Reprinted from The Journal of Physical Chemistry, 1994, 98.

Extracting the Energy Barrier Distribution of a Disordered System from the Instantaneous Normal Mode Density of States: Applications to Peptides and Proteins

John E. Straub* and Ji-Kyung Choi

Department of Chemistry, Boston University, Boston, Massachusetts 02215

Received: June 15, 1994; In Final Form: August 9, 1994[⊗]

A general method for estimating the statistical barrier height distribution in a disordered system is presented. The method is based on the interpretation of the temperature dependence of the instantaneous normal mode density of states. An integral equation is derived which relates the fraction of unstable instantaneous normal modes at a particular temperature to the intrinsic distribution of one-dimensional energy barriers $g(E)$ for the $3N - 6$ internal degrees of freedom in the system. We argue that the overall distribution of multidimensional barriers on the potential energy hypersurface $\Gamma(E)$ is a Gaussian distribution with a mean and variance which can be calculated from the intrinsic barrier height distribution $g(E)$. A technique for solving the integral equation is presented and applied to derive the energy barrier distribution for the isobutyrl-val-ala$_2$-methylamide tetrapeptide, the S-peptide of ribonuclease A, and the bovine pancreatic trypsin inhibitor. We compare our results with the random energy model and suggest how parameters for the statistical Hamiltonian used in that theory might be derived using computer simulation.

I. Introduction

The success of the normal mode picture of solids in the description of the temperature dependence of the heat capacity is well-known.[1] In such a picture, the potential energy of the crystal is approximated by a quadratic function of the coordinates, centered about an equilibrium minimum energy conformation of the crystal. The normal mode frequencies are real, and motion consists entirely of stable oscillations of the solid atoms about well defined minimum energy positions. Of course, for a disordered system such as a liquid or biomolecule there may be many important minima contributing to the thermodynamics and this simplest normal mode picture breaks down.

Fortunately, the thermodynamics of such disordered systems has been elegantly described by Stillinger and Weber.[2] In their inherent structure theory of liquids, the conformation space is divided into regions called basins of attraction. All conformations lying within a basin of attraction will be mapped by a steepest descent quench onto a single point which is the energy minimum of the basin. The partition function is written as a sum over the contributions of all basins. This theory represents a profitable way of thinking about liquids and has provided important insight into problems involving the melting of solids and liquid state dynamics.

One picture of liquid state dynamics based on the inherent structure model was presented by Zwanzig.[3] In that picture, the dynamics of a liquid consists of periodic oscillations in a well with occasional transitions between wells. The parameters of the Zwanzig hopping model are the density of vibrational states for the wells in which the system modes oscillate and the hopping frequency of transition between wells. This theory assumes that the hopping frequency is small compared with the frequency of oscillation in a well. As such, it is most appropriate for supercooled liquids. Keyes and co-workers[4,5] and others[6] have used the instantaneous normal mode density of states[7-9] to determine the parameters of the Zwanzig model and probe the potential energy hypersurface. Their work has provided considerable insight into the dynamics of supercooled liquids and atomic and molecular clusters.

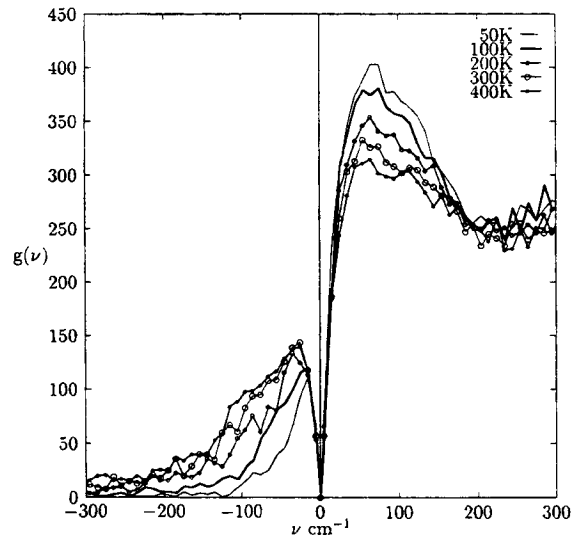

Figure 1. Instantaneous normal mode vibrational density of states shown for the protein BPTI at a variety of temperatures.

The instantaneous normal modes are computed using an equilibrium MD trajectory where one occasionally takes an instantaneous configuration and diagonalizes the force constant matrix to determine the eigenfrequencies. Since the system need not be at a mechanically stable configuration, some of the eigenvalues may be negative and the associated eigenfrequencies imaginary. The instantaneous normal mode density of states is calculated by repeating this procedure for many configurations along the trajectory and averaging until convergence. Representative spectra are shown in Figure 1 for a globular protein as a function of temperature. As the temperature increases, the relative number of imaginary modes increases.

The resulting density of states is an intensive thermodynamic property for the system. As such, we can use equilibrium classical statistical mechanics to define the fraction of unstable modes at a temperature T, $f_u(T)$, in terms of an average over the multidimensional potential surface $U(\bar{r})$. We can attribute the regions of negative curvature to the presence of barriers and anharmonic regions of the potential. To a first approxima-

[⊗] Abstract published in *Advance ACS Abstracts,* September 15, 1994.

0022-3654/94/2098-10978$04.50/0

tion we can assume that all of the unstable modes are related to motion in the barrier regions. What remains is to extract the distribution of barrier heights from the fraction of unstable modes.

Recently, Straub and Thirumalai proposed that the fraction of unstable modes for the multidimensional system, $f_u(T)$, can be written as the fraction of unstable modes for an individual coordinate with barrier height E, $\bar{f}_u(T,E)$, averaged over the intrinsic distribution of barriers for the $3N - 6$ independent modes, $g(E)$.[10] The result is

$$f_u(T) = \int_0^\infty dE\, g(E)\bar{f}_u(T,E) \tag{1}$$

which is a Fredholm integral equation of the first kind.[11] By proposing a model for $\bar{f}_u(T,E)$, it was possible to find a $g(E)$ which fitted the fraction of unstable modes of the S-peptide of ribonuclease A. The solution was based on an informed guess at the functional form of $g(E)$ consisting of a constant distribution of low-energy barriers (less than 0.2 kcal/mol) and a Poisson distribution of higher energy barriers. This distribution is the intrinsic barrier height distribution for the set of one-dimensional coordinates which can become unstable at high enough temperature. An alternative method for determining the barrier height distribution based on the temperature dependence of the fraction of unstable modes has been developed by Keyes and applied to a supercooled Lennard-Jones fluid.[12]

In this paper we derive this integral equation starting from the inherent structure theory of complex systems. We also present a numerical method for solving eq 1 for $g(E)$ given data for $f_u(T)$ and using a reasonable assumption for the form of $\bar{f}_u(T,E)$. We show that, for the simplest case of a single intrinsic barrier height, the distribution of barriers on the multidimensional potential energy hypersurface, $\Gamma(E)$, is approximately Gaussian. For the general case, when there is a continuous distribution of intrinsic barrier heights, the total distribution of barrier heights $\Gamma(E)$ is also Gaussian with a mean and variance which are proportional to the first and second moments of the intrinsic barrier height distribution $g(E)$ of the $3N - 6$ individual coordinates.

There has been impressive progress in developing methods to isolate reaction paths and energy barriers in complex many body systems.[13-15] However, it remains a very difficult problem to measure the statistical distribution of barriers. Our method provides a relatively simple and straightforward way to obtain a realistic approximation to this important property of the potential energy hypersurface. Results for the barrier height distributions of a tetrapeptide, the S-peptide of ribonuclease A, and the bovine pancreatic trypsin inhibitor protein are presented. A connection is made between the distribution of barrier heights predicted in this model and that of the random energy model (REM) of Bryngelson and Wolynes (which is also Gaussian). The result is one means of deriving parameters for statistical Hamiltonians using computer simulation.

II. Derivation of Eq 1 Using the Inherent Structures Theory

We derive eq 1 from the inherent structures theory of Stillinger and Weber making two simplifying assumptions. The partition function for the ensemble of states for N distinguishable particles is

$$Q(T) = \left(\prod_{k=1}^{N}\Lambda_k^d\right)^{-1}\int d\vec{r}\; e^{-\beta U(\vec{r})} = \left(\prod_{k=1}^{N}\Lambda_k^d\right)^{-1}Z(T) \tag{2}$$

where $\Lambda_k = (h^2/2\pi m_k k_B T)^{1/2}$ is the thermal de Broglie wavelength for a particle of mass m_k, $U(\vec{r})$ is the potential energy, and $Z(T)$ is the configuration integral. Following Stillinger and Weber,[2] we think of our dN dimensional configuration space in terms of a set of basins of attraction. Any point in configurational space (disregarding the sets of zero measure consisting of saddle points and ridges) will be mapped by a steepest descent quench to a minimum on the potential surface. Labeling the basins α and the regions of configuration space $R(\alpha)$ which form basins of attraction draining to the αth minimum of energy E_α, the partition function may be rewritten

$$Z(T) = \sum_\alpha \int_{R(\alpha)} d\vec{r}\; e^{-\beta U(\vec{r})} \tag{3}$$

Furthermore, the potential energy in the αth basin may be written exactly as $U(r) = E_\alpha + \Delta_\alpha U(\vec{r})$, leading to

$$Z(T) = \sum_\alpha e^{-\beta E_\alpha}\int_{R(\alpha)} d\vec{r}\; e^{-\beta\Delta_\alpha U(\vec{r})} \tag{4}$$

This is an exact expression for the partition function. To develop an expression for the fraction of unstable modes, which is a thermodynamic quantity, we define an operator which counts the number of unstable modes and whose ensemble average is $f_u(T)$. It is

$$\hat{F}_u = \frac{1}{3N-6}\sum_k \Theta[-\omega_k^2(\vec{r})] \tag{5}$$

where Θ is a Heaviside function and ω_k is the eigenfrequency associated with the kth instantaneous normal mode coordinate q_k. The operator $\hat{F}_u$ acts as follows. At a given position $\vec{r}$ the potential is expanded quadratically and the $3N - 6$ normal modes are determined. The Heaviside function $\Theta[-\omega_k^2(\vec{r})]$ counts as one the kth normal mode if the eigenvalue ω_k^2 is negative. The sum over k of the Heaviside function counts the total number of imaginary modes at the point $\vec{r}$ normalized to the total number of normal modes. The operator $\hat{F}_u$ averaged over all $\vec{r}$ is the fraction of unstable modes

$$f_u(T) = \langle\hat{F}_u\rangle = \frac{1}{Z(T)}\sum_\alpha e^{-\beta E_\alpha}\int d\vec{r}\;\hat{F}_u e^{-\beta\Delta_\alpha U(\vec{r})} \tag{6}$$

The fraction of unstable modes $f_u(T)$ can be measured in an equilibrium molecular dynamics or Monte Carlo simulation. To make a connection between this expression for the fraction of unstable modes and the integral equation eq 1 we make a few key approximations.

(1) We assume that while there exists a distribution of basins α with energy minima E_α, the potential energy as seen from the minimum of any basin will be identical to any other. This is what Madan and Keyes call the "equivalent minima" assumption,[5] which amounts to assuming that $\Delta_\alpha U(\vec{r}) = \Delta U(\vec{r})$ is independent of α. As a result we find

$$f_u(T) = \sum_\alpha e^{-\beta E_\alpha}\int d\vec{r}\;\hat{F}_u e^{-\beta\Delta_\alpha U(\vec{r})}\Big/\sum_\alpha e^{-\beta E_\alpha}\int d\vec{r}\; e^{-\beta\Delta_\alpha U(\vec{r})} \tag{7}$$

$$= \int d\vec{r}\;\hat{F}_u e^{-\beta\Delta U(\vec{r})}\Big/\int d\vec{r}\; e^{-\beta\Delta U(\vec{r})} \tag{8}$$

Our expression for $f_u(T)$ is now free of the sum over basins, since we assume that the fraction of unstable modes calculated in any basin will equal that in any other. (Later we will see that this has important consequences for the energy landscape.)

(2) We assume that within a basin we can define a set of orthogonal coordinates which are good throughout the basin but may be anharmonic. The coordinates q_j with associated eigenvalues $\omega_j^2(q_j)$ may be positive or negative (real or imaginary frequency ω_j). We can write the fraction of unstable modes as

$$f_u(T) = \int d\vec{q} \frac{1}{3N-6}\sum_k \Theta[-\omega_k^2(q_k)]\prod_j e^{-\beta\Delta U_j(q_j)} / \int d\vec{q} \prod_j e^{-\beta\Delta U_j(q_j)} \quad (9)$$

where $d\vec{q} = dq_1 dq_2 ... dq_{3N-6}$. Because each $\Theta[-\omega_k^2(q_k)]$ is a function of one coordinate only, for each term in the summation we can integrate over all other coordinates $j \neq k$ and are left with a ratio of integrals over a single coordinate

$$f_u(T) = \frac{1}{3N-6}\sum_k [\int dq_k \Theta[-\omega_k^2(q_k)] e^{-\beta\Delta U_k(q_k)} / \int dq_k e^{-\beta\Delta U_k(q_k)}] \quad (10)$$

Inserting the identity operator $\Theta[-\omega^2] + \Theta[\omega^2] = 1$ in the denominator of each term in the sum, we can rewrite this result as

$$f_u(T) = \frac{1}{3N-6}\sum_k \left[\frac{p(T,E_k)}{1+p(T,E_k)}\right] \quad (11)$$

where the function

$$p(T,E_k) = \int dq_k \Theta[-\omega_k^2(q_k)] e^{-\beta\Delta U_k(q_k)} / \int dq_k \Theta[\omega_k^2(q_k)] e^{-\beta\Delta U_k(q_k)} \quad (12)$$

represents the probability of a point along q_k being found in a region of imaginary frequency ω_k relative to the probability of it being found in a region of real frequency and E_k is the characteristic barrier height for motion along coordinate q_k. Defining the function

$$\bar{f}_u(T,E_k) = \frac{p(T,E_k)}{1+p(T,E_k)} \quad (13)$$

as the absolute probability that the kth coordinate will be found in a region of negative curvature of the potential, we can rewrite eq 11 as

$$f_u(T) = \frac{1}{3N-6}\sum_k \bar{f}_u(T,E_k) \quad (14)$$

where the sum extends over the $3N-6$ normal modes of the system. If we approximate the sum as an integral

$$\frac{1}{3N-6}\sum_k \rightarrow \int_0^\infty dE\, g(E) \quad (15)$$

where $g(E)$ is the distribution of barrier heights for the $3N-6$ independent coordinates. Not all coordinates can become unstable. Some modes, such as local bond stretches modeled as harmonic oscillators, are stable at all energies. The integral over the distribution is the fraction of degrees of freedom which can, at high enough energies, become unstable. That is

$$\int_0^\infty dE\, g(E) = \frac{N_u}{3N-6} \quad (16)$$

where N_u is the number of degrees of freedom which can become unstable. The final result is

$$f_u(T) = \int_0^\infty dE\, g(E)\bar{f}_u(T,E) \quad (17)$$

which is eq 1 proposed earlier.[10]

To extract $g(E)$ from our knowledge of $f_u(T)$, we must solve eq 1 for a particular model of $\bar{f}_u(T,E)$. We invoke the simplest possible model for $\bar{f}_u(T,E)$ and assume that the potential is periodic and piecewise parabolic where the force constant in the well and barrier are the same. For this one-dimensional coordinate with barrier height E we can calculate the fraction of unstable modes. At a temperature T, $p(T,E)$ of eq 13 is the probability of being in the neighborhood of the barrier (region of negative curvature) relative to the probability of being in the neighborhood of a well (region of positive curvature).[16] For our model it can be expressed as

$$p(T,E) = e^{-2x^2}[\int_0^x dt\, e^{t^2} / \int_0^x dt\, e^{-t^2}] \quad (18)$$

which is the ratio of Dawson's integral and the error function and $x = \sqrt{\beta E/2}$.

To simplify the model further, we propose an accurate approximation to eqs 13 and 18 in the form

$$\bar{f}_u(T,E) = \frac{e^{-(2/3)\beta E}}{1+e^{-(2/3)\beta E}} \quad (19)$$

This approximation is very useful. It provides a starting point for the development of a general method for solving our integral equation eq 1 to any desired accuracy using a series expansion (see the Appendix). It also has the form of a probability of being in the excited state of a two-level system where the two levels correspond to the regions of configurations space where motion is stable and unstable. However, the energy gap is $2E/3$ rather than the barrier height E. To reach the region of negative curvature, a mode must have energy above the inflection point, which is $E/2$ in this model.

III. Calculation of the Total Energy Barrier Distribution for the Potential Hypersurface

First let us consider the simplest case where each internal coordinate has an intrinsic barrier height of E_0[5,10] so that

$$g(E) = \left(\frac{N_u}{3N-6}\right)\delta(E-E_0) \quad (20)$$

where N_u is the number, out of $3N-6$ coordinates, which can, at high enough energies, become unstable. The resulting fraction of unstable modes as a function of temperature is then

$$f_u(T) = \left(\frac{N_u}{3N-6}\right)\frac{e^{-(2/3)\beta E_0}}{1+e^{-(2/3)\beta E_0}} \quad (21)$$

The intrinsic barrier for each coordinate is E_0. What is the total distribution of barrier heights for the multidimensional potential energy hypersurface in this model? To a good approximation, it is a Gaussian distribution. We assume that, for each of the N_u coordinates, over one half of the configuration space, the coordinate is in a region of negative curvature and, over the other half, it is in a region of positive curvature. For the case of two coordinates, there would be four regions of configuration space with equal volume: one region which is a two-

dimensional minimum, two regions with a one-dimensional saddle of barrier height E_0, and a fourth region of a two-dimensional saddle with barrier height $2E_0$. In general, for N_u coordinates of this form the normalized distribution of energy barriers on the N_u dimensional potential hypersurface is given by the binomial coefficients

$$\Gamma(E) = \frac{1}{2^{N_u}} \frac{N_u!}{(N_u - E/E_0)!(E/E_0)!} \tag{22}$$

For large N_u, this distribution of binomial coefficients is well approximated by a Gaussian[17]

$$\Gamma(E) = \frac{1}{\sqrt{2\pi\Delta E^2}} \exp(-(E - \bar{E})^2/2\Delta E^2) \tag{23}$$

with mean and variance

$$\bar{E} = \frac{N_u}{2}E_0 \qquad \Delta E^2 = \frac{N_u}{4}E_0^2 \tag{24}$$

For this most simple model of independent coordinates with a single intrinsic barrier height, the total barrier distribution on the potential hypersurface is to a good approximation Gaussian.

Now consider the general case where the distribution of intrinsic energy barriers for the N_u coordinates is given by the continuous distribution $g(E)$. We divide the N_u dimensional configuration space into a product of r subspaces. Within the kth subspace, we assume there are N_k coordinates and that each has an intrinsic energy barrier of E_k. The sum of the subspace dimensions

$$\sum_{k=1}^{r} N_k = N_u \tag{25}$$

If we further assume that for each barrier energy E_k the number of coordinates N_k is large, then within each subspace k the energy barrier distribution is a binomial distribution (as above) which we approximate as a Gaussian distribution with mean and variance

$$\bar{E}_k = \frac{N_k}{2}E_k \qquad \Delta E_k^2 = \frac{N_k}{4}E_k^2 \tag{26}$$

If we have many such subspaces of configuration space, we can estimate the distribution of intrinsic energy barriers E_k to be the continuous distribution $g(E)$. From the properties of the normal distribution, it follows that the probability of finding a barrier of energy $E \geq 0$ on the total potential energy hypersurface of the N_u coordinates is

$$\Gamma(E) = \frac{1}{\sqrt{2\pi\Delta E^2}} \exp(-(E - \bar{E})^2/2\Delta E^2) \tag{27}$$

with mean and variance

$$\bar{E} = \frac{N_u}{2}\int_0^\infty g(E)E\,dE / \int_0^\infty g(E)\,dE = \frac{N_u}{2}\langle E\rangle_g$$

$$\Delta E^2 = \frac{N_u}{4}\int_0^\infty g(E)E^2\,dE / \int_0^\infty g(E)\,dE = \frac{N_u}{4}\langle E^2\rangle_g \tag{28}$$

where $\langle ...\rangle_g$ signifies an average over the distribution of intrinsic barriers $g(E)$. Therefore, within the approximations given here, which should be quite reasonable for a biomolecule, the distribution of energy barriers is expected to be Gaussian. Recall, however, that in our initial derivation we assumed that the total partition function could be divided into many equivalent basins, and our results pertain to a single basin on the potential hypersurface.

In an equilibrium simulation with temperature T, the most probable barrier height encountered during the simulation will be the maximum of $\Gamma(E)\exp(-\beta E)$, which peaks at

$$E_{mp} = \bar{E} - \beta\Delta E^2 \tag{29}$$

The distribution in eq 27 is only valid for energies $E \geq 0$. As such, for temperatures $T < \Delta E^2/k_B\bar{E}$ the most probable barrier height E_{mp} will be $E \cong 0$. Note that the mean and variance scale as N_u, the number of modes which can become unstable, and that

$$\frac{(\Delta E^2)^{1/2}}{\bar{E}} \simeq 1/\sqrt{N_u} \tag{30}$$

For large N_u the distribution will be strongly peaked at energy E_{mp}.

The non-Arrhenius temperature dependence in the form of $\exp(-\text{const}/T^2)$ has been discussed by Bassler,[18] Zwanzig,[19] and Thirumalai, Mountain, and Kirkpatrick[20] and used to describe the relaxation in glassy systems. This form arises when there is a Gaussian roughness in the potential energy surface. This Gaussian roughness emerges as a generic feature of our simple, independent mode model of the potential energy hypersurface.[21]

IV. Predictions of the Random Energy Model

Bryngelson and Wolynes (BW)[22,23] have applied the REM[24] to study properties of protein folding. In the REM of BW, the probability that a protein of N_0 amino acids each with $\nu + 1$ states will be found in a state with energy between E and $E + dE$ is given by $g_{REM}(E)\,dE$ where $g_{REM}(E)$ is a Gaussian statistical distribution of energies[22]

$$g_{REM}(E) = \frac{1}{\sqrt{2\pi\Delta E_{REM}^2}} \exp(-(E - \bar{E}_{REM})^2/2\Delta E_{REM}^2) \tag{31}$$

with mean

$$\bar{E}_{REM} = N_0[-\epsilon\varrho - L\varrho^2] \tag{32}$$

and variance

$$\Delta E_{REM}(\varrho)^2 = N_0[\Delta\epsilon^2(1 - \varrho) + \Delta L^2(1 - \varrho^2)] \tag{33}$$

ϱ is the fraction of residues in their native state, ϵ and $\Delta\epsilon^2$ fix the monomer or intraresidue energetics, and L and ΔL^2 describe pair interaction or interresidue energetics. The parameters ϵ, $\Delta\epsilon^2$, L, and ΔL^2 must be specified, and this has been done by BW in several ways. In each case, the relations

$$\left(\epsilon - \frac{\Delta\epsilon^2}{2T_s}\right) = -0.2T_s$$

$$\left(L - \frac{\Delta L^2}{2T_s}\right) = 2.6T_s \tag{34}$$

are assumed, where T_s sets the temperature scale for the problem.

At a given energy E and number of amino acids N_0, the probability that the protein has an energy less than E is

$$p^<(E,N_0) = \int_{-\infty}^{\infty} dE\, g_{REM}(E,N_0) \tag{35}$$

For the REM, Madan and Keyes have shown[5] that the fraction

of unstable modes at a given temperature can be written as

$$f_u(T,N_0) = \int_{-\infty}^{\infty} dE\, p^<(E,N_0) g_{REM}(E,N_0) e^{-\beta E} / \int_{-\infty}^{\infty} dE\, g_{REM}(E,N_0) e^{-\beta E} \quad (36)$$

This follows from the fact that the probability of finding a saddle with α directions of downward curvature is equivalent to the probability of finding a point with α neighbors of lower energy and $N_0\nu - \alpha$ neighbors of higher energy, which is $p^<(E,N_0)^\alpha p^>(E,N_0)^{(N_0\nu-\alpha)}$ weighted by the binomial coefficient for the number of possible ways of constructing such a saddle. Averaging the fraction of unstable states for each saddle over this distribution provides the average fraction of unstable states at a given energy E, which is simply $p^<(E,N_0)$. Averaging this result over all energies leads to eq 36. The REM predicts that in the high-temperature limit

$$\lim_{T\to\infty} f_u(T,N_0) = \int_{-\infty}^{\infty} dE\, p^<(E,N_0) g_{REM}(E,N_0) \quad (37)$$

$$= \int_{-\infty}^{\infty} dE\, g_{REM}(E,N_0) \int_{-\infty}^{E} dE'\, g_{REM}(E',N_0) = \frac{1}{2} \quad (38)$$

which can be shown through integration by parts. This is a larger fraction than we find for the peptides and proteins, where not all degrees of freedom can become unstable even at the highest temperatures. However, if we examine the temperature dependence of the scaled function $f_u(T)/f_u(\infty)$, we find that the REM provides a reasonable qualitative description of the simulation data.

What is the distribution of barrier heights on the REM potential hypersurface? At energy E, the relative probability that a state of the protein is a minimum is $p^>(E,N_0)^{N_0\nu}$. The probability that a state of the protein is a barrier is then the relative probability of being a barrier, $1 - p^>(E,N_0)^{N_0\nu}$, multiplied by the probability of having an energy E, which is $g_{REM}(E,N_0)$. It follows that the total distribution of barriers for the potential energy hypersurface described by the REM is

$$\Gamma_{REM}(E) = g_{REM}(E,N_0)(1 - p^>(E,N_0)^{N_0\nu}) \simeq g_{REM}(E,N_0) \quad (39)$$

The approximation follows from the fact that when N_0 is large, the relative probability of a state being a minimum $p^>(E,N_0)^{N_0\nu}$ is small, and most of the potential surface is composed of barriers, as was first pointed out by Keyes and co-workers.[5] The probability of being on a barrier of energy E not too small is to a good approximation the probability of having energy E, which is just $g_{REM}(E,N_0)$. Comparing eqs 27 and 39, we see that the width of the distribution of barriers $\Gamma(E)$ should be approximately equal to the width of the statistical distribution of energy states in the REM.

V. Application to Peptides and Proteins

Our calculations of dynamics and vibrational normal modes were carried out using our version 22 CHARMM molecular simulation program.[25] Studies of the tetrapeptide and the S-peptide employed the version 19 polar hydrogen parameter set, while data for the bovine pancreatic trypsin inhibitor (BPTI) were generated using a preliminary set of the version 22 all atom empirical CHARMM force field. A nonbonded truncation using the atom-based shifting function

$$S(r) = \begin{cases} (1 - (r^2/r_c^2))^2 & 0 < r \le r_c \\ 0 & r_c < r \end{cases} \quad (40)$$

was used throughout, where $r_c = 9$ Å for the tetrapeptide study and $r_c = 12$ Å for the studies of the S-peptide and the BPTI protein. The tetrapeptide and BPTI data were generated using a constant vacuum dielectric of $\epsilon = 1$ while the S-peptide study employed a distance dependent dielectric $\epsilon = 4r$. Calculations for the tetrapeptide and S-peptide were performed on a Titan 3040 computer at Boston University. Calculations on the BPTI protein were completed on the Cray YMP at the NCSA located at the University of Illinois and the Cray C90 at the Pittsburgh Supercomputing Center (PSC).

The numerical method for computing $g(E)$ given $f_u(T)$ by inverting eq 1 is described in detail in the Appendix. To compute $g(E)$ we must first calculate the inverse Laplace transform of the fraction of unstable modes $\mathcal{L}^{-1}[f_u](E)$ as a function of energy E given the data for $f_u(T)$ at a series of temperatures. Inverse Laplace transforms are notorious for their sensitivity to the method chosen. However, we can test our result by inserting $g(E)$ into eq 1 and comparing the predicted value of $f_u(T)$ with the input simulation data.

We first fit the simulation data $f_u(T)$ to a series of functions whose inverse Laplace transform is well defined. In this work we represent the inverse transform of $f_u(T)$ in terms of Poisson distribution functions $P_n(E) = E^n \exp(-E)/n!$. The resulting exact expressions for $g(E)$ are given in the Appendix. There are alternative representations for $f_u(T)$ in terms of, for example, impulse functions. In practice we find that eq A.11 provides a good representation, and it was used throughout the study.

Once the fraction of unstable modes as a function of temperature was determined, the data was fit to eq A.9 using the Levenberg–Marquardt algorithm[26] with $n_l = 0$, 1, and 1. The barrier height distribution was then computed using eq A.11, where the sum was extended to include 2^{20} terms. In practice, fewer terms are needed to determine the low-energy portion of the distribution. The portion of the distribution which is important for the form of $f_u(T)$ when T = 0–300 K is established by the first 2^{12} terms. However, the fraction of unstable modes at very high temperatures (exceeding 1000 K) is sensitive to the details of the high-energy tail of the distribution and required a considerable number of terms to reach convergence.

We note that the force fields used in this study were developed to provide a reasonable description of the protein potential hypersurface at or below room temperature. However, we have used simulation data for temperatures as high as 10 000 K. At these high temperatures we expect that the harmonic approximation to the bond stretching potential will be less accurate (although k_BT remains below most bond dissociation energies), as will approximations concerning the temperature independence of the dielectric constant. However, we do not expect our results for the statistical barrier height distribution, which we argue is dominated by dihedral angle potentials and nonbond potential interactions responsible for hydrogen bonding and hydrophobic interactions, to be very sensitive to these changes. The purpose of our study is to provide information on the nature of the potential energy hypersurface and, in particular, those features which are related to protein stability and function at room temperature. As such, the constraints imposed by harmonic bond stretching potentials are desirable, since they effectively prevent those relatively uninteresting degrees of freedom from contributing to $f_u(T)$.

A. Isobutryl-val-ala$_2$-methylamide Tetrapeptide. We have calculated the instantaneous normal mode density of states for the 28 atom polar hydrogen representation of the isobutryl-val-ala$_2$-methylamide tetrapeptide. This peptide was originally studied by Choi and Elber, who calculated the distribution of barrier heights as part of a general study of helix formation in tetrapeptides.[27,28] We ran equilibrium dynamics trajectories at a number of temperatures. Each dynamics trajectory was run for 100 ps of equilibration and 1000 ps of equilibrium dynamics.

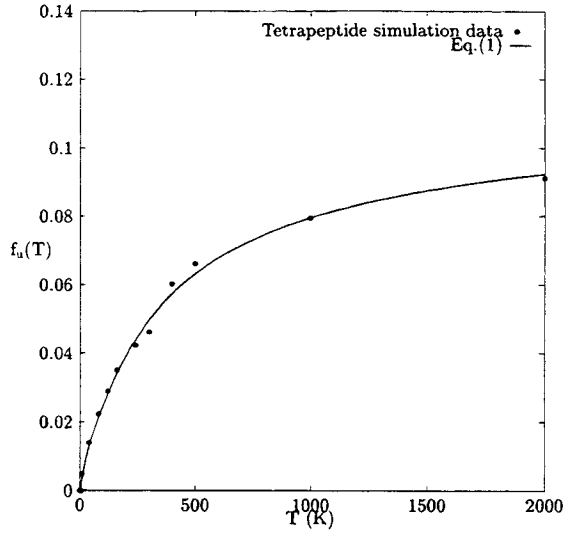

Figure 2. Fraction of unstable modes $f_u(T)$ for the tetrapeptide shown as a function of temperature. Alongside the data is the fit derived from the calculated energy barrier distribution and eq 1.

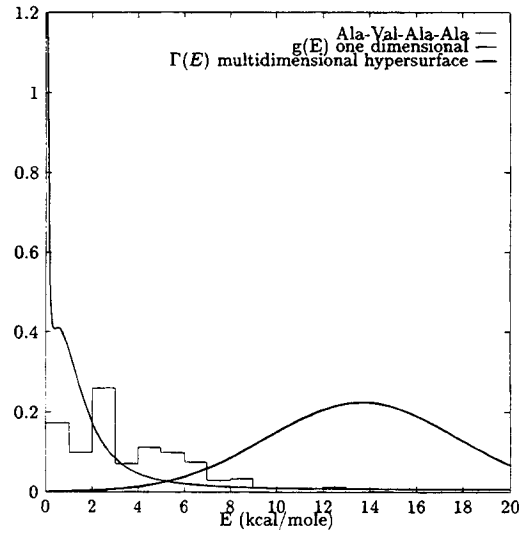

Figure 3. Intrinsic distribution of barrier heights extracted from a fit to the fraction of unstable modes as a function of temperature $g(E)$ and the total barrier height distribution for the multidimensional potential energy hypersurface of the tetrapeptide $\Gamma(E)$. Shown for comparison is the distribution of adiabatic barrier heights measured directly for the tetrapeptide by Choi and Elber.

Coordinates for the tetrapeptide were saved each 1 ps, and the 1000 instantaneous structures were later analyzed to determine the instantaneous normal mode density of states.

The fraction of unstable modes $f_u(T)$ for the tetrapeptide as a function of temperature up to 2000 K is displayed in Figure 2. The reason that data is calculated at such high energy is to determine, at least approximately, the high-temperature limit of $f_u(T)$. A reasonable estimate of $f_u(T=\infty)$ is important to an accurate determination of $g(E)$.

The statistical energy barrier distribution of $g(E)$ consistent with eq 1 is shown in Figure 3. The temperature dependence of the fraction of unstable modes predicted by the calculated barrier height distribution is in excellent agreement with the data up to 2000 K. The distribution consists of a large density of low-energy barriers and a lobe of higher energy barriers centered around 1 kcal/mol and extending to 5 kcal/mol. A narrow tail extends to higher energies and contributes to the fraction of unstable modes at higher temperatures (for $T = 5000$ K, $k_BT \approx 10$ kcal/mol).

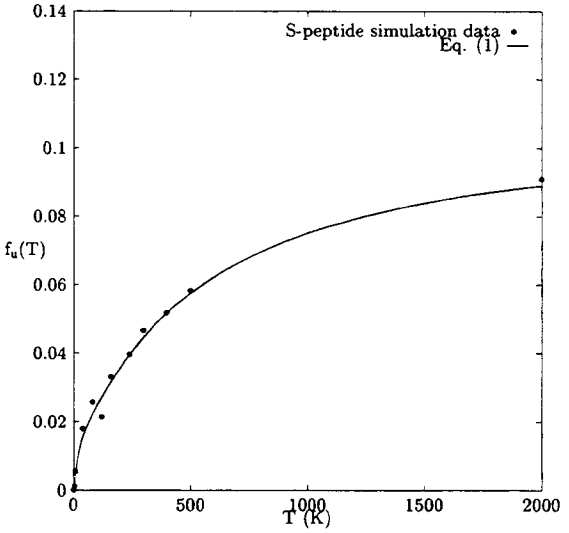

Figure 4. Fraction of unstable modes $f_u(T)$ for the S-peptide shown as a function of temperature. Alongside the data is the $f_u(T)$ predicted by eq 1 using the statistical energy barrier distribution derived in this work.

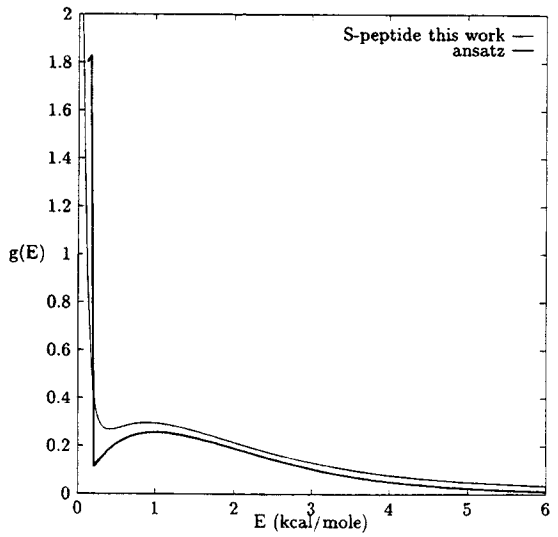

Figure 5. Intrinsic distribution of barrier heights extracted from a fit to the fraction of unstable modes as a function of temperature for the S-peptide. Shown for comparison is the intrinsic distribution of barrier heights computed previously by Straub and Thirumalai based on eq 41.[10,29]

We also show the total barrier height distribution $\Gamma(E)$ for the tetrapeptide. This distribution peaks at higher energy than the distribution calculated by Choi and Elber. That is reasonable in that their calculation is designed to find the minimum energy pathway (ideally that pathway with the lowest intervening barrier) between minima and, as such, is likely to exclude many of the higher energy barriers included in $\Gamma(E)$.

B. S-Peptide of Ribonuclease A. A detailed description of the calculations for the 186 atom polar hydrogen representation of the S-peptide has been presented previously.[10,29] The instantaneous normal mode density of states was calculated at each temperature from 120 instantaneous structures taken from a 75 ps equilibrium trajectory.

The fraction of unstable modes for the S-peptide is presented in Figure 4. We have supplemented the data for the S-peptide up to $T = 500$ K with the high-temperature data determined for the tetrapeptide (up to 10 000 K). The statistical energy barrier distribution is shown in Figure 5. Use of this distribution in eq 1 leads to a predicted form of $f_u(T)$ in excellent agreement with the simulation data through $T = 2000$ K.

In previous work, Straub and Thirumalai[10,29] assumed the following form for the energy barrier distribution

$$g(E) = a\Theta(E_l - E) + bEe^{-E/E_d} \tag{41}$$

consisting of a constant density of low-energy barriers added to a simple Poisson distribution of higher energy barriers with parameters $a = 0.325$ (kcal/mol)$^{-1}$, $E_l = 0.2$ kcal/mol, $b = 0.130$ (kcal/mol)$^{-2}$, and $E_d = 1$ kcal/mol.[10,29] The barrier height distribution integrates to $\int dE\, g(E) = 0.195$ with a contribution of 0.065 from the constant distribution of low barriers and 0.130 from the high-energy Poisson distribution.

In Figure 5 we plot the distribution of eq 41. This *ansatz* is in excellent agreement with the barrier height distribution derived in this work. A difference in the two forms results from the model of the low-energy distribution as a Heaviside function in the previous work and a single exponential in the present work. In the present work, the distribution is calculated by an optimized fit to the fraction of unstable modes followed by an accurate numerical inversion of the Fredholm integral equation. The previous results of Straub and Thirumalai are in good agreement with the results of the general method presented here.

In a recent paper, Straub, Rashkin, and Thirumalai used the intrinsic barrier height distribution computed from simulation to predict the rate of dihedral angle transitions as a function of temperature for the S-peptide.[29] The predictions were compared with direct measurement of the dihedral angle transition rate from molecular dynamics simulation and found to be in good agreement. In this paper we have argued that the total distribution of barrier heights on the potential hypersurface is $\Gamma(E)$ and that $g(E)$ is the distribution of the barrier transitions involving a single coordinate (which might be a dihedral angle). The fact that the transition rate predicted by the barrier height distribution $g(E)$, which describes local transitions of a single coordinate, agrees with the simulation results indicates that the majority of dihedral angle transitions in the simulation are local transitions rather than cooperative motions involving many coordinates.

C. Bovine Pancreatic Trypsin Inhibitor (BPTI). A typical calculation for the 892 all atom representation of the BPTI protein consisted of 50 ps of equilibration and 50 ps of equilibrium dynamics during which a coordinate set was saved each 2.5 ps. The resulting group of 20 instantaneous structures were analyzed to generate the instantaneous normal mode density of states of each temperature. The fraction of unstable modes is displayed as a function of temperature in Figure 6 alongside the results for the other systems.

The inversion of the integral equation eq 1 was carried out as for the tetrapeptide and the S-peptide. The fraction of unstable modes is displayed in Figure 6, and the calculated statistical energy distribution is shown in Figure 7 alongside the data for the tetrapeptide and S-peptide. The statistical character of the distribution is similar to that of the tetrapeptide and S-peptide.[30] However, there are slight differences in the temperature dependence of the fraction of unstable modes which can be related to the differing structures of the three systems (see below).

VI. Discussion

A. Dependence of $f_u(T)$ on System Size. The low-temperature dependence of the fraction of unstable modes for the three systems is presented in Figure 6. The tetrapeptide shows the most rapid increase in $f_u(T)$ followed by the S-peptide which is faster than the BPTI protein. We have previously argued that much of the barrier height distribution can be accounted for in terms of the dihedral angle degrees of freedom.

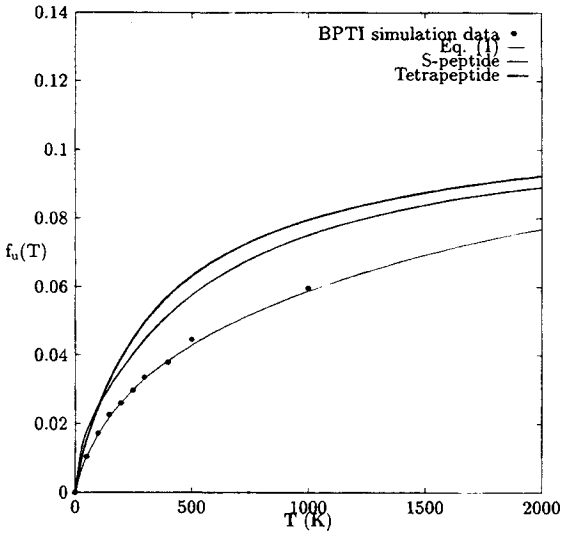

Figure 6. Fraction of unstable modes $f_u(T)$ as a function of temperature calculated for the tetrapeptide, the S-peptide, and the BPTI protein.

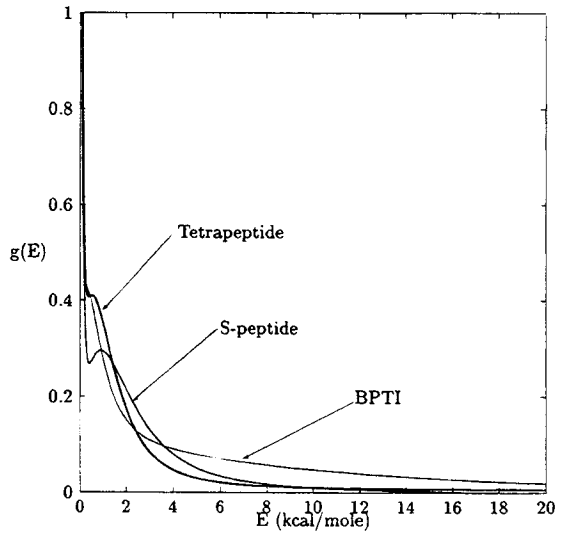

Figure 7. Intrinsic distribution of barrier heights $g(E)$ calculated for the tetrapeptide, S-peptide, and BPTI protein.

For both the tetrapeptide and BPTI, the fraction of unstable modes measured at 10 000 K is approximately 12% (meaning $N_u \simeq 3N/4$). If we assume that the dihedral angle degrees of freedom alone contribute to the fraction of unstable modes, we can estimate that the number of these degrees of freedom varies as the number of atoms N. The total number of modes will be $3N - 6$. In our theory, we assume that at high temperature each mode which can become unstable will have a probability of $^1/_2$ of being in a well and a probability of $^1/_2$ of being in the region of unstable motion. Therefore, an estimate of the high-temperature limit is

$$\lim_{T\to\infty} f_u(T) = \frac{1}{2}\frac{N}{(3N-6)} \approx 0.17 \tag{42}$$

This rough estimate misses the facts that (1) there are a significant number of unstable modes contributed by hydrogen bonding and short-range attractive interactions such as hydrophobic contacts and (2) in our empirical force field many of the dihedral degrees of freedom are treated as improper dihedral angles where the potential is harmonic in the dihedral angle and the motion is stable. As such, the agreement is quite reasonable. Clearly our system, even at 10 000 K, has not reached the limit of $f_u(T=\infty)$ due to the existence of barriers

TABLE 1: Moments of the Intrinsic Barrier Height Distribution $g(E)$, Using Cutoff Energies of (a) $E \leq 10$ kcal/mol for the Tetrapeptide, S-Peptide, and BPTI and (b) $E \leq 25$ kcal/mol for BPTI, and Parameters for the Gaussian Distribution $\Gamma(E)$ for the Barrier Distribution on the Potential Energy Hypersurface (with Energy in kcal/mol)

system	no. residues (N_0)	no. atoms (N)	$N_u \simeq 3N/4$	$\langle E \rangle_g$	$\langle E^2 \rangle_g$	$\sqrt{\Delta E^2}$
tetrapeptide[a]	4	28	21	1.30	3.16	4.1
S-peptide[a]	19	186	140	1.52	4.21	12.1
BPTI protein[a]	58	892	669	1.66	5.24	29.6
BPTI protein[b]				1.95	10.5	41.9

significantly higher than $k_BT = 20$ kcal/mol which will only contribute to $f_u(T)$ at higher temperature.

For the case of the tetrapeptide, the dihedral angles are relatively unhindered while for BPTI in globular form many of the dihedral angles have barriers significantly higher due to close packing around the side chains in the protein interior, as was originally demonstrated by Gelin and Karplus.[31] Therefore, we expect to see a larger fraction of low-energy barriers in the tetrapeptide and S-peptide than in BPTI. This is exactly what is seen in Figure 7. The reduced density of low-energy barriers in BPTI manifests itself in a slower increase in $f_u(T)$ (see Figure 6).

A strong dependence on system size is found for the barrier height distribution for the potential energy hypersurface $\Gamma(E)$ where the mean barrier energy scales as the number of modes which can become unstable N_u. The predictions for the parameters of the barrier height distribution for the tetrapeptide, S-peptide, and BPTI protein are summarized in Table 1.

B. Difficulties in Resolving $g(E)$'s High-Energy Tail. Our numerical method for computing $g(E)$ involves evaluating eq A.8, which is an alternating series. Calculation of the high-energy tail of $g(E)$ consists of the addition and subtraction of many large numbers. The round-off error is rather severe, and in practice, while we can quantitatively reproduce the behavior of $f_u(T)$ through 2000 K from the calculated $g(E)$, above that temperature the computed $f_u(T)$ plateaus prematurely and does not approach the correct high-temperature limit.

The high-temperature limit is given by

$$\int_0^\infty dE\, g(E) = 2f_u(T{=}\infty) \tag{43}$$

which is a normalization condition for $g(E)$. If we take eq A.11 and perform an integral over E before summing the series, followed by summation of the coefficients to 2 (see eq A.14), the normalization condition is well satisfied. However, if we perform the sum first and then numerically integrate the resulting $g(E)$, we no longer find that correct high-temperature limit—information is lost in the numerical summation of the function $g(E)$.

C. Predictions of the Random Energy Model. Comparing the result for the total barrier height distribution of the REM with the value of $\Gamma(E)$ derived from our theory, we can relate the width of the two distributions. Our simulation data for peptides and proteins indicate that $N_u/(3N - 6) \simeq 1/4$, where N is the total number of atoms. In the REM the value of N_0 is taken to be a number of amino acid residues, which is approximately $1/10$ the total number of atoms in the protein in the polar hydrogen atom representation used for the tetrapeptide and S-peptide (and $1/15$ of N for the all atom representation used for BPTI). It follows that $N_u \simeq 15N_0/2$ (or for the all atom model $N_u \simeq 11N_0$). Combining this result with eqs 28 and 33 leads to $\Delta E_{REM}^2 \simeq (2{-}3)N_0\langle E^2 \rangle_g$, which demonstrates the relation between the second moment of the intrinsic barrier height distribution, calculable from the temperature dependence of $f_u(T)$, and the width of the distribution of states in the REM. Through the calculation of $g(E)$ we can parametrize the REM using computer simulation data. Ideally, data could be taken over a range of ϱ and the individual parameters $\Delta\epsilon^2$ and ΔL^2 could be resolved. In this paper we restrict ourselves to a calculation of the width ΔE_{REM}^2.

TABLE 2: Estimates of the Variance in the Distribution of Energies for the Random Energy Model ΔE_{REM}^2 Based on the Literature[22,23] and This Work Using Eq 45 Based on the Distribution of Low-Energy Barriers $g_l(E)$ Alone

system	no. residues (N_0)	N_u	$\sqrt{\Delta E_{REM}^2}$ [22]	$\sqrt{\Delta E_{REM}^2}$ [23]	$\sqrt{\Delta E^2}$
tetrapeptide	4	21	1.4	0.3	0.2
S-peptide	19	140	3.2	0.7	0.5
BPTI protein	58	669	5.5	1.2	1.3

We can recognize three sets of the parameters defined in the work of Bryngelson and Wolynes as characterizing the distribution of states for a protein. They are $\epsilon = \Delta\epsilon = 2.18T_s$,[22] $\Delta\epsilon = \Delta L = 0.3$,[23] and $\epsilon = 0.1T_s$[32] with the assumption that $\Delta\epsilon = \Delta L$. We use the folded state ϱ which BW take to be 0.9. The calculated values of ΔE_{REM}^2 for the three systems are listed in Table 2 and can be compared with the predictions of eq 28 (see Table 1).

The distribution width derived from the simulation data using eq 28 is significantly larger than the value used in the REM. For each system, the ratio $\sqrt{\Delta E^2/\Delta E_{REM}^2} \geq 3$, indicating that the interactions which characterize the high-temperature dependence of $f_u(T)$ are higher in energy than those used by BW in their parametrization of the REM for protein folding. The interactions which contribute to ΔE^2 include hydrogen bonding interactions between residue pairs with a fixed distance in sequence space, which lead to secondary-structure formation, and close contacts between residues which are distant in sequence space, such as short-range hydrophobic interactions, important to tertiary-structure formation. These interactions are typically smaller in magnitude than those arising from dihedral angle transitions which define the energy landscape at high temperatures and the high-temperature dependence of $f_u(T)$.

We noted that there is a natural division of the intrinsic barrier height distribution into independent contributions from dihedral angles $g_d(E)$ and low-energy barriers $g_l(E)$[29]

$$g(E) = g_l(E) + g_d(E) \tag{44}$$

In fact, the dihedral barrier distribution $g_d(E)$ was used to predict the dihedral angle transition frequency as a function of temperature, leading to good agreement with simulation.[29] In our comparison with the REM, it makes sense to concentrate on the low-energy barrier distribution $g_l(E)$. The low-energy barrier distribution, which we imagine is due to anharmonicity in potentials leading to hydrogen bonding and weak short-range interactions, has a width in each system of

$$\Delta E^2 = \frac{N_u}{4}\int_0^\infty g_l(E)E^2\, dE / \int_0^\infty g(E)\, dE = \frac{N_u}{4}\langle E^2 \rangle_g \tag{45}$$

For example, using the parametrization of eq 41 we find $\sqrt{\langle E^2 \rangle_g} = 0.115$ kcal/mol. Using the low-energy distribution in this work, we find values of $\sqrt{\langle E^2 \rangle_g} = 0.098$, 0.086, and 0.098 kcal/mol for the tetrapeptide, S-peptide, and BPTI protein, respectively. The associated values of ΔE^2 calculated from eq 45 for the three systems are listed in Table 2. Comparing these values with the set of parameters derived for calculations of

folding times,[23] the ratio of $\sqrt{\Delta E^2/\Delta E_{REM}{}^2}$ is close to unity for all three systems. In fact, given the approximations in deriving ΔE^2 and the assumptions of the REM, the agreement is a bit too good. However, it does indicate that it is the distribution of lower energy barriers $g_l(E)$ which characterizes the energy landscape in the REM of Bryngelson and Wolynes. More importantly, it suggests that computer simulation can be used to derive parameters for statistical Hamiltonians such as the one used in the REM.

VII. Conclusion

While the instantaneous normal mode density of states is an equilibrium thermodynamic property of the system, its character and temperature dependence are intimately related to dynamical properties of the system through the distribution of barrier heights. We have presented a general method for estimating the statistical distribution of energy barriers in a disordered system. Our method is based on interpreting the temperature dependence of the fraction of unstable modes, determined from instantaneous normal mode calculations. It allows one to derive an approximate form for the barrier height distribution for peptides and even globular proteins of some size.

Application to a tetrapeptide, the S-peptide of ribonuclease A (a helical peptide), and the bovine pancreatic trypsin inhibitor (a globular protein) led to intrinsic energy barrier distributions $g(E)$ with common "universal" features as well as quantitative differences which can be related to the presence or absence of secondary and tertiary structure. The total barrier height distribution $\Gamma(E)$ for the multidimensional potential hypersurface is Gaussian within our model. The mean and variance of the barrier height distribution can be calculated from moments of $g(E)$ and as expected scale as the number of atoms in the protein. Comparison with the REM of Bryngelson and Wolynes allows one to estimate parameters for the statistical energy distribution using computer simulation. There is good agreement between the width of the distribution of low-energy barriers and the distribution of energy states in the REM for all three systems. This implies that the distribution of energy barriers can be used as a probe of interactions important to both thermal stability and folding kinetics in proteins.

Acknowledgment. We thank Dave Thirumalai, Tom Keyes, Jianpeng Ma, and Patricia Amara for helpful discussions. We are grateful to Chyung Choi and Ron Elber for providing us with parameters and their unpublished data on the barrier height distribution of the tetrapeptide and to Oren Becker and Martin Karplus for providing us with the initial coordinates and parameters used in our study of the bovine pancreatic trypsin inhibitor. We thank the Donors of the Petroleum Research Fund, administered by the American Chemical Society, and the National Science Foundation (Grant CHE-9306375) for partial support of this research. We also acknowledge a generous grant of supercomputer time from the National Center for Supercomputing Applications (NCSA) at the University of Illinois at Urbana-Champaign, and the Pittsburgh Supercomputing Center (PSC).

Appendix: Numerical Solution for $g(E)$ by Inversion of Eq 1

We have previously solved eq 1 by chosing a nonunique functional form which incorporates the physical insight provided by the changes in $f_u(T)$ with temperature.[10] There are problems associated with solving eq 1 by postulating a functional form and then finding the best fit of the function. For one thing, it is time consuming to carry out a nonlinear fit of $g(E)$. For another, the procedure is too subjective and the accuracy of the result is limited by the incomplete representation of the function $f_u(T)$. Therefore, it is desirable to solve the integral equation eq 1 numerically, assuming nothing about the form of $g(E)$.

It is possible to do this by developing a perturbation expansion for the dependence of $f_u(T)$ on the integral over the barrier height distribution. The starting point for this procedure is an accurate approximation to eqs 13 and 18 in the form

$$\bar{f}_u(T,E) = \frac{e^{-(2/3)\beta E}}{1 + e^{-(2/3)\beta E}} \tag{A.1}$$

We can expand eq A.1 as a geometric series so that

$$\bar{f}_u(T,E) = \sum_{n=1}^{\infty}(-1)^{n+1}e^{-(2n/3)\beta E} \tag{A.2}$$

Inserting this result in the integral equation, we find

$$f_u(T) = \sum_{n=1}^{\infty}(-1)^{n+1}\int_0^{\infty} dE\, g(E)e^{-(2n/3)\beta E} \tag{A.3}$$

We recognize each term in the sum as the Laplace transform of the barrier height distribution

$$\mathcal{L}[g(E)] = \int_0^{\infty} dE\, g(E)e^{-sE} = \mathcal{G}(s) \tag{A.4}$$

This allows us to compactly rewrite the series as

$$f_u(T) = \sum_{n=1}^{\infty}(-1)^{n+1}\mathcal{L}[g(E)] = \sum_{n=1}^{\infty}(-1)^{n+1}\mathcal{G}(2n\beta/3) \tag{A.5}$$

The series in eq A.5 can be inverted to obtain an expression for $g(E)$ in terms of the inverse Laplace transform $\mathcal{L}^{-1}[f_u(\beta)]$. The general form for the inverted series is

$$\mathcal{G}(2\beta/3) = \sum_{m=1}^{\infty} c_m f_u(m\beta) \tag{A.6}$$

By inspection, we find that the coefficients can be generated from the following formulas

$$c_m = \begin{cases} (-1)^k 2^{n-1} & m = 2^n \prod_{j=1}^{k} p_j, \quad \text{if } n > 0 \\ (-1)^k & m = \prod_{j=1}^{k} p_j, \\ 0 & m = p^n \times \text{anything}, \quad \text{if } n > 1 \end{cases} \tag{A.7}$$

where p and p_j represent the prime numbers greater than 2. Any positive integer m can be factored uniquely as $m = p_1^n p_2^{n_2} \ldots p_k^{n_k}$, where $p_1 > p_2 > \ldots > p_k$ are prime numbers and every $n_i > 0$. Therefore, all possible indices m and coefficients c_m may be generated from eq A.7.

Inverse transformation of eq A.6 leads to the final result

$$g(E) = \sum_{m=1}^{\infty} c_m \mathcal{L}^{-1}[f_u(3m\beta/2)] = \sum_{m=1}^{\infty} c_m \left(\frac{2}{3m}\right) \mathcal{L}^{-1}[f_u](2E/3m) \tag{A.8}$$

which provides a solution for the barrier height distribution in terms of the inverse Laplace transform of the fraction of unstable modes at decreasing temperature.

For example, if the inverse transform of $f_u(T)$ is fit to a series of Poisson distribution functions, we find

$$f_u(T) = \sum_{l=0} \left(\frac{a_l^2}{k_B^{n_l}}\right)\left(\frac{1}{\beta + b_l^2/k_B}\right)^{n_l} \tag{A.9}$$

where a_l and b_l are coefficients and n_l is the order of the Poisson distribution. The inverse transform of $f_u(T)$ is the series

$$\mathcal{L}^{-1}[f_u](E) = \sum_{l=0}\left(\frac{a_l^2}{k_B^{n_l}}\right)\frac{1}{(n_l-1)!}E^{n_l-1}e^{-b_l^2E/k_B} \quad (A.10)$$

Combining this result with eq A.8 leads to the final expression for the energy barrier distribution

$$g(E) = \sum_{m=1} c_m\left[\sum_{l=0}\left(\frac{a_l^2}{k_B^{n_l}}\right)\left(\frac{2}{3m}\right)\frac{1}{(n_l-1)!}\left(\frac{2E}{3m}\right)^{n_l-1}e^{-b_l^2 2E/3mk_B}\right] \quad (A.11)$$

There is a near regular variation in c_m/m as a function of the base 2 logarithm of the index m. For large enough m there is an apparent recurrence in the form of c_m. The coefficient for c_m/m for $m = 2^n$ is always equal to $^1/_2$, indicating that higher order terms in the series describe increasingly higher energy regions of $g(E)$. The first few terms will dominate the lower energy form of $g(E)$ while many higher order terms must be summed to accurately represent the high-energy tail.

By examining the normalization of the barrier height distribution, we can find a sum rule for the coefficients c_m which must be satisfied. If we take the limit as the temperature becomes infinite, for eq 1 we find

$$\int_0^\infty dE\, g(E) = 2f_u(T{=}\infty) \quad (A.12)$$

using the fact that $\bar{f}_u(T{=}\infty,E) = {}^1/_2$. Similarly, the limit of eq A.6 as T approaches infinity is

$$\int_0^\infty dE\, g(E) = \sum_{m=1}^{\infty} c_m f_u(T{=}\infty) \quad (A.13)$$

Combining these results, we find the sum rule for the coefficients c_m

$$\sum_{m=1}^{\infty} c_m = 2 \quad (A.14)$$

In practice we can sum only a finite number of terms. However, we must pay attention that the sum rule is satisfied if the overall normalization of the barrier height distribution is to be correct.

To compute the barrier height distribution, one adds terms to the series eq A.8 until the result for $g(E)$ has converged and the sum rule is satisfied by adjusting the coefficient of the last term in the sum. Because of the nature of the coefficients c_m, we find it necessary to sum a very large number of terms to find convergence in $g(E)$ through high energies. In practice this presents no difficulty or computational burden.

References and Notes

(1) McQuarrie, D. A. *Statistical Mechanics*; Harper and Row: New York, 1976.

(2) Stillinger, F. H.; Weber, T. A. *Phys. Rev. A* **1982**, *25*, 978; **1983**, *28*, 2408; *Science* **1984**, *225*, 983.

(3) Zwanzig, R. *J. Chem. Phys.* **1983**, *79*, 4507.

(4) Seeley, G.; Keyes, T. *J. Chem. Phys.* **1989**, *91*, 5581. Seeley, G.; Madan, B.; Keyes, T. *J. Chem. Phys.* **1991**, *95*, 3847. Madan, B.; Keyes, T.; Seeley, G. *J. Chem. Phys.* **1990**, *92*, 7565; **1991**, *94*, 6762.

(5) Madan, B.; Keyes, T. *J. Chem. Phys.* **1993**, *98*, 3342.

(6) Buch, V. *J. Chem. Phys.* **1990**, *93*, 2631. Adams, J. E.; Stratt, R. M. *J. Chem. Phys.* **1990**, *93*, 1332. Beck, T. L.; Marchioro, T. K., II. *J. Chem. Phys.* **1990**, *93*, 1347. Buchner, M.; Ladanyi, B. M.; Stratt, R. M. *J. Chem. Phys.* **1992**, *97*, 8522.

(7) Rahman, A.; Mandell, M.; McTague, J. P. *J. Chem. Phys.* **1976**, *64*, 1564.

(8) LaViolette, R. A.; Stillinger, F. H. *J. Chem. Phys.* **1985**, *83*, 4079.

(9) Rosenberg, R. O.; Thirumalai, D.; Mountain, R. D. *J. Phys.: Condens. Matter* **1989**, *1*, 2109.

(10) Straub, J. E.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 809.

(11) Tricomi, F. G. *Integral Equations*; Dover: New York, 1985.

(12) Keyes, T. To be published.

(13) Czerminski, R.; Elber, R. *Int. J. Quantum Chem.* **1990**, *24*, 167. Elber, R.; Karplus, M. *Chem. Phys. Lett.* **1987**, *139*, 375.

(14) Czerminski, R.; Elber, R. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 6963. Czerminski, R.; Elber, R. *J. Chem. Phys.* **1990**, *92*, 5580.

(15) Fischer, S.; Karplus, M. *Chem. Phys. Lett.* **1992**, *96*, 5272.

(16) Seeley and Keyes, in their discussion of instantaneous normal modes and self-diffusion in a simple liquid, suggested that a good approximation for $p(T,E)$ might be

$$\tau_b/\tau_v = \frac{\bar{f}_u(T,E)}{1-\bar{f}_u(T,E)}$$

where τ_b is the time spent *crossing* the barrier and τ_v is the time spent oscillating in the well. For a system with barrier height E we can assume that the time spent in the well will be exponentially longer than the time spent crossing the barrier $\tau_b/\tau_v \simeq e^{-\beta E}$. This leads to an estimate

$$\bar{f}_u(T,E) = \frac{e^{-\beta E}}{1+e^{-\beta E}}$$

for the temperature dependence of the fraction of unstable modes.

(17) Olkin, I.; Gleser, L. J.; Derman, C. *Probability Models and Applications*; Macmillan: New York, 1994.

(18) Bassler, H. *Phys. Rev. Lett.* **1987**, *58*, 767.

(19) Zwanzig, R. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 2029.

(20) Thirumalai, D.; Mountain, R. D.; Kirkpatrick, T. R. *Phys. Rev. A* **1989**, *39*, 3563 (see the discussion contained in footnote 45).

(21) In the study of ligand recombination rates, several groups have proposed that the distribution of barriers which describes the recombination kinetics is Gaussian or quasi-Gaussian.[33-36] Using eq 28 it should be possible to estimate the number of degrees of freedom which contribute to those barriers.

(22) Bryngelson, J. D.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 7524.

(23) Bryngelson, J. D.; Wolynes, P. G. *J. Phys. Chem.* **1989**, *93*, 6902.

(24) Derrida, B. *Phys. Rev. Lett.* **1980**, *45*, 79; *Phys. Rev. B* **1981**, *24*, 2613.

(25) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.

(26) Press, W. H.; Teukolsky, S. A.; Flannery, B. P.; Vetterling, W. T. *Numerical recipes: The art of scientific computing*; Cambridge University Press: Cambridge, U.K. 1986.

(27) Choi, C.; Elber, R. *J. Chem. Phys.* **1991**, *94*, 751.

(28) Czerminski, R.; Elber, R. *J. Chem. Phys.* **1990**, *92*, 5580.

(29) Straub, J. E.; Rashkin, A.; Thirumalai, D. *J. Am. Chem. Soc.* **1994**, *116*, 2049.

(30) Using a generic Hamiltonian of the form

$$\mathcal{H} = \kappa_{ij}x_i x_j$$

it is possible to show that the functional form of the high-energy regime of the intrinsic barrier height distribution $g(E)$ arrived at in this work and elsewhere[10] is a reasonable one (Thirumalai, D.; Straub, J. E. To be published.). This fact gives us added confidence in the overall functional form for $g(E)$.

(31) Gelin, B. R.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72*, 2002.

(32) Bryngelson, J. D.; Wolynes, P. G. *Biopolymers* **1990**, *30*, 177.

(33) Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. *Science* **1991**, *254*, 1598. Frauenfelder, H.; Parak, F.; Young, R. D. *Annu. Rev. Biophys. Chem.* **1988**, *17*, 451.

(34) Austin, R. H.; Beeson, K. W.; Eisenstein, L.; Frauenfelder, H.; Gunsalus, I. C. *Biochemistry* **1975**, *14*, 5355.

(35) Agmon, N.; Hopfield, J. J. *J. Chem. Phys.* **1983**, *79*, 2042.

(36) Young, R. D.; Bowne, S. F. *J. Chem. Phys.* **1984**, *81*, 3730.