Protein Science (1998), 7:649–666. Cambridge University Press. Printed in the USA.

Locally accessible conformations of proteins: Multiple molecular dynamics simulations of crambin

LEO S.D. CAVES,[1,3] JEFFREY D. EVANSECK,[1,4] AND MARTIN KARPLUS[1,2]

[1]Department of Chemistry and Clinical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138
[2]Laboratoire de Chimie Biophysique, Université ISIS, Louis Pasteur, 4 Rue Blaise Pascal, 67000 Strasbourg, France
[3]Department of Chemistry, University of York, York, YO1 5DD, United Kingdom

(RECEIVED April 22, 1997; ACCEPTED November 17, 1997)

Abstract

Multiple molecular dynamics (MD) simulations of crambin with different initial atomic velocities are used to sample conformations in the vicinity of the native structure. Individual trajectories of length up to 5 ns sample only a fraction of the conformational distribution generated by ten independent 120 ps trajectories at 300 K. The backbone atom conformational space distribution is analyzed using principal components analysis (PCA). Four different major conformational regions are found. In general, a trajectory samples only one region and few transitions between the regions are observed. Consequently, the averages of structural and dynamic properties over the ten trajectories differ significantly from those obtained from individual trajectories. The nature of the conformational sampling has important consequences for the utilization of MD simulations for a wide range of problems, such as comparisons with X-ray or NMR data. The overall average structure is significantly closer to the X-ray structure than any of the individual trajectory average structures. The high frequency (less than 10 ps) atomic fluctuations from the ten trajectories tend to be similar, but the lower frequency (100 ps) motions are different. To improve conformational sampling in molecular dynamics simulations of proteins, as in nucleic acids, multiple trajectories with different initial conditions should be used rather than a single long trajectory.

Keywords: conformational space; dimensionality reduction; ergodicity; molecular dynamics simulation; multiple minima; potential energy surface; principal components analysis; sensitivity to initial conditions

Proteins are complex systems (Brooks et al., 1988). Although significant progress has been made in establishing the relation of structure, dynamics, and function, much remains to be learned. One approach makes use of computer simulation methods, of which molecular dynamics has been shown to be very effective for studying proteins. Simple arguments suggest that the potential energy surface describing the native state contains multiple minima corresponding to very similar, but slightly different conformations (Cooper, 1976). This has been confirmed by molecular dynamics simulation (Elber & Karplus, 1987). Conformational flexibility, which involves sampling of many of these minima, is essential for the function of many proteins (Perutz, 1989; Karplus & Petsko, 1990; Steinbach et al., 1991; Gerstein et al., 1994). Thus, the success of simulation studies requires an effective exploration of the native potential energy surface of the molecule (Brooks et al., 1988).

Protein simulations, like those of other complex systems, such as glasses (Barrat & Klein, 1991), have to address the problem of ergodicity (Tolman, 1938). This issue arises for any system with a multiminimum character where there are barriers that may not be crossed sufficiently frequently to obtain an equilibrium distribution in accessible computation times. If the time average of some quantity depends on the initial conditions, the system is nonergodic on the time scale and at the temperature of the simulation. Clearly different properties will have different convergence behavior, so that it is necessary to study more than one to determine the ergodicity of a simulation (Mountain & Thirumalai, 1989). Molecular dynamics simulations of ribonuclease-A at 300 K revealed that the equipartition of kinetic energy was achieved in about 50 ps (Straub & Thirumalai, 1993). The non-bonded interactions showed much longer relaxation times (longer than the 75 ps duration of the simulation) reflecting a wide range of interconversion rates between conformational substates (Straub et al., 1994). Molecular dynamics simulations of myoglobin (Kuczera et al., 1990) and thioredoxin (Elofsson & Nilsson, 1993) have revealed a lack of

Reprint requests to: Martin Karplus, Laboratoire de Chimie Biophysique, Institut le Bel, Universite Pasteur, 4 Rue Blaise Pascal, 67000 Strasbourg, France; e-mail: marci@brel.u-strasbg.fr.

[4]Current address: Department of Chemistry, University of Miami, P.O. Box 249118, Green Gables, Florida 33124.

Abbreviations: RMS, root mean square; RMSF, RMS fluctuations; RMSD, RMS displacements; BPTI, bovine pancreatic trypsin inhibitor; MD, molecular dynamics; PES, potential energy surface; PCA, principal components analysis.

convergence of dynamic and average properties, such as the atomic fluctuations, on the 100 ps timescale. In the study of thioredoxin, large variations were found in properties from independent runs with different initial velocity assignments and initial conformations. Thus, by analogy with the coin tossing problem (Feller, 1971), even when frequent barrier crossing occurs in molecular dynamics simulations, initial biases may take a long time to disappear (Smith et al., 1993).

This makes the sampling problem of considerable concern in macromolecular dynamics simulations (Hodel et al., 1993; Caspar, 1995; Clarage et al., 1995; Hunenberger et al., 1995). A number of techniques have been developed for enhancing the underlying sampling characteristics of the molecular dynamics method (Bruccoleri & Karplus, 1990; Elber & Karplus, 1990; Durup, 1991; Mao, 1991; van Schaik et al., 1992; Amadei et al., 1996), but these have yet to be applied extensively to systems the size of proteins; some recent work describes the application of the method proposed by Amadei et al. (1996) to the 85 residue protein HPr in vacuum (de Groot et al., 1996). Another approach to the sampling problem is to exploit the increasing power of computers to perform molecular dynamics simulations that extend into the nanosecond timescale (Daggett & Levitt, 1993; van Gunsteren et al., 1995; Cheatham & Kollman, 1996). However, there have been suggestions that more effective sampling of the conformational space of proteins can be obtained from multiple short trajectories rather than a single, long trajectory (Elofsson & Nilsson, 1993; Straub et al., 1994)

A multiple molecular dynamics simulation approach has been investigated by Auffinger et al. (1995) and Auffinger and Westhof (1997) in studies of a terminally restrained loop fragment of $tRNA^{Asp}$. They report six independent 500 ps simulations that differ only in the initial velocities. An analysis of deviations of the trajectories from the initial structure and pairwise RMSD of the average structures revealed extensive and non-convergent sampling, but a detailed analysis of the how the individual trajectories contributed to the conformational space distribution was not presented. They conclude that multiple trajectories should be used in place of longer trajectories, but unfortunately, no explicit comparison with one or more 3 ns or longer trajectories was made to verify this conclusion. Also, the distinct character of nucleic acid structure requires specialized treatments (Louise-May et al., 1996) and it is clear that proteins warrant independent investigation.

In this paper, we perform a set of ten independent 120 ps molecular dynamics simulations at a temperature of 300 K to sample the conformational space of crambin in the neighborhood of the native structure and we use principal components analysis to interpret the results. The trajectories start with the same structure and differ only in the initial velocity assignment. The nature and variation in the resulting structural averages and fluctuations are analyzed. Although the simulations were performed in vacuum, we expect the analysis of the sampling to provide information that can be carried over to solvated simulations. The latter would require much more computer time due to the significantly increased cost per dynamics step and because conformational transitions appear to be slower in the presence of explicit solvent. Crambin (Fig. 5A) was used for this exploratory study because it is a small system for which both crystal (Hendrickson & Teeter, 1981) and solution (Bonvin et al., 1993) structural data exist. Crambin has also been used as a test system for a variety of simulations (Whitlow & Teeter, 1986; Ornstein, 1990; Teeter & Case, 1990; Garcia, 1992) and is thus a case of general interest. We note that, because crambin contains three disulphide bridges, it is expected to have reduced conformational freedom relative to systems without such cross-links (Tidor & Karplus, 1993).

The different molecular dynamics simulations result in a range of values for various properties, indicating that all of these are not ergodic on the timescale examined. The characteristics of the computed average structures and atomic fluctuations are related to the nature and distribution of the conformational substates sampled by the different trajectories. Particular regions of conformational space are found to be preferentially sampled and these are related to the underlying potential energy surface. Selected trajectories were extended into the nanosecond regime to examine their behavior relative to that of the ten 120 ps trajectories. Through a direct comparison, our analysis confirms that multiple short trajectories are more efficient than a single long trajectory for sampling conformational space. Principal components analysis was used for reducing the dimensionality of the conformational space sampled in the simulations and formed the basis for visualization of the behavior of the multiple trajectories. The overall averages from all ten 120 ps simulations were compared with single trajectory results. It was found that the overall average structure is closer to the reference structure than any of the average structures from individual trajectories. We consider the sensitivity of the trajectories to initial conditions and the implications for the results of studies which use single trajectories for investigating physical phenomena or for testing different parameterizations, methodologies, and protocols.

Results

General features of the structure and energetics of the trajectories are presented first. The average structures and atomic fluctuations of the trajectories are then described at both the molecule level and for the individual residues. The average properties of individual trajectories are compared with the corresponding average over all ten trajectories. The differences among the individual trajectories are then characterized by means of a principal components analysis of the conformational space.

General characteristics of the simulations

Energetics

The average temperature over the ten independent simulations is 305 ± 9 K (Table 1). Thus, essentially the same amount of kinetic energy is available to explore the potential energy surface (PES) in each of the trajectories. The ratio of RMS fluctuations in the total and kinetic energy, averaged over all trajectories is 2.8 ± 0.4%, an indication that the simulations are well behaved (van Gunsteren & Berendsen, 1977). The lowest average potential energy is found in trajectories T4 and T9, the highest in T6 and T2. The values differ by up to 12.5 kcal mol^{-1}, a value comparable to the RMS fluctuation of the potential energy within each trajectory (~10 kcal mol^{-1}, see Table 1).

Evolution from the reference structure

The evolution of the trajectories was monitored via the RMSD of the backbone atoms of transient MD structures with respect to the reference structure (Fig. 1). All of the trajectories exhibit an initial rapid (<5 ps) shift. This is typical of molecular dynamics simulations of proteins in vacuum (Caves et al., 1991) and in solution (Levitt & Sharon, 1988). In the present simulations the

Table 1. *Summary energetics of the ten independent trajectories (T1–T10)*

	Trajectory[a]				Minima	
ID	$\langle T \rangle$[b] (K)	$\langle \Delta T \rangle$ (K)	$\langle V \rangle$[c] (kcal mol^{-1})	$\langle \Delta V \rangle$ (kcal mol^{-1})	$\langle V \rangle$ (kcal mol^{-1})	σV (kcal mol^{-1})
T1	309.2	9.1	8.8	10.8	3.9	3.8
T2	304.6	8.7	12.5	10.4	14.1	1.7
T3	304.3	9.5	3.6	11.2	5.6	2.3
T4	303.2	8.6	0	10.3	2.1	2.1
T5	309.9	9.2	5.3	10.9	2.4	3.5
T6	307.3	9.0	12.0	10.7	9.3	2.4
T7	302.1	8.3	5.5	9.9	9.8	2.3
T8	302.9	9.2	5.8	11	8.9	3.2
T9	302.2	8.9	0	10.6	5.7	3.9
T10	306.7	9.1	1.8	10.8	0.0	2.6

[a] $\langle\ \rangle$ and σ denote mean and standard deviation, respectively. $\langle \Delta T \rangle$ and $\langle \Delta V \rangle$ denote RMS fluctuations of temperature and potential energy, respectively.
[b] T = temperature.
[c] V = potential energy.

initial shift is similar in magnitude for all the trajectories (RMSD of 0.73 to 1.09 Å for the backbone atoms). After the initial 5 ps, the RMSD for each trajectory is more stable, except for T4 and T5 which only stabilize after ~40 ps and ~80 ps, respectively. In the individual trajectories, the fluctuations from the mean RMSD value differ significantly. The fluctuations are relatively small in some cases (e.g., T1, T8, and T10) and large in others (T2 and T7); the magnitude of the fluctuations in trajectory T9 varies significantly during the simulation. Insight into the origin of this range of behavior is provided by the direct examination of the conformational space sampled by the trajectories (see below).

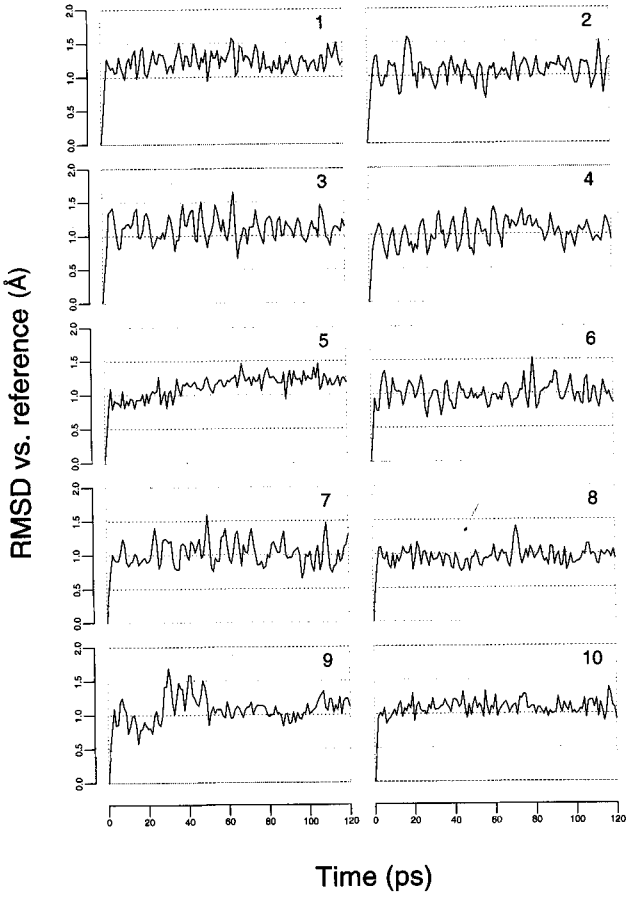

Fig. 1. The backbone atom RMSD of transients vs. the reference structure as a function of time for each of the 120 ps molecular dynamics simulations of crambin (1 ps time intervals).

Overall structural characteristics

Static properties of the average structures and their relationship to the reference and NMR structures

The molecular volume of the individual average structures decreases fairly uniformly by 3–4% relative to the reference structure, though in one case (T2) the change is over 6% (Table 2). The accessible surface area consistently decreases by a similar amount (average of 3.5%), with a wide range of variation (0.5–4%). The relative changes in the backbone atom radius of gyration are smaller, typically within 1–2% of the reference value. In general, there is an expansion along the primary axis, R_X (range of 2.5–4%) and there is a consistent, though variable, reduction (range of 0.8–5.2%) in R_Y (see Materials and methods for definition of terms). In two cases (T5, T10), a significant (>2.8%) contraction in R_X occurs, which is related to lower than average contractions along R_Y. The global average structure maintains the radius of gyration of the reference structure but undergoes a large (~8%) reduction in volume. A large component of the low relative molecular volume of the simulation average structures is attributable to an increase in atomic packing density, as indicated by poor internal geometry and unfavorable van der Waals energy, resulting from the extensive Cartesian coordinate averaging. These effects are readily eliminated by energy minimization. When the bonded geometry is corrected, by 20 steps of steepest descent minimization, the resulting volumes are at worse 1% lower than the reference structure. Further minimization, in which the softer non-bonded terms are improved, results in a range of volumes that are 0.5 to 2% lower than the reference structure (data not shown).

Table 2. *Structural properties of the reference, trajectory-average, and NMR structures*

ID	RMSD vs. back	X (Å) all[a]	Volume (Å^3)	ASA[b] (Å^2)	Radius of gyration[c] (Å) R_g	R_X	R_Y	R_Z	RMSF (Å)
X			4,886	3,033	9.58	7.36	5.18	3.26	0.47
Individual (T1–T10) and global (T) trajectory-average[d] structures									
T1	1.09	1.47	4,727	3,011	9.65	7.60	4.93	3.31	0.59
T2	0.88	1.36	4,591	2,968	9.56	7.56	4.91	3.19	0.62
T3	0.94	1.21	4,702	2,980	9.73	7.66	4.98	3.33	0.60
T4	0.84	1.10	4,712	2,983	9.71	7.65	4.97	3.31	0.61
T5	1.06	1.37	4,739	2,916	9.41	7.12	5.12	3.41	0.52
T6	0.80	1.35	4,723	2,994	9.69	7.66	5.01	3.18	0.59
T7	0.84	1.34	4,705	3,018	9.71	7.67	4.99	3.25	0.62
T8	0.73	1.06	4,687	2,951	9.61	7.55	4.95	3.30	0.60
T9	0.77	1.15	4,666	2,906	9.53	7.33	5.13	3.28	0.78
T10	0.96	1.33	4,728	2,928	9.38	7.15	5.12	3.25	0.51
$\langle T \rangle$[e]	0.89	1.27	4,698	2,965	9.60	7.50	5.01	3.28	0.60
σT[e]	0.12	0.13	43	39	0.13	0.21	0.08	0.07	0.07
T	0.64	0.93	4,488	2,931	9.58	7.49	5.00	3.26	0.85
Individual (N1–N8) and average (N) NMR structures									
N1	1.21	1.92	4,884	2,985	9.40	7.37	4.92	3.14	
N2	1.08	1.67	4,872	2,999	9.40	7.12	5.11	3.40	
N3	1.02	1.72	4,913	3,000	9.49	7.24	5.18	3.31	
N4	1.13	1.81	4,882	3,023	9.37	7.05	5.07	3.53	
N5	1.00	1.62	4,787	2,994	9.42	7.11	5.17	3.38	
N6	0.94	1.83	4,906	3,041	9.50	7.32	4.97	3.47	
N7	1.05	1.76	4,883	3,057	9.50	7.13	5.26	3.43	
N8	1.04	1.73	4,884	3,027	9.40	7.10	5.17	3.36	
$\langle N \rangle$	1.06	1.76	4,876	3,016	9.44	7.18	5.10	3.38	
σN	0.08	0.10	38	25	0.05	0.11	0.11	0.12	
N	0.90	1.56	4,919	3,074	9.45	7.18	5.08	3.44	

[a]Non-hydrogen atoms.
[b]Solvent accessible surface area: 1.6 Å probe.
[c]Backbone atoms only.
[d]Average of transient structures from 20–120 ps at resolution of 0.2 ps.
[e]$\langle\ \rangle$ and σ denote mean and standard deviation, respectively.

The variation in the average structural properties resulting from the different molecular dynamics simulations can be compared with that seen in the eight structural models derived from solution NMR experiments. In general, the NMR structures have volumes that are very similar to that of the reference structure (differences of <0.5%). In one case, however (N5), a significant relative compaction (2%) is found, although this is less than that in any of the average structures. The accessible surface areas also are very similar to, but in general, smaller than that of the reference structure; the range of differences (up to 1.6%) encompass half of the distribution in the average structures. The NMR structures differ in shape from the reference structure. The radius of gyration is smaller (average 1.5%), which results mainly from a reduction in R_X (average 2.5%) and a concomitant increase in R_Z (average 3.5%) in all but one case (N1). Two of the structures (N6 and N7) have the same radius of gyration, but are quite different in shape, as reflected by their different values for R_X and R_Y.

A reduction in molecular volume as a result of molecular dynamics simulations in vacuum is common and the magnitude of the changes found here is comparable to that of other studies of small proteins, whether in vacuum (Caves et al., 1991), with explicit crystallographic waters (Post et al., 1986), or fully solvated (Levitt & Sharon, 1988). The components of the radius of gyration indicate that the observed reduction in volume is not due to a uniform overall compaction, but that more complex changes in the molecular shape occur. This is also reflected in the lack of any simple correlation of these relative structural changes with the RMSD from the reference structure (see below). The variation in these properties (−2.1 to 1.6% and −3.03 to−6.05% for radius of gyration and volume, respectively, Table 2) of the average structures is of the same order as those found in other simulation studies of crambin using different empirical force-fields, molecular environment, and methodological conditions (Whitlow & Teeter, 1986; Jorgensen & Tirado-Rives, 1988; Lii et al., 1989; Ornstein, 1990; Teeter & Case, 1990).

Overall conformational differences

The overall differences between the average structures from the trajectories and the reference structure were examined in terms of

Table 3. *Pairwise RMSD (Å) for the reference (X) and individual trajectory-average (T1–T10), and global trajectory-average structures (T); lower triangle: backbone atoms; upper triangle: all (non-hydrogen) atoms*

ID	X	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T
X		1.47	1.36	1.21	1.10	1.37	1.35	1.34	1.06	1.15	1.33	0.93
T1	1.09		1.49	1.18	1.07	1.82	1.53	1.54	1.12	1.25	1.76	1.05
T2	0.88	0.96		1.23	1.21	1.53	0.88	0.76	0.98	1.19	1.64	0.79
T3	0.94	0.90	0.63		0.47	1.60	1.30	1.12	0.93	1.03	1.66	0.75
T4	0.84	0.87	0.61	0.19		1.54	1.24	1.10	0.80	0.98	1.58	0.66
T5	1.06	1.46	1.16	1.37	1.31		1.81	1.71	1.26	1.06	1.06	1.13
T6	0.80	0.93	0.59	0.60	0.56	1.33		0.60	1.14	1.41	1.77	0.93
T7	0.84	0.94	0.49	0.35	0.35	1.28	0.38		1.12	1.34	1.76	0.84
T8	0.73	0.86	0.51	0.69	0.61	1.05	0.61	0.61		0.92	1.48	0.56
T9	0.77	0.90	0.75	0.81	0.76	0.87	0.87	0.83	0.68		1.15	0.66
T10	0.96	1.44	1.20	1.41	1.33	0.75	1.25	1.28	1.19	0.93		1.19
T	0.64	0.78	0.44	0.53	0.46	0.95	0.50	0.44	0.40	0.47	0.97	

pairwise RMSD values (Tables 2 and 3). The average structures differ from the reference structure by 0.73 to 1.09 Å (average of 0.89 Å) for the backbone atoms and 1.06 to 1.47 Å (average of 1.27 Å) for all atoms. The average structures are, in general, closer to the reference structure than are the NMR structures (average 1.06 Å for the backbone). The individual average structures differ from each other by 0.19 to 1.46 Å in the backbone and by 0.47 to 1.82 Å for all atoms (Table 3). Thus, the average structures differ from each other (in terms of RMSD values) as much as they differ from the reference structure. This suggests that each trajectory explores its own region of conformational space; this important aspect of the analysis is examined in more detail below.

The extent of the conformational space sampled can be seen from by the differences of the average structures computed from the 10–20 ps and 100–120 ps periods of the trajectories (see Table 3B in Electronic supplementary material). The structures do not differ significantly from each other with respect to the reference structure (mean absolute difference in the backbone RMSD of 0.11 Å). However, for each trajectory the structures from these two periods differ significantly (mean RMSD of 0.60 Å). Schematically, the geometry of the accessible conformational space can be regarded as a hypersphere around the reference structure, with each trajectory sampling along a different radial direction (Elber & Karplus, 1987). The bulk of the displacements in a trajectory occur rapidly along the radial direction (average backbone RMSD of ~0.9 Å in <10 ps); additional sampling is then confined to the surface of the hypersphere (i.e., largely maintaining the distance from the reference structure), but is less extensive; i.e., after 10 ps, the RMSD between pairs of structures in a given trajectory varies up to approximately two-thirds of their RMSD from the reference structure. Consequently, a structure representative of all trajectories (the global trajectory-average) should be a better measure of the behavior of the system than any individual average structure. The global trajectory-average structure has an RMSD from the reference structure of 0.64 Å for the backbone atoms and 1.05 Å for all atoms (Table 2). Thus, averaging of structures from a set of independent trajectories, relative to an average structure representative of a single trajectory, improves the agreement with the reference structure by ~30% for the backbone atoms. It is interesting to note that this effect can also be found in the comparison of the individual and average structures obtained from the NMR results (see Table 2); the effect is smaller for the NMR structures because they are more similar to each other.

Atomic motions

The atomic positional distribution functions can be characterized in terms of their moments (Ichiye & Karplus, 1987, 1988). In the individual trajectories, the backbone RMSF have a range of 0.51 to 0.78 Å, with an average of 0.61 Å (see Table 2). The standard deviation of the RMSF is 0.07 Å, about 10% of the average value. Measures of the higher moments (skewness, kurtosis) show a larger range of relative variation. The anisotropy measure A_1 shows the greatest relative variation, although this is influenced by a single trajectory (T9), which has a significantly larger value than the others (see Table 5 in Electronic supplementary material). These results, computed over 100 ps, contrast with data collected over 8 ps intervals, which are very similar (data not shown). Thus, the individual trajectories are characterized by low-frequency motions that are the most anisotropic and anharmonic in nature and can differ significantly from one trajectory to another, while the local (higher-frequency) motions are more similar for the different trajectories. However, it should be noted that the low frequency motions may not have fully developed on the 100 ps timescale used here (Teeter & Case, 1990; Hunenberger et al., 1995). A detailed analysis of the time development of the atomic fluctuations will be given separately.

Local structural characteristics

Figure 2 shows the residue-averaged trends of the atomic displacements (RMSD relative to the reference structure) and the atomic fluctuations (RMSF) for the backbone atoms. These trends are similar, with correlation coefficients in the range 0.20 (T10) to 0.73 (T7) as shown in Figure 2. The RMSD values are approximately twice as large as the RMSF in most cases. However, they can be larger, for example, in T9 and the C-terminal region of T8. The variation in the trends of the RMSD is rather large (Figs. 3, 4A) with correlation coefficients between trajectories in the range of 0.23 to 0.98. The RMSF trends (Figs. 2, 4B) are more consistent across the trajectories, with correlation coefficients in the range 0.56 to 0.98. Thus, the individual trajectories are more similar in their average motion than in their average structure.

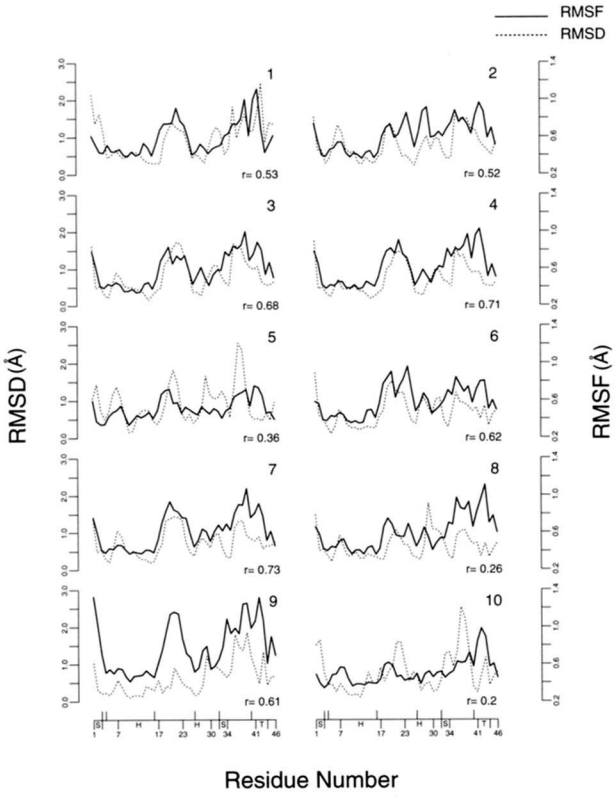

Fig. 2. Residue averaged backbone atom RMSD vs. the reference structure (dashed line) and the residue averaged backbone RMSF (solid line). The value of the correlation coefficient (range 0.20 to 0.73) relating the two for each trajectory is given in the lower right hand corner of each plot. Significant residue numbers are indicated on the abscissae; regions of secondary structure are indicated as follows: H, α-helix; S, β-strand; T, β-turn. The vertical ticks above the abscissae indicate the positions of the cysteine residues.

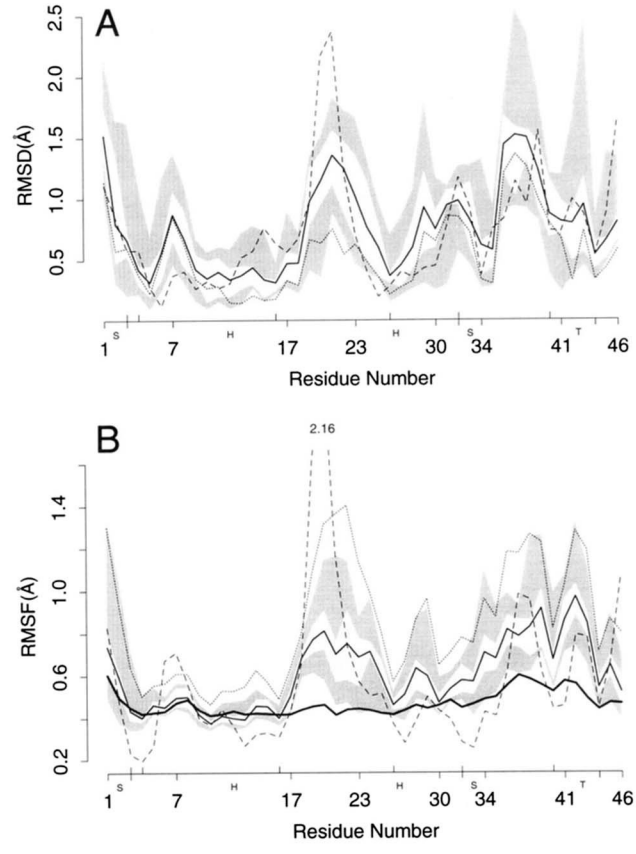

Fig. 3. Summary of the residue-averaged backbone atomic RMSD vs. the reference structure and RMSF data for the individual trajectories. The range of the data is represented as the dark shaded region, with the interquartile range as the lighter shaded region (a quarter of the data lies below the first quartile and a quarter above the third quartile; the interquartile range is the difference between the two), the median as the white line, the mean as the thin solid line. **A:** RMSD data: The RMSD of the global trajectory-average structure (T) is given in the dotted line; the NMR average structure (N) is given in the dashed line. **B:** RMSF data: The RMSF with the global trajectory-average structure as the reference is given as the dotted line; that calculated from the NMR ensemble (N1 to N8) is given in the dashed line. The RMSF derived from the crystallographic *B*-factors is given as the thick solid line.

These general trends are readily interpreted in terms of the topology and secondary structure of the protein (Hendrickson & Teeter, 1981). For example, residues involved in disulfide bridges (see Fig. 5A) or in the α-helices, in general, show relatively smaller displacements from the reference structure and have lower mobility than other regions of the protein. Correspondingly, in regions between secondary structure elements (for example, between the α-helices and between the C-terminal strand and β turn) and toward the chain termini, larger variations in the displacements and motion occurs. There are exceptions to this general behavior, e.g., the regions near the ends of helix 2 can exhibit both large displacements from the reference structure (T8) and high intrinsic mobility (T2) (see Fig. 2). The residue-averaged RMSD of the average NMR structure (N) from the reference structure corresponds well with the average trend of the individual trajectories (correlation coefficient of 0.66) and has a very similar overall magnitude (see Table 2). The lower overall RMSD of the global-trajectory average structure (T) compared with the average displacements of the individual trajectories is clearly evident in Figure 3A, especially in the highly variable inter-helix region.

The residue-averaged RMSF derived from the crystallographic *B*-factors are also shown in Figure 3B. The RMSF of the individual trajectories are consistently larger and show more variability than the *B*-factor results (average of 0.60 vs. 0.47 Å, see Table 2) reflecting the relative rigidity of the crambin crystal structure (Hendrickson & Teeter, 1981). The 8 NMR models (N1–N8) can be considered as a structural ensemble and the resulting residue-averaged backbone RMSF computed (see Fig. 3B): the magnitude (0.58 Å) is close to the average of the trajectory results and the residue trends are also similar (correlation coefficient of 0.55); the solution structure of crambin has particularly high variation in the loop regions between the helices (residues 18–22) and in the C-terminus (residues 35–39 and 41–44) (Bonvin et al., 1993). It is not clear whether this is due to actual mobility or to the lack of sufficient NOE data to provide definition in these regions.

If the atomic fluctuations are calculated with respect to the global trajectory-average structure (that representative of all ten trajectories), rather than the average structure for each individual trajectory, there is a significant increase in the fluctuation magnitudes (0.85 Å, see Table 2 and Fig. 3B). These overall fluctuations represent an equal weighting of data from each of the trajectories; a more realistic treatment of the data would be based on an ap-

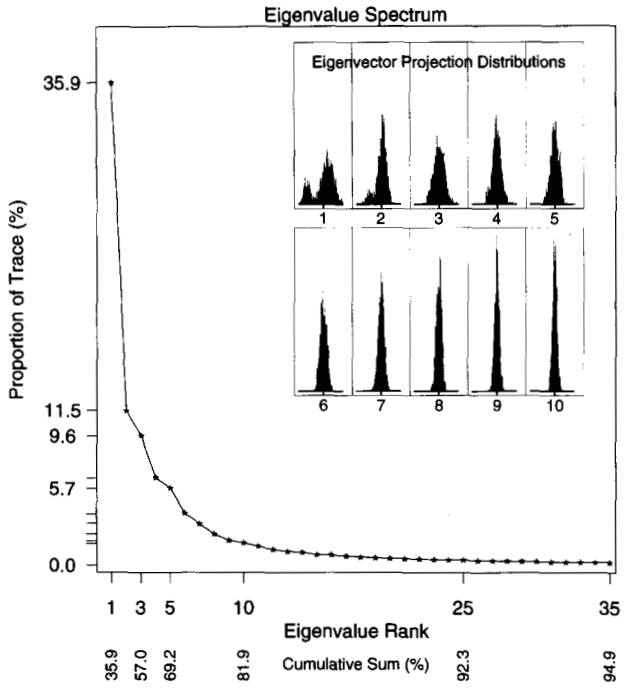

Fig. 4. Summary of the results of the principal components analysis (PCA). The most significant region of the rank ordered eigenvalue spectrum resulting from the diagonalization of the sample covariance matrix of the backbone atomic coordinates of the structures from the 10 trajectories (transients taken at 1 ps resolution) and the reference structure (total of 1,201 structures). The full dimensionality of the conformational space is 408 (3N-6, N = 138 backbone atoms). The magnitude of the eigenvalues is expressed as a percentage of their overall sum (which corresponds to the mean-square atomic fluctuation for all backbone atoms, averaged over the sample of structures). The proportion of the total variance (mean-square coordinate fluctuations) accounted for in 3 dimensions is 57%, with 95% accounted for in 35 dimensions. Inset are the histograms resulting from the projection of the sample structures onto the first ten principal components (the eigenvectors corresponding to the ten largest eigenvalues). For ease of comparison, each histogram has the same scale; the abscissa ranges from −1.28 to 1.00 Å and the maximum value on the ordinate is 96. The number of bins used in the histograms was 100; the sample size was 1,201.

propriate weighting of the individual trajectories (see Discussion). The consideration of fluctuations about an overall average structure resulting from many independent samples of conformational space is analogous to the interpretation of X-ray crystallographic *B*-factors, which takes into account extensive temporal and spatial averaging (Ringe & Petsko, 1985).

Global structural characteristics of the trajectories

Visualization of the conformational space sampled in the trajectories

To gain insight into the origin of the variation in structural properties described above, the conformational space of the molecule was examined by means of PCA (see Materials and methods). The Cartesian coordinates of the 138 (N) backbone atoms were used to define a conformational space of 408 dimensions (3N–6) after net rotational and translational degrees of freedom are removed by coordinate superposition. Transient structures from the 120 ps trajectories were taken at 1 ps intervals which, with the addition of the reference structure, constituted a sample size of 1,201 data points. PCA was then performed by diagonalizing the coordinate covariance matrix. The rank ordered eigenvalue spectrum (Fig. 4) suggests an inherent dimensionality of 5 due to the lower and monotonically decreasing magnitude of the eigenvalues after this point (Jackson, 1991). The first 3 components account for 57% (35.9, 11.5, and 9.6%, respectively); 5 components account for 69%, whereas 10 account for only an additional ~12%. A total of 95% of the variance is contained in 35 dimensions. The distribution of the sample along each of the first 10 principal components was examined. The first, and to a lesser extent the second, components have a bimodal distribution (Fig. 4, insert), whereas the remaining components are unimodal and become progressively sharper, reflecting their decreasing variance. For practical purposes, further analysis of the conformational space is restricted to the first 3 components which provides a very useful description of the system.

Structural characterization of the principal components

The principal components (eigenvectors) can be characterized directly as atomic displacements; stereo diagrams are presented in Figure 5. The first principal component (pc1) corresponds to a bending about the molecular cleft (see Fig. 5A). It involves the concerted displacement of the interhelix region (residues 19–23) and the C-terminal helix (residues 23–28) with the extended region between the C-terminal strand and β-turn (residues 36–38). This, in turn, is coupled to the N-terminus (residues 1 and 2) through the (residue 3–40) disulfide bridge. The second principal component (pc2) predominantly involves the C-terminus, especially the region around the β-turn (residues 42–43), which pivots with respect to the rest of the molecule. The displacements in this region are well characterized by changes to the mainchain dihedral angles of Gly 42. In addition, there are smaller displacements in other regions of the molecule. The third principal component (pc3), like the first, is delocalized and centered about the molecular cleft, but has the character of a twisting displacement.

It is interesting to examine the relation of the principal components with respect to the principal geometric axes of the molecule (see Fig. 5). The atomic displacements corresponding to pc1 are distributed predominantly along the principal axes, *X* and *Z* (mean-square displacements of 52% and 37%, respectively), where a positive step along the principal component corresponds to an expansion along *X* and a contraction along *Z*. For pc2, the character is more localized along *Z* (61%), with a positive step corresponding to a contraction along *Z*. For pc3, the displacements are similar to that of pc1, but with opposite phase. This suggests that the coordinates of the structures in the principal component space can be used as an aid to interpreting the values for the components of R_g with which they are associated. For example, for the individual trajectory average structures, R_X shows the highest standard deviation (Table 2), in accordance with its large component along pc1, the direction of maximum variance in the conformational space distribution. The average structures, T5 and T10, represent the largest positive and negative displacements along pc1 (Fig. 7, center panel) and correspondingly show the highest and lowest values of R_X, respectively (Table 2). Similarly, average structures T1 and T2 represent the positive and negative extremes of pc2, and show a corresponding contraction and expansion in R_Z, respectively. Thus, through PCA, the distribution of the components of the radius of gyration in the trajectory-average structures can be

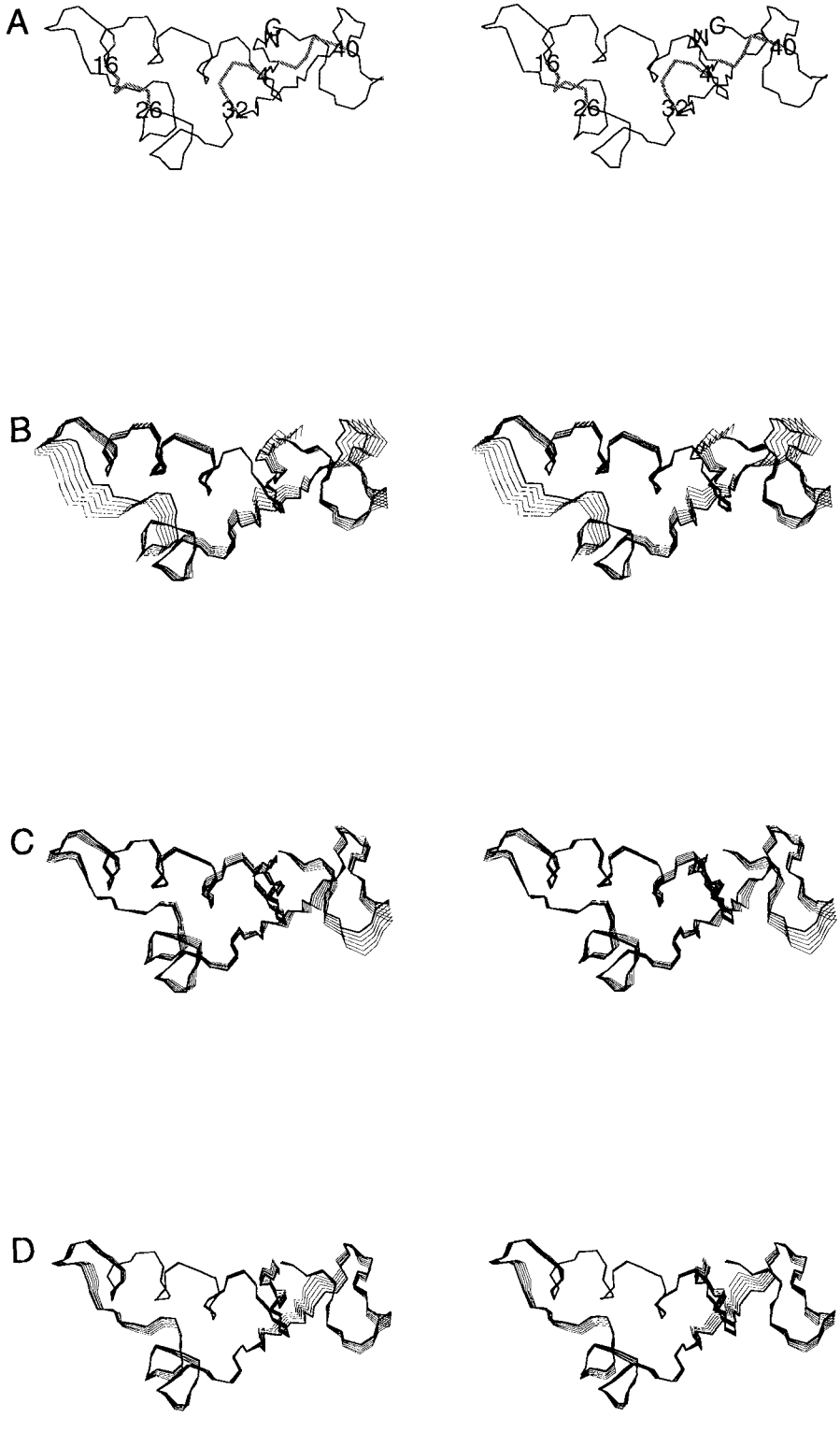

Fig. 5. Stereo figures (wall-eyed) of the first three principal components, represented as atomic displacements from the reference structure. The scale factor for the displacement vector is three times the standard deviation along the principal component (the square root of the associated eigenvalue). **A:** Crambin (from the seed of crambe abyssinica), a small hydrophobic protein homologous to several plant toxins, consists of 46 residues and has 3 disulphide bridges (Hendrickson & Teeter, 1981). Crambin has a range of secondary structural elements: an antiparallel pair of α-helices (residues 7–17 and 23–30) and antiparallel β-sheet (strands at 1–3 and 32–34) and a type I β–turn (41–44). The figure shows the reference structure with backbone (N,Cα,C) atoms shown as thin lines and the three disulfide bridges (residues 3 to 40, 4 to 32, and 16 to 26) as thick grey lines. The molecule is oriented in the *XZ* plane of the principal geometric axes defined by the backbone atom coordinates (with *X* horizontal and and *Z* vertical). Positions of key residues and the N- and C-termini are indicated. **B, C, D:** Atomic positions corresponding to equidistant displacements along the positive direction of principal components 1 to 3. Six structures are drawn in each case; the reference structure is drawn in the heavier line. The maximum amplitude is three times the RMS displacement (square root of the eigenvalue) along each respective direction. The disulfide bridges are not drawn as they were not included in the analysis.

associated with distinct locations in conformational space and specific concerted atomic displacements. Further, for these data, an effective three-dimensional visualization of the conformational distribution can be obtained through the use of the components of the radius of gyration (see Table 2) as structural coordinates.

Classification of distinct conformational regions

Projection of the structures onto the first three principal components expresses their coordinates in the most significant conformational subspace (see Materials and methods). The resulting coordinates provide the basis for visualizing the distribution of the structures in the planes spanned by the first three principal components (Fig. 6A). In these plots, the error in projecting the structures into two dimensions (see Electronic supplementary material for a definition of the error function) varies from 0.51 to 0.78 Å (see Fig. 6A). Other (oblique) views of the principal conformational subspace were also examined. However, the views described here suffice to reveal the essential structure of the data. The different trajectories clearly sample different parts of conformational space. The overall distribution in the principal plane (pc2 vs. pc1) is roughly characterized by four broad regions (A to D, see Fig. 7, center panel) with the reference structure near the center.

To quantify the qualitative (visual) classification of different regions, the structures in the sample were subjected to a hierarchical cluster analysis (Gordon, 1981). The classification was based on the average RMSD between clusters, where the coordinates of each structure used in the distance calculation are those derived from the projection onto the first three principal components (i.e., the coordinates used in Fig. 6A). Inspection of the resulting clusters revealed a good correspondence with the visual classification

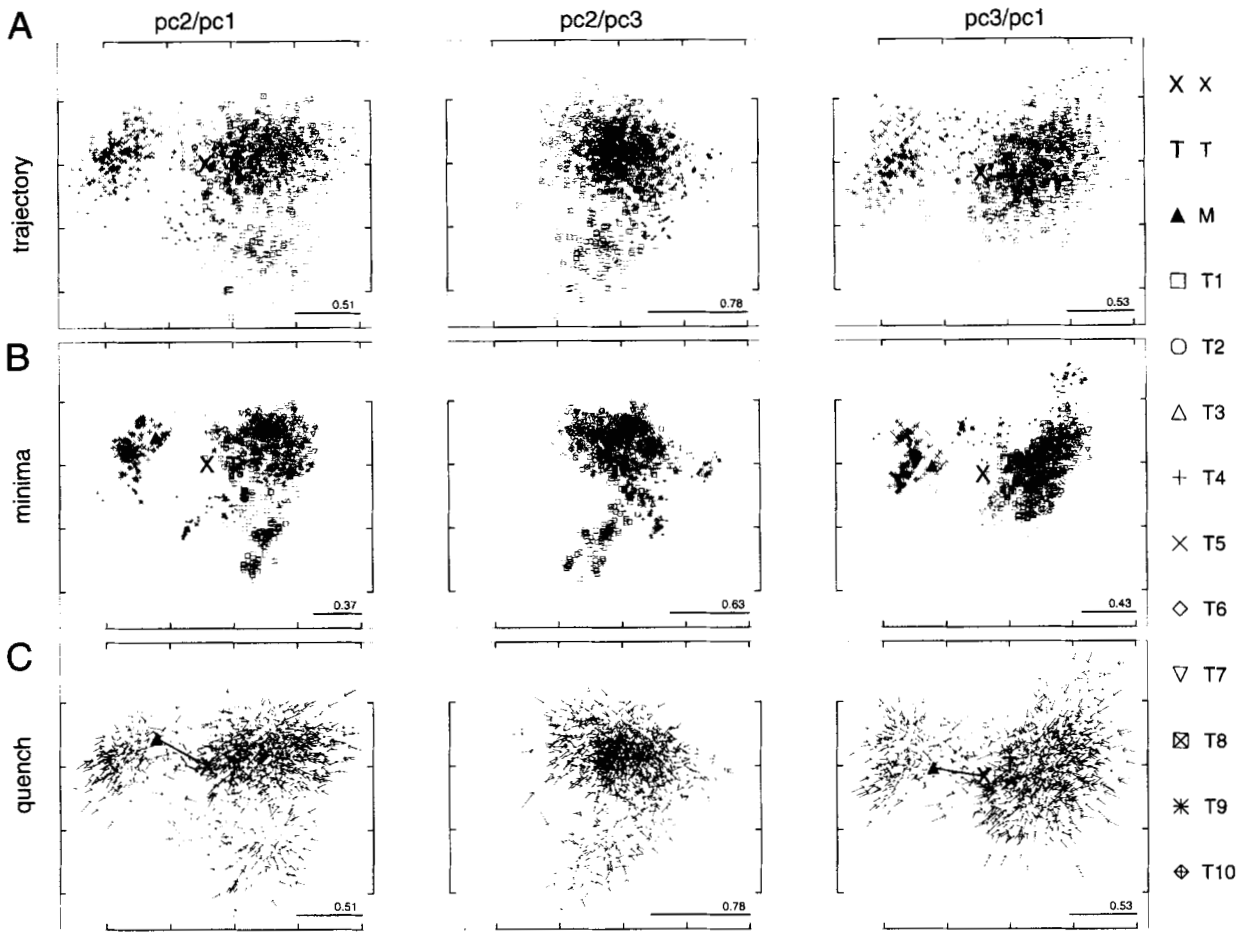

Fig. 6. Projection of transient MD structures onto the principal planes spanned by the three most significant principal components (pc). Row wise the panels represent the planes for pc2 vs. pc1, pc2 vs. pc3, and pc3 vs. pc1. The axes are scaled to represent the RMSD of the structures and tick marks are drawn at 0.5 Å intervals. The position of the global trajectory average structure (T) is at the origin (not shown). The legend for the symbols representing the structures is drawn on the right hand side of the plot; the reference structure, its minimum and the global average structure (which marks the origin) are indicated by the large, bold symbols X, M, and T, respectively. The error (in Å) of each projection is given in the lower right hand side of the plot. (For additional details see Electronic supplementary material). **A:** Transients from the ten 120 ps trajectories taken at 1 ps intervals, plus the reference structure. **B:** Local energy minima corresponding to the structures in A. The large filled black triangle represents the local energy minimum of the reference structure. **C:** The relationship of the transient structures to their corresponding local minima. Arrows are drawn along the vector connecting the transients to their corresponding minima. For clarity, the arrows are scaled by a factor of 0.2, so that the gross direction of movement upon minimization is visible from the position of the transient structure. The movement of the reference structure is indicated by a heavy line which connects it to its local minimum.

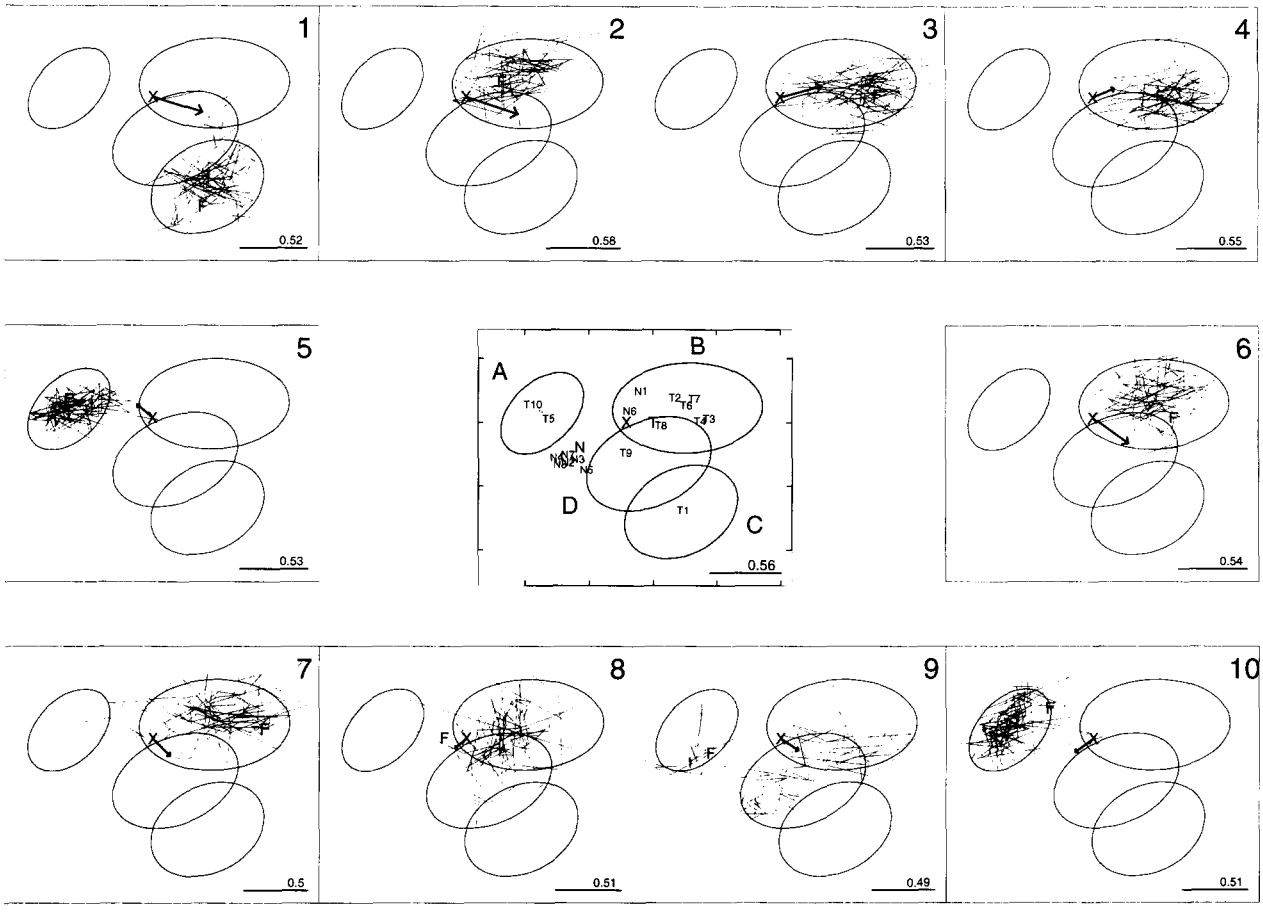

Fig. 7. Classification and extent of the conformational regions sampled by the trajectories. Center panel: Summary of the projections of the MD transients onto the principal conformational plane (pc2 vs. pc1 cf. Fig. 6A, panel 1). The extent of each of the four major conformational regions (A to D) classified by cluster analysis is indicated. Each ellipse represents ~98% of the conformational distribution within each region, if it is considered Normal (see Results). The projection of individual structures is indicated by the following symbols: X, reference structure; T1–T10, trajectory-average; N1–N8, NMR transients; N, NMR average; T, the global trajectory average, marks the origin. Note that the labels T5 and T10 are offset for clarity and their actual positions indicated by small dots. For each axis, ticks are placed at intervals of 0.5 Å. The error of this projection (in Å) is given in the lower right hand side of the plot. Outer panels: Time-evolution of the individual trajectories projected onto the principal plane spanned by the first two principal components (see also Fig. 6A, panel 1). The plot of the trajectory connects transient structures at 1 ps intervals. The initial (reference) and final structures are indicated by the labels X and F, respectively. The ellipses which circumscribe the major conformational regions (A to D) are shown (see center panel). The initial velocity vector ($t = 0$) for each trajectory is represented by an arrow in a heavy line; it has been scaled (arbitrarily) by a factor of 10 to make it easily visible. The error in each projection (in Å) is given in the lower right hand side of the plot.

(data not shown). The members of the four major clusters were assigned to the specific regions identified in the visual classification. In this way, the extent of the different conformational regions could be quantified and summarized. The ellipses in Figure 7, center panel, represent the extent of the regions (at ~98% probability assuming a Normal distribution) in the principal conformational plane.

The relationship of the reference, MD-average, and NMR structures when projected onto the principal planes is shown in Figure 7, center panel. The NMR structures lie between region A and D, except for NMR models N1 and N6 which lie within region B. This representation of the conformational sampling clearly reveals why the overall average structure is more similar to the reference structure than any of the average structures from the individual trajectories.

Evolution of individual trajectories

The contribution of individual trajectories to the overall distribution in the conformational space described above was examined by tracing their paths in the principal plane (Fig. 7). All trajectories are characterized by a fast initial movement away from the reference structure to a local conformational region which, in most cases, is sampled for the rest of the trajectory (Fig. 7). This initial displacement reflects the behavior seen in Figure 1. Several trajectories sample the same local region. Trajectories T5 and T10 sample region A (refer to Fig. 7, center panel), whereas trajectory T1 shows a fast initial movement to region C that is not sampled by any other trajectory; the remaining trajectories (T2, T3, T4, and T7) sample region B. Only trajectories T8 and T9, particularly the latter, sample more than one region to a significant extent.

For the first 50 ps, trajectory T9 resides in region B. During this period, the trajectory makes some extensive, yet rapid movements. For example, between 26 and 30 ps the transient structures develop an RMSD of 1.2 Å. This corresponds to a traversal of ~0.9 Å across region B along a direction that bisects the first two principal components (Fig. 7); the one dimensional character of this displacement is reflected by the 0.85 Å increase in RMSD from the reference structure during this period (Fig. 1). At 48 ps, the trajectory starts to move back toward the reference structure and between 50 and 51 ps makes a transition to another region (region D, see Fig. 7, center panel and Fig. 9). The trajectory samples region D for 45 ps before making another transition to region A at 96 ps, where it stays for the remainder of the simulation.

Four of the original 120 ps trajectories were extended: three to 1 ns (denoted T2x, T5x, and T9x) and one to 5 ns (T1x). Trajectories T1, T2, and T5 were chosen because they sample each of the three major conformational regions C, B, and A, respectively. Trajectory T9 was chosen because it was unique, in that it samples both regions A and B and a further region, D. The path of the extended trajectories was examined in the principal plane described above and is shown in Figure 8. Although this may not be the optimal projection for the extended trajectories, it is useful to retain the frame of reference used for exhibiting the multiple short trajectories. Trajectories T5x and T9x remain trapped in region A which they had already reached after 100 ps (see Fig. 7), for the entire 1 ns period of the extended trajectory (Fig. 8). Trajectory T1x undergoes a minor transition at ~800 ps to a nearby (~0.5 Å RMSD) local region, which it samples for the remainder of the 5 ns duration of the simulation. Exceptional behavior is seen in trajectory T2x, where a transition (at ~400 ps) marks the beginning of a lengthy excursion to a distant conformational region. It is interesting to note that this locally unfolded state of the crambin struc-

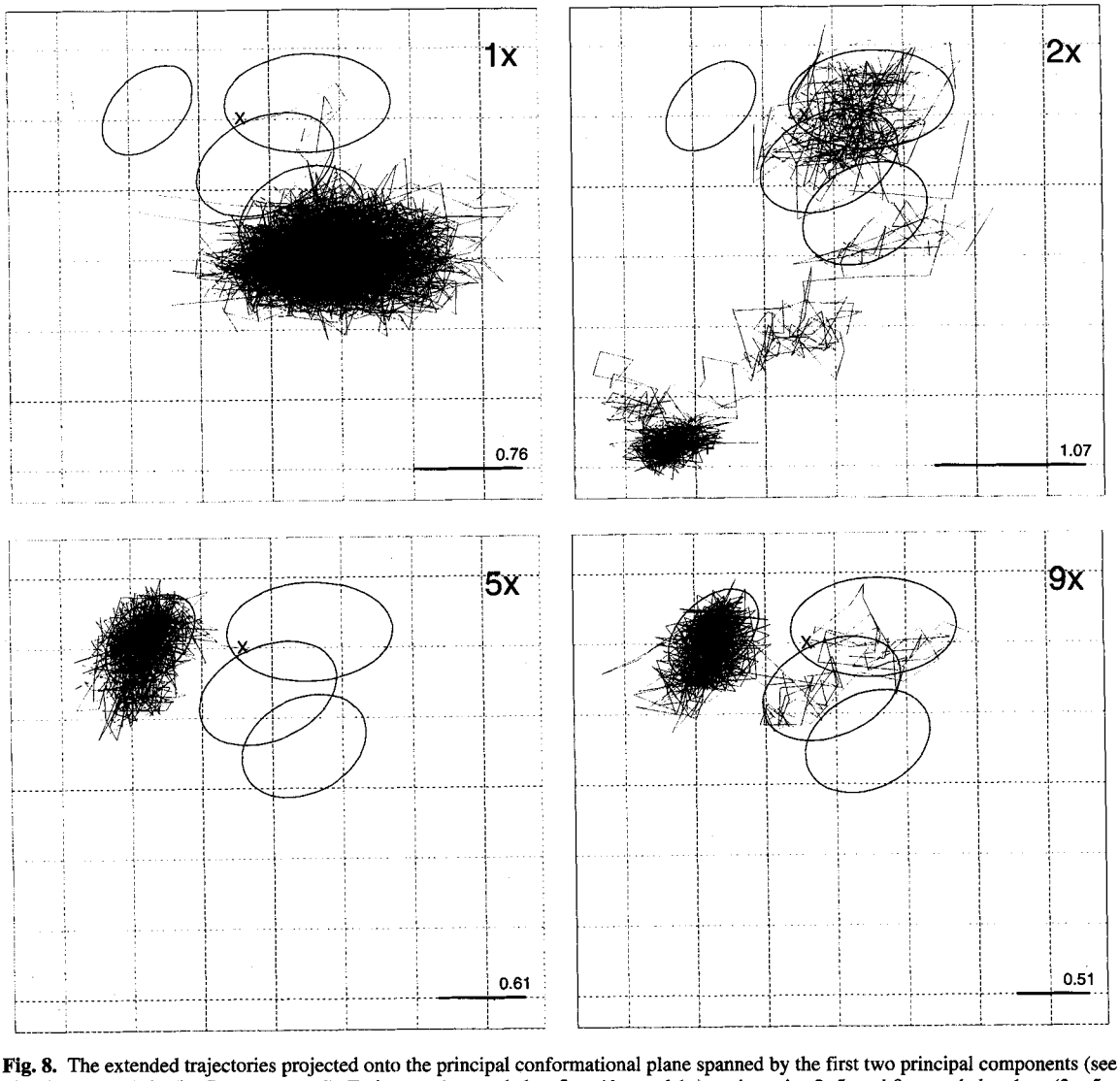

Fig. 8. The extended trajectories projected onto the principal conformational plane spanned by the first two principal components (see Fig. 6A, panel 1 & Fig. 7, center panel). Trajectory 1 extended to 5 ns (denoted 1x); trajectories 2, 5, and 9 extended to 1 ns (2x, 5x, 9x). The initial (reference) and final structures are indicated by the labels X and F, respectively. The ellipses which circumscribe the major conformational regions (**A–D**) are shown (see Fig. 7, center panel). The path taken by individual extended trajectories includes the initial 120 ps portion. The grid marks out areas of 0.25 $Å^2$. The error in each projection (in Å) is given in the lower right hand side of the plot.

ture (backbone RMSD of 3.9 Å from the reference) is stable for the last 400 ps of the trajectory. Further details on the behavior of this trajectory are given in Electronic supplementary material.

The underlying potential energy surface

Thus far, we have considered the extent of the conformational space that is sampled by the molecular dynamics trajectories at a temperature of 300 K. Information about the features of the underlying potential energy surface (PES) is provided by quenching the trajectories to remove their kinetic energy (Stillinger & Weber, 1984; Levitt, 1985; Elber & Karplus, 1987; Ohmine, 1995; Troyer & Cohen, 1995). This technique probes the minima on the adiabatic PES. For each of the ten 120 ps trajectories, the transient structures at 1 ps intervals were energy minimized (see Materials and methods). The distribution of the energies for the minima of the individual trajectories and for the different conformational regions is summarized in Tables 1 and 4, respectively. The overall range of the energies of the individual minima is 43 kcal mol^{-1}. The range of the average energies of the minima from the individual trajectories is 14.1 kcal mol^{-1}. The distribution of minima within each trajectory is smaller with standard deviations in the range 1.7 to 3.9 kcal mol^{-1}. Within each region (A to D) the distributions of the energies of minima are relatively similar to each other, though wider than that found in any individual trajectory, with standard deviations in the range 3.7 to 5.2 kcal mol^{-1}. The region (B) which is the most populated (Figs. 6, 8; Table 4) has the minima with the highest corresponding average potential energy; the least populated region (C) has the lowest energy minimum of the sample. The potential energy distributions of the minima are suggestive of a significant structure to the underlying PES (see below). Clearly, the relative energies of the minima are not sufficient to explain the distribution of the conformations found at 300 K in the present simulations.

The relationship of the thermalized structures to their local minima (and additionally, the reference structure and its local minimum) was examined by projection into the principal conformational subspace defined by the 10 individual 120 ps trajectories at 300 K (Figs. 6, 8A). The resulting view of the conformational space occupied by the minima reveals a significant condensation of their corresponding thermal distribution (Fig. 6B vs. Fig. 6A). The maximum extent of the conformational space encompassed by the MD transient structures, as measured by the backbone atom RMSD, is 2.5 Å vs. 2.0 Å for the corresponding minima. The average displacement of the transients upon minimization is 0.6 ± 0.1 Å with a range 0.3 to 1.1 Å. The location of the broad conformational regions (A to D) identified in the thermal distribution is maintained, but they are distinguished more easily. The reference structure is displaced significantly by minimization (backbone RMSD of 0.92 Å) to a position in region A (see Fig. 6C).

Different initial velocity assignments result in trajectories that explore regions other than the local minimum nearest to the reference structure. A simple projection of the initial velocity vector onto the principal conformational plane reveals an approximate correlation of its direction and the subsequent path of the trajectory (Fig. 7). These results suggest that, as the initial conformation (the reference structure) lies on a slope of the (multidimensional) potential energy surface (i.e., it is not an energy minimum), the impulse of the initial velocities are important in determining the direction of the descent of the trajectories toward lower energy regions. Clearly, the magnitude and direction of the initial velocities are sufficient to divert the trajectory from the path of steepest descent (gradient of the PES) which leads to region A. The location of the most populated region (region B, see Figs. 6A, 8A; Table 4) suggests that it is kinetically the most accessible from the reference structure (effectively incorporating the latter in its basin of attraction). There are no local minima in the immediate vicinity of the reference structure with confining barriers that are significant at 300 K; such minima are found ~0.6–0.9 Å away, in the major conformational regions (see Figs. 6A, 7, center panel).

Upon quenching, the transient structures have a tendency to move toward the center of the conformational region in which they are located, but they are trapped in the nearest local minimum (Fig. 6B). Some fragmentation of the major conformational regions is evident upon minimization (for example, there is a split in region C, see Fig. 6B), which implies that the broad potential well defining the regions has a rugged "floor," whose local barriers are easily surmounted at 300 K, but become evident upon minimization. More information is available from the quenching process than just the positions of the local minima. The paths connecting the transient and minimized conformers give an indication of the gradient of the PES (Fig. 6C). When projected onto the conformational subspace under consideration, such a map reveals the approximate slope of the PES. Structures on the periphery of the conformational regions generally show larger displacements upon minimization. This reflects the steeper slope of the walls of the potential energy basins that serve as the underlying attractors for the trajectories. Some transient structures that are spatially close, diverge upon quenching, dropping into different local minima

Table 4. *Summary energetics of the minima corresponding to the four major regions (A to D) identified in the room temperature conformational distributions*

Region[a]	Number of structures[b]	Proportion of total	V_{min}[c]	V_{max}[c]	$\langle V \rangle$[c,d]	σV[d]
A	254	0.21	3.6	29.3	11.3	4.9
B	744	0.62	3.6	43.4	18.0	5.2
C	98	0.08	0.0	21.8	12.7	3.9
D	104	0.09	7.4	25.5	15.2	3.7

[a]For extent of regions see Figure 7, center panel.
[b]Total of 1,200 minima.
[c]V = potential energy in kcal mol^{-1} offset from the minimum energy of the sample of −561.5 kcal mol^{-1} found in region C.
[d]$\langle\ \rangle$ and σ denote mean and standard deviation, respectively.

(Elber & Karplus, 1987). For example, transient structures that lie between regions A and B (Fig. 6A, panel 1) minimize into either of these regions (see Fig. 6C, panel 1) so that the intermediate region is relatively unpopulated. The pattern of divergence of the transients reveals the approximate position of the potential energy barriers (ridges) that separate the major conformational regions. An example can be found in the transient structures of trajectory T9 in the part of region D that is sampled en route from region C to region A (see Fig. 7). The ridges that separate this (sub-)region from regions C and B are well defined; there is a pattern of displacement vectors from the transients to the minima that serve to circumscribe the boundary of the local minima (Fig. 6C, panel 1).

In this analysis, we have performed a classification and clustering of minima based on the structural similarity of the 300 K conformational distributions. It should be noted that the close proximity of minima in conformational space (based on the RMSD metric) does not necessarily imply that they are kinetically accessible (Becker & Karplus, 1997). To establish the connectivity of the minima requires an extensive knowledge of the topology of the PES. A more detailed study of the topography and topology of the PES sampled in these trajectories, including a quantitative estimation of the barriers separating individual minima, will be presented separately (J.D. Evanseck, L.S.D. Caves, and M. Karplus, in prep.).

Discussion

Multiple molecular dynamics simulations of crambin were performed by starting from a single conformation and using different random distributions for the initial velocity assignment. The coordinate space evolution of the different simulations was examined through low-dimensional projections (via principal components analysis) of the conformational space sampled by the trajectories. The different initial conditions resulted in very different trajectories. In the sample of ten room temperature trajectories, the analysis of the distributions sampled in the principal conformational plane (that containing the maximum variance) showed that four major regions of conformational space play an important role. It is likely that these four regions represent only a portion, albeit an important one, of the total conformational space accessible to room temperature vacuum trajectories of crambin.

The picture that emerges from the analysis is that most trajectories rapidly move away from the reference position to sample a single, localized region of conformational space. Different trajectories can sample the same local region but follow distinct paths. Only two of the 120 ps trajectories (T8 and T9) and one of the extended trajectories (T2x, 1 ns duration) sample more than one of the major local regions for significant time periods. Thus, extending the trajectories (at the same temperature) for a limited time period does not guarantee more extensive sampling, as a trajectory can remain trapped in a particular region of conformational space for a long time. However, as a simulation is extended in time it does become more probable for a transition to another region of conformational space to occur; this may be a region that is not accessible from the starting structure on shorter timescales. This is the case during the extension of trajectory T2x, but it must be contrasted with the extended trajectory T1x (5 ns duration), which is relatively unproductive in terms of additional sampling of conformational space over that achieved in the initial 120 ps period. This suggests that the conformations generated in a single trajectory (even if very long) are unlikely to be representative of the full range of conformations thermally accessible to the system; i.e., such trajectories are effectively non-ergodic.

Our results confirm that structural properties computed from molecular dynamics simulations of standard length (100 ps) or even considerably longer (1 ns) show a significant dependence on the initial conditions (Elofsson & Nilsson, 1993; Smith et al., 1993; Auffinger et al., 1995). Analysis of atomic positional distribution functions indicate that short timescale (high frequency) motions (<8 ps) are very similar for the different trajectories. Longer timescale (~100 ps) motions show greater variation among the trajectories, reflecting the difference in the character of the potential energy surface in the neighborhood of the various minima. Thus, care has to be used in interpreting the convergence of a property. It may indicate only that equilibrium has been reached in a local sense (Straub et al., 1994).

The average structures resulting from the independent trajectories can be very different. This raises questions about the use of single trajectories to evaluate force-fields, the role of the environment or the simulation protocol (York et al., 1993; Daura et al., 1996; Schiffer & van Gunsteren, 1996; Yasri et al., 1996). The present study demonstrates that it can be inappropriate to assess the quality of a potential energy function, for example, based on properties of the average structure and (to a lesser extent) of the atomic fluctuations of a single molecular dynamic simulation. Considerable care has to be taken to determine whether or not the conformational sampling is sufficient (Brooks, 1995). This is a difficult problem. Its relevance in the context of using MD simulations to interpret X-ray scattering data has been noted by Caspar (1995). Moreover, the need for more complete sampling has to be evaluated for each particular property, since some are less sensitive to structural changes than others. There are clear implications from the present analysis for the computation of thermodynamic quantities by use of free energy simulations (Hodel et al., 1993), and for the comparison of simulations with data from solution NMR experiments (de Groot et al., 1996).

This study provides insights for increasing the effectiveness of sampling within the current framework of protein molecular dynamics simulations. It is recommended that many shorter trajectories be performed versus a single (or few) longer trajectories. A single trajectory is likely to become "bound" to a single local basin of attraction and repeatedly (and therefore redundantly) explore a small region of conformational space. This is useful for estimating the entropy associated with the basin (Levy et al., 1984), but not for exploring the accessible conformational space. This point has been noted previously (Levy et al., 1982; Elofsson & Nilsson, 1993; Smith et al., 1993) but has not been examined in detail for proteins. In simulations of tRNA, multiple short molecular dynamics simulations have been used recently to improve sampling (Auffinger et al., 1995; Auffinger & Westhof, 1996, 1997).

The role of multiple (short) trajectories in relation to other methods of enhanced conformational sampling in biomolecular systems (Turner et al., 1993; Amadei et al., 1996; Berne & Straub, 1997 for a review) will need to be investigated. It is anticipated that multiple initial conditions for trajectories will form part of any methodology for efficient conformational sampling. For example, simulations based on the "essential dynamics" method have been shown to yield a factor of two greater sampling (for each of the principal components) over standard methods on the ~1 ns timescale (de Groot et al., 1996). In the present study, the use of ten different 100 ps trajectories also increases the sampling relative to that of a single trajectory by about a factor of two; over 1 ns periods the enhancement could be greater, but this will require further studies.

An illustration of the effects of insufficient sampling is found in the comparison of trajectory-average structures. Compared to the average structures from single trajectories, a significant reduction in the RMSD with the reference structure is found when data from multiple independent trajectories are averaged together to produce a single "global" average structure, representative of the sampling achieved by all the simulations. This result is due entirely to more effective sampling of conformational space and was not achieved by adjustments of the potential function or simulation protocol. The essential factor is that each trajectory samples its own region of space, which can be significantly different from that sampled by any other trajectories and that the various regions are displaced in different directions from the reference structure.

Most MD simulations utilize as initial conditions structures in the neighborhood of those determined by X-ray or NMR experiments, which are time and space averages and in addition, for macromolecules, are generally underdetermined by the experimental data (Ringe & Petsko, 1985; Petsko, 1996), although recent technical advances have greatly increased the resolution attainable for macromolecular crystal structures (Dauter et al., 1995). Thus, most structures are models representing a compromise between the experimental data and restraints reflecting prior knowledge about the system (Brunger & Karplus, 1991). In addition, the environment of the experiment (e.g., crystal contacts, solvent or salt content) may not be well reproduced or even explicitly considered in the model used in simulations. Thus, an experimentally determined structure is not necessarily at a minimum; it can be situated on a slope or peak of the empirical potential energy surface. When allowed to evolve from this point (e.g., in molecular dynamics or energy minimization), the structure moves down into one or another of the surrounding low energy regions. Which region is determined in the present case by the initial velocity assignments. Thus, the initial conditions had the effect, through the use of multiple trajectories, of increasing the conformational sampling. The use of partially minimized structures, or even distorted structures obtained for example by high temperature dynamics (Bruccoleri & Karplus, 1990), from which to initiate trajectories may be useful in extending the sampling. By contrast, if the structure is at a minimum, the sensitivity to the initial velocities is expected to be diminished due to the confining effects of the local curvature of the potential surface. This issue, which needs further investigation, has implications for the protocol of room temperature protein molecular dynamics simulations, which often start from well minimized structure. It also raises the question as to the best method or combination of methods to achieve reliable converged sampling of conformational space.

In the present work, the trajectories were quenched to examine the structure of the underlying potential energy surface (Stillinger & Weber, 1984; Elber & Karplus, 1987; Troyer & Cohen, 1995). The results provide insight into the origin of the conformational distributions seen in the room temperature trajectories, in terms of the location of the minima which serve as the basins of attraction. When the kinetic energy is removed, there is a significant contraction in the volume of conformational space sampled and the distinct conformational regions become more evident. The paths connecting the transients and their corresponding minima provide an indication of the approximate location and slope of the barriers which serve to delineate the major conformational regions. The clustering of conformations upon minimization also reveals additional structure within each of the major conformational regions, reflecting the presence of low barriers, which are not significant at room temperature. A more detailed analysis of the topological and topographical features of the potential energy surface (Czerminski & Elber, 1990; Becker & Karplus, 1997), including the barriers between minima, will be presented separately (J.D. Evanseck, L.S.D. Caves, and M. Karplus, in prep.).

One complication that arises in using multiple trajectories for analysis concerns the weight given to the individual trajectories for determining average values of properties. Molecular dynamics yields Boltzmann sampling of the available conformational space by individual trajectories. To combine data from different trajectories we have given equal weight to trajectories of equal length, corresponding to the assumption of equal *a priori* probabilities for trajectories with the same total energy (the microcanonical ensemble). Since the trajectories are sampling different regions of the potential energy surface it is possible to divide the configurational space into basins (Becker & Karplus, 1997) with a certain energy and configurational entropy (Stillinger & Weber, 1984). From the present results, the four basins identified in the room temperature conformational distributions have underlying minima that vary in potential energy by 43 kcal mol^{-1}. Future work will address the estimation of the relative free energy of the different minima sampled in this study, using the configurational entropy derived from the harmonic or quasiharmonic partition functions (McQuarrie, 1973; Levy et al., 1984). The weighted histogram analysis method (Kumar et al., 1992) can also be used to combine trajectory data to yield the free energies and, thus, the relative probabilities of different conformational states (Boczko & Brooks, 1995; Kumar et al., 1995).

The results from the present study on crambin can be compared with vacuum simulations of another small protein, BPTI. Given that the definition of conformational regions is similar, the residence times in distinct conformational regions can be compared. For trajectories which do sample more than one conformational region, the period of ~40–50 ps, agrees well with behavior seen in MD simulations of BPTI (Levitt, 1985; Hayward et al., 1994). The ~1 ns timescale suggested by the Monte Carlo simulations of BPTI (Noguti & Go, 1989) is in line, with the fact that, once a minimum has been found, the majority of vacuum trajectories remain there on the 100 ps to 1 ns timescale. A comparison of vacuum and solution simulations of BPTI (Hayward et al., 1993) indicated that the solute-solvent interactions introduce many more minima on the potential energy surface. This results in the sampling of many smaller conformational clusters that are typically occupied for periods of ~20 ps. An MD study of crambin in aqueous solution (Garcia, 1995) reports a mean residence time in distinct conformational regions of 40–70 ps, in agreement with our results obtained in vacuum. Apparently, the vacuum potential energy surface represents the overall envelope of the solution free energy surface, but solvation results in more, but shallower minima that involve solvent rearrangements, which renders the motion more diffusive (Hayward et al., 1993).

Biomolecular dynamics simulations provide models of complex, many-body dynamical systems with a rugged, multiple minimum potential energy surface (Frauenfelder et al., 1991). Central to the connection between averages over the simulations and phase space averages is the ergodic hypothesis (Tolman, 1938). For effective ergodicity, the simulations must correctly sample the allowed regions of phase space of the system during the trajectory (Mountain & Thirumalai, 1989). Simple convergence of a property with time is not sufficient, because parts of the phase space may not be sampled adequately (Straub et al., 1994). This problem is

exacerbated by the complexity of biomolecular systems with their range of properties on widely different spatial and temporal scales (Frauenfelder & Wolynes, 1994; Frauenfelder, 1995). The tools of dynamical systems theory (Hilborn, 1994) offer a way to increase our understanding of the phase space behavior of these systems and how best to deal with the analysis, as described here. For example, in this study we have considered aspects of the phase space behavior of a conservative (Hamiltonian) system. The simple use of a thermostat, common in biomolecular dynamics simulations, would result in a dissipative system, which in principal, has very different phase space behavior. The influence of factors that control the degree of dissipation (such as the time constants coupling a system to its heat bath) on the phase space behavior of biomolecules, has yet to be investigated. Conformational subspace projection (via PCA) is an important tool for such analyses, analogous to the role of phase portraits in characterizing the complex phase space behavior of non-linear dynamical systems (Caves et al., 1991; Clarage et al., 1996).

The standard theory of complex dynamical systems (Ott, 1993) applied to biomolecular dynamics has revealed that proteins dynamics is chaotic (K. Kuczera & M. Karplus, unpubl. obs.; Zhou & Wang, 1996). In examining chaotic behavior, typically small perturbations to the initial conditions (such as the variation in the initial position or velocity by say, 1 part in 10^7) are made and the subsequent exponential divergence in the trajectories is analyzed (Ruelle, 1991; Hilborn, 1994). However, many molecular simulation studies involve much larger changes to the initial conditions and thus, the existence of such chaotic behavior does not, by itself, explain the effect of initial conditions that have been reported in biomolecular dynamics simulations in this and other studies (Elofsson & Nilsson, 1993; Smith et al., 1993; Auffinger et al., 1995). Auffinger et al. argue that the degree of divergence of trajectories is proportional to errors in the force field or in the simulation protocol which accumulate at each time step and that these inaccuracies may become significant after several hundred picoseconds (Auffinger & Westhof, 1996). However, it has been shown that molecular dynamics trajectories diverge exponentially with characteristic timescales of 2 to 100 ps even with stable integrators (Allen & Tildesley, 1987; Watanabe & Karplus, 1995; Zhou & Wang, 1996). Thus, the divergence of trajectories is an inherent property of biomolecular dynamics simulations and a manifestation of the underlying multiminimum potential energy surface.

It is clear that much has been learned about the nature and origin of the internal dynamics of proteins. To extend our knowledge, new tools and more powerful computers are required. As the field of biomolecular dynamics simulations comes of age (Berendsen, 1996), improved methods that are being developed (Turner et al., 1993; Derremaux & Schlick, 1995) are likely to play an increasing role. Investigations of the barriers, as well as the minima on the potential energy surface, are needed to provide a more complete description of the origin of the conformational sampling described in this study. An analysis of the barriers on the energy landscape in the region of the native state of crambin will be presented subsequently (J.D. Evanseck, L.S.D. Caves, & M. Karplus, in prep.).

Materials and methods

Atomic coordinates

Atomic coordinates of crambin were obtained from the Brookhaven Protein Data Bank (PDB, Bernstein et al., 1977), corresponding to a 1.5 Å resolution crystal structure (Hendrickson & Teeter, 1981, PDB entry 1crn). For the *B*-factors we use a set from a 0.945 Å room temperature structure of crambin (M. Teeter, pers. comm.); the differences from the 1.5 Å set are minor, having ~3% higher magnitude and similar spatial trends (correlation coefficient of 0.96) for the RMS fluctuations of the backbone atoms. A structure of crambin in aqueous solution determined by NMR (Bonvin et al., 1993) was used for comparison: the 8 individual model structures (PDB entry 1ccm) and the minimized average structure (PDB entry 1ccn) are denoted N1 to N8 and N respectively, in this study.

Potential energy function

The energy, minimization, and molecular dynamics calculations were performed with the program CHARMM (version 22g5) (Brooks et al., 1983). The polar-hydrogen representation was used, defined by the CHARMM parameter and residue topology files PARAM19 and TOPH19, respectively (Neria et al., 1996). Details of the model are given in Electronic supplementary material.

Simulation protocol

Energy minimization

To ensure that the crambin crystal structure was free of strongly repulsive non-bonded contacts and other geometric distortions inconsistent with the potential energy function, mild energy minimization was performed: 80 steps of steepest descent followed by 20 steps of the conjugate gradients method resulted in a reduction of the internal forces (as measured by the RMS magnitude of the gradient vector, the GRMS) from 18.8 to 0.9 kcal mol^{-1} $Å^{-1}$. The minimization resulted in only small deviations from the crystal structure coordinates with an RMSD (see Methods of analysis) of 0.13 and 0.18 Å for backbone (N,Cα,C) atoms and all (non-hydrogen) atoms, respectively. This partially minimized structure is referred to as the reference structure and is denoted X. Minimization was also performed on transient structures from the molecular dynamics trajectories. For details, see Electronic supplementary material.

Molecular dynamics

In the molecular dynamics simulations, integration of the Newtonian equations of motion was performed with the Leapfrog algorithm (Allen & Tildesley, 1987) and a step size of 0.001 picoseconds (ps). Ten 120 ps MD simulations (denoted T1 to T10) were performed. All trajectories were started from the same reference conformation (X) and differed only in the initial assignment of the atomic velocities. For details, see Electronic supplementary material.

Methods of analysis

Structural properties

The molecular structural properties considered in this study are based on the atomic Cartesian position vectors ($\mathbf{r}_i = X_i, Y_i, Z_i$), for an atom i). A spatial reference frame was established by superposition of each structure on the backbone (N,Cα,C) atoms of the reference structure (X) (Kabsch, 1978). This operation removes coordinate differences due to overall translation and rotation of the

molecule. For each individual trajectory, an average structure was generated by averaging the Cartesian coordinates from the last 100 ps at a resolution of 200 fs (500 frames). A single average structure representative of all ten 120 ps trajectories was generated in an analogous way (5000 frames). To examine changes in molecular shape, the radius of gyration (R_g) was decomposed into three orthogonal spatial components, corresponding to the principal geometric axes of the molecule in the standard reference frame (see above) and are denoted R_X, R_Y, and, R_Z, respectively, in order of decreasing magnitude. For further details on the structural analysis methods, see Electronic supplementary material.

Principal components analysis (PCA)

To understand the relationship of the structures considered in this study it is useful to examine their distribution in the conformational space of the molecule. Because the conformational space is of high dimension (for the internal Cartesian coordinates of a molecule with N atoms, the dimension is 3N-6), its direct examination and visualization is impracticable. Thus, this study focuses on important features that can be determined by restricting the analysis to low dimensional aspects of the high dimensional distributions. Many methods of multivariate analysis have been described for reducing dimensionality (Alt, 1990; Everitt & Dunn, 1991). Principal components analysis (PCA, Jackson, 1991) focuses on the variance of the distribution and essentially proceeds by computing the second moment of the sample multivariate distribution and its principal axes (the principal components) (Green & Carrol, 1976). By examining the distribution of data along the principal components, the analysis can be restricted to a new set of variables (that are linear combinations of the original variables) along which the sample variance is largest. In favourable cases, a significant fraction of the variance of the multidimensional distribution can be accounted for in a few dimensions. Full discussions of PCA and its relationship to other methods of dimensionality reduction are available elsewhere (Krzanowski, 1988; Jackson, 1991). Some additional discussion is given in Electronic supplementary material. Data manipulation, PCA, and visualization were performed with the statistical programming system S (Becker et al., 1988).

Electronic supplementary material

Supplementary tables (2): Table 3B (RMSD of 10-20ps and 110-120ps structures) and Table 5 (Properties of the atomic fluctuations) in tab-delimited ASCII text files TAB3B.TXT and TAB5.TXT, respectively. Supplementary figures (2): Schematic representations of crambin structure and Details of trajectory 2x are in GIF format bitmap files 1CRN.GIF and T2X.GIF, with legend text in the ASCII text files 1CRN.TXT and T2X.TXT, respectively. Supplementary details of the Materials and methods are available in Microsoft Word format file METHOD.DOC. These files are available in the Electronic Appendix to Protein Science.

Acknowledgments

Discussions with Themis Lazaridis, Oren Becker, Arnaud Blondel, Alex MacKerrell, Chandra Verma, and Rod Hubbard are gratefully acknowledged. Martha Teeter is thanked for providing unpublished data on the temperature factors of crambin. LSDC thanks the BBSRC for support of the computing infrastructure in the Protein Structure Research Group at York. This work was supported in part by a grant from the National Institutes of Health.

References

Allen MP, Tildesley DJ. 1987. *Computer simulation of liquids*. Oxford: Oxford University Press.

Alt M. 1990. *Exploring Hyperspace*. New York: McGraw-Hill.

Amadei A, Linssen ABM, de Groot BL, van Aalten DMF, Berendsen HJC. 1996. An efficient method for sampling the essential subspace of proteins. *J Biomol Struct Dynam 13*:615–625.

Auffinger P, Louise-May S, Westhof E. 1995. Multiple molecular-dynamics simulations of the anticodon loop of tRNA(ASP) in aqueous-solution with counterions. *J Am Chem Soc 117*:6720–6726.

Auffinger P, Westhof E. 1996. H-bond stability in the tRNA(ASP) anticodon hairpin: 3 ns of multiple molecular dynamics simulations. *Biophys J 71*:940–954.

Auffinger P, Westhof E. 1997. RNA Hydration: Three ns of multiple molecular dynamics simulations of solvated tRNA(ASP) anticodon hairpin. *J Mol Biol 269*:326–341.

Barrat JL, Klein ML. 1991. Molecular dynamics simulations of supercooled liquids near the glass transition. *Ann Rev Phys Chem 42*:23–53.

Becker OM, Karplus M. 1997. The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics. *J Chem Phys 106*:1495–1517.

Becker RA, Chambers JM, Wilks AR. 1988. *The new S language. A programming environment for data analysis and graphics*. Pacific Grove, California: Wadsworth and Brooks/Cole.

Berendsen HJC. 1996. Biomolecular dynamics comes of age. *Science 271*:954–955.

Berne BJ, Straub JE. 1997. Novel methods of sampling phase space in the simulation of biological systems. *Curr Opin Struct Biol 7*:181–189.

Bernstein FC, Koetzle TF, Williams GJB, Meyer EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The protein data bank: A computer-based archival file for macromolecular structures. *J Mol Biol 122*:532–542.

Boczko EM, Brooks CL. 1995. First principles calculation of the folding free energy of a 3-helix bundle protein. *Science 269*:393-396.

Bonvin AMJJ, Rullmann JAC, Lamerichs RMJN, Boelens R, Kaptein R. 1993. Ensemble iterative relaxation matrix approach—A new NMR refinement protocol applied to the solution structure of crambin. *Proteins Struct Funct Genet 15*:385–400.

Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. 1983. CHARMM—A program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem 4*:187–217.

Brooks CL III. 1995. Methodological advances in molecular dynamics simulations of biological systems. *Curr Opin Struct Biol 5*:211–215.

Brooks CL III, Karplus M, Pettitt BM. 1988. *Proteins: A theoretical perspective of dynamics, structure and thermodynamics*. New York: Wiley Interscience.

Bruccoleri RE, Karplus M. 1990. Conformational sampling using high-temperature molecular-dynamics. *Biopolymers 29*:1847–1862.

Brunger AT, Karplus M. 1991. Molecular dynamics simulations with experimental restraints. *Acc Chem Res 24*:54–61.

Caspar D. 1995. Problems in simulating macromolecular movements. *Structure 3*:327–329.

Caves LSD, Nguyen DT, Hubbard RE. 1991. Conformational variability of insulin: A molecular dynamics analysis. In: Goodfellow JM, ed. *Molecular dynamics: Applications in molecular biology*. London: Macmillan Press. pp 27–68.

Cheatham TE, Kollman PA. 1996. Observation of the A-DNA to B-DNA transition during unrestrained molecular-dynamics in aqueous-solution. *J Mol Biol 259*:434–444.

Clarage JB, Romo T, Andrews BK, Pettitt BM, Phillips GN. 1995. A sampling problem in molecular-dynamics simulations of macromolecules. *Proc Natl Acad Sci USA 92*:3288–3292.

Clarage JB, Romo TD, Andrews BK, Pettitt BM, Phillips GN. 1996. Using phase-portraits to visualize sub-states, sampling, and chaos in protein dynamics. *Biophys J 70*:MPMC1–MPMC1.

Cooper A. 1976. Thermodynamic fluctuations in protein molecules. *Proc Natl Acad Sci USA 73*:2740–2741.

Czerminski R, Elber R. 1990. Reaction path study of conformational transitions in flexible systems—Applications to peptides. *J Chem Phys 92*:5580–5601.

Daggett V, Levitt M. 1993. Realistic simulations of native-protein dynamics in solution and beyond. *Ann Rev Biophys Biomol Struc 22*:353–380.

Daura X, Oliva B, Querol E, Aviles FX, Tapia O. 1996. On the sensitivity of MD trajectories to changes in water-protein interaction parameters: The potato carboxypeptidase inhibitor in water asa test case for the GROMOS force field. *Proteins Struct Funct Genet 25*:89–103.

Dauter Z, Lamzin VS, Wilson KS. 1995. Proteins at atomic resolution. *Curr Opin Struct Biol 5*:784–790.

de Groot BL, Amadei A, Scheek RM, van Nuland NAJ, Berendsen HJC. 1996. An extended sampling of the configuration space of HPr from *E. coli*. *Proteins Struct Funct Genet 26*:314–322.

Derremaux P, Schlick T. 1995. Long timestep dynamics of peptides by the dynamics driver approach. *Proteins Struct Funct Genet 21*:282–302.

Durup J. 1991. Protein molecular dynamics constrained to slow modes: Theoretical approach based on a hierarchy of modes with a set of holonomic constraints—The methods and its tests on citrate synthase. *J Phys Chem 95*:1817–1829.

Elber R, Karplus M. 1987. Multiple conformational states of proteins—A molecular-dynamics analysis of myoglobin. *Science 235*:318–321.

Elber R, Karplus M. 1990. Enhanced sampling in molecular-dynamics—Use of the time-dependent hartree approximation for a simulation of carbon-monoxide diffusion through myoglobin. *J Am Chem Soc 112*:9161–9175.

Elofsson A, Nilsson L. 1993. How consistent are molecular dynamics simulations? Comparing structure and dynamics in reduced and oxidized *E. coli* thioredoxin. *J Mol Biol 233*:766–780.

Everitt BS, Dunn G. 1991. *Applied multivariate data analysis*. London, UK: Edward Arnold.

Feller W. 1971. *Introduction to probability theory and its applications*. New York: Wiley.

Frauenfelder H. 1995. Complexity in proteins. *Nature Struct Biol 2*:821–823.

Frauenfelder H, Sligar SG, Wolynes PG. 1991. The energy landscape and motions of proteins. *Science 254*:1598–1603.

Frauenfelder H, Wolynes PG. 1994. Biomolecules: Where the physics of complexity and simplicity meet. *Physics Today 47*:58–64.

Garcia AE. 1992. Large-amplitude nonlinear motions in proteins. *Phys Rev Lett 68*:2696–2699.

Garcia AE. 1995. Multi-basin dynamics of a protein in aqueous solution. In: Peyrard M, ed. *Nonlinear excitations in biomolecules*. New York: Springer. pp 191–206.

Gerstein M, Lesk AM, Chothia C. 1994. Structural mechanisms for domain movements in proteins. *Biochemistry 33*:6739–6749.

Gordon AD. 1981. *Classification*. London: Chapman and Hall.

Green PE, Carrol JD. 1976. *Mathematical tools for applied multivariate data analysis*. London: Academic Press.

Hayward S, Kitao A, Go N. 1994. Harmonic and anharmonic aspects in the dynamics of BPTI—A normal mode analysis and principal component analysis. *Protein Sci 3*:936–943.

Hayward S, Kitao A, Hirata F, Go N. 1993. Effect of solvent on collective motions in globular protein. *J Mol Biol 234*:1207–1217.

Hendrickson WA, Teeter MM. 1981. Structure of the hydrophobic protein crambin determined directly from the anomalous scattering of sulfur. *Nature 290*:107–113.

Hilborn RC. 1994. *Chaos and nonlinear dynamics*. Oxford: Oxford University Press.

Hodel A, Simonson T, Fox RO, Brunger AT. 1993. Conformational substates and uncertainty in macromolecular free energy calculations. *J Phys Chem 97*:3409–3417.

Hunenberger PH, Mark AE, van Gunsteren WF. 1995. Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular-dynamics simulations. *J Mol Biol 252*:492–503.

Ichiye T, Karplus M. 1987. Anisotropy and anharmonicity of atomic fluctuations in proteins: Analysis of a molecular-dynamics simulation. *Proteins Struct Funct Genet 2*:236–259.

Ichiye T, Karplus M. 1988. Anisotropy and anharmonicity of atomic fluctuations in proteins: Implications for X-ray-analysis. *Biochemistry 27*:3487–3497.

Jackson EJ. 1991. *A user's guide to principal components*. London: John Wiley & Sons.

Jorgensen WL, Tirado-Rives J. 1988. The OPLS potential functions for proteins—Energy minimizations for crystals of cyclic-peptides and crambin. *J Am Chem Soc 110*:1666–1671.

Kabsch W. 1978. A discussion of the solution for the best rotation to relate two sets of vectors. *Acta Cryst A34*:827–828.

Karplus M, Petsko GA. 1990. Molecular dynamics simulations in biology. *Nature 347*:631–639.

Krzanowski WJ. 1988. *Principles of multivariate analysis: A user's perspective*. Oxford: Oxford University Press.

Kuczera K, Kuriyan J, Karplus M. 1990. Temperature-dependence of the structure and dynamics of myoglobin—A simulation approach. *J Mol Biol 213*:351–373.

Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM. 1992. The weighted histogram analysis method for free-energy calculations on biomolecules. 1. The method. *J Comput Chem 13*:1011–1021.

Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA. 1995. Multidimensional free energy calculations using the weighted histogram analysis method. *J Comput Chem 16*:1339–1350.

Levitt M. 1985. Molecular dynamics of native protein: II Analysis and nature of motion. *J Mol Biol 168*:621–657.

Levitt M, Sharon R. 1988. Accurate simulation of protein dynamics in solution. *Proc Natl Acad Sci USA 85*:7557–7562.

Levy RM, Karplus M, Kushick J, Perahia D. 1984. Evaluation of the configurational entropy for proteins—Application to molecular-dynamics simulations of an alpha-helix. *Macromolecules 17*:1370–1374.

Levy RM, Perahia D, Karplus M. 1982. Molecular-dynamics of an alpha-helical polypeptide: Temperature-dependence and deviation from harmonic behavior. *Proc Natl Acad Sci USA 79*:1346–1350.

Lii JH, Gallion S, Bender C, Wikstrom H, Allinger NL, Flurchick KM, Teeter MM. 1989. Molecular mechanics (MM2) calculations on peptides and on the protein crambin using the CYBER-205. *J Comput Chem 10*:503–513.

Louise-May S, Auffinger P, Westhof E. 1996. Calculations of nucleic acid conformations. *Curr Opin Struct Biol 6*:289–298.

Mao B. 1991. Mass-weighted molecular dynamics simulations and conformational analysis of polypeptide. *Biophys J 60*:611–622.

McQuarrie DA. 1973. *Statistical thermodynamics*. New York: Harper & Row.

Mountain RD, Thirumalai D. 1989. Measures of effective ergodic convergence in liquids. *J Phys Chem 93*:6975–6979.

Neria E, Fischer S, Karplus M. 1996. Simulation of activation free-energies in molecular-systems. *J Chem Phys 105*:1902–1921.

Noguti T, Go N. 1989. Structural basis of hierarchical multiple substates of a protein—Introduction. *Proteins Struct Funct Genet 5*:97–103.

Ohmine I. 1995. Liquid water dynamics: Collective motions, fluctuation and relaxation. *J Phys Chem 99*:6767–6776.

Ornstein RCJ. 1990. Using molecular dynamics simulations of crambin to evaluate the suitability of different continuum dielectric and hydrogen-atom models for protein simulations. *J Biomol Struct Dynam 7*:1019–1041.

Ott E. 1993. *Chaos in dynamical systems*. Cambridge: Cambridge University Press.

Perutz MF. 1989. Mechanisms of cooperativity and allosteric regulation in proteins. *Quart Rev Biophys 22*:139–236.

Petsko GA. 1996. Not just your average structures. *Nature Struct Biol 3*:565–566.

Post CB, Brooks BR, Karplus M, Dobson CM, Artymiuk PJ, Cheetham JC, Phillips DC. 1986. Molecular-dynamics simulations of native and substrate-bound lysozyme—A study of the average structures and atomic fluctuations. *J Mol Biol 190*:455–479.

Ringe D, Petsko GA. 1985. Mapping protein dynamics by X-ray diffraction. *Prog Biophys Mol Biol 45*:197–235.

Ruelle D. 1991. *Chance and chaos*. Princeton, New Jersey: Princeton University Press.

Schiffer CA, van Gunsteren WF. 1996. Structural stability of disulphide mutants of basic pancreatic trypsin inhibitor: A molecular dynamics study. *Proteins 26*:66–71.

Smith PE, Pettit BM, Karplus M. 1993. Stochastic dynamics simulations of the alanine dipeptide using a solvent modified potential energy surface. *J Phys Chem 97*:6907–6913.

Steinbach PJ, Ansari A, Berendzen J, Braunstein D, Chu K, Cowen BR, Ehrenstein D, Frauenfelder H, Johnson JB, Lamb DC, Luck S, Mourant JR, Nienhaus GU, Ormos P, Phillipp R, Xie A, Young RD. 1991. Ligand binding to heme proteins: Connection between dynamics and function. *Biochemistry 30*:3988–4001.

Stillinger FH, Weber TA. 1984. Packing structures and transitions in liquids and solids. *Science 225*:983–989.

Straub JE, Rashkin AB, Thirumalai D. 1994. Dynamics in rugged energy landscapes with applications to the S-peptide and ribonuclease A. *J Am Chem Soc 116*:2049–2063.

Straub JE, Thirumalai D. 1993. Theoretical probes of conformational fluctuations in S-peptide and RNase A/3′-UMP enzyme product complex. *Proteins Struct Funct Genet 15*:360–373.

Teeter MM, Case DA. 1990. Harmonic and quasiharmonic descriptions of crambin. *J Phys Chem 94*:8091–8097.

Tidor B, Karplus M. 1993. The contribution of cross-links to protein stability—A normal mode analysis of the configurational entropy of the native-state. *Proteins Struct Funct Genet 15*:71–79.

Tolman R. 1938. *The principles of statistical mechanics*. Oxford: Oxford University Press.

Troyer JM, Cohen FE. 1995. Protein conformational landscapes: Energy minimization and clustering of a long molecular dynamics trajectory. *Proteins Struct Funct Genet 23*:97–110.

Turner JD, Weiner PK, Chun HM, Lupi V, Gallion S, Singh UC. 1993. Variable reduction techniques applied to molecular dynamics simulations. In: van Gunsteren W, ed. *Computer simulation of biomolecular systems*. Leideu: ESCOM. pp 535–555.

van Gunsteren WF, Berendsen HJC. 1977. Algorithms for macromolecular dynamics and constrained dynamics. *Mol Phys 34*:1311–1327.

van Gunsteren WF, Hunenberger PH, Mark AE, Smith PE, Tironi IG. 1995. Computer simulation of protein motion. *Comput Phys Commun 91*:305–319.

van Schaik RC, van Gunsteren WF, Berendsen HJC. 1992. Conformational search by potential-energy annealing—Algorithm and application to cyclosporin A. *J Comput-Aided Mol Design 6*:97–112.

Watanabe M, Karplus M. 1995. Simulations of macromolecules by multiple time-step methods. *J Phys Chem 99*:5680–5697.

Whitlow M, Teeter MM. 1986. An empirical-examination of potential-energy minimization using the well-determined structure of the protein crambin. *J Am Chem Soc 108*:7163–7172.

Yasri A, Chiche L, Haiech J, Grassy G. 1996. Rational choice of molecular dynamics simulation parameters through the use of three-dimensional autocorrelation method: Application to calmodulin flexibility study. *Protein Eng 9*:959–976.

York DM, Darden TA, Pedersen LG. 1993. The effect of long-range interactions in the simulations of macromolecular crystals: A comparison of the Ewald and truncated list methods. *J Chem Phys 99*:8345–8348.

Zhou H, Wang L. 1996. Chaos in biomolecular dynamics. *J Phys Chem 100*:8101–8105.