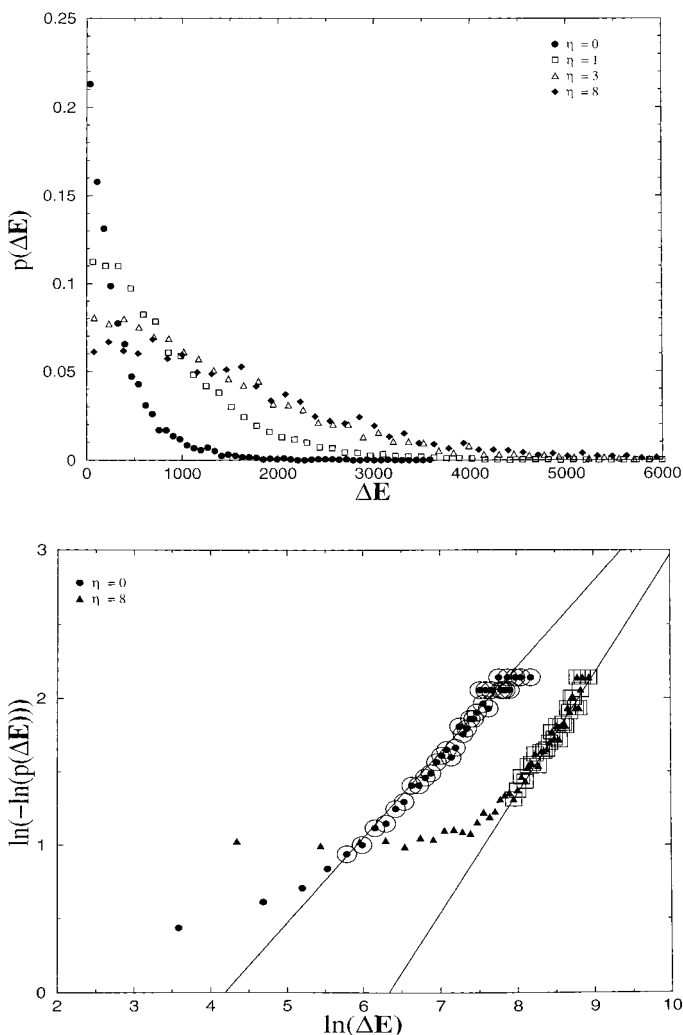
Fig. 6. (a) The distribution of energy steps along a trajectory in phase space according to the η-dynamics of the $N = 48$ chain, for different η. The last 5000 steps along a 10000 steps trajectory were considered. (b) The fits, for $\eta = 0$ (left) and for $\eta = 8$ (right) to the stretched exponential form $\sim \exp(-\Delta E^c)$, for $c = 0.58$ and $c = 0.81$ respectively, in the large ΔE limit.

**Table III.   The Parameters c Used for Fitting the Distribution of
Energy Steps to a Stretched Exponential in the Form
$P(\Delta E) \sim \exp[\Delta E)^c]$, Together with the Correlation Coefficients r^2**

η	c	r^2 (corr. coef.)
-8	0.50	0.89
-6	0.49	0.97
-4	0.54	0.98
-2	0.54	0.98
0	0.58	0.97
1	0.74	0.95
2	0.73	0.96
3	0.81	0.95
4	0.73	0.96
6	0.85	0.95
8	0.81	0.95

negative or positive, and that they obeyed a stretched exponential distribution (for positive ΔE),

$$P(\Delta E) \sim \exp[-(\Delta E)^c)]. \tag{11}$$

The distribution of energy steps for different η are given in Fig. 6a. The plots of $\ln(-\ln(P(\Delta E)))$ versus $\ln(\Delta E)$ for $\eta = 0$ and $\eta = 8$ are shown in
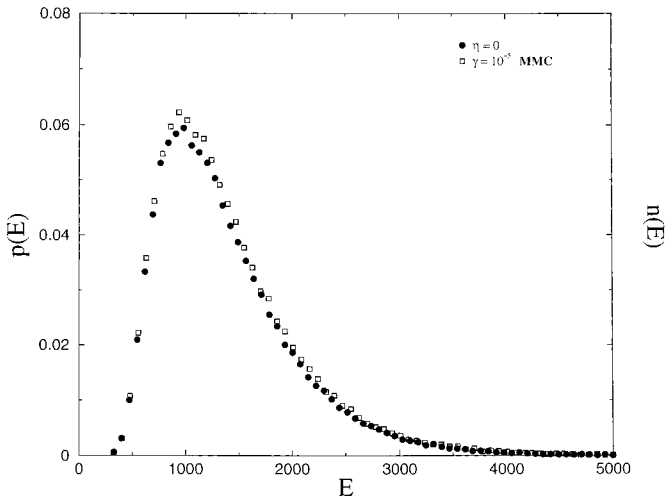

Fig. 7.   Energy histogram for $\eta = 0$ and the density of states, $n(E)$, for $\gamma = 10^{-5}$ obtained for η-dynamics and the Metropolis Monte Carlo respectively.

Fig. 6b. The fits for other η values are equally good. The values of the exponent c are shown in Table III, where it can be seen that c starts from small values for $\eta = -8$ and seems to tend to 1 as η becomes large and positive, again exhibiting a decorrelation effect as the energy landscape is probed with larger and larger η. The correlation coefficients showing the goodness of fit for each η are also given in Table III.

Finally, it is interesting to make a comparison between the energy states explored under the "η"-dynamics and Metropolis Monte Carlo. Since

$$P(E) = n(E) \exp(-\gamma E) \tag{12}$$

$n(E)$ should become identical to $P(E)$ in the limit of $\gamma \to 0$. In Fig. 7 we show our results for $n(E)$ with $\gamma = 10^{-5}$, and $P(E)$ for $\eta = 0$.

IV. 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 (9) or the inverse Gaussian (10) 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.[12] It can easily be seen that such a process for a single particle in one dimension would be described by the Langevin equation,

$$\frac{dx}{dt} = -\frac{1}{\zeta} gx + \xi(t) \tag{13}$$

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

Since there is no explicit time dependence of E, we have

$$\frac{dE}{dt} = \sum_i \frac{\partial E}{\partial \mathbf{r}_i} \cdot \frac{\partial \mathbf{r}_i}{\partial t}. \tag{14}$$

Substituting from (7) we get,

$$\frac{dE}{dt} = -\frac{1}{\zeta_r} \sum_i \left(\frac{\partial E}{\partial \mathbf{r}_i} \right)^2 + \sum_i \frac{\partial E}{\partial \mathbf{r}_i} \cdot \xi_i(t). \tag{15}$$

From (3) we may compute that

$$\sum_i \left(\frac{\partial E}{\partial \mathbf{r}_i}\right)^2 = \frac{2KE}{\zeta_r} + \sum_{\substack{i,\,j,\,k \\ i \neq j}} c_{ik} c_{jk} (\mathbf{r}_i - \mathbf{r}_k) \cdot (\mathbf{r}_j - \mathbf{r}_k). \tag{16}$$

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. (15) 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,

$$\frac{dE}{dt} = -\frac{2KE}{\zeta_r} + \xi_E(t). \tag{17}$$

This stochastic equation is equivalent[27] to the Fokker–Planck equation

$$\frac{\partial P(E, t)}{\partial t} = -\frac{\partial}{\partial E}\left[-\frac{\partial \tilde{\Phi}(E)}{\partial E} P(E, t) - D \frac{\partial P(E, t)}{\partial E} \right] \tag{18}$$

for the probability distribution of E, where $D = 2\langle \xi_E^2 \rangle$ and $\tilde{\Phi}(E) = \int_0^E (2K/\zeta_r)\, x\, dx = bE^2/2$, with $b = 2K/\zeta_r$. The constraints we have placed on our configurational degrees of freedom (see Eq. (4)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 ∞ 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 ∞ then lead to a stationary solution $P(E)$,

$$P(E) = a e^{-\Phi(E)} \tag{19}$$

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

$$\Phi(E) = \tfrac{1}{2} b(E - E_0)^2 - \ln(E - E_0) \tag{20}$$

and a is a normalization constant. Substituting (20) in (19) leads to the Wigner formula (9).

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.,

$$\Phi(E) = \frac{A}{2B^2E}(E-B)^2 + \frac{3}{2}\ln E \qquad (21)$$

This also leads to reflecting boundary conditions, at $E=0$ and $E \to \infty$, and a point of equilibrium at $E=B$. As the stationary solution we obtain the inverse Gaussian distribution (10), as can be seen from direct substitution into (19).

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[13] 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. (13), with a repulsive barrier at $C_{\min}$ and at ∞. This may be modelled by the same Fokker Planck equation (18), and potentials Φ, 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 ref. 12,

$$f(t) = \frac{2yd}{\pi^{1/2}\sigma}\left(\frac{\rho}{1-y^2}\right)^{3/2} e^{-(\rho y^2 d^2)/(\sigma^2(1-y^2))} \qquad (22)$$

where $\rho = g/\zeta$, g is the spring constant, and $y = \exp(-\rho t)$. We see that (22) goes over, in the limit of large times, i.e., $y \ll 1$, to

$$f_W(y) = \frac{2d\rho^{3/2}}{\pi^{1/2}\sigma} y e^{-\rho d^2 y^2/\sigma^2} \qquad (23)$$

On the other hand, for very small times, (22) becomes, to leading order,

$$f_{IG}(t) = \frac{2\pi\, d\sigma^2}{(2\pi\sigma^2 t)^{3/2}} e^{-(d-vt)^2/2\sigma^2 t} \qquad (24)$$

where we have defined $\rho d = v$. It should be noted that these functions (23 and 24) have the same form, as functions of y and t, respectively, as the Wigner and inverse Gaussian distributions which we have found above, and numerically are very similar to each other.

V. DISCUSSION

In this paper we have explored certain features of random heteropolymers undergoing a folding process driven by hydrophobic interactions. Our aim is to seek generic models for those coarse grained properties of real proteins that may be rather universal, and therefore depending only on the statistical properties of random chains, and independent of the particular sequence of amino acids making up the protein chain.

We have adopted the premise that the "native state" should be considered as a dynamical attractor, rather than a single static state, and that therefore the interesting statistical quantities should depend on the way that the phase space is explored under a given dynamics. Modeling hydrophobic interactions with effective Hookean springs, we have proposed a dissipative dynamics with a tunable parameter, revealing different degrees of fine- or coarse-graining of the energy landscape.

We have been able to fit the histograms of the energy states encountered along such dissipative folding trajectories, with distributions of the Wigner or inverse Gaussian form. We observe that the distribution of first passage times in rather general global optimization problems[13] also falls on similar curves. We have provided an explanation for this fact on the basis of the quadratic (Hookean) form of the Hamiltonian.

It is very intriguing that the seemingly universal behaviour which ben-Avraham finds for the density of vibrational states[1] and the ultraviolet absorption spectra reported by Mach *et al.*[28] for various proteins, also display very similar curves. Thus there seems to be a striking universality[3] not only between different protein-like structures, but also between different ranges of length and energy scales. It is very surprising that the density of vibrational states should have a behaviour similar to energy histograms under our dynamics, since the former involve inertial degrees of freedom, while the latter arise from dissipative dynamics.[29]

We would like to make a few concluding remarks about a remaining puzzle, which is the connection between the Wigner distribution, which arises as the distribution of eigenvalue spacings for Gaussian orthogonal matrices,[8–10, 30] and the present limiting form we have found for the distribution of energy states. The puzzle promises to be fruitful, because we believe that the current state of the art of understanding the energy landscapes of proteins is very similar to that of the study of the energy

spectra of very large nuclei in the '50s, when it was realized that the problem might be very advantageously treated as a statistical one. For real, large nuclei, the level spacing distribution, properly scaled with the average density of states for different energies, exhibits a remarkable invariance over the entire energy range.[9] It was then found that the energy levels and nearest-neighbor spacings are governed to a large extent by the statistics of eigenvalues and eigenvalue spacings of orthogonal matrices with a Gaussian distribution of matrix elements.[8–10, 30] The statistics of nuclear energy level spacings approximately obey the so called Wigner "surmise" (2). This is the distribution of eigenvalue spacings for 2×2 Gaussian orthogonal matrices which can be also obtained by considering the square root of the sum of the squares of two independent but identically distributed Gaussian random variables of zero mean.

It is of great interest to note that the so-called Wigner distribution arises also in "quantum chaos"[31–33] and in the energy spectra of large atomic clusters.[34] Thus, there seems to be a universality to the spectral fluctuations of confined systems of sufficient complexity.[9, 35]

It is very intriguing to compare the results for $\eta \geqslant 0$ (Fig. 2) with the numerically obtained nth neighbor spacing distributions of the eigenvalues for Gaussian orthogonal matrices, as reported by Porter,[11] where the identical shift of the peak and tendency to a symmetric Gaussian distribution is found. This we interpret as reinforcing our observation that larger η dynamics results in a more coarse-grained sampling of the energy landscape. Nevertheless, the connection to nth neighbor spacing distributions of eigenvalues of random matrices and the Ornstein–Uhlenbeck process still remains to be understood.

ACKNOWLEDGMENTS

REFERENCES

1. D. ben-Avraham, *Phys. Rev. B* **47**:14559 (1993).
2. 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).
3. M. M. Tirion, *Phys. Rev. Lett.* **77**:1905 (1996).

4. T. Haliloglu, I. Bahar, and B. Erman, *Phys. Rev. Lett.* **79**:3090 (1997).
5. B. Erman and K. Dill, *J. Chem. Phys.* **112**:1050 (2000).
6. B. Erman, Hydrophobic collapse of proteins into their near-native configurations, unpublished.
7. E. Tüzel and A. Erzan, *Phys. Rev. E* **61**:1040 (2000).
8. E. Wigner, *Proc. Cambridge Phil. Soc.* **47**:790 (1951); *Ann. Math.* **62**:548 (1955).
9. T. Brody, J. Flores, J. B. French, P. A. Mello, A. Pandey, and S. S. S. Wong, *Rev. Mod. Phys.* **53**:385 (1981).
10. C. E. Porter, *Statistical Theories of Spectra: Fluctuations* (Academic Press, New York, 1965).
11. C. E. Porter, *J. Math. Phys.* **4**:1039 (1963).
12. W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. II (Wiley, New York, 1957), p. 332ff.
13. C. N. Chen, C. I. Chou, C. R. Hwang, J. Kang, T. K. Lee, and S. P. Li, *Phys. Rev. E* **60**:2388 (1999).
14. H. Frauenfelder, S. G. Sligar, and P. G. Wolynes, *Science* **254**:1598 (1991).
15. J. L. Green, J. Fan, and C. A. Angell, *J. Phys. Chem.* **98**:13780 (1994).
16. B. Erman and I. Bahar, *Macromol. Symp.* **133**:33 (1998).
17. J. Colmenero, A. Arbe, and A. Algera, *Phys. Rev. Lett.* **71**:2603 (1993).
18. A. Yu. Grosberg, *J. Stat. Phys.* **38**:149 (1985).
19. We thank the referee for clarifying this point.
20. P. J. Flory, *Statistical Mechanics of Chain Molecules* (Interscience, New York, 1969).
21. T. Halpin Healy and Y. C. Zhang, *Physics Reports* **254**:215-414 (1995).
22. J. Krug, P. Meakin, and T. Halpin-Healy, *Phys. Rev. A* **45**:638 (1992).
23. A. Erzan, E. Veermans, R. Heijungs, and L. Pietronero, *Phys. Rev. B* **41**:11522 (1990).
24. E. Veermans, A. Erzan, R. Heijungs, and L. Pietronero, *Physica A* **166**:447 (1990).
25. G. Parisi and L. Pietronero, *Physica A* **179**:16 (1991).
26. N. D. Socci and J. N. Onuchic, *J. Chem. Phys.* **103**:4732 (1995).
27. H. Risken, *The Fokker–Planck Equation* (Springer, Berlin, 1984).
28. H. Mach, D. B. Volkin, C. J. Burke, and C. R. Middaugh, Ultraviolet absorption spectroscopy, B. A. Shirley, ed., *Methods in Molecular Biology*, *Vol. 40: Protein Stability and Folding* (Humana Press, Totowa, New Jersey, 1995), pp. 91–114.
29. We are indebted to Nazmi Postacıoğlu for this remark.
30. M. L. Mehta, *Random Matricies and the Statistical Theory of Energy Levels* (Academic Press, New York, 1967).
31. M. V. Berry and M. Robnik. *J. Phys. A* **17**:2413 (1984).
32. E. Yurtsever and J. Brickman, *Phys. Rev. A* **38**:1027 (1988).
33. E. Yurtsever and J. Brickman, *Phys. Rev. A* **41**:6688 (1990).
34. D. Wales, private communication.
35. D. Bohigas, M. J. Giannoni, and C. Schmidt, *Phys. Rev. Lett.* **52**:1 (1984).