Reprinted from the Journal of the American Chemical Society, 1994, *116*.

Copyright © 1994 by the American Chemical Society and reprinted by permission of the copyright owner.

Dynamics in Rugged Energy Landscapes with Applications to the S-Peptide and Ribonuclease A

John E. Straub,*,† Alissa B. Rashkin,†,§ and D. Thirumalai‡

Contribution from the Department of Chemistry, Boston University, Boston, Massachusetts 02215, and Department of Chemistry and Biochemistry, Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20742

Received June 3, 1993®

Abstract: A simple means of computing the rate of conformational space sampling and energy transfer in computer simulations of biomolecules using replica molecular dynamics is described. The method is based on the idea that in an ergodic system trajectories should be self-averaging—properties measured over two independent trajectories must average to the same result. Replica molecular dynamics simulation is used to calculate the generalized ergodic measure and the rate of self-averaging for the force and potential energy for the S-peptide and RNase A enzyme over a range of temperatures from 40 to 400 K. The results clearly demonstrate that even on a short time scale on the order of 10 ps, several distinct conformational states are sampled. The ergodic measures are used to obtain quantitative estimates of the rate at which conformational substates separated by relatively small barriers (on the order of a few kcal/mol) are sampled. Examination of the ergodic measure for nonbonded and dihedral angle forces proves that the time required for effective conformational space sampling is long (especially motions involving long length scales) compared to realizable computational times at all temperatures. The atomic force ergodic measure is evaluated for a harmonic system of normal modes and shown to provide a direct means of calculating the second moment of the vibrational density of states for the protein using a short dynamics trajectory. Finally, the instantaneous normal mode spectrum is calculated for the S-peptide as a function of temperature. A simple model of the potential energy hypersurface is developed and used to interpret the fraction of unstable modes in terms of the distribution of energy barriers separating the various peptide conformational substates. The distribution of energy barriers has a constant density of low-energy barriers and a Poisson distribution of high-energy barriers. The resulting energy barrier distribution is used to calculate the number of dihedral angle transitions expected in a dynamic trajectory, and the results are in good agreement with those found in the simulations. This study contains the first semianalytic method for extracting the distribution of barrier heights in systems with complex energy landscapes. The implications of our study for biomolecular simulations are discussed.

I. Introduction

It is widely believed that "complexity" in the inherent energy landscape is responsible for the rich dynamical behavior in proteins. The rugged energy surface arises because of the presence of many energy scales in proteins, i.e., due to the intrinsically heterogeneous nature of the systems. The equilibrium and dynamical properties of proteins are thought to be determined by a temperature-independent multidimensional potential hypersurface consisting of many minima (conformational substates), maxima, and saddle points (see Figure 1). In this picture, the distribution of energies for the minima, the frequencies of vibration about the minima, and the distribution of barrier heights separating these conformational substates determine the thermodynamics and dynamics of the system. The qualitative aspects of the conformational substates (CSs) model are *certainly beyond doubt*—there exist many minima, corresponding to unique protein conformations, with a distribution of energies. This point has been confirmed by the disorder seen in X-ray crystallographic studies and in the wide distribution of time scales for protein motion seen in the ligand photodissociation/rebinding experiments of Frauenfelder and co-workers on heme proteins.[1] However, only a few studies have provided the quantitative information needed to describe the distribution of minimum energy conformations, the rate of exploration of these conformations, the barriers separating these CSs, and the relation to measurable properties such as free energies and relaxation.

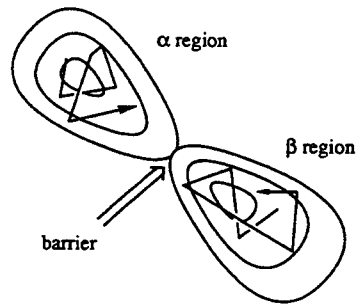

Figure 1. Schematic of the energy surface of a protein. The solid contours correspond to a fixed value of the total energy. Isoenergetic conformers belonging to the minima labeled *a* and *b* are separated by the potential barrier ΔE_{ab}. In general there are numerous minima and a distribution of barrier heights.

In this paper we describe general numerical methods that can be used in conjunction with standard computer simulation techniques to characterize the minimum energies, vibrational frequencies, and barriers for the potential hypersurfaces of peptides and proteins. Specifically we address two broad questions. (1) How can the existence of the conformational substates and the rate at which they are sampled be determined using finite time molecular dynamics trajectories? One of the important conclusions of our study is that, even in tens of picoseconds, *several distinct CSs* are sampled.[2] Moreover, we find that there is a distribution of rates of sampling of the CS,

† Boston University.
‡ University of Maryland.
§ Present address: Department of Biophysics, University of Pennsylvania, Philadelphia, PA 19104-6089.
• Abstract published in *Advance ACS Abstracts*, February 1, 1994.

(1) 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.

(2) Straub, J. E.; Thirumalai, D. *Proteins* **1993**, *15*, 360.

0002-7863/94/1516-2049$04.50/0 © 1994 American Chemical Society

implying that many barriers of differing heights are involved. (2) Can instantaneous normal mode spectra of the protein be used to obtain a distribution of barrier heights $g(E_B)$ in proteins? We introduce a novel method based on a simple caricature of the protein's energy surface that leads to an integral equation relating $g(E_B)$ and the fraction of unstable modes at a given temperature, $f_u(T)$. Our results suggest that the Poisson distribution characterizing $g(E_B)$ may indeed be a generic feature of proteins and perhaps of other heterogeneous systems as well. A summary of the major results has been presented in a recent communication.[3]

There have been a few numerical studies which have demonstrated that there are several thermally accessible minima in the neighborhood of the native structure of any protein. Elber and Karplus have used quench structures generated from a high-temperature (300 K) molecular dynamics (MD) trajectory of myoglobin to describe the distribution of energies of the distinct minima.[4] They argued that the structural differences between the *distinct minima* of the various conformational substates correspond to relative orientations of the helices which seem to be initiated by side-chain rearrangements. In addition, studies in model proteins have also shown that there are distinct pockets in the energy landscape in which structures are similar but with differing energies.[5] It has been argued that these distinct minima correspond to tier 0 in the conformational substates model of Frauenfelder *et al.* On the other hand, the minima found by Elber and Karplus correspond to substates within a single tier 0.

Reaction path studies of Czerminski and Elber have enumerated the minima available in a tetrapeptide and the distribution of barrier heights connecting the many minima.[6] The barriers separating the minima range from tenths of a kcal/mol to 5 or 10 kcal/mol. Using this detailed knowledge of the minima and their connectivity, Czerminski and Elber divided the minima into disjoint sets where the minima forming each set were connected by barriers less than or equal to a given cutoff energy. At low energies, the number of sets is the total number of minima; at high cutoff energies, all the minima are connected, forming a single set. In the neighborhood of a cutoff energy of 5 kcal/mol, Czerminski and Elber found a sharp decrease in the number of clusters of minima, which they suggested to be a lakes-to-oceans percolation transition.[7] Time scales for motions which depend on activated transitions at room temperature will range from picoseconds (transitions between similar structures confined to a small set of minima separated by barriers on the order of k_BT) to seconds (where activated transitions over large barriers are required and before minima of disjoint sets can be bridged). To date, the study of Czerminski and Elber represents the most complete characterization of the 0 K potential energy surface of a peptide or protein. Through these computational studies,[2-6] an increasingly detailed picture of the nature of the conformational substates of proteins and peptides has emerged.

In addition to introducing numerical methods for analyzing the distribution of energy minima and barrier heights in proteins, we also want to address the technical question of the adequacy of sampling the distinct conformations of proteins in the computation of thermodynamic averages using molecular dynamics. It is well appreciated that the time scales involved in the folding of a protein from a random coil configuration using molecular dynamics is beyond the current limits of computer simulation. In fact, the number of conformations increases roughly exponentially with the number of atoms, thus making an exhaustive search of the conformation space prohibitive for all but the smallest systems.[8] For the same reason, in an all-atom representation it is currently impossible to calculate thermodynamic averages for the unfolded state of a protein. Thus, coarse grained models have been useful in providing a general picture of the folding transition.[5,9-12] In the folded state of a protein, the number of conformations available is greatly reduced from that in the unfolded state. X-ray diffraction structures indicate that the majority of residues are in specific conformations and that there is limited, although often important,[13] disorder in the dihedral angles of the backbone and side chains. Similar observations have been made for carcinogen adducts in DNA.[14]

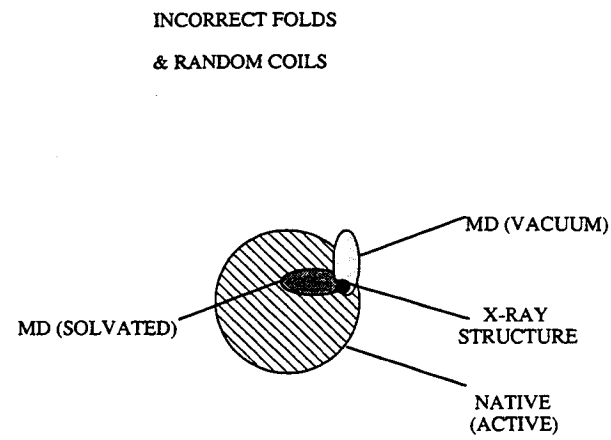

Figure 2. Schematic diagram of the conformational space available to a protein and the portions of this space which represent native, biologically active conformations and those which are inactive, being incorrectly folded or unfolded. Also indicated are regions which are sampled in a molecular dynamics calculation *in vacuo* and a more realistic treatment with solvent.

Comparison of average structure and atomic fluctuations of atoms in computer simulations with those derived from experiment indicate that for many residues the conformations available in the folded state are being sampled adequately.[15] However, this is not always the case. In Figure 2, we schematically represent our view of the conformation space available to a protein. This space can be divided into two regions containing (1) the native, biologically active conformers and (2) incorrectly folded or unfolded random coil configurations, which are inactive. The X-ray structure is represented as a single conformation within the space of possible native-like structures.

Superimposed on this division of the conformation space is a representation of the conformational space explored by a molecular dynamics trajectory originating from the X-ray structure both *in vacuo* and in solvent. The temperature is assumed to be one for which the native protein would be thermodynamically stable for days or more, meaning that an accurate trajectory for the solvated protein would move within the space of native-like states of the protein. Simulations which ignore solvent may sample conformations of the protein outside the space of native-like states, within which an accurate simulation of the protein would be expected to remain. Another important point is that calculations of picosecond or nanosecond duration will not explore the full

(3) Straub, J. E.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 809.
(4) Elber, R.; Karplus, M. *Science* **1987**, *235*, 318.
(5) Honeycutt, J. D.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 3526; *Biopolymers* **1992**, *32*, 695.
(6) Czerminski, R.; Elber, R. *J. Chem. Phys.* **1990**, *92*, 5580.
(7) Simon, S. H.; Dobrosavljevic, V.; Stratt, R. M. *J. Chem. Phys.* **1991**, *94*, 7360.

(8) It is not clear that all of these conformations will, in fact, be sampled in any kinetic process. Excluded volume interactions and other energetic considerations will, in general, eliminate many of these conformations from being sampled. For a discussion of this point, see: Dill, K. *Biochemistry* **1985**, *24*, 1501.
(9) Kolinski, A.; Skolnick, J.; Yaris, R. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 7267.
(10) Shakhnovich, E.; Farztdinov, G.; Gutin, A. M.; Karplus, M. *Phys. Rev. Lett.* **1991**, *67*, 1665.
(11) Chan, H. S.; Dill, K. A. *Macromolecules* **1989**, *22*, 4559.
(12) Chan, H. S.; Dill, K. A. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 6388.
(13) Haydock, K.; Lim, C.; Brünger, A.; Karplus, M. *J. Am. Chem. Soc.* **1990**, *112*, 3826.
(14) Loechler, E. *Carcinogenesis* **1991**, *12*, 1693. Dosanjh, M. K.; Loechler, E. L.; Singer, B. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 3983.
(15) Brooks, C. L.; Karplus, M.; Pettitt, M. *Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics*; John Wiley and Sons: New York, 1988.

conformation space of the native-like states. To do so would require sampling fluctuations over a broad range of time scales extending to milliseconds or longer.

In light of the above picture, the question arises, with the wide distribution of time scales present, as to whether one can hope to obtain a reasonable estimate of thermodynamic averages from a computer simulation of a protein. In particular, can standard MD methods be suitably modified to obtain relatively accurate values for thermodynamic properties?

Some of the issues raised above are tackled in this study using replica molecular dynamics simulations of the 19 amino acid S-peptide and full RNase A/3′-UMP enzyme/product complex as illustrative examples. The rest of the paper is organized as follows. In section II we present a powerful method for measuring the convergence of averages over dynamics trajectories using replica molecular dynamics—the generalized ergodic measure (GEM). We have recently presented a related analysis using this technique to examine the rate of sampling kinetic energy and atomic force and found it to be a useful analytical tool for investigating time scales for energy equipartitioning and conformational space sampling.[2,3] In this study we calculate the rate of self-averaging for the atomic force and the nonbonded and dihedral potential energy using replica molecular dynamics and the generalized ergodic measure. The force metric is evaluated analytically for a system of normal modes, and the rate of convergence for the total force is shown to be related to the second moment of the vibrational density of states for the system. The details of this calculation are presented in the Appendix. In section III, we further analyze the dynamics of the S-peptide by calculating the instantaneous normal mode spectrum as a function of temperature. Employing a simple model, we are able to extract an estimate of the distribution of barrier heights encountered during the simulation by solving an integral equation. The computational details are presented in section IV, and the results of our study are discussed in section V. The paper is concluded in section VI with a few additional remarks.

II. Generalized Ergodic Measures (GEMs)

In this section we present the definition of the force metrics $d(t)$ and the associated fluctuation metrics measuring the approach to the equilibrium values of physical observables. The motivation for introducing these quantities for probing the conformational states in heterogeneous systems is given elsewhere.[2] Finally, we present a detailed analysis of the total force metric for a harmonic system with a general frequency distribution and for the particular case of the Debye model.

A. The Fluctuation Metric $\Omega(t)$. Suppose we have an observable $F_j(t)$ for the jth atom of a system of N atoms, for example, for the total force, $F_j(t) = m_j \mathrm{d}v_j/\mathrm{d}t$. We write the time average of $F_j(t)$ as

$$f_j(t) = \frac{1}{t}\int_0^t \mathrm{d}s\, F_j(s) \tag{1}$$

Further, writing the average of $f_j(t)$ over all N atoms of the system as

$$\bar{f}(t) = \frac{1}{N}\sum_{j=1}^{N} f_j(t) \tag{2}$$

we can define the mean square difference of the individual $f_j(t)$ values from the average $\bar{f}(t)$ as

$$\Omega(t) = \frac{1}{N}\sum_{j=1}^{N} [f_j(t) - \bar{f}(t)]^2 \tag{3}$$

This is known as the fluctuation metric.[2] It can be shown that after a short time the function $\Omega(t)$ decays to zero with $1/t$ or that[16]

$$\Omega(0)/\Omega(t) \simeq Dt \tag{4}$$

The slope is proportional to a diffusion constant for the exploration of the range of values (space) accessible to the variable $F(t)$. The power law decay of $\Omega(t)$ to zero at long times implies that the system is "self-averaging"—that the time-averaged property $f_j(t)$ of the jth atom converges at long times to the same value $\bar{f}(t)$ for all N atoms of the system. This is a necessary, but not a sufficient, condition for the system dynamics to be ergodic. For example, imagine that phase space is divided into two regions A and B by an impassable barrier. Within either region, given enough time, any trajectories will explore all of the allowed phase space. For a set of trajectories started in region A, $\Omega(t)$ will decay to zero and the property $F(t)$ will appear to be self-averaging. However, unless we start one of the trajectories in region B, we cannot know that the partition exists and that the system is not ergodic. Therefore, the decay of $\Omega(t)$ to zero is a necessary, but not a sufficient, condition for ergodicity. It is, however, a readily calculable measure of the convergence of a given property in a numerical simulation. This is particularly true in a liquid or protein, where the calculation of alternative measures of ergodicity (or stochasticity), such as Lyapunov exponents,[17] is considerably more involved and not as obviously relevant to the convergence of thermodynamic properties as the ergodic measure.

B. The Force and Energy Metrics $d(t)$. The fluctuation metric $\Omega(t)$ provides a measure of the rate of averaging a given property, such as the kinetic energy, over a single trajectory. An alternative is to calculate averages over two independent trajectories and to measure the rate at which these two independent averages converge to the same value, as they must for an ergodic system. This becomes necessary when the underlying energy landscape has a multivalley structure, as is the case in proteins. We have extended this idea to examine the rate of convergence for the various contributions to the atomic force (bonds and angles, dihedrals, and nonbonded van der Waals and Coulombic interactions) in peptides and proteins.[2]

In this section we define the force metric and evaluate it for the special case of a harmonic system. The equations for the energy metric are found by substituting the scalar energy into expressions where the corresponding force vector is found. The averages of the force on the jth atom is defined as for the fluctuation metric as

$$\bar{f}_j^a(t) = \frac{1}{t}\int_0^t \mathrm{d}s\, \bar{F}_j^a(s) \tag{5}$$

where a indicates that the average is calculated over the ath independent trajectory. Given two independent trajectories a and b, starting from independent initial configurations, we define the difference between the averages calculated over each trajectory as the metric

$$d(t) = \frac{1}{N}\sum_{j=1}^{N} |\bar{f}_j^a(t) - \bar{f}_j^b(t)|^2 \tag{6}$$

In an ergodic system, at long times the averages of each trajectory converge to the same value and the metric decays to zero as $d(t) \simeq 1/Dt$.[16] Conversely, if $d(t)$ does not decay as a power law, then it is clear that there must be a bottleneck separating the initial conformations, i.e., a and b must belong to two distinct CSs. This method of computing $d(t)$ is termed *replica molecular dynamics* because in practice one generates two replicas of the system, corresponding to two independent initial conditions.

(16) Thirumalai, D.; Mountain, R. D. *Phys. Rev. A* **1990**, *42*, 4574.
(17) Lichetenberg, A. J.; Lieberman, M. A. *Regular and Stochastic Motion*; Springer: New York, 1983.

A useful decomposition of the force metric shows the relationship to the fluctuation metric for each trajectory a and b. By expanding the norm in eq 6, we can rewrite the metric as[2,16]

$$d(t) = \Omega^a(t) + \Omega^b(t) - X^{ab}(t) \tag{7}$$

in terms of the fluctuation metrics for trajectories a and b

$$\Omega^\alpha(t) = \frac{1}{N}\sum_j^N |\vec{f}_j^{\,a}(t) - \vec{\bar{f}}_j^{\,a}(t)|^2 \qquad \alpha = a \text{ and } b \tag{8}$$

and a cross term measuring the correlation or overlap between the two trajectories

$$X^{ab}(t) = \frac{2}{N}\sum_j^N \vec{f}_j^{\,a}(t)\cdot\vec{f}_j^{\,b}(t) \tag{9}$$

For the metric to decay to zero, either (1) the cross terms and the fluctuation metrics must decay to zero or (2) the cross term must equal the sum of the fluctuation metrics, meaning the individual atomic averages must be equal for the two trajectories, $\vec{f}_j^{\,a}(t) = \vec{f}_j^{\,b}(t)$, for all atoms. For proteins and peptides where the relaxation of the atomic force or energy requires activated conformational transitions over energy barriers of many kcal/mol, the rate of convergence is slow. As such, the rate of convergence of $d(t)$ provides insight into the barrier height distribution and the approximate times needed for overcoming the bottlenecks. In a previous study we showed that the rate of convergence for the nonbonded force is tied to the relaxation of the dihedral angle degrees of freedom and is a very slow process at room temperature.[2]

1. Calculation of $d_{FT}(t)$ for an Ideal System of Normal Modes. When the fluctuations of a system consist of small excursions about a well-defined average structure, it is often a good approximation to model the dynamics of the system of N coupled 3-dimensional Cartesian coordinates on $3N$ uncoupled 1-dimensional normal mode coordinates. Zero temperature normal mode calculations are now commonly used to analyze the distribution of vibrational frequencies in proteins and peptides about a potential minimum.[15] While the exact values of the lowest frequency vibrations are sensitive to the particular method used to truncate the long-range potential, the characters of the modes described by the eigenvectors have provided valuable insight into global fluctuations in proteins[18-20] and smaller helicies.[21] With a knowledge of the normal mode vibrational density of states, the thermodynamics and dynamics of the system are completely solved, provided anharmonic motions are not significant. Here we evaluate the force metric $d(t)$ within the normal mode approximation.

In a zero temperature normal mode approximation, the potential is expanded in a Taylor series of the $3N$ coupled coordinates $\vec{r}^N$ about a mechanically stable equilibrium position $\vec{r}_0^{\,N}$

$$U(r^N) = U_0(r_0^{\,N}) - \sum_j^N \vec{F}_j\cdot\vec{x}_j + {}^1/_2\sum_j^N\sum_i^N \vec{x}_i\cdot\kappa_{ij}\cdot\vec{x}_j \tag{10}$$

where the potential energy at the equilibrium position is $U_0(r_0^N)$, $\vec{x} = (\vec{r}^N - \vec{r}_0^{\,N})$, $\vec{F} = -\nabla_i U(r^N)$ is the $3N$-dimensional force vector (which is zero about the equilibrium position), and κ is the ($3N \times 3N$)-dimensional force constant matrix or matrix of second derivatives about the equilibrium position whose eigenvalues are the square of the eigenfrequencies ω_i for the $3N-6$ normal modes of vibration and the three translational and three rotational modes of zero frequency representing the rigid body motion of the protein.

It is possible to evaluate the total force metric in the harmonic approximation. The details are provided in the Appendix. The final result is

$$d_{FT}(t) = \frac{12k_BT}{t^2}\left[M - \frac{1}{3N}\sum_i^{3N}\langle m\rangle_i \cos(\omega_i t)\right] \tag{11}$$

We have defined the "average mass of the ith normal mode" as

$$\langle m\rangle_i = \sum_j^N m_j|a_{ji}|^2 \tag{12}$$

using the normalization condition, for the eigenvectors a_{ji} for the ith normal mode, that

$$\sum_j^N |a_{ji}|^2 = 1 \tag{13}$$

Calculation of $\langle m\rangle_i$ using the density of states for a number of proteins has shown that the average mass is fairly independent of the mode number for low-frequency modes due to the delocalized nature and the participation of all atoms in the system.[22] This approximation is examined in the case of the S-peptide in section V.

The initial value of the metric determined from eq 11 can be expressed

$$d_{FT}(0) = 6k_BT\frac{1}{3N}\sum_i^{3N}\langle m\rangle_i\omega_i^2 = 6k_BT\langle\langle m\rangle\omega^2\rangle \tag{14}$$

The outer angular brackets in the rightmost equation indicate an average over all normal modes. Combining eqs 11 and 14, we find a remarkably simple form for the asymptotic limit of the force metric in the normal mode representation

$$d_{FT}(0)/d_{FT}(t) = \frac{t^2}{2M}\langle\langle m\rangle\omega^2\rangle \tag{15}$$

If we approximate the average mass of each normal mode as the average mass of an atom in the system $\langle m\rangle_i \simeq M$ (see Figure 7 and surrounding discussion), we find the useful approximate result

$$d_{FT}(0)/d_{FT}(t) \simeq {}^1/_2t^2\langle\omega^2\rangle \tag{16}$$

where $\langle\omega^2\rangle$ is the second moment of the density of states. For a system undergoing harmonic motion, the inverse force metric will converge in time as t^2. More importantly, the curvature will provide an estimate of the second moment of the vibrational density of states for the system.

2. Evaluation of $d_{FT}(t)$ for the Debye Model. For properties which are dominated by the fluctuation of global, low-frequency vibrations, a continuum approximation such as the Debye model can provide a good estimate of the vibrational density of states. For a system of N atoms, the Debye distribution is given by

$$g(\omega) = \begin{cases} 9N\omega^2/\omega_D{}^3 & 0 < \omega \le \omega_D \\ 0 & \omega_D < \omega \end{cases} \tag{17}$$

where $g(\omega)$ is normalized as

$$\int_0^{\omega_D} d\omega\, g(\omega) = 3N \tag{18}$$

The zero-time value of the force metric in the Debye model is simple to calculate. The leading order term in time, which is the

(18) Brooks, B.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 6571.
(19) Levitt, M.; Sander, C.; Stern, P. S. *J. Mol. Biol.* **1985**, *181*, 423.
(20) Gō, N. T.; Noguti, T.; Nishikawa, T. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 3696.
(21) Roux, B.; Karplus, M. *Biophys. J.* **1988**, *53*, 297.
(22) Straub, J. E., unpublished results.

only one to survive after an initial short time decay, is

$$d_{FT}(0)/d_{FT}(t) \simeq ({}^3/_{10})\omega_D{}^2t^2 \qquad (19)$$

This result provides a means of calculating the approximate Debye frequency for the system from the curvature of the total force metric.

It has been suggested that for glasses the low-frequency behavior of the density of states is a power law $g(\omega) \propto \omega^\alpha$ with $\alpha < 2$. It may be that for proteins this is also the case. Our calculation of the Debye frequency is simply meant to provide a convenient measure of the scale of frequencies in $g(\omega)$.

III. Instantaneous Normal Mode Spectra

The normal mode model has provided valuable insight into the nature of fluctuations in proteins.[18-20] Normal mode calculations are typically carried out at zero temperature by expanding the potential about a mechanically stable energy minimum of the potential surface, where the force on all atoms is zero. Recently, a number of studies in liquids composed of spherical particles have proved the usefulness of calculating normal mode frequencies for conformations chosen randomly from a dynamics trajectory. The instantaneous normal modes spectrum of eigenfrequencies consists of a lobe of real frequencies corresponding to stable, bounded motion and a lobe of imaginary frequencies corresponding to unstable motion over regions of the potential surface with negative curvature.[23] By performing instantaneous normal mode calculations for dynamics trajectories at a series of temperatures, it is possible to monitor the change in character of the potential surface traversed at increasingly higher temperatures. The percentage of imaginary frequencies has been shown to be intimately related to the process of barrier crossing and the onset of self-diffusion in the transition from solid or glass to liquid.[24-26]

It is straightforward to extend such calculations to proteins. At low enough temperatures, we expect to see an instantaneous normal mode spectrum consisting of stable motion about a single potential minimum. As the temperature is increased, barriers will be traversed, primarily by dihedral angle fluctuations of side chains, which will show up as unstable, imaginary frequencies. It is also expected that the anharmonicity of the protein potential surface will manifest itself in a significant number of imaginary frequencies not associated with dihedral angle transitions or barrier crossing. This is the case for water clusters, where Buch has found a significant percentage of imaginary frequencies at temperatures where there is no self-diffusion.[27] Similar observations have been made for Lennard-Jones systems by Beck.[28]

Given a general form for the potential, one can, in principle, calculate the equilibrium percentage of unstable modes at a particular temperature. For an ergodic system, this could also be accomplished using a dynamics trajectory or Monte Carlo since $f_u(T)$ is an intensive equilibrium thermodynamic property of the system. However, if we assume that the system of N nonlinearly coupled three-dimensional Cartesian coordinates can be accurately described by the $3N$ one-dimensional normal coordinates, the problem is greatly simplified. If we further assume that the potential for each of the normal coordinates is periodic piecewise parabolic and defined solely by a barrier height E_B (which may be different for each mode), it is straightforward to calculate the fraction of unstable modes $\bar{f}_u(T,E_B)$ as a function of temperature. For the $3N$ normal coordinates, there will be a distribution of barrier heights $g(E_B)$ which is normalized to unity.

Assuming a continuous distribution of barrier heights, the fraction of unstable modes for the system measured at a particular temperature will be[3]

$$f_u(T) = \int_0^\infty dE_B\, g(E_B)\bar{f}_u(T,E_B) \qquad (20)$$

For a periodic piecewise parabolic potential

$$U(x) = \begin{cases} {}^1/_2\kappa x^2 & 2n\Delta x \le x + {}^1/_2\Delta x \le (2n+1)\Delta x \\ E_B - {}^1/_2\kappa x^2 & (2n+1)\Delta x \le x + {}^1/_2\Delta x \le 2n\Delta x \end{cases} \qquad (21)$$

where $\Delta x = 2(E_B/\kappa)^{1/2}$, the fraction of unstable modes can be calculated for a particular value of T and E_B

$$\bar{f}_u(T,E_B) = \frac{e^{-2\xi^2}\int_0^\xi dx\, e^{x^2}}{\int_0^\xi dt\, e^{-t^2} + e^{-2\xi^2}\int_0^\xi dx\, e^{x^2}} \qquad (22)$$

where $\xi = (\beta E_B/2)^{1/2}$. By calculating $f_u(T)$ from a dynamics trajectory at a series of temperatures, one can, in principle, solve eq 20 and extract the distribution of barrier heights $g(E_B)$ for the system. It is possible to derive other approximations for the kernel $\bar{f}_u(T,E_B)$ by assuming that the well frequencies are different. This leads to additional complications that are not worth describing, given the simplicity of the caricature of the potential energy surface.

IV. Computational Details

We carried out calculations on the S-peptide of bovine pancreatic ribonuclease A [KETAAAKFERQHMDSSTSA] and the RNase A/3′-UMP enzyme/product complex in vacuum.[2] The initial conditions for each system were taken from the X-ray structure of the RNase A/3-UMP complex. The protein parameters are based on the version 19 CHARMM parameter set. The inhibitor and histidine parameters were developed separately and are presented elsewhere.[29]

The dynamics were performed at constant energy using a modified version 20 of the CHARMM simulation program.[30] The Verlet algorithm with a time step of 0.5 fs was employed to avoid the use of rigid constraints on bonds involving hydrogen atoms. A distance-dependent dielectric of the form $\epsilon(r) = 4r$ was employed in our vacuum simulations. Nonbonded interactions were truncated by multiplying the potential due to charge and van der Waals interactions by the atom-based shifting function of the form

$$S(r) = \begin{cases} (1 - (r^2/r_c{}^2))^2 & 0 < r \le r_c \\ 0 & r_c < r \end{cases} \qquad (23)$$

For our study, $r_c = 12$ Å. Calculations on myoglobin have shown that truncation of the long-range interactions using the shifting function were superior to truncation using group- or atom-based switching functions and provide a good approximation to calculations performed with no cutoff.[31]

Trajectories were run at a series of temperatures ranging from 40 K to 400 K. Equilibration was carried out for the *a* structures over a 50-ps run using velocity rescaling every 0.5 ps and starting from the X-ray structure. (An exception was the equilibration at 40 K for 150 ps.) Calculation of the force metric requires an independent *b* trajectory at the same temperature T_1. To generate independent initial conditions for the *b* trajectory, we ran 50 ps of equilibration dynamics at a higher temperature T_2 to "amnesiate" (induce a state of amnesia in) the molecule and then quenched and reequilibrated the trajectory at the previous temperature T_1 for 50 ps. The trajectories with $T_1 = 40$ and 80 K were amnesiated using a $T_2 = 120$ K; for the $T_1 = 120$, 160, and 240 K trajectories, $T_2 = 300$ K; for the $T_1 = 300$ K trajectory,

(23) Rahman, A.; Mandell, M.; McTague, J. P. *J. Chem. Phys.* **1976**, *64*, 1564.
(24) LaViolette, R. A.; Stillinger, F. H. *J. Chem. Phys.* **1985**, *83*, 4079.
(25) Rosenberg, R. O.; Thirumalai, D.; Mountain, R. D. *J. Phys.: Condens. Matter* **1989**, *1*, 2109.
(26) Seeley, G.; Keyes, T. *J. Chem. Phys.* **1989**, *91*, 5581.
(27) Buch, V. *J. Chem. Phys.* **1990**, *93*, 2631.
(28) Beck, T. L.; Marchioro, T. L., II. *J. Chem. Phys.* **1990**, *93*, 1347.
(29) Straub, J. E.; Lim, C.; Karplus, M. *J. Am. Chem. Soc.*, in press.
(30) Brooks, B. R.; Bruccoleri, R. D.; Olafson, B. O.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(31) Loncharich, R. J.; Brooks, B. R. *Proteins* **1989**, *6*, 32.

$T_2 = 400$ K. For the enzyme, the second trajectory was initiated from the equilibrated 300 K structure and reequilibrated for 25 ps at $T_1 = 40$ K and 12.5 ps at $T_1 = 120$ and 240 K. For $T_1 =$ 240 and 300 K, the trajectories were amnesiated for 12.5 ps at $T_2 = 400$ K and then reequilibrated at T_1 for 12.5 ps.

Instantaneous normal mode frequencies were calculated from the 75-ps *a* trajectory at each temperature. Structures were chosen every 0.625 ps (1250 time steps), and 120 structures were examined at each temperature. All calculations were performed on a Stardent 3030 computer at Boston University.

V. Results and Discussion

To examine the rate of conformational space sampling and energy equipartitioning using replica molecular dynamics, we ran a series of trajectories at temperatures ranging from 40 to 400 K. At each temperature, two "independent" trajectories were generated using the methods described in section IV. In this section we report the results for the rate of force and energy averaging for the S-peptide and RNase A systems. This is followed by the results of our instantaneous normal mode analysis for the S-peptide as a function of temperature. We begin by characterizing the changes in structure of the S-peptide as a function of temperature for two independent trajectories *a* and *b*.

A. Conformational Characteristics of the S-Peptide. For the S-peptide we have characterized the conformations of two independent trajectories *a* and *b* generated at each temperature. In Figure 3 we show the average potential and total energy for each trajectory as a function of temperature. The energy of the *a* trajectory increases monotonically and linearly with temperature, while the energy of the *b* trajectory is noticeably higher relative to the *a* trajectory at 40, 80, and 160 K. The same trend is seen in the total energy.

The *a* trajectories were heated directly from the X-ray structure, and the linearly increasing average potential and total energy are typical of a harmonic system. The higher temperature equilibration and quench carried out in generating the initial conditions for the *b* trajectory allow the structure to make transitions away from the initial conformation. At some temperatures (40, 80, and 160 K) this leads to the localization of the *b* trajectory in a higher lying substate, while at other temperatures (120, 240, and 300 K) the potential energy is annealed and lower than that of the *a* trajectory.

Experimentally, the S-peptide portion of the ribonuclease A enzyme is helical at most through residue 13 in the X-ray structure.[32,33] Plots of the (ϕ,ψ) angles for the interior residues 3–13 calculated from the average structure are shown in Figure 4. We see that for the *a* trajectory the peptide is strongly helical below 160 K and that at higher temperatures the helix unravels. We find that the unraveling starts from the C-terminus. The (ϕ,ψ) maps of the *b* trajectory at low temperatures show a predominance of helical residues at 40 and 80 K which is absent at 120 K and above. This indicates that the heating involved in the preparation of initial conditions for the generation of the *b* trajectories leads to an unraveling of the helix which is not restored in the quench.

The change between full and partial helical character of the *a* trajectories as the temperature is increased beyond 240 K manifests itself as a clear decrease in the radius of gyration R_g, defined for a system of N atoms as

$$\langle R_g^2 \rangle = \frac{1}{N^2} \sum_{i<j} \langle r_{ij}^2 \rangle \qquad (24)$$

Results for the radius of gyration are shown in Figure 5 for the S-peptide as a function of temperature. For reference, we have noted the R_g for a model-built "perfect" helix where (ϕ,ψ) =

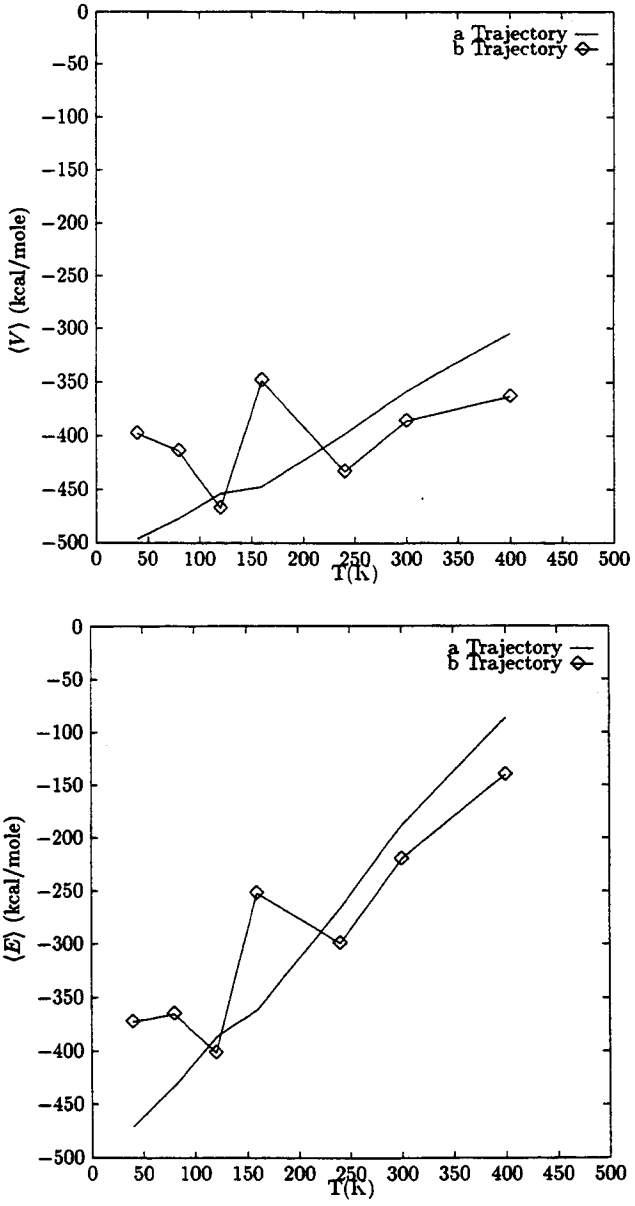

Figure 3. Average (top) potential energy and (bottom) total energy for the *a* and *b* trajectories of the S-peptide as a function of temperature.

(–60°, –60°) for each residue. At low temperatures (40 and 80 K), the S-peptide is frozen in its extended helical configuration with large R_g. At 120 K, the peptide helix folds on itself, preserving the helical (ϕ,ψ) angles but leading to a compact structure with small R_g. Similar behavior is found at 160 K. At temperatures above 240 K, the helix unravels and entropic effects lead to an extended random coil with increased R_g.

For the *b* trajectory at 40 and 80 K, R_g is small and nearly identical to that of the *a* trajectory at 120 K, which was its precursor. At 120, 160, and 240 K, R_g of the *b* trajectory is similar to that of the precursor 300 K *a* trajectory. As expected, both the *a* and *b* trajectories show similar R_g values at high temperatures. Overall, the radius of gyration is well correlated with the extent of helical conformation as a function of temperature and provides a useful measure of the compactness of the structure.

Simulations of 30-mers of polyglycine and polyalanine both in vacuum and in solution have shown that the α-helix tends to be destabilized by solvation.[34] In the presence of water, it has been observed that the opening of a N—H···C=O hydrogen bond is accompanied by insertion of a water molecule to form a three-

(32) Kim, P. S.; Baldwin, R. L. *Nature* **1984**, *307*, 329.
(33) Nelson, J. W.; Kallenbach, N. *Proteins* **1986**, *1*, 211.
(34) DiCapua, F. M.; Swaminathan, S.; Beveridge, D. L. *J. Am. Chem. Soc.* **1991**, *113*, 6145.

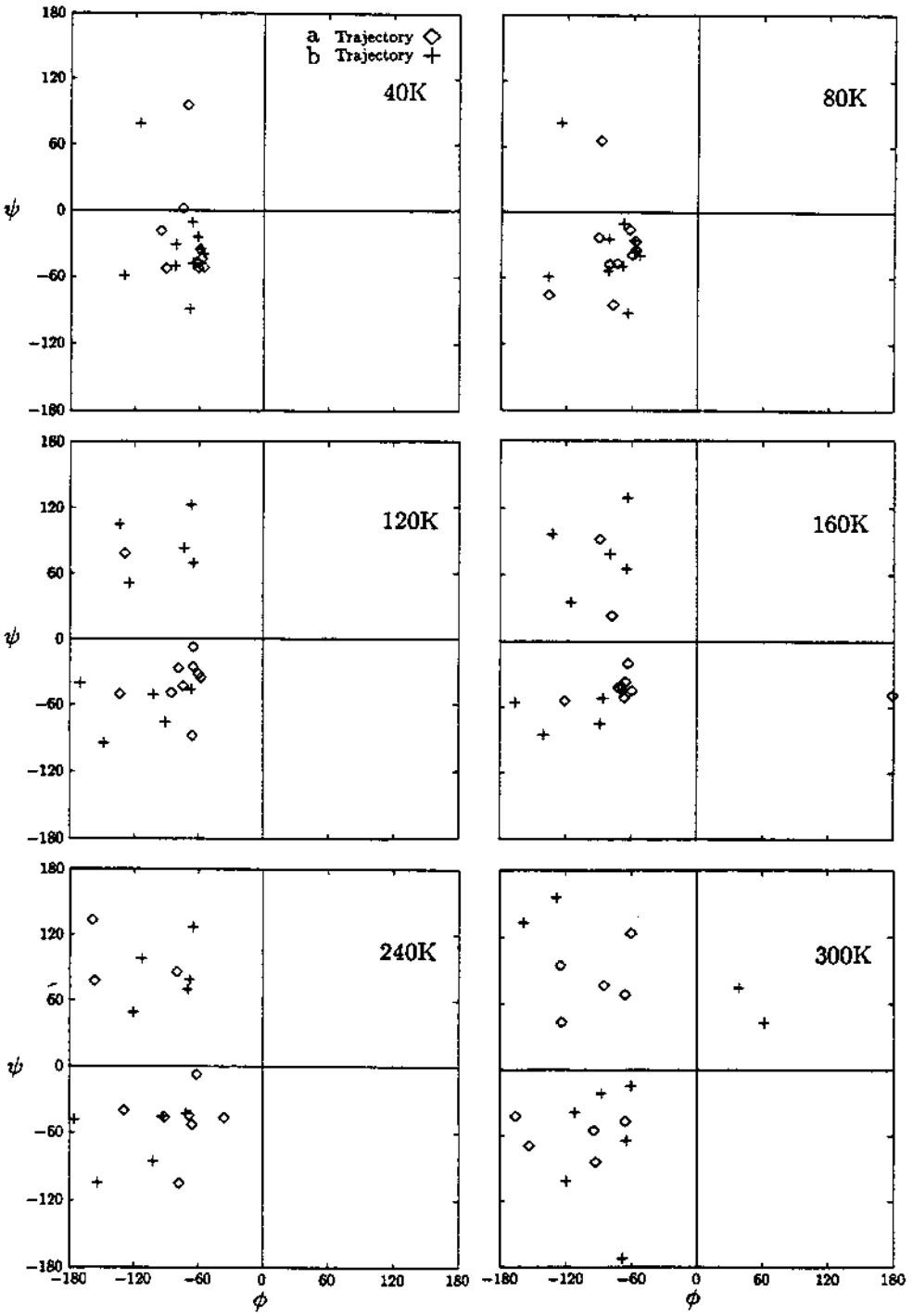

Figure 4. Series of ϕ, ψ plots showing the conformations of the inner 3–13 residues of the S-peptide at a number of temperatures.

centered hydrogen bond which functions as an intermediate to helix unfolding. Support for this mechanism has been lent by simulations of an S-peptide analogue in solution, where it was observed that main-chain helical hydrogen bonds would often open through an $\alpha \rightleftharpoons 3_{10} \rightleftharpoons$ no hydrogen bond mechanism.[35] In the later simulation, the S-peptide analogue in solution was found to be stable over a 300-ps trajectory at 278 K but was observed to partially unravel during a 500-ps trajectory at 358 K. This behavior correlates well with experimental evidence for the helix stability.

While our simulations were performed in vacuum, it is worth noting the important side-chain contacts which persisted over the length of each simulation. It has been proposed, and seems quite plausible to us, that contacts formed between side chains along the α-helix can act to stabilize the α-helix formed between residues 3 and 13. Maximum helix formation is seen near pH 5, where the His 12 residue would be protonated and charged. In our simulations, titratable residues were protonated to simulate neutral pH, and the His 12 residue was neutral. It has been suggested that a salt bridge formed by Glu 9$^-$ and His 12$^+$ at low pH acts to stabilize the α-helix.[36] No contact between Glu 9$^-$ and His 12^0 was observed in our simulations at any temperature.

A salt bridge formed by His 12$^+$ and Asp 14$^-$ has been suggested as a possible helix stop signal by Kim and Baldwin.[32] In our simulations, the His 12 was neutral and no contact was observed at any temperature between His 12^0 and Asp 14$^-$. In agreement with experimental evidence, no helix formation was observed beyond residue 13.

(35) Tirado-Rives, J.; Jorgensen, W. *Biochemistry* **1991**, *30*, 3864.

(36) Bierzynski, A.; Kim, P. S.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 2470.

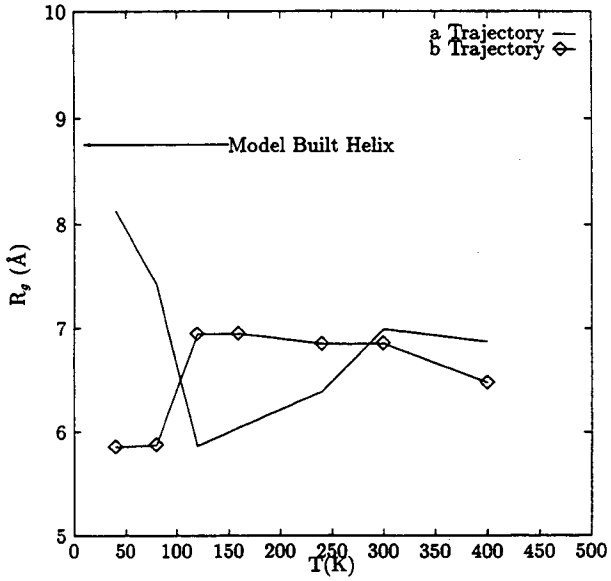

Figure 5. Radius of gyration R_g for the a and b trajectories of the S-peptide as a function of temperature.

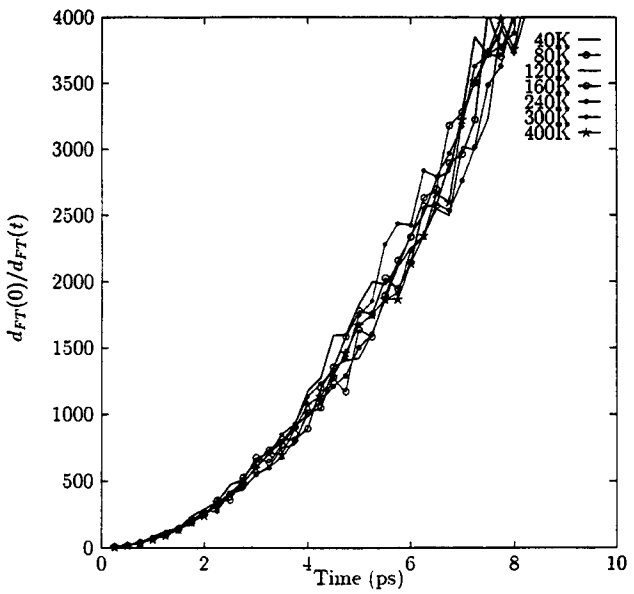

Figure 6. Reciprocal of the total force metric, $d_{FT}(0)/d_{FT}(t)$, as a function of time for the S-peptide at seven temperatures. The plot demonstrates that the inverse of the normalized force metric is identical at all temperatures.

In our lowest temperature simulation, a hydrogen-bonding quadruplet was observed between Lys 7^+···Gln 11^0···His 12^0···C-terminal COO^- along one side of the α-helix. At higher temperatures, the contact between Lys 7^+ and Gln 11^0 was maintained (through 400 K). Similarly the contact between His 12^0 and the C-terminal carboxylate was observed at 40, 160, 300, and 400 K. Tirado-Rives and Jorgensen have noted that in their simulation of the S-peptide analogue at 278 K, the His 12 residue turned toward the C-terminus.[35] At 240 K, a salt bridge was formed between the C-terminal carboxylate and Lys 7^+. The initial contact between Gln 11^0 and His 12^0 was lost above 120 K.

The salt bridge formed by a contact between Glu 2^- and Arg 10^+ was observed at every temperature but 80 K. In the simulation of Tirado-Rives and Jorgensen of the solvated S-peptide analogue at 278 K, it was observed that the contact ion pair was replaced by a solvent-separated ion pair.

We have noted that at higher temperatures the α-helix observed at low temperatures between residues 3 and 13 unravels from the C-terminal end. Previous studies of solvent effects on helix stability might lead us to believe that the S-peptide helix would be more stable in vacuum (as in our simulations) than in solvent (as in experiment).[34] However, a possible contribution to the instability of the helix in our simulations is that we employed a neutral His 12 residue as we might expect to find at pH 7. While a salt bridge between Glu 9^- and His 12^+ has been suggested as a helix stabilizer, we saw no favorable contact between Glu 9^- and His 12^0 in our simulations. Further simulations of the S-peptide at neutral pH and pH 5 could be used to test this hypothesis.

B. Force and Energy Averaging. We calculated the force metric for the total force acting on atoms of the S-peptide as a function of temperature using trajectories a and b, and the results are presented in Figure 6. At all temperatures the inverse total force metric $d_{FT}(0)/d_{FT}(t)$ shows a clear t^2 dependence with a curvature that is independent of temperature as predicted by our analysis (see section II and the Appendix).

We expect that the total force on atoms in the peptide is dominated by local harmonic bond and angle terms due to the highly connected nature of peptides and proteins. Out of the $3N - 6$ vibrational degrees of freedom, there are $N - 1$ bonds and $N - 2$ angles, in addition to many dihedral angles with high barriers, where the motion at room temperature is effectively harmonic. Therefore, the majority of degrees of freedom will be harmonic, and the normal mode picture, particularly at short times and high frequencies, will provide a good approximation to the peptide motion. This implies that algorithms designed to exploit short time oscillatory dynamics in near-harmonic systems can potentially lead to significant savings in computer time for molecular dynamics of biomolecules.[37]

Analysis of the total force metric for a harmonic system led to the result

$$d_{FT}(0)/d_{FT}(t) = \langle\omega^2\rangle t^2/2$$

(see Section II.B.1), where we have defined the average mass of the ith normal mode as

$$\langle m\rangle_i = \sum_j^N m_j |a_{ji}|^2 \qquad (25)$$

How sensitive a function of the details of the normal mode frequency and eigenvector is $\langle m\rangle_i$? Previous results have shown that with the exception of the hydrogen stretches, $\langle m\rangle_i$ is fairly independent of the details of the specific normal mode and can be reasonably approximated by the average mass of the protein atom, M.[22] The result for the specific case of the S-peptide is shown in Figure 7. For all but the highest frequency localized modes, $\langle m\rangle_i$ is reasonably independent of mode number. This is expected for the lower frequency global fluctuations, where the participation of all atoms is nearly equal. To a good approximation, we can replace the average mass for the ith normal mode, $\langle m\rangle_i$, with the average atomic mass for the peptide or protein atoms, M. For the S-peptide in the polar hydrogen model, M = 11.62 g/mol. For the higher frequency modes localized in dihedral angle out-of-plane bends and carbonyl stretches, the insensitivity of $\langle m\rangle_i$ to mode number is likely due to the few atom types and limited range of atomic masses for the protein heavy atoms (mostly C, N, and O).

Within the approximation that $\langle m\rangle_i = M$, the total force metric is

$$d_{FT}(0)/d_{FT}(t) \simeq {}^1/_2\langle\omega^2\rangle t^2$$

Thus a short dynamics trajectory is all that is needed to calculate $d_{FT}(t)$, providing a straightforward means of calculating the second moment of the vibrational density of states $\langle\omega^2\rangle$ for the S-peptide

(37) Tuckerman, M.; Martyna, G. J.; Berne, B. J. *J. Chem. Phys.* **1990**, *93*, 1287.

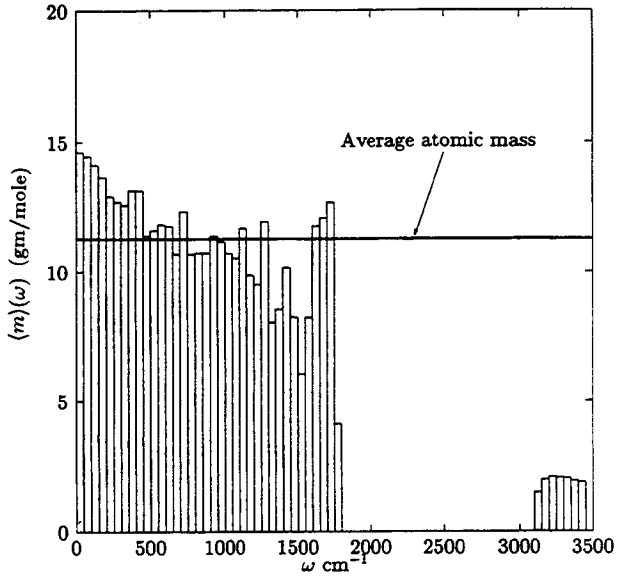

Figure 7. Histogram of the average mass of the ith normal mode of the S-peptide as defined by eq 12 as a function of the mode frequency. The plot demonstrates that, with the exception of the highest frequency normal modes, the average mass, $\langle m \rangle_i$, is relatively independent of frequency and is reasonably well approximated by the average atomic mass M.

Table 1. Summary of Results for the Moments of the Imaginary and Real Lobes of the Instantaneous Normal Mode Density of States at Each Temperature[a]

	imaginary			real	
T (K)	$f_u(T)$	$\langle\omega\rangle$	$\langle\omega^2\rangle^{1/2}$	$\langle\omega\rangle$	$\langle\omega^2\rangle^{1/2}$
0	≃0	-8.5	9.5	661.9	852.9
5	0.12	-13.7	18.2	664.5	854.4
10	0.55	-19.1	23.2	664.2	854.2
40	1.79	-39.3	56.5	668.6	858.5
80	2.57	-54.3	75.5	675.0	862.7
120	2.14	-66.7	91.4	677.3	863.7
160	3.31	-88.0	130.9	680.6	867.0
240	3.97	-95.2	135.4	688.1	872.9
300	4.66	-116.9	176.1	691.4	875.6
400	5.18	-130.1	204.3	693.3	877.7
500	5.83	-143.3	213.8	698.0	881.4

[a] The moments were calculated using the distribution $g(\omega)$ shown in Figure 12 by restricting the limits on frequency to be $|\omega| \leq 2000$ cm^{-1}.

from the curvature of the inverse total force metric. From the data in Figure 6 for the total force metric, we estimate that $\langle\omega^2\rangle_{FT} = 1.4 \times 10^5$ cm^{-2}, roughly independent of the temperature. In the Debye model, the density of states is assumed to be quadratic, which will be a good approximation at low frequencies and a poor one at high frequencies. Inspection of normal mode spectra for a number of proteins of approximately 100 residues indicates that the vibrational density of states is Debye-like up to 40 cm^{-1}.[22] From eq 19 and our data for the total force metric, we estimate the Debye frequency to be $\omega_D = 480$ cm^{-1}, which is slightly larger than the direct estimate of the root-mean-square frequency $\langle\omega^2\rangle_{FT}^{1/2} = 370$ cm^{-1} based on the total force metric. These results are in disagreement with the second moment of the density of states directly calculated from the normal mode data for the S-peptide, which is typically $\langle\omega\rangle_{exact}^{1/2} = 860$ cm^{-1} (see Table 1 where the frequencies considered were restricted to be $|\omega| \leq 2000$ cm^{-1} to exclude the hydrogen stretches (≃3000–3400 cm^{-1}), which have a low effective mass and do not contribute substantially to $\langle\omega^2\rangle_{FT}$).

The magnitudes of the three calculated values of the second moment of the density of states are ordered $\omega_{FT} < \omega_D < \omega_{exact}$. The ordering $\omega_D < \omega_{exact}$ is due to the fact that the Debye model assumes a stronger frequency dependence $g(\omega) \simeq \omega^2$ than is observed in the calculated density of states which consists of a bunch of modes below 1800 cm^{-1} followed by very little until the hydrogen stretches above 3000 cm^{-1}. The Debye form must count these bunches of high frequency modes, and it does so by collecting them at lower frequency, leading to a smaller estimate for the mean-square width of the distribution.

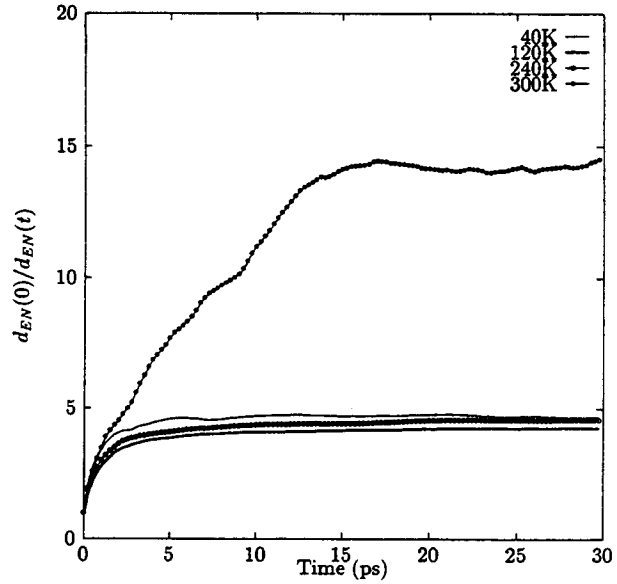

Figure 8. Normalized nonbonded energy metric, $d_{EN}(0)/d_{EN}(t)$, as a function of time for the enzyme/product complex.

The discrepancy between $\langle\omega^2\rangle_{FT}$ and $\langle\omega^2\rangle_{exact}$ is likely due to the anharmonicity of the protein motion, which is ignored in the normal mode approximation made in the derivation of $d_{FT}(0)/d_{FT}(t)$. The data in Figure 7 demonstrate that the assumption that the average effective mass of each normal mode is the same is justified. Returning to eq 15, we have calculated the quantities $\langle\langle m\rangle\omega^2\rangle$ and $M\langle\omega^2\rangle$. For the configuration tested we found the square root of the ratio, $(\langle\langle m\rangle\omega^2\rangle/M\langle\omega^2\rangle)^{1/2}$, to be 0.74 for all modes and 0.97 when hydrogen stretches are excluded. Thus, once we restrict the average for $\langle\omega^2\rangle_{exact}$ to frequencies less than 2000 cm^{-1}, we can expect eq 16 to provide an accurate result within the harmonic approximation. Nevertheless, the value predicted by the harmonic model for the total force metric is significantly smaller than the exact value computed directly from the density of states. The difference in $\langle\omega\rangle_{FT}$ and $\langle\omega^2\rangle_{exact}$ must result from the assumption that the potential energy hypersurface is harmonic, even at the lowest temperatures. Anharmonicity in the vibrations, particularly at lower frequency, leads to a significant reduction in the value of $\langle\omega^2\rangle_{FT}$ below the exact harmonic result. *We conclude that anharmonicity is essential to the accurate description of the protein dynamics even for very short times of 1–5 ps.* It may be possible to employ a quasiharmonic analysis to treat these differences.[15]

Using the fluctuation metric, we have shown that the kinetic energy is equipartitioned on a timescale of picoseconds.[2] The longer time relaxation associated with conformational transitions in the peptide is best explored using the metric of the nonbonded (Coulombic and van der Waals) potential energy $e_j^\alpha(t)$ for the jth atom in the αth trajectory

$$d_{EN}(t) = \frac{1}{N}\sum_{j=1}^{N}[e_j^a(t) - e_j^b(t)]^2 \qquad (26)$$

The results are displayed in Figure 8.

In the nonbonded energy metric, we see a rapid initial convergence followed by a slow, long time decay. The initial convergence is significantly greater at 300 K than in the lower temperature trajectories. The plateau in the reciprocal metric is reached within 3 ps for $T < 240$ K. This indicates that at lower temperatures the peptide motion is confined to fluctuations about a single CS without significant conformational transitions on a

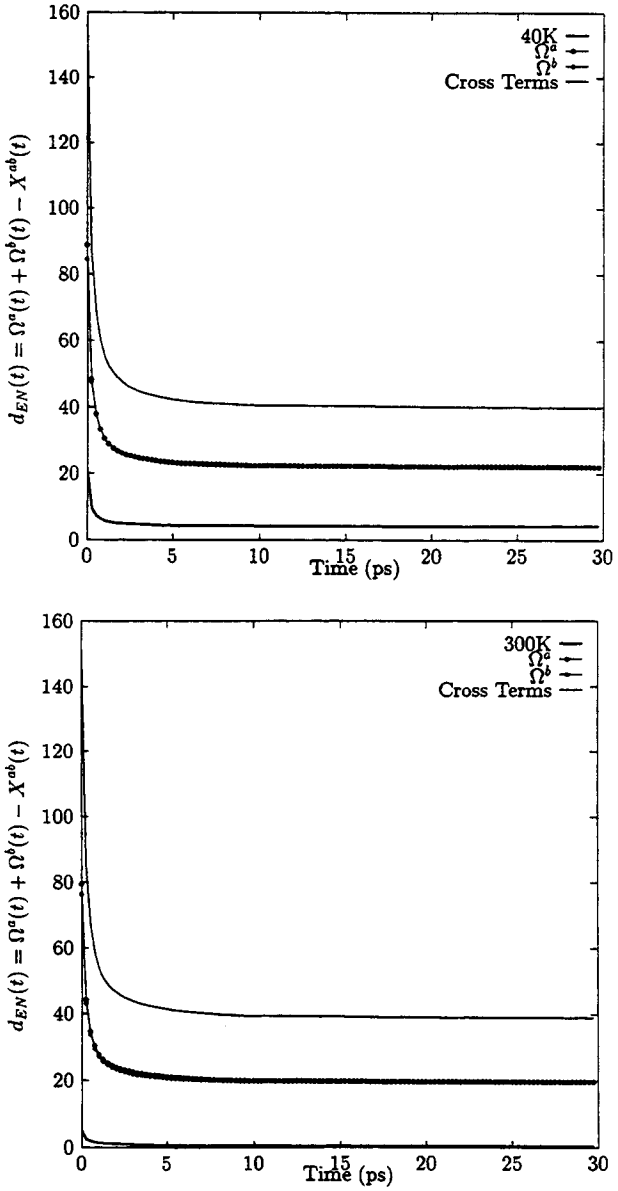

Figure 9. Plots of the nonbonded energy metric $d_{EN}(t)$, $\Omega^a(t)$, $\Omega^b(t)$, and $X^{ab}(t)$ as a function of time for the RNase A enzyme/product complex at (top) 40 and (bottom) 300 K. The thin solid line represents $X_{ab}(t)$, the open circles correspond to $\Omega^a(t)$, the closed circles to $\Omega^b(t)$. The thick solid line is $d_{EN}(t) = \Omega^a(t) + \Omega^b(t) - X^{ab}(t)$.

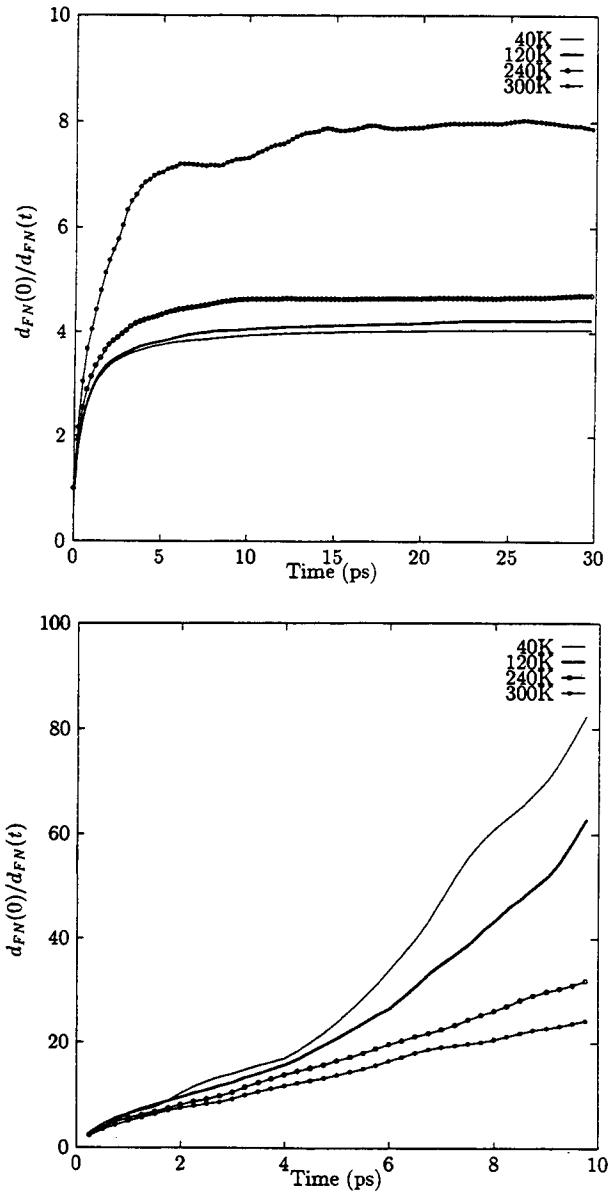

Figure 10. Normalized nonbonded force metric, $d_{FN}(0)/d_{FN}(t)$, as a function of time for the enzyme/product complex generated from (top) two independent trajectories and (bottom) two 25-ps portions of a single 50-ps trajectory.

75-ps time scale. At 300 K, there is a significant region of linear convergence, followed by a plateau beyond 15-ps. This behavior clearly indicates the presence of a wide distribution of time scales for the protein motion. The significant point is that *even within this short time scale, several CSs are sampled*, as indicated by the change in the slope of $d_{EN}(t)$. We expect that the initial convergence is due to two mechanisms: (1) atoms rapidly exploring the local potential and (2) the similarity in the *a* and *b* trajectories of interactions along the main chain which are largely independent of the protein conformation. The longer time relaxation is related to infrequent barrier crossing, largely in the form of dihedral angle transitions (discussed below) and the diffusive motion of subdomains of the protein, which may shift in relative orientation.[4] We expect these motions to occur on much longer time scales.

Decomposition of the nonbonded energy metric into the fluctuation metric and its cross terms provides an interesting contrast between the low- and high-temperature dynamics. The corresponding decomposition, into fluctuation metrics and cross terms according to eq 7 is shown in Figure 9 at 40 and 300 K. At both 40 and 300 K, the fluctuation metrics are nearly identical, indicating that the fluctuations in the two structures are similar. The similarity in the average structure of the *a* and *b* trajectories at 40 K manifests itself in the presence of a significant cross term $X^{ab}(t)$, indicating that the nonbonded energy is correlated between the two structures and does not decay over 30 ps. At 300 K, the nonbonded energy metric is initially smaller and decays to near zero in approximately 10 ps, indicating that there are dihedral transitions allowing the system to sample independent CSs within the 30 ps trajectory.

A dramatic demonstration of the presence of multiple CSs in proteins can be made by comparing the force metric calculated using *two independent* (*a* and *b*) trajectories and the force metric calculated by dividing a *single trajectory* (just *a*) into two portions. In Figure 10 we show the reciprocal of the nonbonded force metric $d_{FN}(t)$ for the full enzyme calculated using (1) two independent trajectories and (2) two portions of a single 50-ps trajectory. The data calculated using the independently generated trajectories are similar in form to the results for the S-peptide.[2] These results suggest that on the time scale of 50 ps the system explores distinct CSs and that much longer times are needed for transitions between CSs explored by the two independent trajectories. However, the data calculated from a single trajectory show that at low

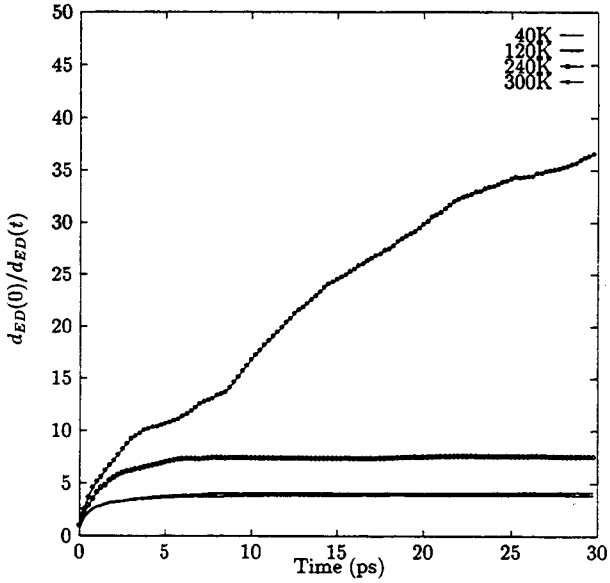

Figure 11. Reciprocal of the dihedral energy metric, $d_{ED}(0)/d_{ED}(t)$ as a function of time for the enzyme/product complex.

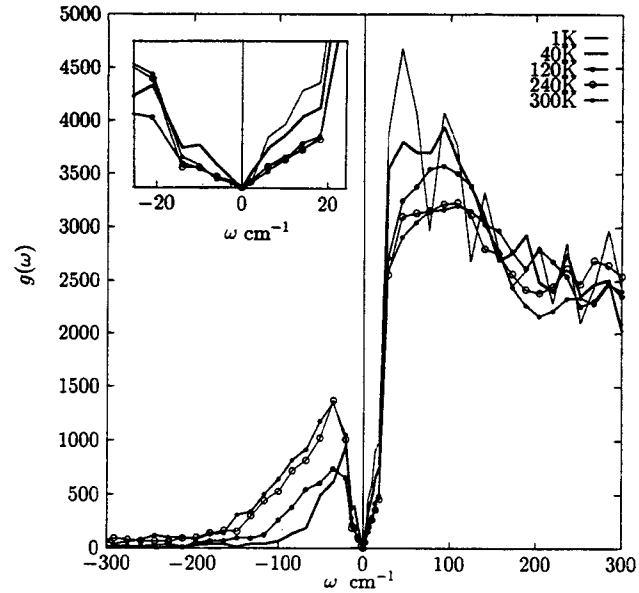

Figure 12. Instantaneous normal mode density of states at several temperatures calculated over a 75-ps constant energy trajectory of the S-peptide. Imaginary frequencies are plotted on the negative frequency axis.

temperatures there is rapid convergence of the force metric. This indicates that on the time scale of our simulation at low temperatures the single equilibrium trajectory tends to fluctuate about *one particular substate*. Interestingly, as the temperature is increased, the rate of convergence *decreases*, indicating that the protein is making transitions between nonequivalent substates, slowing the convergence of the metric. A similar comparison made for the S-peptide using two 75-ps portions of a 150-ps trajectory leads to qualitatively identical results. This calculation demonstrates that averaging over a single long trajectory may lead to artificially small fluctuations and a false sense of the accuracy of a given calculated property since the calculation includes an average over a single CS. *Thus, running an ensemble of independent, short trajectories of* 10–15 *ps will provide better averaging than a single long trajectory.*

Since the work of Flory, it has been appreciated that significant changes in the conformation of peptides and proteins are tied to transitions of dihedral angles. We have developed a useful means of measuring the progress in conformational sampling through dihedral angle transitions using the dihedral angle energy metric $d_{ED}(t)$.[2,38] The results for $d_{ED}(t)$ of the S-peptide are shown in Figure 11. It is expected that the nonbonded energies depend parametrically on the dihedral angles and that relaxation of the nonbonded interactions requires dihedral angles transitions.[2] At temperatures <240 K, the behavior of the dihedral energy metric closely follows that of the nonbonded energy $d_{EN}(t)$ (see Figure 8). At 300 K, there is more rapid self-averaging in the energy of the dihedral angles, which shows a roughly linear convergence over the full 30 ps. However, it is important to note that there are several slopes in the 300 K trajectory, indicating the presence of multiple time scales, presumably due to the distribution of rates of transition for the various dihedral angles.

There is an important distinction between convergence of $d_{EN}(t)$ and $d_{ED}(t)$. The dihedral angle potential energy is a local function of the dihedral angle between four atoms, and local transitions in the dihedral angles can lead to significant convergence in $d_{ED}(t)$. However, the nonbonded potential energy is a nonlocal function, and the convergence of $d_{EN}(t)$ requires global, cooperative fluctuations in addition to localized dihedral angle transitions. Therefore, convergence of $d_{ED}(t)$ is a necessary but not sufficient condition for the convergence of $d_{EN}(t)$.

C. Instantaneous Normal Mode Spectra. Our results for the force and energy metrics show that the long time scale relaxation is strongly tied to infrequent events such as dihedral angle transitions and activated barrier crossing. To further explore the nature of the relaxation processes, we have calculated the instantaneous normal mode spectrum for the S-peptide from the *a* trajectories at eight temperatures ranging from 40 to 500 K. The resulting normal mode density of states is shown in Figure 12 at a few select temperatures. At sufficiently low temperatures, the eigenfrequencies are real and the density of states agrees with the 0 K normal mode spectrum. As the temperature is increased, the number of imaginary eigenfrequencies increases. In fact, there is a significant percentage of unstable modes evident even at 40 K. As the temperature is increased further, the number of imaginary frequencies increases and the center of the imaginary frequency lobe is pushed to higher (imaginary) frequency. These trends are summarized in Table 1.

The inset of Figure 12 shows in detail the low-frequency behavior of the instantaneous normal mode distribution. For both the imaginary and real frequency lobes, the initial increase at low frequencies, $\omega < 20$ cm^{-1}, is linear with a slope which is independent of temperature for $T \geq 120$ K. At lower temperatures the initial slope is obviously steeper. Keyes and Madan have noticed a similar difference in slope for the density of states of a Lennard-Jones system above and below the glass transition temperature.[39] For the S-peptide, the change in low-frequency behavior is coincident with the onset of a significant number of dihedral angle transitions, just as the change in slope for the Lennard-Jones system is coincident with the onset of self-diffusion.

The linear frequency dependence for $g(\omega)$ translates into a predicted low-temperature heat capacity which increases as $C_v(T) \approx T^2$. It is well known that the low-temperature heat capacity of glasses varies approximately linearly with temperature. This is often interpreted in terms of a constant density of tunneling states. However, at higher (but still low) temperatures, where phonons make some contribution, one can derive a $C_v(T) \approx T$ by assuming a constant density of vibrational states. Data of Hutchens, Cole, and Stout[40] for the heat capacity of insulin and bovine chymotrypsinogen A in the range of 10 K $\leq T \leq$ 20 K could be interpreted, with some ambiguity, to be increasing as

(38) The related dihedral angle energy *fluctuation* metric has been employed recently in an interesting study relating the uncertainty in biomolecular free energy calculations to the existence of conformational substates; see: Hodel, A.; Simonson, T.; Brünger, A. T. *J. Phys. Chem.* **1993**, *97*, 3409.

(39) Keyes, T., private communication.

(40) Hutchens, J. O.; Cole, A. G.; Stout, J. W. *J. Biol. Chem.* **1969**, *244*, 26.

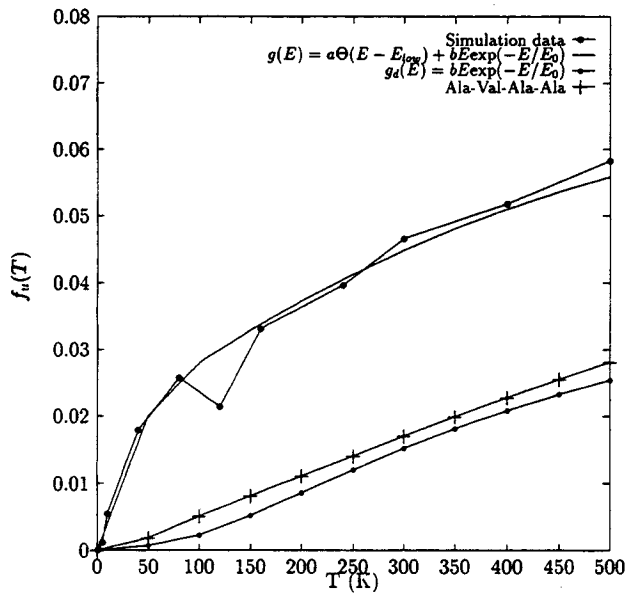

Figure 13. Fraction of unstable modes $f_u(T)$ as a function of temperature calculated from a 75-ps constant energy trajectory of the S-peptide. Shown for comparison is the fit to $f_u(T)$ using eqs 20, 22, and 27 and that calculated from the distribution of barriers computed for the tetrapeptide Ala-Val-Ala-Ala by Choi and Elber.

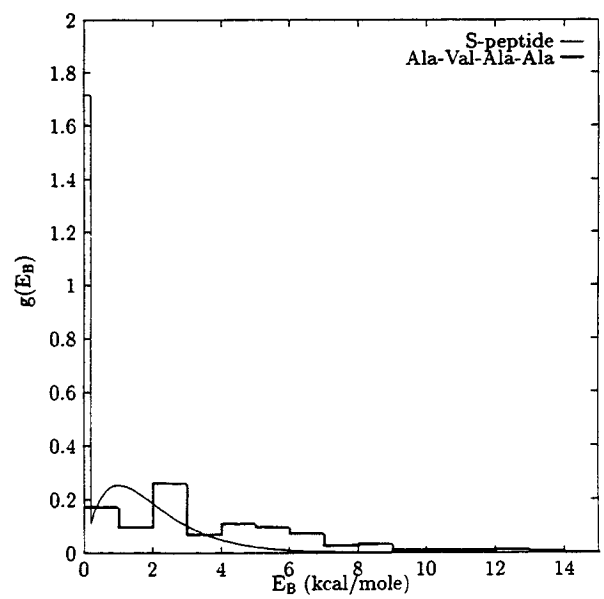

Figure 14. Distribution of barrier heights extracted from a fit of the fraction of unstable modes as a function of temperature. Shown for comparison is the distribution of adiabatic barrier heights measured directly for the tetrapeptide Ala-Val-Ala-Ala by Choi and Elber.

T^2. However, data at lower temperatures would be required to remove this ambiguity.

In Figure 13 we show the fraction of unstable modes $f_u(T)$ as a function of temperature for the S-peptide in vacuum based on the data in Figure 12. The fraction of unstable modes is zero at 0 K and then rises quickly through 100 K, at which point there is a break in slope, leading to a slower increase through higher temperature. At 120 K there is a noticeable dip in the fraction of unstable modes.

The behavior of $f_u(T)$ is readily interpreted in terms of the model presented in section III. According to this model, we represent the rugged energy landscape by a piecewise harmonic potential with an arbitrary distribution of barrier heights $g(E_B)$. A large number of modes in the peptide, such as localized harmonic bond and angle vibrations, will remain stable even at the highest temperatures. Other modes, such as those localized in dihedral angles or the low-frequency, continuum-like, global fluctuations, can become unstable. Physically, we know that the quick increase in $f_u(T)$ even at very low temperatures indicates significant anharmonicity in the lowest frequency vibrational modes and the presence of many low barriers where $E_B/k_B \simeq 100$ K. In our model, this rapid rise in $f_u(T)$ at low temperature is attributed to the presence of many low barriers. The slower rise in $f_u(T)$ at higher temperatures indicates the presence of many high barriers, and the rate of increase in $f_u(T)$ at high temperatures may depend sensitively on the particular distribution $g(E_B)$.

The distribution function $g(E_B)$ can be obtained by solving the homogeneous Fredholm integral equation of the first kind (cf. eq 20) using our caricature of the true rugged energy landscape. In general, there are two methods for solving such equations. The solution may be obtained by expanding $g(E_B)$ in terms of known functions and adjusting the coefficients to minimize the square of the deviation between the computed $f_u(T)$ and the assumed $f_u(T)$. Alternatively, a variational ansatz for $g(E_B)$ can be chosen, and the coefficients in the ansatz can be adjusted to obtain the best fit. We adopted the later procedure and chose a (nonunique) functional form for $g(E_B)$ that incorporates the physical insight provided by the changes in $f_u(T)$ with temperature:

$$g(E_B) = a\theta(E_{low} - E_B) + bE_B e^{-E_B/E_0} \qquad (27)$$

In our model, $g(E_B)$ consists of a constant density of low-energy barriers for $E < E_{low}$ (written in terms of the Heaviside function $\Theta(E)$) and a Poisson distribution of higher energy barriers (with a maximum at $E_{max} = E_0$. A fit to our data for $f_u(T)$ using a symmetric piecewise parabolic potential energy leads to $a = 0.325$ (kcal/mol)$^{-1}$, $E_{low} = 0.2$ kcal/mol, $b = 0.130$ (kcal/mol)$^{-2}$, and $E_0 = 1$ kcal/mol. The distribution divided by the proportionality constant 0.65 (kcal/mol)$^{-2}$, which was used as a fitting factor, was reported previously.[3]

The barrier height distribution integrates to $\int dE\, g(E) = 0.195$, of which 0.065 is due to the constant distribution of low barriers with $E_B < 0.2$ kcal/mol while the Poisson distribution of higher energy barriers contributes 0.13. For our model of each normal mode moving over a symmetric piecewise parabolic potential, at high enough temperature the probability of being in a well or barrier region will be $\bar{f}_u(E_B,T) = 1/2$. It follows that in the limit of high temperatures, the fraction of unstable modes will approach $f_u(T = \infty) = 0.0975$, meaning that roughly 20% of the instantaneous normal modes are expected to become unstable.

Our variational ansatz for $g(E_B)$ with the above choice of parameters solves the integral equation adequately over the entire temperature range. This is shown in Figure 13, where a comparison between $f_u(T)$ computed by MD simulations and that obtained by evaluating the right-hand side of eq 20 using eqs 22 and 27 is given. The fit is good over the full range of temperatures examined.

The form of $g(E_B)$ is shown in Figure 14 with a histogram of the distribution of energy barriers calculated for the tetrapeptide Ala-Val-Ala-Ala by Choi and Elber. The tetrapeptide data are plotted with a resolution of 1 kcal/mol and show a large number of low-energy barriers, where $E_B < 1$ kcal/mol, in qualitative agreement with our result for $g(E_B)$. However, the tetrapeptide has many high-energy barriers which are not present in our form of $g(E_B)$.

Figure 13 shows the result of using the distribution of barriers calculated by Choi and Elber for the tetrapeptide in eq 38 to predict $f_u(T)$. We also show two other plots, the first being the contribution to the fraction of unstable modes due to the distribution of the higher energy barriers based on eq 27 for $g(E_B)$, where $a = 0$:

$$g_d(E_B) \simeq bE_B e^{-E_B/E_0} \qquad (28)$$

The resulting $f_u(T)$ predicted by (1) the high-energy portion of our distribution $g_d(E_B)$ and (2) the tetrapeptide data of Choi and Elber are in reasonable agreement. Therefore, the difference

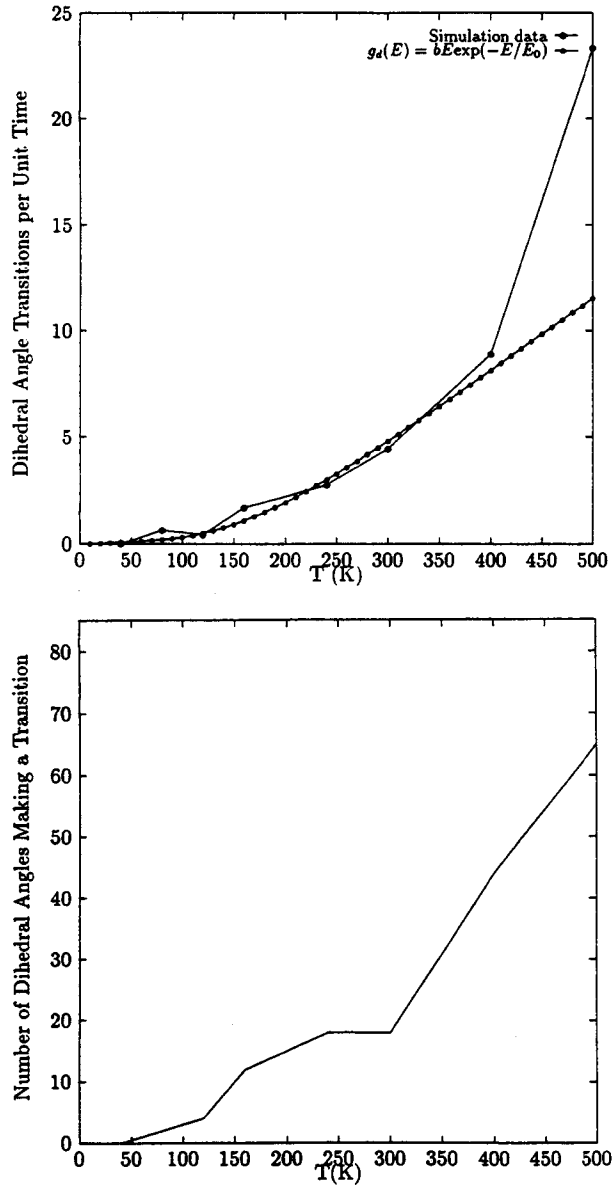

Figure 15. Number of (top) dihedral angle transitions per unit time (1 ps) and (bottom) dihedral angles making transitions as a function of temperature for the 75-ps constant energy trajectories for the S-peptide at a series of temperatures. In the bottom part we show our theoretical estimate of the number of dihedral angle transitions based on the distribution of barrier heights derived from our model for $f_u(T)$.

between $f_u(T)$ at large E_B calculated using $g_d(E_B)$ for the tetrapeptide and our model cannot be resolved by a measurement of $f_u(T)$ through 500 K. To obtain information on the high-energy portion of $g(E_B)$, one must simulate $f_u(T)$ at much higher temperatures so that $\bar{f}(T,E_B)$ is large enough to contribute to $f_u(T)$.

Long-length scale changes in the peptide structure necessary for effective sampling of conformation space will require many dihedral angle transitions. Based on our model of the increase in $f_u(T)$ as a function of temperature and our analysis of the nonbonded metrics, we expect to see an increase in the number of dihedral angle transitions as the temperature is increased. In Figure 15 we show (a) the number of dihedral angles making transitions and (b) the total number of dihedral angle transitions as a function of temperature. The increase in dihedral transitions is strongly correlated with the increase in the fraction of unstable modes described by the high-energy portion of $g(E_B)$ (see Figure 14). At 40 K and below, no dihedral angle transitions were measured. It follows that the significant fraction of unstable modes measured at 40 K and modeled by the constant low-energy term in $g(E_B)$ does not result from dihedral angle transitions but is more likely associated with collective fluctuations or stable, anharmonic motion.

To examine this proposal more carefully, we can estimate the total number of dihedral angle transitions expected during a 75-ps trajectory of the S-peptide based on $g(E_B)$ and the assumption that the high-energy Poisson distribution corresponds to the distribution of barriers for the peptide dihedral angles. For a single dihedral angle with barrier E_B, the total number of transitions which occur in a simulation at temperature T will be approximately $k(E_B,T)t_{sim}$—the rate of transition $k(E_B,T)$ per unit time times the simulation time t_{sim}. If we average the number of transitions expected for a single dihedral with characteristic barrier height E_B over the distribution of barriers $g_d(E_B)$ and multiply by the total number of vibrational degrees of freedom $3N$, we can estimate the total number of dihedral angle transitions per unit time (taken to be 1 ps) to be

$$n_d(T) = 3N\int_0^\infty dE_B\, g_d(E_B)k(E_B,T) \qquad (29)$$

For the distribution of barrier heights, we use the high-energy Poisson distribution $g_d(E_B)$ derived from the instantaneous normal mode analysis (see eq 28).

For the S-peptide, the number of atoms in the polar hydrogen representation is $N = 186$. From the normalization of $g_d(E_B)$, we see that the number of dihedral angles which can make a transition will be $3N\int dE_B\, g_d(E_B) \approx 73$. This number agrees surprisingly well with the data of Figure 14 at 500 K, where the number of dihedral angles making a transition is 64. For the transition rate, we assume a simple transition-state theory equilibrium rate constant for a harmonic system with well frequency ω_0 in which the barrier height $E_B \gg k_BT$:

$$k(E_B,T) = \frac{\omega_0}{2\pi}e^{-\beta E_B} \qquad (30)$$

A more general form of the rate constant for conformational transitions in the protein can be derived from the theory of Helfand and Skolnick.[41] Equation 30 is a special case of their theory in the transition-state limit for the case of a one-dimensional periodic potential.

The frequency ω_0 was approximated by assuming that the well minima are separated by π radians and that the moment of inertia of the dihedral angle is 25 gÅ²/mol. Combining eqs 29, 28, and 30, the well frequency is $\omega_0 = (4/\pi)(418.4E_B/I)^{1/2}$, where E_B is in kcal/mol and I is in gÅ²/mol, and the total number of dihedral angle transitions as a function of temperature can be estimated. The results are displayed in Figure 15. Through 300 K, our theoretical estimate of the number of dihedral angle transitions agrees well with the simulation data. We emphasize that the number of dihedral angle transitions was predicted from the barrier height distribution derived from $f_u(T)$ and includes no adjustable parameters other than the moment of inertia (a multiplicative constant). It follows that the agreement in Figure 15a is an independent confirmation of the reasonableness of our predicted barrier height distribution $g(E_B)$. Certainly our model has many approximations, including the use of an approximate moment of inertia and the asymptotic ($E_B \gg k_BT$) form of the transition state theory rate constant for a harmonic system. As such, it is not surprising that the increase in transition number at 500 K is not followed by the model. Nevertheless, the good agreement lends supports to our claim that the instantaneous normal mode calculation may be used as a powerful probe of the barrier height distribution in peptides and proteins.

(41) Helfand, E. *J. Chem. Phys.* **1971**, *54*, 4651. Skolnick, J.; Helfand, E. *J. Chem. Phys.* **1980**, *72*, 5489. Skolnick, J.; Mattice, W. L. *Macromolecules* **1981**, *14*, 292.

VI. Conclusion

The ultimate goal of theoretical modeling of proteins is to understand, at a microscopic level, the relation between their motions and biological activity. It is becoming clear that this involves the knowledge of how proteins explore the complex energy landscape. In this study we have presented quantitative ways of probing the multivalley structure in heterogeneous systems in general and in proteins in particular. The major results of this study are summarized below.

(1) We have used replica molecular dynamics, in which the properties of two molecular dynamics trajectories generated from independent initial configurations are compared, as a means of exploring the complex energy landscape in proteins. This method is used to compute the force metric and energy metric from the dynamics of conformational state sampling in the S-peptide and ribonuclease A.

The most important conclusion of our study is that the conformational substates picture, which has been found to be necessary to rationalize a series of experiments on the dynamics of binding of small ligands to proteins, already becomes relevant when describing the motion of proteins on the time scale of 10 ps. This suggests that CSs are present on all length scales, and therefore even moderate rearrangement of segments in proteins necessary for folding from denatured states will be a highly cooperative process involving transitions over several barriers separating the different conformational states. The metrics also provide a microscopic relationship between the different motions that are involved as the structure of the protein changes with temperature. In particular we find that there appears to be a close connection between the relaxation of nonbonded interactions in proteins and the dynamics of dihedral angle transitions. The former large-scale motion which is responsible for the folding of proteins is in fact mediated by a sequence of local dihedral angle transitions.

(2) Since there are numerous CSs that lead to diversity in the time scale for protein motion, it becomes necessary to provide a statistical description of the energy landscape. To this end we have also introduced a novel method to compute the distribution of barrier heights in the complex energy landscape. This is accomplished by exploiting the properties of the eigenvalue distribution of the normal modes of the system. A simple caricature of the complex energy landscape is used to derive an integral equation relating the distribution of barrier heights to the fraction of unstable normal modes. The picture of the landscape used to obtain the integral equation is quite general and should be applicable to any systems whose underlying landscape is rough. The application of this methodology to the S-peptide shows that the distribution of barrier heights obeys a Poisson function. Evidence from theoretical studies of a tetrapeptide and experiments suggest that this may be a general feature of proteins. The Poisson distribution for barrier heights becomes necessary to account for kinetics of rebinding of small ligands in which the binding is dictated by the complex energy landscape of proteins.[1] The analysis of the normal mode spectrum of proteins also leads to the number of dihedral angle transitions that are expected to occur on a given time scale. We have used the Poisson distribution for the barrier heights to calculate the number of dihedral angle transitions as a function of temperature that are expected to occur on a given time scale. The results of this calculation are in very good agreement with the simulations, further validating our model of the complex energy landscape.

(3) From a technical point of view, our study has very important implications for the simulation of biological molecules in which the interplay between several energy scales leads to a complex energy landscape. We have shown that in order to obtain reliable results using molecular dynamics simulations, a straightforward application may not suffice. Our analysis suggests that the following strategy, which has been applied in another context,[42] should be followed when simulating systems dominated by the presence of several distinct conformational states separated by large barriers. One should generate an ensemble of independent initial conditions at high enough temperatures where the relevant barrier heights become irrelevant. Starting from these distinct initial conditions, the system can be quenched to the desired low temperature. The various initial configurations, in all likelihood, would map on to distinct conformational substates. The conformations belonging to a single CS can be sampled effectively at a rate that can be monitored using the fluctuation metric. The average over all the conformational substates would provide an accurate result for the quantity to be computed. Thus it is more accurate to average over several short trajectories each (presumably) sampling a distinct conformational substate than to average over one long trajectory.

The possibility of obtaining converged results in simulations of proteins has been the subject of a mounting discussion.[2,38,43,44] In this and previous work we have presented an easily implemented quantitative measure of the extent of convergence in simulation averages. The question of whether convergence in complex systems can be achieved using present computing resources remains to be answered in future work.

Acknowledgment. J.E.S. thanks Bruce Tidor, Bhupinder Madan, and Tom Keyes for helpful discussions. D.T. is grateful to Ray Mountain and Zhuyan Guo for their help. We thank Ron Elber for providing unpublished results on the tetrapeptide. J.E.S. is grateful to the donors of the Petroleum Research Fund, administered by the American Chemical Society, and to the National Science Foundation (CHE-9306375) for partial support of this research. D.T. acknowledges partial support from the National Science Foundation, the Air Force Office of Scientific Research, and the Camille and Henry Dreyfus Foundation.

Appendix: The Total Force Metric

Here we derive the form of the total force metric for a general dynamics and for the specific case of a harmonic system. We begin with an analysis of the fluctuation metric. The time average of the total force, $\vec{f}_j(t)$, on the jth atom is

$$\vec{f}_j(t) = \frac{1}{t}\int_0^t \vec{F}_j(s)\,ds = \frac{m_j}{t}[\vec{v}_j(t) - \vec{v}_j(0)] \qquad (A.1)$$

where we have used the fact that $\vec{F}_j = m_j\, d\vec{v}_j/ds$. Assuming conservation of linear momentum, we find

$$\bar{f}(t) = \frac{1}{N}\sum_j^N \vec{f}_j(t) = \frac{1}{Nt}\sum_j^N m_j[\vec{v}_j(t) - \vec{v}_j(0)] = 0 \qquad (A.2)$$

Combining the last two equations with eq 3, we find the following form for the total force fluctuation metric:

$$\Omega_{FT}(t) = \frac{1}{Nt^2}\sum_j^N m_j^2[v_j^2(t) + v_j^2(0) - 2\vec{v}_j(t)\cdot\vec{v}_j(0)] \qquad (A.3)$$

After a short time we expect that the kinetic energy will equipartition, and we can replace the first two terms on the right-hand side with the thermal average $\langle m_j v_j^2\rangle = 3k_BT$. The final term averaged over all atoms can be written as the momentum time correlation function:

$$\langle\vec{p}(t)\cdot\vec{p}(0)\rangle = \frac{1}{N}\sum_j^N m_j^2\vec{v}_j(t)\cdot\vec{v}_j(0) \qquad (A.4)$$

The final result is

(42) Bader, J. S.; Kuharski, R. A.; Chandler, D. *J. Chem. Phys.* **1990**, *93*, 230.
(43) Hermans, J. *Proteins* **1993**, *17*(1), ii.
(44) Shi, Y.-Y.; Mark, A. E.; Wang, C.-X.; Huang, F.; Berendsen, H. J. C.; van Gunsteren, W. F. *Protein Eng.* **1993**, *6*, 289.

$$\Omega_{FT}(t) = \frac{2}{t^2}[3k_BTM - 2\langle\vec{p}(t)\cdot\vec{p}(0)\rangle]^2 \qquad (A.5)$$

where we have defined the average atomic mass

$$M = \frac{1}{N}\sum_j^N m_j \qquad (A.6)$$

The total force metric is defined in terms of two independent trajectories labeled a and b as

$$d_{FT} = \frac{1}{N}\sum_j^N [\vec{f}_j^{\,a}(t) - \vec{f}_j^{\,b}(t)]^2 \qquad (A.7)$$

Equation 7 relates the total force metric to the fluctuation metric. Substituting our result for the fluctuation metric eq A.5 into eq 7 leads to the result

$$d_{FT}(t) = \frac{2}{t^2}[6k_BTM - \langle\vec{p}^{\,a}(t)\cdot\vec{p}^{\,a}(0)\rangle + \langle\vec{p}^{\,b}(t)\cdot\vec{p}^{\,b}(0)\rangle] \qquad (A.8)$$

where we assume that the velocities are Gaussian variables and the cross term

$$\sum_j^N (\vec{p}_j^{\,a}(t) - \vec{p}_j^{\,a}(0))\cdot(\vec{p}_j^{\,b}(t) - \vec{p}_j^{\,b}(0)) = 0 \qquad (A.9)$$

Equation A.8 is a general result for the time dependence of the total force metric. To evaluate it, we must calculate the momentum autocorrelation function. In general, this is a difficult problem. However, we now focus on the special case of a harmonic system, where this is a simple matter.

In a zero temperature normal mode approximation, the potential is expanded in a Taylor series of the $3N$ coupled coordinates r^N about a mechanically stable equilibrium position r_0^N:

$$U(r^N) = U_0(r_0^N) - \sum_j^N \vec{F}_j\cdot\vec{x}_j + 1/2\sum_j^N\sum_i^N \vec{x}_i\cdot\kappa_{ij}\cdot\vec{x}_j \qquad (A.10)$$

where the potential energy at the equilibrium position is $U_0(r_0^N)$, $\vec{x} = (\vec{r}^N - \vec{r}_0^N)$, $\vec{F}$ is the $3N$-dimensional force vector (which is zero about the equilibrium position), and κ is the $(3N \times 3N)$-dimensional force constant matrix or matrix of second derivatives about the equilibrium position. The normal mode transformation is defined in the usual way. To summarize, the equations of motion for small displacements of the jth atom about the equilibrium position are

$$m_i\ddot{\vec{x}}_i = -\sum_j^N \kappa_{ij}\cdot\vec{x}_j \qquad (A.11)$$

It is convenient to transform from Cartesian to mass-weighted coordinates $\vec{y}_j = (m_j)^{1/2}\vec{x}_j$. The equations of motion become

$$\ddot{\vec{y}}_i = -\sum_j^N \omega^2_{ij}\cdot\vec{y}_j \qquad (A.12)$$

The symmetric frequency matrix ω^2 may be diagonalized by a unitary transformation defined by the matrix a to give the equations of motion for the normal mode coordinates q_i where $\ddot{q}_i = -\omega_i^2 q_i$. The eigenvalues ω_i^2 of the frequency matrix are the eigenfrequencies squared of the $3N$ independent normal mode coordinates. The original Cartesian coordinates are related to the normal mode coordinates by

$$\vec{x}_j = \frac{1}{(m_j)^{1/2}}\sum_i^N a_{ji}q_i \qquad (A.13)$$

where a_{ji} are the projections of the ith normal mode on the mass-weighted coordinates of the jth atom.

To evaluate the total force metric, we need to evaluate the momentum correlation function. Since

$$\vec{p} = m_j\dot{\vec{y}}_j = m_j\dot{\vec{x}}_j = (m_j)^{1/2}\sum_i^N a_{ji}\dot{q}_i \qquad (A.14)$$

it follows that

$$\vec{p}_j(t)\cdot\vec{p}_j(0) = m_j\sum_{i,k} a_{ji}a^*_{jk}\dot{q}_i(t)\dot{q}_k(0) \qquad (A.15)$$

Using the condition of orthonormality of the normal mode eigenvectors

$$\vec{p}_j(t)\cdot\vec{p}_j(0) = m_j\sum_i |a_{ji}|^2\dot{q}_i(t)\dot{q}_i(0) \qquad (A.16)$$

The solution for the time evolution of a given normal mode is

$$q_i(t) = q_i(0)\cos(\omega_i t) + \frac{\dot{q}_i(0)}{\omega_i}\sin(\omega_i t) \qquad (A.17)$$

and

$$\dot{q}_i(t) = -\omega_i q_i(0)\sin(\omega_i t) + \dot{q}_i(0)\cos(\omega_i t) \qquad (A.18)$$

To evaluate eq A.16, we first approximate the autocorrelation of the velocity of a given normal mode as

$$\dot{q}_i(t)\dot{q}_i(0) = \dot{q}_i^{\,2}(0)\cos(\omega_i t) \simeq k_BT\cos(\omega_i t) \qquad (A.19)$$

where we have performed a thermal average over the initial conditions. Collecting eqs A.16 and A.19, the momentum correlation function of eq A.4 is

$$\langle\vec{p}(t)\cdot\vec{p}(0)\rangle = \frac{k_BT}{N}\sum_j^N m_j\sum_i^{3N}|a_{ji}|^2\cos(\omega_i t) = \frac{k_BT}{N}\sum_i^{3N}\langle m\rangle_i\cos(\omega_i t) \qquad (A.20)$$

We have defined the "average mass of the ith normal mode" as

$$\langle m\rangle_i = \sum_j^N m_j|a_{ji}|^2 \qquad (A.21)$$

using the normalization condition that

$$\sum_j^N |a_{ji}|^2 = 1 \qquad (A.22)$$

Inserting eq A.20 for the momentum correlation function into the general form for the total force metric eq A.8, we find the final result for the total force metric in the normal mode approximation:

$$d_{FT}(t) = \frac{12k_BT}{t^2}\left[M - \frac{1}{3N}\sum_i^{3N}\langle m\rangle_i\cos(\omega_i t)\right] \qquad (A.23)$$