

Analysis of Calculated Normal Modes of a Set of Native and Partially Unfolded Proteins

Herman W. T. van Vlijmen[†] and Martin Karplus*

Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, and Laboratoire de Chimie Biophysique, Institut le Bel, Université Louis Pasteur, 4, Rue Blaise Pascal, 67000 Strasbourg, France

Received: August 21, 1998; In Final Form: February 2, 1999

To determine the effect on the normal modes of the multimimum surface of the protein native state, calculations have been done for a series of conformations and different crystal structures of bovine pancreatic trypsin inhibitor (BPTI) and hen egg white lysozyme (HEWL). The conformations were generated by molecular dynamics calculations. The results indicate that the conformational space spanned by the 75 (BPTI) and 130 (HEWL) low-frequency normal modes, which account for approximately 90% of the atomic fluctuations, is relatively invariant with respect to different conformations in the native state manifold. However, the space spanned by the first three normal modes of the conformation is more variable. The overlap of the low-frequency normal modes of the native conformations with two (partially) unfolded HEWL structures is considerably lower. It is shown that averaging over the normal-mode results from an ensemble of native conformations smoothens the density of states curve and improves the agreement with inelastic neutron scattering experiments. Quantum-mechanical configurational entropies of the different structures were calculated from the normal-mode frequencies. The variation in entropy values between different crystal structures is similar to the variation between the different molecular dynamics structures. For the partially unfolded HEWL structures the entropy increased as the protein was further unfolded. To avoid biasing normal-mode calculations to a particular crystal structure and to provide error bounds, normal mode results should be averaged over a set of native conformations which can be generated by molecular dynamics.

Introduction

Since the first molecular dynamics (MD) simulations of proteins,¹ computer simulations of the motions of macromolecules have advanced significantly, largely because of improvements in methodology and the availability of faster computers.^{2,3} The study of the dynamics of proteins is important since the internal motions are related to their function.^{4,5} An alternative approach to the internal motions is provided by harmonic (or normal mode) analysis.³ In normal-mode calculations,⁶ the potential function is approximated by a sum of quadratic terms in the displacements, usually around a (local) minimum energy conformation. The calculation consists of a diagonalization of the mass-weighted second derivative matrix of the system; the eigenvectors are the normal modes. Within the harmonic approximation an essentially exact solution for the dynamics of the system is obtained in this way. Thus, quantities associated with the dynamics can be calculated without statistical errors. In addition, normal modes provide a direct method for including quantum effects in the calculation of thermodynamic quantities, such as the heat capacity and absolute entropy.³ Normal-mode analyses were applied first to proteins in 1983,^{7–9} and have been reviewed recently.^{10–12}

Normal modes have also been used to analyze molecular dynamics or Monte Carlo simulations. In such a quasiharmonic (or principal component) treatment, the second derivative matrix is constructed from the second moment matrix of the atomic displacements obtained from the simulation.^{13–15} This method implicitly includes some aspects of the anharmonicity by

approximating the potential by a quadratic surface that gives the same second moments of the displacements as the simulation. The method of essential dynamics^{16–19} uses projections of a regular MD simulation onto selected (non-mass-weighted) quasiharmonic modes to extract large scale motions of interest. A disadvantage of such approaches is that a trajectory has to be calculated first, so it cannot be used as an alternative to an MD simulation. Moreover, the statistical errors of the simulation are carried over to the normal-mode results.

The harmonic approximation to the potential energy can, in principle, introduce systematic errors into the description of dynamic properties. Elber and Karplus²⁰ have shown from molecular dynamics calculations of myoglobin that the native protein potential energy surface has a multimimum character. This was confirmed by Gō and Noguti²¹ for the bovine pancreatic trypsin inhibitor (BPTI). Recently, Jařezić et al.²² made a detailed comparison of harmonic and quasiharmonic dynamics of BPTI and found that only 42% of the root mean square atomic fluctuations of the protein in an MD simulation were obtained by harmonic dynamics. They concluded that the remaining 58% is due to anharmonic motions, such as transitions between conformational substates. Hayward et al.^{23,24} also calculated harmonic and quasiharmonic modes for BPTI, the latter based on Cartesian space MD simulations, and found that the first quasiharmonic mode contributed 43% to the total mean square of atomic fluctuations. It was suggested that this quasiharmonic mode corresponded to transitions between different energy minima based on projecting the MD trajectory onto the plane of the first two quasiharmonic modes. The atomic fluctuations on a residue by residue basis due to all but the first quasiharmonic mode were shown to have the same relative

[†] Present address: Biogen, Inc., 14 Cambridge Center, Cambridge, MA 02142.

values as the fluctuations calculated from normal modes. The correspondence was not exact, however, since the fluctuations from the quasiharmonic modes were, on average, a factor of 1.4 greater than the fluctuations resulting from the normal modes. By calculating the projections of the five²² or ten²³ lowest quasiharmonic modes onto the five or ten lowest normal modes, it was shown that the subspace spanned by the quasiharmonic modes can be described to a significant extent (up to ~50%) by the subspace spanned by the normal modes.

Hayward et al.²³ proposed that isolated protein normal modes would agree better with MD (or quasiharmonic) calculations of proteins in crystals. They based this proposal on the argument that the motion represented by the anharmonic modes is frozen out in the crystal, leaving only the harmonic contributions to the atomic fluctuations. Results of the normal-mode *B*-factor refinement method^{25,26} support this proposal. By refining *B* factors for a lysozyme crystal structure using a subspace of low-frequency normal modes of the isolated protein, Kidera et al.²⁶ showed that there was a quantitative agreement between the atomic fluctuations based on the normal modes of the molecule and the fluctuations based on *B* factors, excluding uniform rotation and translation of the molecule.

Smith et al.²⁷ compared the density of states from a BPTI powder inelastic neutron scattering experiment with densities of states calculated from normal-mode analyses of the minimized BPTI crystal structure. They observed a strong dependence of the results on the method of truncation of the nonbonded electrostatic interactions. In particular, the “switch” truncation method of the CHARMM program²⁸ gave better agreement with the experimental spectra, but the “shift” method agreed better with a calculation with no electrostatic truncation. Recently, Jañezič and Brooks²⁹ studied the effect of electrostatic truncations on normal-mode calculations and also found better agreements between the shift method and the method without truncation. Normal modes calculated with the shift method had a significantly higher density of states at the lower end of the frequency spectrum ($<50\text{ cm}^{-1}$), which are not present in the observed spectrum and are probably overdamped.²⁷ Also, the spectra resulting from any of the normal-mode calculations contained much more “structure” than the experimental spectrum. It was suggested that this is due to the fact that the normal modes were calculated for a single conformation.³⁰

It is of considerable interest to investigate the uniformity of the normal modes over a set of conformations in the multi-minimum subspace of the native state. Jañezič et al.²² did an extensive analysis of BPTI normal modes of many minimized MD snapshots and calculated projections of the lowest three normal modes of a particular trajectory frame onto the corresponding modes of a frame later in the simulation. They found some decrease in projection as the time difference between the frames increased. However, even at large time differences ($>100\text{ ps}$), the projections remained constant at a relatively high level (85%), indicating that the spatial character of the lowest frequency normal modes was relatively similar for the different structures obtained from the MD simulation.

Here, we analyze the normal modes of a set of native subspace conformations of BPTI and hen egg white lysozyme (HEWL) produced by MD simulations. We study the spatial similarity of the normal modes by calculating projections of the low-frequency modes subspace of one time frame onto others. The subspaces considered by us are significantly larger than the ones analyzed by Jañezič et al.²² (75 or 130 modes vs 3 modes), so as to include all modes that contribute significantly to the atomic fluctuations. For 10 different conformations we

calculate residue-based atomic fluctuations, cumulative atomic fluctuations, and density of states, and analyze the variability of these quantities. We also compare the results to the normal modes calculated from several available crystal structures of BPTI and HEWL. For BPTI we calculate simulated inelastic neutron scattering spectra as described by Smith et al.³¹ and determine that averaging of the spectra of individual conformations leads to a smoothing of the curve that would improve the agreement with experiment.²⁷ In addition to the native subspace conformations, we calculate normal modes and entropies for two partially unfolded HEWL conformations which were taken from a high-temperature unfolding simulation (Buck, M.; Karplus, M., manuscript in preparation). The description of the HEWL hinge bending motion by the low-frequency normal modes is analyzed for the native subspace conformations, the crystal structures, and the partially unfolded conformations.

Methods

MD Simulations. MD simulations of 100 ps were calculated for BPTI and HEWL in a bath of explicit water molecules with the stochastic boundary method.³² The program CHARMM²⁸ with the all-hydrogen parameter set 22 was used to set up the systems and calculate the trajectories. Initial coordinates for BPTI and HEWL were taken from crystal structure entries 4pti³³ and 1lyz³⁴ of the Brookhaven Protein Data Bank (PDB),³⁵ respectively. Hydrogen atoms of the protein and crystal waters were added with the HBUILD facility in CHARMM. The proteins were then solvated in a sphere of equilibrated TIP3 water molecules of radius 32 and 37 Å for BPTI and HEWL, respectively. These water sphere sizes ensured that the proteins had a solvation shell of at least 15 Å. Waters with their oxygen atom closer than 2.8 Å to a protein or crystal water heavy atom were discarded. A stochastic boundary was added to the system, with a reaction region of radius 29 Å for BPTI and 34 Å for HEWL, and a buffer region shell of 3 Å for both proteins. Stochastic boundary molecular dynamics propagates atoms within the reaction region according to regular Newtonian dynamics, while atoms in the buffer region are propagated by Langevin dynamics.³⁶ At the edge of the buffer region a boundary force is applied that prevents the water molecules from escaping. With the protein fixed, we did a short (50 steps) minimization of the waters, followed by 10 ps of MD to let the waters relax around the protein structure. The water overlay, minimization, and dynamics sequence was repeated three times, with the TIP3 sphere rotated 90° around the *X*, *Y*, and *Z* axis, respectively. These multiple solvent overlays were done to make sure that no water-sized holes remained in the system. To speed up the dynamics calculations, we multiplied the hydrogen masses by a factor of 10, which enabled us to use a time step of 2.5 fs. This mass scaling does not affect equilibrium properties of the trajectory and has been used to improve conformational sampling.³⁷ The systems were heated to 300 K and equilibrated for 10 ps. After the equilibration, we collected trajectory snapshots over a period of 100 ps, at times 10, 20, ..., 100 ps. These snapshots were used for calculating normal modes.

For comparison, we also calculated normal modes for several minimized crystal structures of BPTI and HEWL. For BPTI, two structures from the PDB were used: 4pti and 1bpi (Parkin, S., to be published). For HEWL, four structures were used: 1lyz, 1hel,³⁸ 1lza³⁹ (all are tetragonal crystal forms), and 2lzt⁴⁰ (triclinic crystal form).

In addition, we obtained three HEWL conformations from a trajectory in which the protein gradually unfolded (Buck, M.;

Karplus, M., to be published). The unfolding trajectory was calculated using the all-hydrogen CHARMM parameter set 22 with the HEWL structure solvated in a 34 Å sphere of explicit water molecules. Unfolding was induced by protonating all acidic side chains (simulating a low pH condition), and raising the temperature. All disulfide bonds were present. Stochastic boundary dynamics were run for 200 ps at 300 K and 250 ps at 600 K, followed by a temperature increase to 900 K. The three structures we obtained from this trajectory were called 0-unf, 100-unf, and 200-unf, for the minimized crystal structure, and snapshots taken after 100 and 200 ps of the 900 K simulation, respectively.

Normal-Mode Calculations. All normal modes discussed in this study were calculated with the program CHARMM,²⁸ using a full Cartesian basis set. In this basis set, the following secular equation is solved:¹¹

$$|\nabla^2 E - \lambda \mathbf{M}| = 0 \quad (1)$$

where $\nabla^2 E$ is the second derivative (Hessian) matrix and \mathbf{M} is the diagonal mass matrix. By root-mass-weighting the Cartesian coordinates we get

$$|\mathbf{M}^{-1/2}(\nabla^2 E)\mathbf{M}^{-1/2} - \lambda \mathbf{I}| = 0 \quad (2)$$

where \mathbf{I} is the identity matrix. In this coordinate system only the first matrix term in eq 2 needs to be diagonalized. The resulting eigenvalues are proportional to the squares of the vibrational frequencies, and the eigenvectors are the normal modes in mass-weighted Cartesian coordinates.

Mean square atomic fluctuations in the classical limit are given by

$$\langle \Delta x_k^2 \rangle = k_B T \sum_{i=1}^n \frac{|u_{ki}|^2}{M_k \omega_i^2} \quad (3)$$

where i is summed over the n modes u of interest, and k corresponds to one of the $3N$ Cartesian degrees of freedom of the system. Since only the low-frequency portions of the normal-mode spectra contribute significantly to the atomic fluctuations, quantum corrections can be neglected at room temperature.³

The quantum mechanical expression for the vibrational entropy at temperature T only requires knowledge of the vibrational frequencies and can thus be easily applied to calculated normal modes:^{41,42}

$$-TS(\text{vib}) = \sum_{i=1}^{3N-6} \left[\frac{-hc\omega_i}{e^{hc\omega_i/k_B T} - 1} + k_B T \ln(1 - e^{-hc\omega_i/k_B T}) \right] \quad (4)$$

where h is Planck's constant, c is the speed of light, and ω_i is the frequency in cm^{-1} . Index i is summed over all (nontranslation/nonrotation) vibrational frequencies.

The character of normal modes varies from local, where the fluctuations involve only a particular part of the protein, to global, e.g., in the case of large-scale concerted motions. Methods for defining the character of modes include the calculation of centers and radii of localization,⁴³ the determination of the number of heavy atoms necessary to account for 95% of the mean square fluctuations,⁴⁴ and the calculation of local and lack-of-global character indicators.¹¹ We use the latter two in this study. The local character indicator is defined as

$$\sum_{j=1}^{3N} u_{ji}^4 \quad (5)$$

where u_{ji} corresponds to the j th entry of normal-mode vector i . Because the normal-mode vectors form a complete orthonormal basis, the local character indicator values range from $(3N)^{-1}$ to 1, and are large for modes i with significant local character, i.e., modes in which the motion is dominated by a small number of atoms. The lack-of-global character indicator is given by

$$\left(\frac{\sum_{j=1}^{3N} |u_{ji}|}{(3N)^{1/2}} \right)^{-4} \quad (6)$$

which ranges from 1 to $9N^2$, and is large for modes without global character, since the sum in the numerator is largest for modes in which all atoms are moving. Both local and lack-of-global indicators are necessary to identify modes with mixing of local and global character.¹¹

For comparison of the normal modes of different structures, we calculate the inner product between the normal-mode vectors, as described by Hayward et al.²³ Because the normal modes are expressed in Cartesian coordinates, we need to do a mass-weighted least-squares superposition of the two structures for which the modes are to be compared. This removes any global translational and rotational differences between the two structures. The same rotation is then applied to the corresponding set of normal modes. We define \mathbf{u}_i to be the i th normal-mode vector of calculation 1 and \mathbf{v}_j to be the j th normal-mode vector of calculation 2. The degree f to which n modes of calculation 2 overlap m modes of calculation 1 is defined as

$$f(m,n) = \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^n [\mathbf{u}_i \cdot \mathbf{v}_j]^2 \quad (7)$$

Since both the \mathbf{u}_i and \mathbf{v}_j vectors form a complete orthonormal set, the maximum value of f (complete overlap) is 1.

The density of states $g(\omega)$ for a particular normal-mode calculation is determined by counting the number of eigenfrequencies in a window of 5 cm^{-1} width, centered on integer frequency values (1, 2, ... cm^{-1}), followed by normalization (division by 5).

For BPTI, calculated densities of states can be compared to experimental densities of states from inelastic neutron scattering data.^{31,27} To compare the results of the normal modes with experiment, we convoluted the discrete normal-mode frequency spectrum with a Gaussian resolution function³¹

$$g_c(\omega) = \sum_i \frac{1}{(2\pi\sigma_{\omega_i}^2)^{1/2}} \exp \left[\frac{-(\omega - \omega_i)^2}{2\sigma_{\omega_i}^2} \right] \quad (8)$$

where $\sigma_{\omega_i}^2$ corresponds to the variance of the resolution function of the instrument with which the experiments were done (see Smith et al.³¹ for a resolution function plot).

For computational efficiency we used the CHARMM polar hydrogen parameter set param19⁴⁵ for the normal-mode calculations. Since the MD simulations were done with an all-hydrogen representation, we first removed all hydrogens from the MD conformations and added polar hydrogens with the HBUILD utility. All explicit water molecules were removed from the system. The protonation states of (partially) unfolded structures

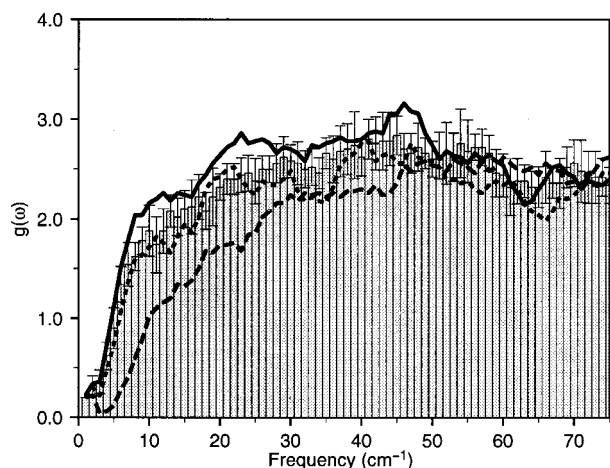


Figure 1. Calculated average densities of states $g(\omega)$ of BPTI, based on normal-mode calculations of 10 different structures; rmsd values are shown as error bars. Histogram: results for calculation with $\epsilon = 2r$ and no cutoff; solid line: $\epsilon = 2r$ and shift cutoff; long-dashed line: $\epsilon = 2r$ and switch cutoff; short-dashed line: $\epsilon = r$ and shift cutoff.

0-unf, 100-unf, and 200-unf were reset to the standard states at pH 7. The systems were minimized by taking 2000 steepest descent steps followed by ABNR steps until the energy gradient fell below 10^{-4} kcal/mol/Å. A distance-dependent dielectric constant was used, equal to 2 times the distance in Å ($\epsilon = 2r$). For BPTI we applied several electrostatic nonbonded cutoff schemes (switch, shift, no cutoff) to check for consistency with earlier studies.^{27,29} For HEWL we only used the shift cutoff scheme, which was found to agree best with the calculations without any cutoff.

Results

BPTI. The minimized snapshot structures had a backbone root mean square deviation (rmsd) from the 4pti crystal structure between 0.75 and 1.36 Å. Backbone rmsd values between minimized snapshots ranged from 0.15 to 1.51 Å. The minimized (without nonbonded cutoff) 4pti and 1bpi crystal structures differed by a backbone rmsd of 1.25 Å. Figure 1 shows the calculated density of states $g(\omega)$ of BPTI, averaged over the 10 different structures. The bars represent average $g(\omega)$ values for the calculation without nonbonded cutoffs and a distance-dependent dielectric constant of 2 times the distance ($\epsilon = 2r$); error bars indicate the rmsd from the average. Results for the shift and switch cutoff methods with $\epsilon = 2r$ (solid and long-dashed lines, respectively), and for the shift cutoff with $\epsilon = r$ (short dashed line) are also shown. The shift method agrees better with the no cutoff method than the switch method, which confirms results reported by others.^{31,29} The differences in densities of states are significantly smaller for frequencies higher than ~ 200 cm^{-1} (Figure 2). The $g(\omega)$ spectrum for the $\epsilon = r$ calculation shows smaller values over the full low-frequency spectrum than for $\epsilon = 2r$, indicating steeper energy wells caused by stronger electrostatic interactions. The smaller number of low-frequency modes for the switch cutoff compared to the shift cutoff is probably caused by the fact that electrostatic forces around the cutoff distance are smaller for the shift cutoff.³¹

All averaged spectra are relatively smooth compared to the spectra of some individual structures. Figure 3 shows the spectra for the MD structures at 30, 50, and 70 ps, calculated with the switch method and $\epsilon = 2r$. On the basis of the 70 ps structure, for instance, the peaks in the spectrum at 14, 23, 35, and 45 cm^{-1} appear significant, while a comparison with the averaged spectrum (Figure 1) shows that the significance of the peaks is

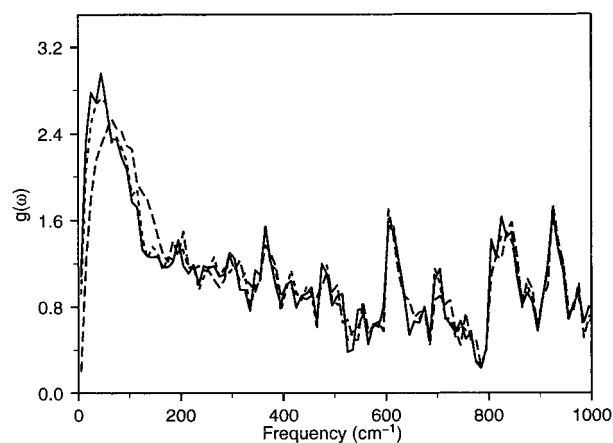


Figure 2. As in Figure 1, for the spectra up to 1100 cm^{-1} . Solid line: $\epsilon = 2r$ and shift cutoff; long-dashed line: $\epsilon = 2r$ and switch cutoff; short-dashed line: $\epsilon = 2r$ and no cutoff.

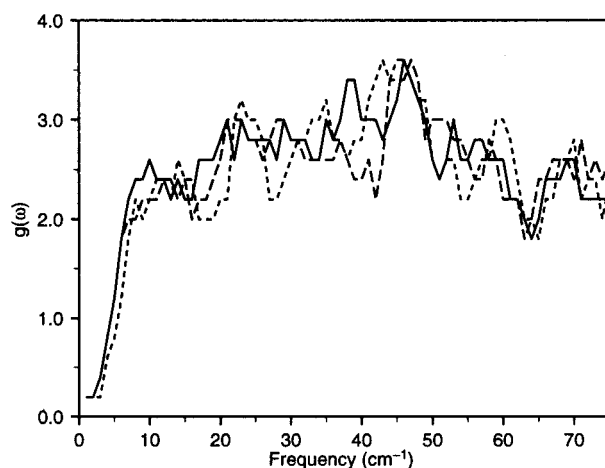


Figure 3. Calculated density of states $g(\omega)$ of BPTI, based on normal-mode calculation of structures at $t = 30$ (solid), 50 (long-dashed), and 70 ps (short-dashed). Minimizations and normal-mode calculations were done with the switch cutoff method and $\epsilon = 2r$.

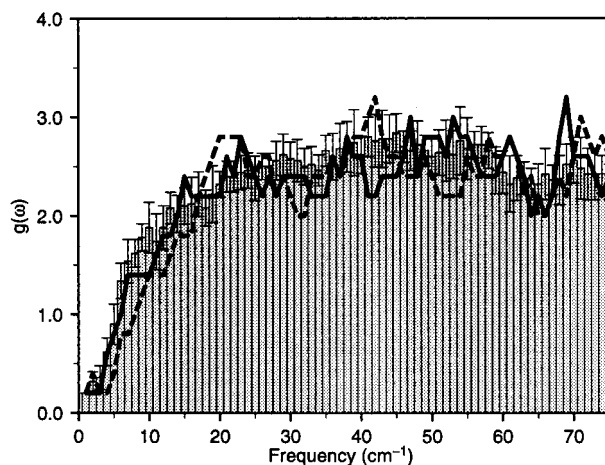


Figure 4. Calculated density of states $g(\omega)$ of BPTI, based on the average of the 10 MD structures (histogram), and the minimized crystal structures of PDB entries 1bpi (solid) and 4pti (dashed). All results in this plot are based on minimizations without electrostatic cutoff and $\epsilon = 2r$.

questionable for an ensemble of conformations. Figure 4 compares the average density of states obtained from the 10 MD structures, with the densities of states calculated for the minimized crystal structures 1bpi and 4pti, respectively. The

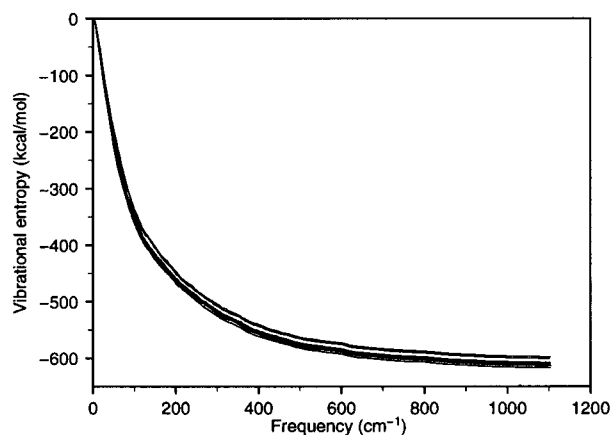


Figure 5. Cumulative free energy ($-TS(\text{vib})$) due to vibrational entropy for the 10 BPTI MD structures at $T = 293$ K.

crystal structure spectra have more pronounced peaks than the averaged spectrum and are comparable in appearance to the spectra of the individual MD structures (Figure 3).

Small differences in density of states at low frequencies affect the entropy of the system, since low frequencies make larger contributions to the total entropy (eq 4). In Figure 5, we show the cumulative free energy ($-TS(\text{vib})$) due to vibrational entropy for the 10 MD structures (shift method). The entropy values are converged at ~ 1000 cm^{-1} , as was previously shown for the crystal structure by Tidor and Karplus.⁴² This was also the case for the other electrostatic cutoff schemes (results not shown). For the first 1194 vibrational modes of BPTI ($\epsilon = 2r$) the average calculated value of $-TS(\text{vib})$ was -609.47 kcal/mol with an rmsd of 6.03 kcal/mol for the 10 MD structures. The largest difference in entropy between two structures was between the 60 and 80 ps snapshot, which had entropy values of -617.16 and -597.42 kcal/mol, respectively (difference of 19.74 kcal/mol). The first 294 modes account for 87% of that entropy difference (-411.92 and -394.71 kcal/mol, respectively). Because the variation in entropy values among the 10 native subspace BPTI structures is relatively large, on the order of 6 kcal/mol, it seems unwise to calculate entropy values based on one conformation. The calculated entropy values for the minimized crystal structures were -599.18 kcal/mol for 1bpi and -593.39 kcal/mol for 4pti. The value for 4pti is ~ 4 kcal/mol less than the range of the 10 MD structures.

Cusack³⁰ suggested that the presence of more structure in the calculated $g(\omega)$ spectrum, as compared with experiment, was likely to be due in part to the single conformation used for the calculation, as initially done by Smith et al.³¹ By convoluting our spectra with a spectrometer-dependent resolution function (see eq 8) and averaging the resulting 10 spectra, we can see if including multiple conformations reduces the presence of peaks in the spectrum. Figure 6 shows the calculated densities of states for all 10 MD conformations, using the shift method and $\epsilon = 2r$. Although there is a high variability in the magnitudes of the peaks at 10 and 45 cm^{-1} , three peaks are still observed in the averaged spectrum at 8.9, 26.5, and 43.5 cm^{-1} (Figure 7). The peak at 8.9 cm^{-1} has a large rmsd and is therefore less certain. The convoluted spectra show more structure at low frequencies than the corresponding densities of states (Figures 1 and 7). This is caused by the narrow shape of the resolution function at low frequencies (Gaussian σ of 1.0 cm^{-1} at frequency of 10 cm^{-1}). For frequencies higher than 50 cm^{-1} , the σ of the resolution function exceeds 3.7 cm^{-1} , resulting in smooth spectra for all conformations. Similar behavior was observed for the averaged calculated spectra obtained with the switch cutoff and

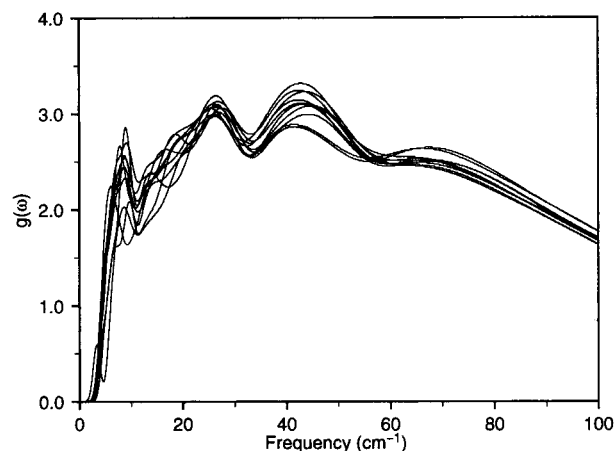


Figure 6. Calculated densities of states $g(\omega)$ of BPTI, based on normal-mode calculations of 10 different structures, using the shift method and $\epsilon = 2r$. The curves are convoluted with a machine-dependent resolution function.

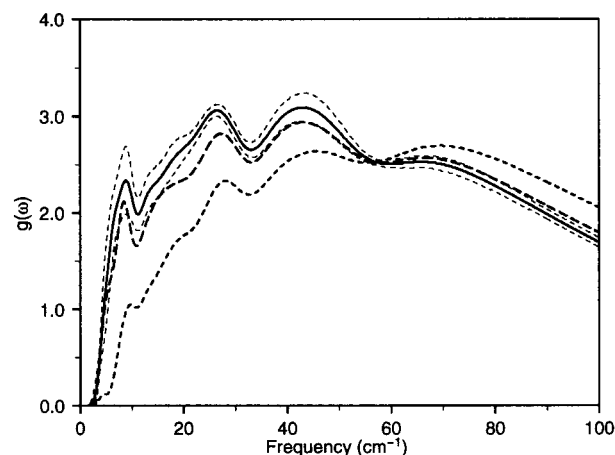


Figure 7. Calculated average density of states $g(\omega)$ of BPTI, based on the 10 curves in Figure 6. The average plus and minus the rmsd are drawn with thin dashed lines. Average results for the calculation without cutoffs are represented by the long-dashed line. The short-dashed line shows the results with the switch cutoff method.

the no cutoff methods. The peaks at low frequencies of the individual normal-mode calculations were smoothed out by the averaging procedure and the spectra show small, relatively broad peaks around 27 and 44 cm^{-1} (Figure 7). Compared to the results of Smith et al.,²⁷ all resulting average curves are smoother and are therefore in better agreement with experiment. The switch method results in lower values of $g(\omega)$ at low frequencies than the shift or no cutoff methods and is closer to the experimental curve, which confirms previous results.^{30,27} Figure 8 shows the local character indicators for normal modes up to 75 cm^{-1} of 5 of the 10 BPTI structures. On the basis of the local and global (not shown) indicators, and confirmed by visual inspection, the modes that have frequencies around 27 and 44 cm^{-1} , which cause the peaks in the spectra, have no particularly local or global character.

The calculated rms fluctuations of the C^α atoms of BPTI are similar to previous calculations.^{9,23,29} A comparison of the calculated fluctuations with those derived from experimental B factors of two crystal structures (PDB entries 4pti and 1bpi), showed that the best agreement was found with fluctuations based on averaged B factors (Figures 9 and 10). In the 1bpi structure, B factors are remarkably low for the loop formed by residues 25–28. For these residues, both the calculated fluctuations and the 4pti B factors are significantly higher.

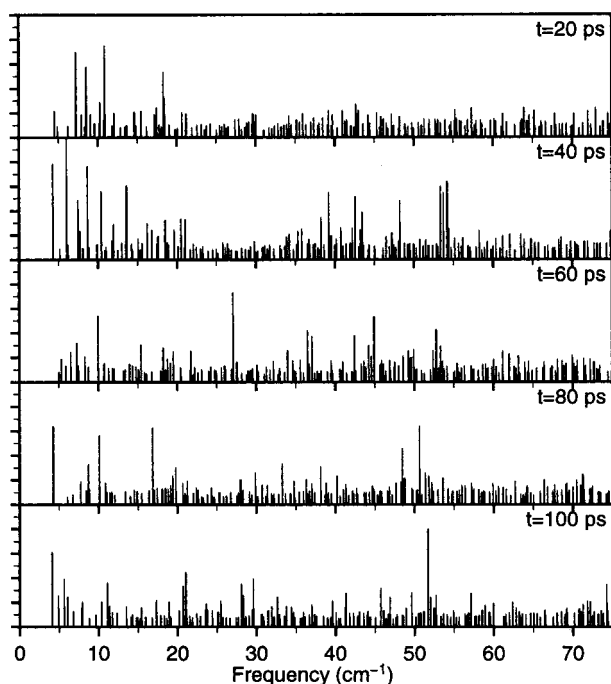


Figure 8. Local character indicators (eq 5) for normal modes calculated with the shift cutoff method and $\epsilon = 2r$. The MD snapshots used are indicated in the upper right-hand corner. The scale of the y-ordinate is from 0 to 0.05.

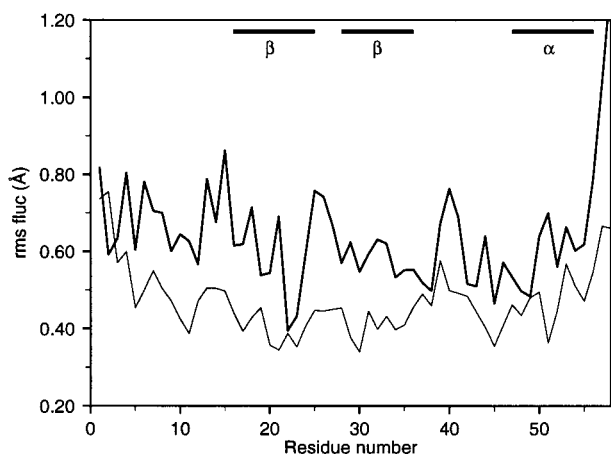


Figure 9. A comparison of rms fluctuations of C α atoms of BPTI, based on the B factors of the 1bpi (thick) and 4pti (thin) crystal structures.

It is already evident from Figure 8 that there is a substantial variation in the frequencies for normal modes calculated for different structures. This is in accord with results reported previously²² for structures that were taken at 10 ps intervals from an MD simulation. Because such a variation in frequencies does not necessarily imply a variation in the space spanned by the low-frequency normal modes, we projected every set of normal modes onto each other, as described in the Methods section. Jačević et al.²² calculated the projection of the three lowest modes of each structure onto the others. To obtain a more meaningful comparison, we consider the normal modes required to account for $\sim 90\%$ of the total mean square of atomic fluctuations. Table 1 lists that number of modes for the 10 BPTI structures. It shows that the mode number at which 90% of the fluctuations are covered is higher for the switch cutoff calculations than for the shift cutoff. This is because the switch cutoff resulted in fewer low-frequency modes and thus lower total atomic fluctuations (see eq 3). Somewhere between 75 and 92

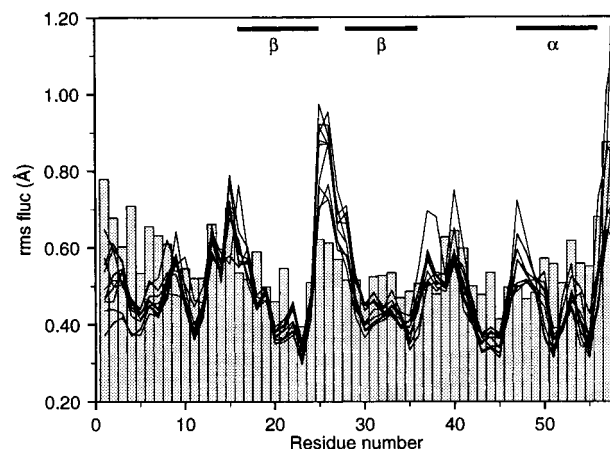


Figure 10. Histogram: rms fluctuations in Å of the C α atoms of BPTI, based on the averaged B factors of the 1bpi and 4pti X-ray structures (conversion: $\text{rms fluc} = (3B/8\pi^2)^{1/2}$, where B is the B factor of the atom). The thin lines show the rms fluctuations based on the first 294 vibrational modes of BPTI, calculated for the 10 different structures. Secondary structure assignments are made according to information in the PDB file.

TABLE 1: Mode Numbers and Frequencies at Which 90% of the Total Mean Square Atomic Fluctuations Are Accounted for^a

t (ps)	mode no. shift	freq (cm ⁻¹)	mode no. switch	freq (cm ⁻¹)
10	78	35.44	123	64.23
20	75	34.29	116	62.05
30	80	35.61	110	57.58
40	78	35.33	123	66.55
50	76	34.70	111	63.39
60	77	34.32	118	65.35
70	92	41.41	125	66.65
80	92	43.51	123	67.25
90	86	40.37	123	66.32
100	78	36.02	100	56.10
1bpi	98	44.96	117	65.56
4pti	97	45.90	138	75.61

^a Columns 2 and 3 list values for the calculations with the shift cutoff method; columns 4 and 5 contain the results of the switch cutoff method. Rows labeled 1bpi and 4pti refer to the minimized crystal structures.

TABLE 2: Overlap Values (see Eq 7) $f(75,75)$ (upper triangle) and $1/2(f(3,5) + f(5,3))$ (lower triangle) for the 10 BPTI Structures, and the Two Minimized Crystal Structures

t (ps)	10	20	30	40	50	60	70	80	90	100	4pti	1bpi
10	1.00	0.77	0.72	0.71	0.71	0.72	0.78	0.72	0.74	0.77	0.63	0.59
20	0.63	1.00	0.71	0.70	0.68	0.69	0.71	0.66	0.71	0.75	0.61	0.53
30	0.41	0.49	1.00	0.81	0.93	0.90	0.75	0.70	0.75	0.75	0.61	0.53
40	0.38	0.40	0.48	1.00	0.78	0.76	0.73	0.68	0.71	0.70	0.61	0.53
50	0.40	0.48	0.97	0.46	1.00	0.95	0.74	0.71	0.77	0.76	0.60	0.54
60	0.38	0.47	0.95	0.46	0.98	1.00	0.76	0.71	0.79	0.76	0.60	0.54
70	0.45	0.50	0.65	0.39	0.64	0.65	1.00	0.83	0.89	0.80	0.61	0.57
80	0.51	0.42	0.44	0.32	0.46	0.43	0.55	1.00	0.82	0.76	0.58	0.55
90	0.54	0.61	0.70	0.39	0.74	0.75	0.74	0.54	1.00	0.83	0.59	0.54
100	0.46	0.62	0.65	0.40	0.67	0.67	0.68	0.49	0.88	1.00	0.63	0.57
4pti	0.69	0.67	0.51	0.43	0.50	0.46	0.49	0.53	0.50	0.45	1.00	0.55
1bpi	0.41	0.40	0.36	0.29	0.33	0.31	0.40	0.40	0.36	0.33	0.48	1.00

modes are needed to reach 90% coverage. The number of normal modes we used for both sets (m and n in eq 7) was 75. This is 4% of the total number of vibrational modes (which is 1698). We present the projection results for the calculations with the shift cutoff method ($\epsilon = 2r$), since they are most consistent with the no cutoff results.

Table 2 shows the overlap values $f(75,75)$ (eq 7) for the 10 BPTI structures in the upper triangle. The lower triangle shows $1/2(f(3,5) + f(5,3))$ values, which were reported by Jačević et

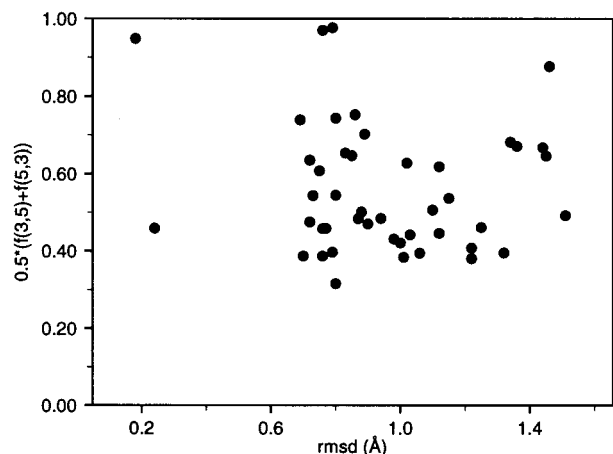


Figure 11. $\frac{1}{2}(f(3,5) + f(5,3))$ versus backbone rmsd for all pairwise comparisons between 10 minimized MD structures of BPTI (shift cutoff).

al.²² They used the symmetrized value to be less subject to changes in the ordering of the three lowest-frequency modes. The values for $f(75,75)$ are relatively high and constant (average 0.758, rmsd 0.064), indicating a relative constancy of the space spanned by the important low-frequency modes. On the other hand, the lower triangle of the table shows that $\frac{1}{2}(f(3,5) + f(5,3))$ is generally much lower and has a higher degree of variability (average 0.562, rmsd 0.165). This indicates that the space spanned by the lowest three normal modes can differ substantially between different BPTI native subspace structures. There is no correlation between the structural similarity (backbone rmsd) of any two MD structures and the normal-mode overlap values $\frac{1}{2}(f(3,5) + f(5,3))$ or $f(75,75)$. Figure 11 shows the rmsd and the $\frac{1}{2}(f(3,5) + f(5,3))$ value for all pairwise comparisons of the MD structures. This suggests that small structural differences, such as those between similar crystal structures, can result in significantly different low-frequency normal modes and that only quantities that depend on sums over a sizable number of modes are meaningful.

To further examine the sensitivity of the results to small structural differences. We used dihedral angle rather than Cartesian coordinate fluctuations. We calculated the fluctuations of all ϕ , ψ , and χ^1 angles due to the first three modes. It was found that in dihedral space there are significant differences between the different snapshots.

The above results indicate that some care is required in applications where a small number of low-frequency normal modes were of interest. Ideally, sets of normal modes from different structures should be compared and checked for consistency. Such applications include normal-mode refinement²⁵ (~40–100 normal modes used), calculation of normal modes of symmetric multimers⁴⁶ (frequencies $<100\text{ cm}^{-1}$), or estimation of reaction coordinates⁴⁷ (frequencies $<5\text{ cm}^{-1}$). The results also show that the degree of overlap between the two crystal structures is comparable to the overlaps between any two MD structures separated by less than 100 ps.

HEWL. The 10 HEWL structures from the regular MD simulation were conformationally similar, with an average backbone rmsd between different snapshots of 0.86 Å (1.01 Å after minimization). The backbone rmsd with respect to the 1lyz crystal structure ranged from 0.74 to 1.08 Å. The 0-unf structure still belongs to the native subspace, with an average backbone rmsd from the 10 MD structures of 0.90 Å (1.21 Å after minimization). Structures 100-unf and 200-unf are in the process of unfolding and have average backbone rmsd's from the 10

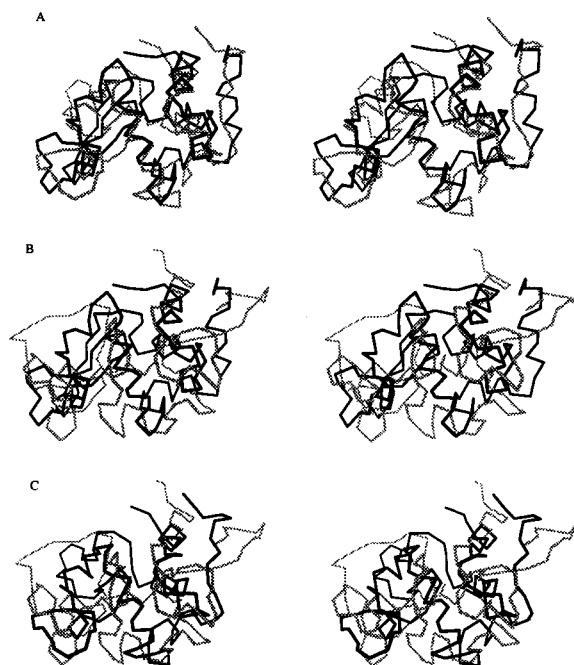


Figure 12. Stereo comparisons between C^α traces of (a) 0-unf (black) and 100-unf (gray), (b) 0-unf (black) and 200-unf (gray), and (c) 100-unf (black) and 200-unf (gray). Structures are optimally superimposed with their C^α atoms. The C-terminus is located in the upper right-hand corner, the active site cleft is at the bottom, left of center.

MD structures of 5.17 Å (4.97 Å) and 8.70 Å (8.52 Å), respectively. Another indicator of (partially) unfolded structures is the radius of gyration. The value of the crystal structure is 13.98 Å, the average value for the 10 MD snapshots is 14.49 Å (\pm rmsd of 0.09 Å), and the values for 0-unf, 100-unf, and 200-unf, are 14.50, 15.42, and 17.71 Å, respectively. The difference is somewhat smaller after minimization: 13.79 Å for the crystal structure, 13.82 ± 0.07 Å for the 10 MD snapshots, and 14.03, 14.30, and 15.21 Å for 0-unf, 100-unf, and 200-unf, respectively. Figure 12 shows the C^α traces of the unminimized 100-unf and 200-unf structures superimposed on the 0-unf structure. In the 100-unf structure, most of the helices are partially unfolded, and the β -strands are fully unfolded. In the 200-unf structure, only a few remnants of helical structure are present. A complete analysis of the properties of the unfolded structures will be presented separately (Buck, M.; Karplus, M., to be published).

The average density of states of the first 294 vibrational modes of HEWL is shown in Figure 13. Averaging was done over normal-mode spectra of the 10 different HEWL structures. For HEWL we used only the shift cutoff method and $\epsilon = 2r$, because this method agreed best with the no cutoff method for BPTI. In comparison with BPTI, the $g(\omega)$ of HEWL has less structure. This is only to a small extent due to averaging, since the histograms for the individual structures also contain less structure than those for BPTI (Figure 14). Figure 13 also shows the $g(\omega)$ curves for the three HEWL structures from the unfolding trajectory. The solid line represents the native conformation 0-unf. Its density of states falls within the limits of the average curve and error bars at virtually every point. The long-dashed curve is of the partially unfolded structure 100-unf. It has a higher density of states for frequencies below 15 cm^{-1} , indicative of a larger number of low-frequency modes. For the further unfolded structure 200-unf $g(\omega)$ is higher still than for 100-unf. The higher density of states for frequencies $<10\text{ cm}^{-1}$ is expected to increase the calculated atomic fluctuations of the most unfolded form. Figure 15 compares the

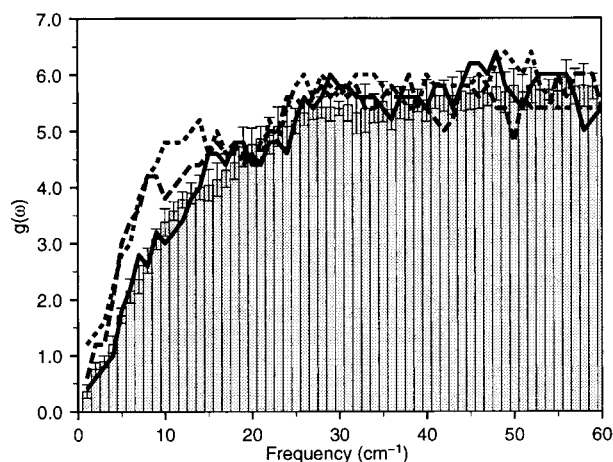


Figure 13. Histogram: calculated averaged density of states $g(\omega)$ of HEWL, based on normal-mode calculations of 10 different structures; rmsd values are shown as error bars; solid line: $g(\omega)$ of structure 0-unf; long-dashed line: $g(\omega)$ of structure 100-unf; short-dashed line: $g(\omega)$ of structure 200-unf.

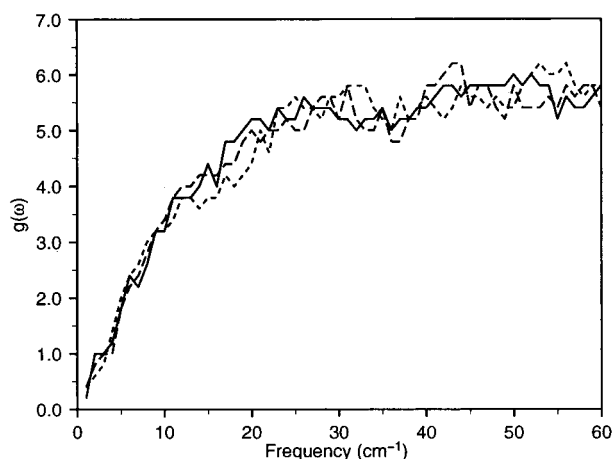


Figure 14. Calculated density of states $g(\omega)$ of HEWL, based on normal-mode calculation of structures at $t = 30$ (solid), 50 (long-dashed), and 70 ps (short-dashed).

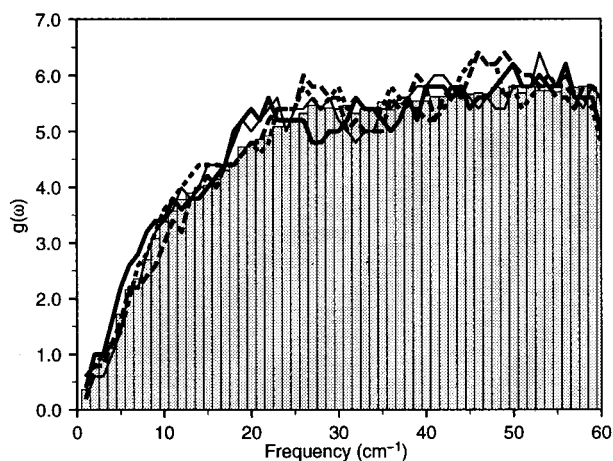


Figure 15. Histogram: as in Figure 13. Densities of states based on normal-mode calculations of minimized crystal structures 1hel, 1lza, 2lzt, and 1lyz are shown as thick solid, thin solid, long-dashed, and short-dashed lines, respectively.

calculated $g(\omega)$ curves for the four minimized HEWL crystal structures with the average curve of the 10 MD structures. There is significant variation between the crystal structure curves, but in the low-frequency domain they are all relatively similar to

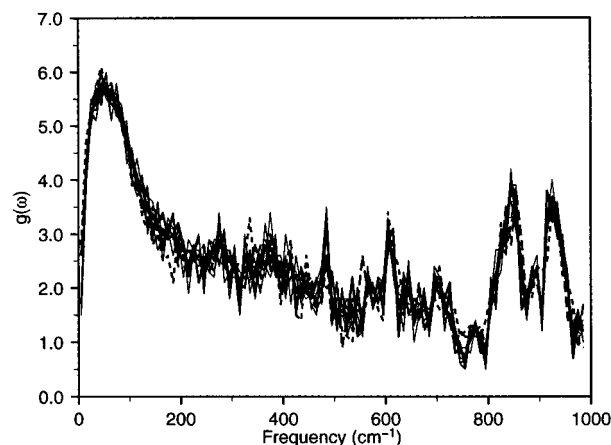


Figure 16. As Figure 13, for the spectra up to 1000 cm^{-1} . The spectra of the 10 MD structures are shown as thin solid lines. The spectra of the 0-unf, 100-unf, and 200-unf structures are shown as thick solid, long-dashed, and short-dashed lines, respectively. The spectra of the four minimized crystal structures are shown as thin dashed lines.

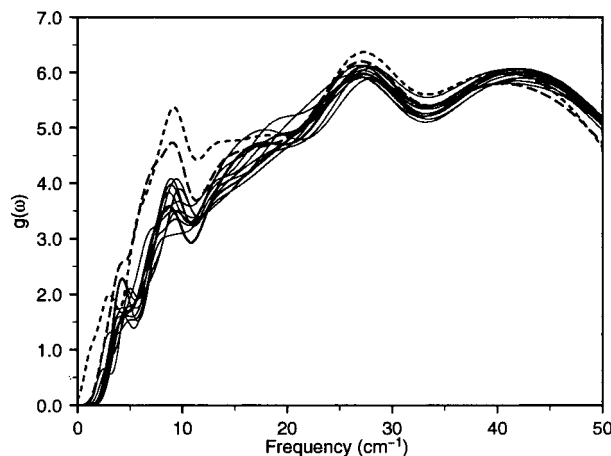


Figure 17. Calculated densities of states of 10 different HEWL structures, convoluted with a machine resolution function. Structures from the unfolding trajectory 0-unf, 100-unf, and 200-unf, are represented by the thick solid line, the long-dashed line, and the short-dashed line, respectively.

the average curve. The frequency spectra up to 1000 cm^{-1} are shown in Figure 16. The general profile of the spectra of the 10 MD structures and the 0-unf, 100-unf, and 200-unf structures is relatively similar, although there is significant variation between the individual curves. On this frequency scale, the spectra of the partially unfolded structures fall within the variations of the 10 native subspace structures. However, as we have already seen (Figure 13), at low frequencies the density of states is significantly higher in 100-unf and 200-unf as compared with the native state structures.

Figure 17 shows the calculated density of states of the low-frequency spectrum, convoluted with the same instrumental resolution function used for BPTI (see Figure 7). We see the same peaks in this curve as in the BPTI curves (at ~ 9 and $\sim 28 \text{ cm}^{-1}$), even though the density of states (Figure 13) shows no significant peaks. The peaks are not present in the density of states curves because we average over a window of 5 cm^{-1} (see Methods), whereas the convoluted spectrum represents every individual normal-mode frequency by a Gaussian of variable width (eq 8).

To compare the rms fluctuations of the C^α atoms with experimental B factors we used the average values of PDB entries 1hel and 1lza (both are tetragonal crystal forms), and

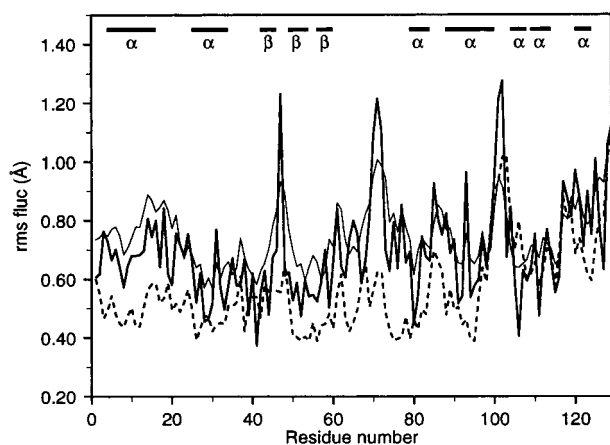


Figure 18. Rms fluctuations in Å of the C α atoms of HEWL, based on the *B* factors of PDB entries 1hel (thick), 1lza (thin), and 2lzt (dashed). Secondary structure assignments are made according to information in the 1hel PDB file.

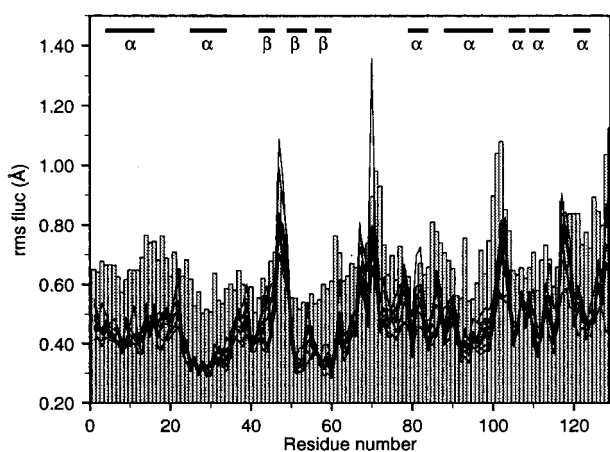


Figure 19. Histogram: rms fluctuations in Å of the C α atoms of HEWL, based on the averaged *B* factors of PDB entries 1hel, 1lza, and 2lzt. The thin lines show the rms fluctuations values at 293 K, based on the first 294 vibrational modes of HEWL, calculated for the 10 different structures.

2lzt (triclinic). There is significant variation in *B* factor values between the three crystal structures (Figure 18). No *B* factors are reported for PDB entry 1lyz. The comparison with the calculated *B* factors (Figure 19) shows a picture similar to that for BPTI. The fluctuations are underestimated, in general, except in certain loop regions and at the C terminus (Figure 19). Phillips⁴⁸ found the best agreement between atomic fluctuations from MD simulations and *B* factor values for myoglobin when he used the maximum *B* factor values of the two myoglobin crystal structures. Here, we find better agreement with the averaged *B*-factor values.

There is a substantial variation between the curves of the 10 normal-mode calculations in the peaks at residues 47, 70, 102, 117, and the C terminus. Figure 20 shows the average of the 10 curves of Figure 19 as histogram and the rms fluctuations of the three HEWL structures from the unfolding simulation. The thick solid line corresponds to the 0-unf structure and agrees well with the average curve. The (partially) unfolded structures 100-unf and 200-unf (medium and thin solid lines, respectively) have significantly larger fluctuations for residues 1–20, a segment that contains an α -helix (res. 4–16) in the native conformation, but not in the unfolded structures.

Because the partially unfolded structures have more low-frequency modes (Figure 13), their total atomic fluctuations are

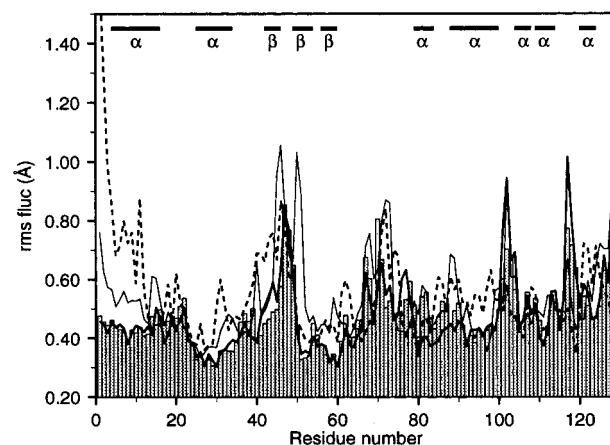


Figure 20. Histogram: Average rms fluctuations in Å at 293 K of the C α atoms of HEWL, based on the 10 curves shown in Figure 19. Rms fluctuations values for structures 0-unf, 100-unf, and 200-unf, are shown as thick, thin, and dashed lines, respectively.

TABLE 3: Normal-Mode Frequencies (<10 cm⁻¹) of the 50 ps MD Structure, and the 0-unf, 100-unf, and 200-unf Structures of HEWL

<i>t</i> = 50 ps	3.22	3.59	4.41	4.92	5.45	6.50	6.94	7.25	7.33	7.72
	8.02	8.38	8.88	9.15	9.46	9.85				
0-unf	3.37	4.02	4.70	5.25	5.89	6.56	6.63	7.12	7.42	7.68
	7.96	8.44	8.70	8.87	9.24	9.42	9.91			
100-unf	2.38	3.44	3.66	4.01	4.27	4.80	5.56	5.60	5.79	5.93
	6.55	6.84	6.89	7.25	7.34	7.35	7.65	8.35	8.50	8.62
	8.94	9.12	9.35	9.51	9.72	9.98				
200-unf	1.09	2.04	2.66	3.07	3.30	4.31	4.85	5.34	5.54	5.79
	5.85	6.35	6.72	6.97	7.06	7.16	7.73	7.95	8.13	8.40
	8.66	8.83	8.94	9.16	9.45	9.63	9.73	9.98		

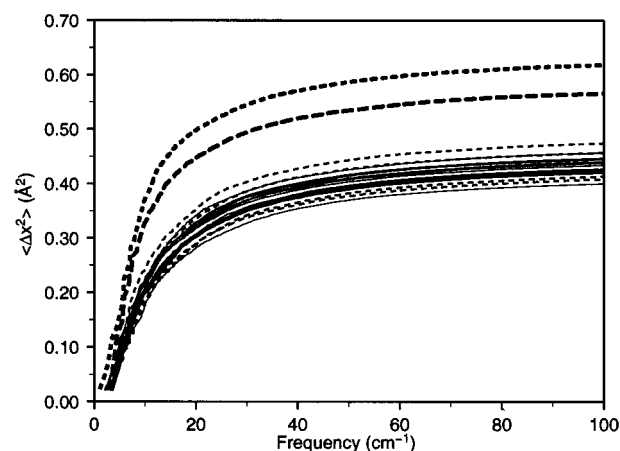


Figure 21. Cumulative mean square total atomic fluctuations of HEWL at 293 K, as a function of normal-mode frequency. The curves for the 10 MD structures are shown as thin solid lines. The structures 0-unf, 100-unf, and 200-unf from the unfolding trajectory are represented by the thick solid line, the long-dashed line, and the short-dashed line, respectively. The results of the four crystal structures are shown as thin dashed lines.

larger. Table 3 lists the normal-mode frequencies of less than 10 cm⁻¹ of the 50 ps MD structure, and the 0-unf, 100-unf, and 200-unf structures. Figure 21 shows the cumulative total mean square atomic fluctuations as a function of normal-mode frequency. At a frequency of ~100 cm⁻¹ the fluctuations have converged. The curves from the 10 structures of the standard MD simulation, the four crystal structures, and the curve of the 0-unf structure cluster around approximately the same value, whereas the 100-unf and 200-unf curves are significantly higher. The higher number of low-frequency modes for the (partially)

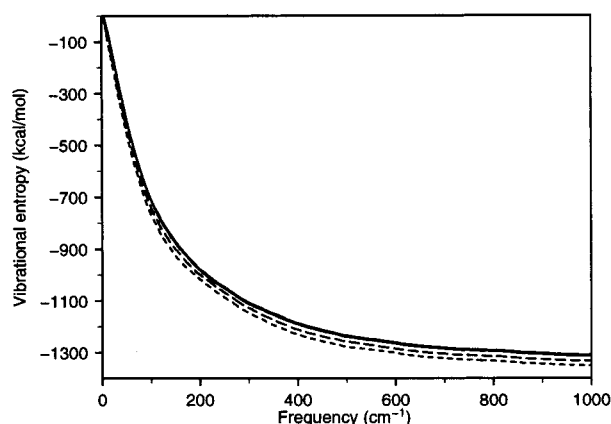


Figure 22. Cumulative free energy ($-TS(\text{vib})$) due to vibrational entropy for the 10 HEWL MD structures at $T = 293$ K (solid lines). The structures 0-unf, 100-unf, and 200-unf from the unfolding trajectory are represented by the thick solid line, the long-dashed line, and the short-dashed line, respectively.

TABLE 4: Overlap values (see Eq 7) $f(130,130)$ (upper triangle) and $\frac{1}{2}(f(3,5) + f(5,3))$ (lower triangle) for the 10 HEWL Structures

t (ps)	10	20	30	40	50	60	70	80	90	100
10	1.00	0.77	0.72	0.69	0.77	0.76	0.70	0.77	0.69	0.68
20	0.78	1.00	0.79	0.74	0.79	0.79	0.68	0.76	0.71	0.65
30	0.74	0.73	1.00	0.70	0.83	0.77	0.69	0.74	0.70	0.63
40	0.66	0.66	0.67	1.00	0.74	0.80	0.72	0.71	0.69	0.70
50	0.82	0.80	0.84	0.64	1.00	0.81	0.74	0.81	0.72	0.68
60	0.85	0.84	0.81	0.64	0.87	1.00	0.78	0.75	0.72	0.69
70	0.77	0.61	0.72	0.58	0.76	0.78	1.00	0.73	0.71	0.71
80	0.78	0.70	0.78	0.59	0.89	0.79	0.77	1.00	0.75	0.67
90	0.64	0.49	0.68	0.53	0.64	0.67	0.71	0.68	1.00	0.75
100	0.51	0.39	0.52	0.44	0.52	0.53	0.62	0.53	0.56	1.00

unfolded structures also increases their entropies. The average value of the vibrational entropy $-TS(\text{vib})$ of the 10 MD structures was calculated according to the quantum mechanical formalism (eq 4) and was equal to -1312.34 kcal/mol (rmsd 2.25 kcal/mol) at 293 K (first 2494 modes). For the four minimized crystal structures, we calculated $-TS(\text{vib})$ values of -1319.25 (1hel), -1313.28 (1lza), -1312.15 (2lzt), and -1319.74 (1lyz) kcal/mol. The results for the 0-unf, 100-unf, and 200-unf structures were -1313.78 , -1333.23 , and -1351.10 kcal/mol, respectively. The convergence of the total vibrational entropy as a function of normal-mode frequency is shown in Figure 22. The entropy converges at ~ 1000 cm^{-1} , which is similar to the result for BPTI (Figure 5). For HEWL, this means that ~ 2500 modes have to be included in the entropy calculation, compared to ~ 1000 modes for BPTI.

The average mode number at which 90% of the mean square atomic fluctuations are accounted for is equal to 130.5 (at an average frequency of 33.8 cm^{-1}) for the 10 MD HEWL structures. To compare the spaces spanned by the normal modes from the different calculations, we projected the 130 first vibrational modes of each structure onto those of all other structures. The 130 modes amount to 3% of the total number (3786) of vibrational modes of the system. Results of the projection calculations are summarized in Table 4. This table also contains the projections of the first three modes, as used by Jařezic et al.²² in the lower triangle. The values in this table are comparable to the corresponding values for BPTI (Table 2). Overlap values $f(130,130)$ are relatively high and have a small standard deviation (average: 0.731, rmsd: 0.045). The $\frac{1}{2}(f(3,5) + f(5,3))$ values have a similar average value (0.678), but have a much larger variability (rmsd 0.122). This was also

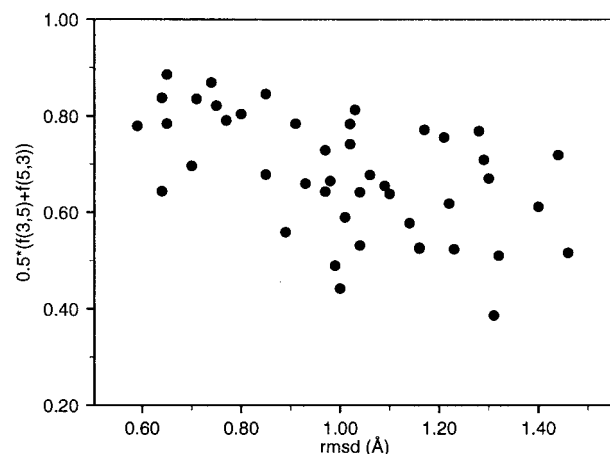


Figure 23. $\frac{1}{2}(f(3,5) + f(5,3))$ versus backbone rmsd for all pairwise comparisons between 10 minimized MD structures of HEWL.

the case for BPTI, and it indicates that the space spanned by the lowest three modes of a particular HEWL structure is not necessarily spanned by the lowest five modes of a different (but also in the native subspace) HEWL structure. By analyzing the overlaps between the first 10 modes of all MD structures, we found that the corresponding normal-mode coordinates interchange between different structures. The average separation (in mode number) between the best overlap of any of the first 10 modes with any of the 10 modes of a different structure was 2.4. Such behavior is not surprising because of the quasidegeneracy of the normal modes in such large systems. Figure 23 shows the backbone rmsd and the $\frac{1}{2}(f(3,5) + f(5,3))$ value for all pairwise comparisons of the folded MD structures, analogous to Figure 11 for BPTI. The figure suggests a weak correlation, but confirms what was found for BPTI, namely that structural similarity does not necessarily imply similarity in normal-mode coordinates.

We calculated the quasiharmonic modes of HEWL using CHARMM, including all frames from the 100 ps MD trajectory. Our HEWL results agreed with the conclusions from the BPTI results of Hayward et al.;²³ i.e., the space spanned by any of the 10 lowest frequency quasiharmonic modes can be described to a significant degree (up to $\sim 50\%$) by the 10 lowest frequency normal modes of any of the 10 MD snapshots. The rms fluctuations due to the first 10 normal modes of the MD snapshots are qualitatively and quantitatively similar in terms of their magnitudes and relative values to the fluctuations due to the first 10 quasiharmonic modes (Figure 24). In the region involving residues 70–90, there is a significant difference for the normal-mode results of some of the structures.

Table 5 shows $f(130,130)$ overlap values for the normal modes of the 10 MD snapshots with the four crystal structures and the three conformations from the unfolding simulation. The overlap values with the crystal structures and the 0-unf structure fall essentially within the range observed in Table 4 (lowest value in Table 4, 0.631). The overlap values of the 100-unf and 200-unf structures are significantly lower (averages of 0.392 and 0.321, respectively). In all cases, the 200-unf overlap values are lower than the 100-unf values, indicating that during the unfolding trajectory the potential energy surface of the structure is becoming increasingly different from the surface of the native structure. Overlap values between the different crystal structures are similar to those calculated for the 10 MD structures (results not shown).

Along the unfolding trajectory the protein structure undergoes large-scale conformational changes that alter the distances and

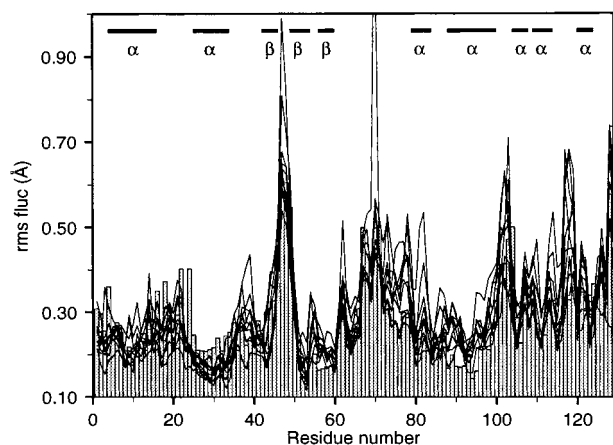


Figure 24. Histogram: rms fluctuations in Å of the C α atoms of HEWL, based on the first 10 quasi-harmonic modes. The thin lines show the rms fluctuations values at 293 K, based on the first 10 vibrational modes of HEWL, calculated for the 10 different structures.

orientations of α -helices and β -strands with respect to each other (Figure 12). It is conceivable that individual secondary structure units and/or protein domains in (partially) unfolded conformations have internal motions that are more similar to those in the fully folded conformations than does the entire structure. The calculated overlap values of the full protein could be low, because the normal modes are "misaligned" by superimposing the unfolded protein onto a folded protein structure. For example, an α -helix that is rotated by 90° in an unfolded structure but still has the same "internal" normal-mode space would add virtually nothing to the total overlap value because its motions (in Cartesian space) would be perpendicular to those in the folded structure. To investigate this possibility, we recalculated the overlap values based on different ways of orienting the structures. Instead of rotating the normal modes based on the all-atom structural superposition between the folded and unfolded structure, we superimposed individual parts of the HEWL structure and rotated the corresponding parts of the normal modes accordingly. Three different structure partitions were used: each residue individually (129 parts), division into segments of secondary structure and coils (PDB definition, 18 parts), and according to the two domains that characterize the hinge-bending motion (domain 1: residues 1–40 and 88–129; domain 2: residues 41–87⁴⁴). None of the alternative mode reorientation methods resulted in better overlaps between the modes of the unfolded and folded structures: average $f(130,130)$ overlap values for 100-unf (200-unf) with the 10 MD structures were 0.392 (0.321) for the all-atom superposition, 0.286 (0.228) for the individual residue superposition, 0.351 (0.287) for the secondary structure superposition, and 0.388 (0.314) for the hinge domains superposition, respectively. These results indicate that the low-frequency normal modes of the (partially) unfolded structures and the folded structures share more similarity on a global scale than on a local scale.

Three previous studies have been concerned with describing the hinge bending motion in HEWL on the basis of normal modes.^{9,44,49} In the studies, the hinge bending motion was described by the first normal mode of the system, which had a frequency of ~ 3 cm⁻¹. Brooks and Karplus⁴⁹ defined the hinge axis passing through Thr 40 and the disulfide bond between Cys 76 and Cys 94, with domain 1 constituted by residues 1–39 and 89–129, and domain 2 by residues 40–88. Levitt et al.⁹ defined domain 1 by residues 1–39 and 87–129, and domain 2 by residues 40–86. Gibrat and Gö⁴⁴ used a hinge axis that passed through C α 55 and C α 76, and defined domain 1 by

residues 1–40 and 88–129, and domain 2 by residues 41–87. Their definition of the hinge axis was based on a search for the axis and angle that gave the best agreement with conformational changes resulting from the first 10 normal modes. They found that the axis and angle based on the first mode only was very similar to the one based on the first 10 modes. We will call the hinge axis and domain definitions of Brooks and Karplus and Gibrat and Gö hinge1 and hinge2, respectively. Because we found that the space spanned by the first three normal modes of a particular conformation is not necessarily covered to a large extent by the modes of a different conformation, we investigated the projections of the low-frequency modes on the hinge bending motion for all MD snapshots. Projections of the normal modes were obtained by calculating the inner product between each individual mode and the normalized difference vector between the minimized conformation and the same conformation with the second domain rotated as a rigid body around the hinge axis by 1.6° (followed by a least-squares superposition on the minimized conformation to remove any translational and rotational differences). Table 6 lists the three normal modes of each conformation with the largest projections on the hinge bending motions defined by hinge1 and hinge2. The first mode has a substantial projection on the hinge motion for most conformations, but does not have the largest projection in many cases, particularly for the hinge1. The total squared projections of the first 294 vibrational modes on the hinge bending motion are relatively constant (average total squared projections of 93.0% and 96.3% for hinge1 and hinge2, respectively), and there are only minor differences between the results for the two alternative hinge axes. There are large differences between the different MD snapshots for the modes with the largest contributions to hinge bending. In addition, the mode numbers vary substantially between the different hinge axis definitions. For certain conformations one mode has the major contribution to the total projection, e.g., at $t = 90$ for hinge1 and $t = 50$ and $t = 80$ for hinge2. In other cases, the three modes with the largest projections contribute a comparable percentage, e.g., $t = 70$ for hinge1 and $t = 40$ for hinge2. The three modes with the largest projections are usually not exclusively the three lowest frequency modes (i.e., other modes including modes 4–10 are involved in most cases), unlike the results of the normal mode studies of the HEWL crystal structure by Levitt et al.⁹ and Gibrat and Gö,⁴⁴ who found that the lowest frequency mode contained the dominant hinge bending motion. Weighting the normal modes according to their frequencies before projection affected only slightly the ordering of the modes. As in the calculations without frequency weighting, the lowest frequency mode was not the dominant hinge bending mode for most structures (results not shown).

The results for the 0-unf, 100-unf, and 200-unf structures agree well with the other 10 native subspace conformations. This would indicate that the hinge bending motion of partially unfolded HEWL is similar to the motions of the native conformation, even though the backbone rmsd is more than 5 Å.

Discussion

The calculation of normal modes for an ensemble of native conformations obtained from molecular dynamics simulations and of several minimized crystal structures provides useful information on the significance of calculations based on a single conformation. In most normal-mode calculations, the minimized crystal structure is used. It was shown for BPTI and HEWL that averaging of the density of states ($g(\omega)$) over different

TABLE 5: Overlap Values (see Eq 7) $f(130,130)$ of the Three Snapshot Structures from the HEWL Unfolding Trajectory and the Four Minimized Crystal Structures, with Each of the MD Structures

t (ps)	10	20	30	40	50	60	70	80	90	100	0-unf	100-unf	200-unf
0-unf	0.69	0.68	0.65	0.63	0.67	0.65	0.64	0.70	0.64	0.63	1.00	0.39	0.32
100-unf	0.39	0.40	0.40	0.39	0.40	0.39	0.39	0.39	0.39	0.39		1.00	0.34
200-unf	0.32	0.32	0.32	0.32	0.33	0.33	0.33	0.32	0.31	0.32			1.00
1hel	0.65	0.65	0.65	0.61	0.66	0.63	0.64	0.67	0.62	0.60	0.72	0.41	0.33
1lza	0.68	0.67	0.69	0.64	0.71	0.67	0.68	0.69	0.64	0.63	0.69	0.39	0.33
2lzt	0.66	0.62	0.60	0.60	0.62	0.63	0.63	0.64	0.64	0.62	0.63	0.41	0.32
1lyz	0.75	0.75	0.69	0.67	0.72	0.70	0.67	0.75	0.77	0.71	0.69	0.39	0.31

TABLE 6: Total Squared Projection Percentages of the First 294 Vibrational Modes on the Hinge Bending Motions Defined by Brooks and Karplus⁴⁹ (Hinge1) and by Gibrat and Gō⁴⁴ (Hinge2); Mode Number, Frequency (in cm^{-1}), and the Squared Projection Percentage of the Three Modes with the Largest Projections

t (ps)	total	no.	freq	proj	no.	freq	proj	no.	freq	proj
Hinge1										
10	92.36	5	5.76	20.06	1	3.61	13.29	4	5.43	10.71
20	92.91	4	5.00	29.97	25	12.43	7.58	8	7.10	7.46
30	91.96	5	5.32	22.52	2	4.14	22.08	6	5.89	8.09
40	93.98	4	4.61	32.12	12	8.48	10.78	6	5.75	7.38
50	92.73	4	4.92	37.23	6	6.50	12.34	1	3.22	9.99
60	93.02	4	5.13	34.96	5	5.76	14.85	2	4.10	6.47
70	92.98	4	5.10	22.66	3	4.72	17.03	1	2.81	15.53
80	93.36	3	4.44	19.25	2	4.29	11.99	6	6.37	8.87
90	93.48	4	4.77	40.58	3	4.11	8.05	12	8.63	7.01
100	92.79	2	3.67	29.64	3	4.48	9.82	7	6.92	7.56
0-unf	92.39	3	4.69	36.79	10	7.68	5.95	1	3.37	5.29
100-unf	93.94	10	5.93	12.49	9	5.79	9.17	3	3.66	6.83
200-unf	93.30	11	5.85	12.90	4	3.07	8.68	24	9.16	6.54
1hel	95.24	3	3.89	48.14	4	4.18	7.88	8	6.43	7.49
1lza	92.59	3	4.87	35.90	1	3.57	13.79	5	5.94	8.63
2lzt	94.72	5	6.18	12.14	9	7.88	11.27	2	3.76	10.54
1lyz	93.48	5	5.17	59.36	23	11.76	5.19	1	2.08	3.06
Hinge2										
10	96.54	1	3.61	43.28	3	4.64	33.10	4	5.43	4.35
20	96.33	5	5.12	41.32	1	2.42	30.91	2	4.25	2.68
30	96.20	3	4.22	28.55	4	4.39	20.97	2	4.14	11.88
40	96.64	3	4.13	24.89	6	5.75	24.48	2	3.80	21.72
50	96.51	1	3.22	60.36	2	3.57	18.54	3	4.41	3.90
60	96.36	2	4.10	49.18	1	3.37	14.98	7	6.44	4.73
70	96.37	2	4.07	41.38	1	2.81	26.51	4	5.10	7.30
80	96.23	2	4.29	59.96	1	3.28	14.75	7	6.57	4.60
90	95.69	3	4.11	43.83	4	4.77	24.32	1	2.56	11.93
100	95.97	1	3.37	47.41	4	4.80	27.79	2	3.67	3.82
0-unf	95.61	1	3.37	53.45	2	4.02	21.07	3	4.69	2.69
100-unf	94.67	3	3.66	20.02	2	3.44	12.56	7	5.56	7.15
200-unf	89.98	3	2.66	30.06	4	3.07	9.90	1	1.09	7.23
1hel	96.62	1	3.22	39.93	2	3.86	32.78	7	6.15	4.02
1lza	96.58	2	4.12	42.42	1	3.57	29.28	3	4.87	8.82
2lzt	95.67	2	3.76	44.85	1	3.55	27.58	3	4.02	7.29
1lyz	95.97	2	2.95	44.46	6	5.89	17.03	1	2.08	12.76

structures results in a smoother curve. In the case of BPTI, the smoother curve agrees better with experiment; for HEWL there are no experimental $g(\omega)$ data. The harmonic approximation to the configurational entropy is another quantity that is likely to have smaller errors when ensemble averaging is done. For BPTI the largest entropy difference between two conformations is 19.74 kcal/mol. This is a relatively large value and indicates that the calculation of configurational entropy based on one conformation should be treated with caution. The largest entropy difference between native subspace conformations of HEWL was smaller (7.70 kcal/mol), but this is still much larger than $k_B T$. Thus, the differences could have a significant impact on calculations that compare the free energy of different protein or nucleic acid states (e.g., Tidor and Karplus,⁴² and Irikura et al.⁵⁰). The inclusion of an ensemble of conformations gives a distribution of entropy values whose average provides a more well-defined value and error bounds. The entropies calculated

for BPTI and HEWL using only the first 17% of the vibrational modes showed comparable relative differences with a set of calculations including all vibrational modes. This confirms that the major differences between calculated entropy values are contributed by the low-frequency modes.

The two partially unfolded HEWL structures 100-unf and 200-unf had more low-frequency modes (Figure 17) and larger entropy values than the native subspace structures. The difference of the vibrational entropy contribution to the free energy ($-T\Delta S(\text{vib})$) of -38.8 kcal/mol between the native (average of 10 structures) and the partially unfolded (200-unf) conformation (difference for the 100-unf conformation: -20.9 kcal/mol) can be compared with the values for lysozyme reported by Makhatadze and Privalov.⁵¹ They list total entropies of protein unfolding in vacuo calculated by empirical estimates and by theoretical models. The empirical estimates resulted in a $-T\Delta S$ of -522 kcal/mol at 298 K; the theoretical models, consisting of separate main chain, side chain, and disulfide bond terms, resulted in a $-T\Delta S$ of -312 kcal/mol at 298 K. The value determined by the theoretical models did not include vibrational contributions. Makhatadze and Privalov reasoned that the discrepancy between the empirical and theoretical estimates could be partially caused by the exclusion of vibrational contributions to the theoretical estimate. The order of magnitude difference between the unfolding entropies of Makhatadze and Privalov and our values for the 200-unf structure suggest that the main contribution to the entropy of unfolding is from the multiplicity of very different conformations that exist in the unfolded state. However, it is important to note that a normal-mode calculation on a minimized, fully extended structure of HEWL yielded an entropy difference with the folded state of about -332 kcal/mol; this large entropy difference arises from the presence of 30 normal modes with a frequency of <1 cm^{-1} in the extended state. Although such an extended polypeptide is not a good model for the unfolded state (Lazaridis and Karplus, submitted for publication), the calculations indicate that a more detailed analysis of the configurational contribution to the entropy of the unfolded state is required.

The results in Tables 2 and 4 indicate that the space spanned by the low-frequency modes is relatively invariant among the different native subspace conformations. Hayward and Gō²³ calculated overlaps of normal modes with quasiharmonic modes and considered the overlap $f(1,10)$ for the first quasiharmonic mode with the first 10 normal modes to be significant at a value of 0.30; no reason for this choice was given. Here we find overlaps of 0.758 and 0.731 for the normal-mode spaces of BPTI and HEWL, respectively. The number of modes used in the overlap calculations was chosen such that 90% of the atomic fluctuations would be accounted for. The high overlap values suggest that the low-frequency normal modes of any native subspace conformation is representative of the corresponding space spanned by any other native subspace conformation. This would mean that the results of a single normal-mode calculation would be sufficient for applications in which the modes are used

as a basis set, such as normal-mode refinement,²⁵ reaction coordinate estimation,^{47,52} or the calculation of normal modes of symmetric multimers.⁴⁶ The conformational space spanned by the low-frequency normal modes of the partly unfolded HEWL structures 100-unf and 200-unf is clearly different from the native subspace (Table 5), with average $f(130,130)$ values of 0.392 and 0.321, respectively.

Although the space spanned by the first 130 normal modes of HEWL is relatively invariant, the space spanned by the first three normal modes is much more variable. This is indicated by the large variation in $1/2(f(3,5) + f(5,3))$ values in Table 4, and also in the differences in projection of the low-frequency modes on the proposed hinge bending motion (Table 6). The hinge motion projection results suggest that it is unrealistic to talk about a single hinge bending mode, in general. However, they do not invalidate the use of normal modes to study the hinge bending motion: it is still a very useful simplification to describe such a complex motion by a small number of degrees of freedom.

Averaging normal-mode results over an ensemble of native subspace conformations improves the results for properties such as the density of states and entropy. Normal-mode calculations based on a single minimized crystal structure can vary significantly, depending on which crystal structure is used. Experimental observations also reflect an average over different native subspace conformations of the protein. However, from a computational perspective, averaging seems to defy the purpose of normal-mode calculations, which is in part to avoid the calculation of computationally expensive MD trajectories. To average, one needs an ensemble of native subspace conformations that have to be generated by calculating an MD or Monte Carlo trajectory. However, there still is an advantage, since time-averaged quantities such as covariances between atoms do not have to be converged. The only function of these simulations is the generation of an ensemble of (nonidentical) native subspace conformations.

In conclusion, the utility of normal-mode calculations for analyzing structural and thermodynamic properties of macromolecules is increased by calculating them for several structures obtained from short simulations and averaging the results. This is expected to provide more meaningful values and makes it possible to determine error bars.

Acknowledgment. We thank Andrej Šali for making the plotting program ASGL available to us. The research was supported in part by a grant from the National Science Foundation.

References and Notes

- (1) McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature* **1977**, *267*, 585–590.
- (2) McCammon, J. A.; Harvey, S. C. *Dynamics of proteins and nucleic acids*; Cambridge University Press: Cambridge, UK, 1987.
- (3) Brooks, C. L., III; Karplus, M.; Pettitt, B. M. *Proteins: a theoretical perspective of dynamics, structure, and thermodynamics*; Adv. Chem. Phys. Vol. LXXI; Wiley & Sons: New York, 1988.
- (4) Karplus, M.; McCammon, J. A. *Annu. Rev. Biochem.* **1983**, *53*, 263–300.
- (5) Karplus, M.; Petsko, G. A. *Nature* **1990**, *347*, 631–639.
- (6) Wilson, E. B., Jr.; Decius, J. C.; Cross, P. C. *Molecular vibrations. The theory of infrared and Raman vibrational spectra*; McGraw-Hill: New York, 1955.
- (7) Brooks, B. R.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 6571–6575.
- (8) Gō, N.; Noguti, T.; Nishikawa, T. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 3696–3700.
- (9) Levitt, M.; Sander, C.; Stern, P. S. *J. Mol. Biol.* **1985**, *181*, 423–447.
- (10) Case, D. A. *Curr. Opin. Struct. Biol.* **1994**, *4*, 285–290.
- (11) Brooks, B. R.; Jañezić, D.; Karplus, M. *J. Comput. Chem.* **1995**, *16*, 1522–1542.
- (12) Hayward, S.; Gō, N. *Annu. Rev. Phys. Chem.* **1995**, *46*, 223–250.
- (13) Karplus, M.; Kushick, J. N. *Macromolecules* **1981**, *14*, 325–332.
- (14) Levy, R. M.; Perahia, D.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 1346–1350.
- (15) Levy, R. M.; Karplus, M.; Kushick, J.; Perahia, D. *Macromolecules* **1984**, *17*, 1370–1374.
- (16) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. *Proteins* **1993**, *17*, 412–425.
- (17) van Aalten, D. M. F.; Amadei, A.; Linssen, A. B. M.; Eijssink, V. G. H.; Vriend, G.; Berendsen, H. J. C. *Proteins* **1995**, *22*, 45–54.
- (18) van Aalten, D. M. F.; Amadei, A.; Bywater, R.; Findlay, J. B.; Berendsen, H. J. C.; Sander, C.; Stouten, P. F. *Biophys. J.* **1996**, *70*, 684–692.
- (19) Groot, B. L. D.; Amadei, A.; van Aalten, D. M. F.; Berendsen, H. J. C. *J. Biomol. Struct. Dyn.* **1996**, *13*, 741–751.
- (20) Elber, R.; Karplus, M. *Science* **1987**, *235*, 318–321.
- (21) Gō, N.; Noguti, T. *Chem. Scr.* **1989**, *29A*, 151–164.
- (22) Jañezić, D.; Venable, R. M.; Brooks, B. R. *J. Comput. Chem.* **1995**, *16*, 1554–1566.
- (23) Hayward, S.; Kitao, A.; Gō, N. *Protein Sci.* **1994**, *3*, 936–943.
- (24) Hayward, S.; Kitao, A.; Gō, N. *Proteins* **1995**, *23*, 177–186.
- (25) Kidera, A.; Gō, N. *J. Mol. Biol.* **1992**, *225*, 457–475.
- (26) Kidera, A.; Inaka, K.; Matsushima, M.; Gō, N. *J. Mol. Biol.* **1992**, *225*, 477–486.
- (27) Smith, J.; Cusack, S.; Tidor, B.; Karplus, M. *J. Chem. Phys.* **1990**, *93*, 2974–2991.
- (28) Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.
- (29) Jañezić, D.; Brooks, B. R. *J. Comput. Chem.* **1995**, *16*, 1543–1553.
- (30) Cusack, S. *Chem. Scr.* **1989**, *29A*, 103–107.
- (31) Smith, J.; Cusack, S.; Pezzeca, U.; Brooks, B.; Karplus, M. *J. Chem. Phys.* **1986**, *85*, 3636–3654.
- (32) Brooks, C. L., III; Karplus, M. *J. Mol. Biol.* **1989**, *208*, 159–181.
- (33) Marquart, M.; Walter, J.; Deisenhofer, J.; Bode, W.; Huber, R. *Acta Crystallogr., Sect. B* **1983**, *39*, 480–490.
- (34) Diamond, R. *J. Mol. Biol.* **1974**, *82*, 371–391.
- (35) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535–542.
- (36) Brooks, C. L., III; Karplus, M. *J. Chem. Phys.* **1983**, *79*, 6312–6325.
- (37) Pomès, R.; McCammon, J. A. *Chem. Phys. Lett.* **1990**, *166*, 425–428.
- (38) Wilson, K. P.; Malcolm, B. A.; Matthews, B. W. *J. Biol. Chem.* **1992**, *267*, 10842–10849.
- (39) Maenaka, K.; Matsushima, M.; Song, H.; Sunada, F.; Watanabe, K.; Kumagai, I. *J. Mol. Biol.* **1995**, *247*, 281–293.
- (40) Ramanadham, M.; Sieker, L. C.; Jensen, L. H. *Acta Crystallogr., Sect. B* **1990**, *46*, 63–69.
- (41) McQuarrie, D. A. *Statistical mechanics*; Harper & Row: New York, 1976.
- (42) Tidor, B.; Karplus, M. *Proteins* **1993**, *15*, 71–79.
- (43) Nishikawa, T.; Gō, N. *Proteins* **1987**, *2*, 308–329.
- (44) Gibrat, J.; Gō, N. *Proteins* **1990**, *8*, 258–279.
- (45) Neria, E.; Fischer, S.; Karplus, M. *J. Chem. Phys.* **1996**, *105*, 1902–1921.
- (46) Simonson, T.; Perahia, D. *Biophys. J.* **1992**, *61*, 410–427.
- (47) Thomas, A.; Field, M. J.; Mouawad, L.; Perahia, D. *J. Mol. Biol.* **1996**, *257*, 1070–1087.
- (48) Phillips, G. N., Jr. *Biophys. J.* **1990**, *57*, 381–383.
- (49) Brooks, B. R.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 4995–4999.
- (50) Irikura, K.; Tidor, B.; Brooks, B. R.; Karplus, M. *Science* **1985**, *229*, 571–572.
- (51) Makhataadze, G. I.; Privalov, P. L. *Adv. Prot. Chem.* **1995**, *47*, 307–425.
- (52) Basu, G.; Kitao, A.; Hirata, F.; Gō, N. *J. Am. Chem. Soc.* **1994**, *116*, 6307–6315.