MOLECULAR DYNAMICS OF MYOGLOBIN AT 298°K

Results from a 300-ps Computer Simulation

RONALD M. LEVY,* ROBERT P. SHERIDAN,* JOE W. KEEPERS,* G. S. DUBEY,*
S. SWAMINATHAN,‡ AND MARTIN KARPLUS‡
*Department of Chemistry, Rutgers University, New Brunswick, New Jersey 08903; and ‡Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138

ABSTRACT We have carried out a very long (300 ps) molecular dynamics simulation of the protein myoglobin. This trajectory is approximately three times longer than the longest previous molecular dynamics simulation of a protein, and ten times longer than protein simulations of comparable size (1,423 atoms in our model). Here we report results from this long simulation concerning the average structure, the mean square fluctuations of atoms about the average structure, and the nuclear magnetic resonance order parameters for various groups in myoglobin. The results demonstrate that the average coordinates change very slowly during the simulation. The relative atomic mobilities are well described by the simulation. For both the mean square atomic fluctuations and the order parameters, however, there are significant quantitative differences when values calculated using shorter portions of the trajectory are compared with results obtained for the entire 300-ps simulation. The implications of this result for obtaining converged properties from protein molecular dynamics simulations for comparison with experiment are discussed.

INTRODUCTION

Computer simulations of protein dynamics constitute the most detailed theoretical approach available for studying their internal motions (1, 2). The simulations provide a large number of data which can be compared with laboratory experiments designed to detect protein motions. Several publications that emphasize the relationship between molecular dynamics simulations of proteins and experimental measurements (e.g., nuclear magnetic resonance [NMR], fluorescence depolarization, x-ray scattering, and vibrational spectroscopy) have appeared (3–13). The intensive nature of these calculations necessitated that the first simulations be limited to short trajectories of small proteins. Several recent developments make it possible to carry out more complex simulations of larger proteins for longer simulation periods in the presence of solvent. These developments include the greatly increased availability of super-minicomputers dedicated to small numbers of users, the use of array processors for these machines, and the increasing access to supercomputers. We have carried out a 300-ps simulation of myoglobin at 298°K in vacuo on a VAX 11/780 computer (Digital Equipment Corp., Marlboro, MA). This trajectory is approximately three times longer than the longest previous molecular dynamics simulation of a protein, and ten times longer than protein simulations of comparable size (1,423 atoms in our model). The length of the simulation permits us to carefully investigate the convergence of molecular properties. Here we report results concerning the average structure, the mean square fluctuations of atoms about the average structure, and order parameters for various groups. The mean square fluctuations calculated from the simulation can be related to experimental Debye-Waller factors (temperature factors) and the order parameters to NMR relaxation times. We explore the extent to which these properties are converged on a 300-ps time scale.

Address all correspondence to Dr. R. M. Levy.

METHODS

The initial coordinates for the myoglobin dynamics studies were obtained from G. Petsko (Massachusetts Institute of Technology, Cambridge, MA); they correspond to a metmyoglobin crystal structure refined to 1.5-Å resolution at 250°K. Since we are using these trajectories to study the barriers to rotation of methyl groups in proteins (Levy, R. M., R. Sheridan, J. Keepers, G. S. Dubey, and M. Karplus, manuscript submitted for publication), we have included the hydrogen coordinates corresponding to each of the methyls of the leucine and isoleucine side chains in the simulation. The hydrogen coordinates were generated from the x-ray structure in the staggered conformation. The system considered contains 1,423 atoms (1,217 protein heavy atoms + 162 methyl hydrogens + 43 heme atoms + 1 water). The empirical energy function used with the exception of the explicit hydrogen parameters was the one contained in the program package CHARMM, which has been described previously (14). The force constants and equilibrium geometries for the explicit hydrogen atoms are: K_b = 314 kcal/mol-Å^2, b_o = 1.09 Å, K_θ = 3.75 kcal/mol-rad^2, θ_o = 1.911 rad, K_ϕ = 1 kcal/mol; the small torsional barrier for the methyl group was used to increase the rotation rate (Levy, R. M., R. Sheridan, J. Keepers, G. S. Dubey, and M. Karplus, manuscript submitted for publication). The partial charges for the methyl carbons and attached hydrogens were −0.089 electrostatic units (e.u.) and +0.025 e.u., respectively. The partial atomic charges were reduced according to their proximity to the surface of the molecule (11) and in

addition, an $\epsilon = 1/R$ dielectric was used (14). A nonbonded interaction cutoff of 9.0 Å produced ~100,000 nonbonded atom pairs; a hydrogen bond cutoff of 4.5 Å and 45° produced ~160 hydrogen bonds. The protein was energy minimized with 200 steepest descent steps after construction of the hydrogen coordinates. The stepsize used to integrate the equations of motion with the Verlet algorithm was 0.001 ps. The bond constraint algorithm SHAKE was used to fix the length of the methyl CH bonds. Starting at 50°K, the protein was slowly heated to 298°K in increments of 25°K, with each heating followed by 250 integration steps. After heating, the velocities of each of the atoms were rerandomized corresponding to a Boltzmann distribution at 298°K every 200 steps for a period of 2 ps. An additional 6 ps of simulation was carried out without any intervention. Following the 10.5 ps of heating and equilibration, the 300-ps portion of the trajectory, which was saved for analysis, was begun. Coordinates and velocities were saved every 50 steps (0.05 ps). 1 ps simulation of the 1,423 atom system required 4 central processor unit (CPU) hours on a VAX 11/780 computer (Digital Equipment Corp.) equipped with floating point accelerator.

RESULTS

Average Structures

In Table I we compare the root mean square (rms) differences between various structures averaged over different portions of the 300-ps simulation. The average deviations between each of the compared structures in Table I have also been calculated and are ~15% smaller than the rms values reported in Table I. We have used both the x-ray structure and a structure obtained from the initial part of the trajectory following 10.5 ps of heating and equilibration to compare with structures obtained by averaging coordinates over 100-ps intervals of the simulation. The rms difference between the x-ray structure and the dynamically averaged structure is 2.03 Å for all atoms after the first picosecond, and increases slightly to 2.12 Å over the first 100 ps. When the x-ray structure is compared with the dynamically averaged coordinates obtained from the entire 300-ps simulation, the rms difference for all atoms is 2.42 Å. For the backbone atoms in the α-helices, the rms differences between the x-ray and dynamically averaged structures are considerably smaller; after 1, 100, 200, and 300 ps they are, respectively, 1.44, 1.45, 1.58 and 1.66 Å. For the backbone atoms, the rms deviations are <1.0 Å between structures averaged over successive 100-ps portions of the trajectory (0.95 Å for 0–100 ps compared with 100–200 ps; 0.77 Å 100–200 ps compared with 200–300 ps). The rate at which the coordinates drift from the x-ray structure decreases somewhat during the simulation; i.e., the rms difference between structures averaged over the first two 25-ps segments is 0.83 Å for all atoms, whereas it is 0.69 Å for the last two 25-ps segments. It is apparent from the results that most of the deviation from the x-ray values arises during the 10.5 ps of heating and equilibration. After that the average molecular dynamics coordinates change much more slowly, e.g., the deviation between the coordinates obtained from the first picosecond of the trajectory (the 000–001 coordinates in Table I and the average coordinates for the entire 300-ps simulation is 1.37 Å for the backbone atoms and 1.65 Å for all the atoms).

The accessible surface area and radius of gyration corresponding to different average structures generated during the trajectory are listed in Table II. As in previous vacuum simulations (11), there is a contraction of both the

TABLE I
ROOT MEAN SQUARE DIFFERENCE BETWEEN STRUCTURES AVERAGED OVER DIFFERENT PORTIONS OF THE 300-ps SIMULATION*

Compared structures	All atoms	Backbone	Backbone (α-helices)
X ray vs. (000–100)‡	2.12	1.81	1.45
(100–200)	2.58	2.23	1.18
(200–300)	2.82	2.45	1.95
(000–200)	2.29	1.98	1.58
(000–300)	2.41	2.09	1.66
X ray vs. (000–001)	2.03	1.76	1.44
(000–001) vs. (000–100)	1.19	0.94	0.85
(100–200)	1.84	1.51	1.38
(200–300)	2.25	1.91	1.74
(000–200)	1.44	1.16	1.07
(000–300)	1.65	1.37	1.26
(000–100) vs. (100–200)	1.01	0.95	0.82
(100–200) vs. (200–300)	0.94	0.77	0.67
(000–100) vs. (200–300)	1.67	1.43	1.23
(000–025) vs. (025–050)	0.83	0.56	0.53
(250–275) vs. (275–300)	0.69	0.54	0.40

*The rms difference is $\{[\langle \vec{r}_1 \rangle - \langle \vec{r}_2 \rangle]^2\}^{1/2}$ where $\vec{r}_i$ is the vector of coordinates of all the atoms, the time average is denoted by $\langle\ \rangle$, and $\{\ \}$ indicates an average over all atoms. The average deviations $\{|\langle r_2 \rangle - \langle r_2 \rangle|\}$ between each of the compared structures have also been calculated and are about 15% smaller than the rms values reported in the table.
‡Coordinates averaged over the interval indicated. The first picosecond of the simulation saved for analysis was preceded by 10.5 ps of heating and equilibrium.

TABLE II
ACCESSIBLE SURFACE AREA AND RADIUS OF GYRATION FOR CONFORMATIONS GENERATED DURING THE 300-ps SIMULATION

Structure	Accessible surface area‡		Radius of gyration§
	All atoms	Main chain	
X-ray	2,275	172	15.04
(000–001)*	1,889	143	14.15
(000–100)	1,876	131	14.07
(100–200)	1,841	132	13.94
(200–300)	1,830	126	13.88
(000–300)	1,851	134	14.00

*Coordinates averaged over the interval indicated.
‡Accessible surface area (Å^2) calculated according to the algorithm of Lee and Richards, reference 35. The probe radius was 1.4 Å.
§The radius of gyration (Å), calculated from the backbone N, C$^\alpha$, C, 0 coordinates. $R_g = \sqrt{\Sigma_i m_i |(\vec{r}_i - \vec{r}^{\,cm}_i)|^2 / \Sigma_i m_i}$.

accessible surface area and radius of gyration. Part of this reduction is due to the folding of surface side chains, which in the crystal make intermolecular contacts. The accessible surface area of the (0–300 ps) average structure is 17% smaller than the x-ray structure and the radius of gyration calculated using only the backbone atoms has decreased by 7% from 15.04 to 14.0 Å. Over 95% of the contraction occurs during the heating and equilibration, although both the accessible surface area and radius of gyration continue to decrease slightly during the course of the simulation. Solvent exerts an average attractive external force field that counters the attractive force between the atoms on the protein surface and those in the core of the protein, so that with the present potential parameters, solvent must be included in the simulation to maintain an atomic density corresponding to the x-ray structure (15). For vacuum molecular dynamics simulations of protein, a small increase in the van der Waals radii for the atoms has been found to correct for density changes due to neglect of solvent (Brooks, B. R., and M. Karplus, unpublished results).

We have examined representative structures from the 300-ps myoglobin simulation on a graphics system (PS 300; Evans & Sutherland, Salt Lake City, UT). Stereo pictures of the myoglobin backbone for the x-ray structure and after 100 and 300 ps are shown in Fig 1. The contraction corresponding to the decrease in the radius of gyration is apparent. Myoglobin is over 80% α-helical, consisting of eight α-helices that pack against each other over extended regions. The overall folding pattern, and the α-helical structure are maintained throughout the entire 300-ps simulation. From examination of pictures of the trajectory on the PS 300, it appears that over tens of picoseconds the α-helices move as semi-rigid bodies with respect to each other. A detailed analysis of the motions and packing changes of the α-helices is underway.

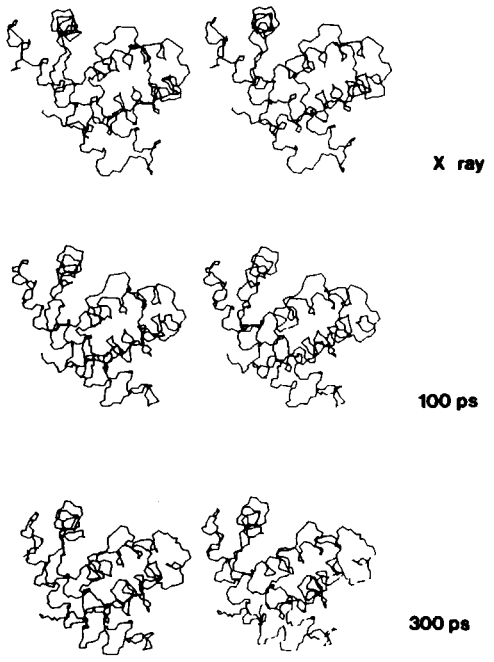

FIGURE 1 C^{α} backbone structures of myoglobin from the x-ray coordinates and after 100 and 300 ps of the simulation.

With respect to deviation from the x-ray structure the results of the current 300-ps simulation are of similar magnitude as earlier (and much shorter) simulations of pancreatic trypsin inhibitor (PTI) (550 heavy atoms) in vacuo (16–19) and simulations of cytochrome C (11). A simulation of the PTI unit cell including four PTI molecules lasting 20 ps has been reported (20). The average structure obtained from the PTI crystal simulation is somewhat closer to the x-ray coordinates than the present myoglobin results, although a direct comparison is difficult because of differing equilibration periods. The rate of drift of the PTI coordinates from the x-ray values during the portions of the simulations used for analysis appears somewhat greater than the myoglobin results. The present results suggest that the drift of coordinates from an average structure for a protein with one thousand or more atoms is very slow so that long crystal simulations will be required to determine the effect of crystal forces on average structures.

Mean Square Fluctuations

The earliest attempt to relate the results of molecular dynamics simulations of proteins with experiment was through comparison of positional fluctuations of atoms calculated from the simulation with values derived from experimental x-ray temperature factors $\langle(\Delta R)^2\rangle = (3B/8\pi^2)$, where B is the x-ray temperature factor and $\langle(\Delta R)^2\rangle$ is the mean square displacement for an atom. Although temperature factors have long been used to obtain information concerning atomic motion for small molecules (21), it is only recently that attempts have been made to correlate atomic mobility with temperature factors in proteins (22–25). The highly anisotropic and anharmonic nature of atomic motions in proteins makes it more difficult to interpret temperature factors quantitatively. Methods based on molecular dynamics simulations are being developed to aid in the interpretation of temperature factors extracted from macromolecular refinement programs. (26; Kuriyan, S., M. Karplus, R. M. Levy, and G. Petsko, manuscript submitted for publication; Westhof, E., and R. M. Levy, manuscript submitted for publication). We are currently using the myoglobin simulations to test the validity of the approximations concerning atomic motions contained in protein refinement programs and their effects on the calculated temperature factors (Kuriyan, S., M. Karplus, R. M. Levy, and G. Petsko, manuscript submitted for publication).

The mean square displacements for the backbone atoms averaged over different portions of the simulation are presented in Figs. 2 and 3 and the rms displacements listed in Table III. The average values for the rms displacements are the same (0.83 Å) during the first and second 100-ps

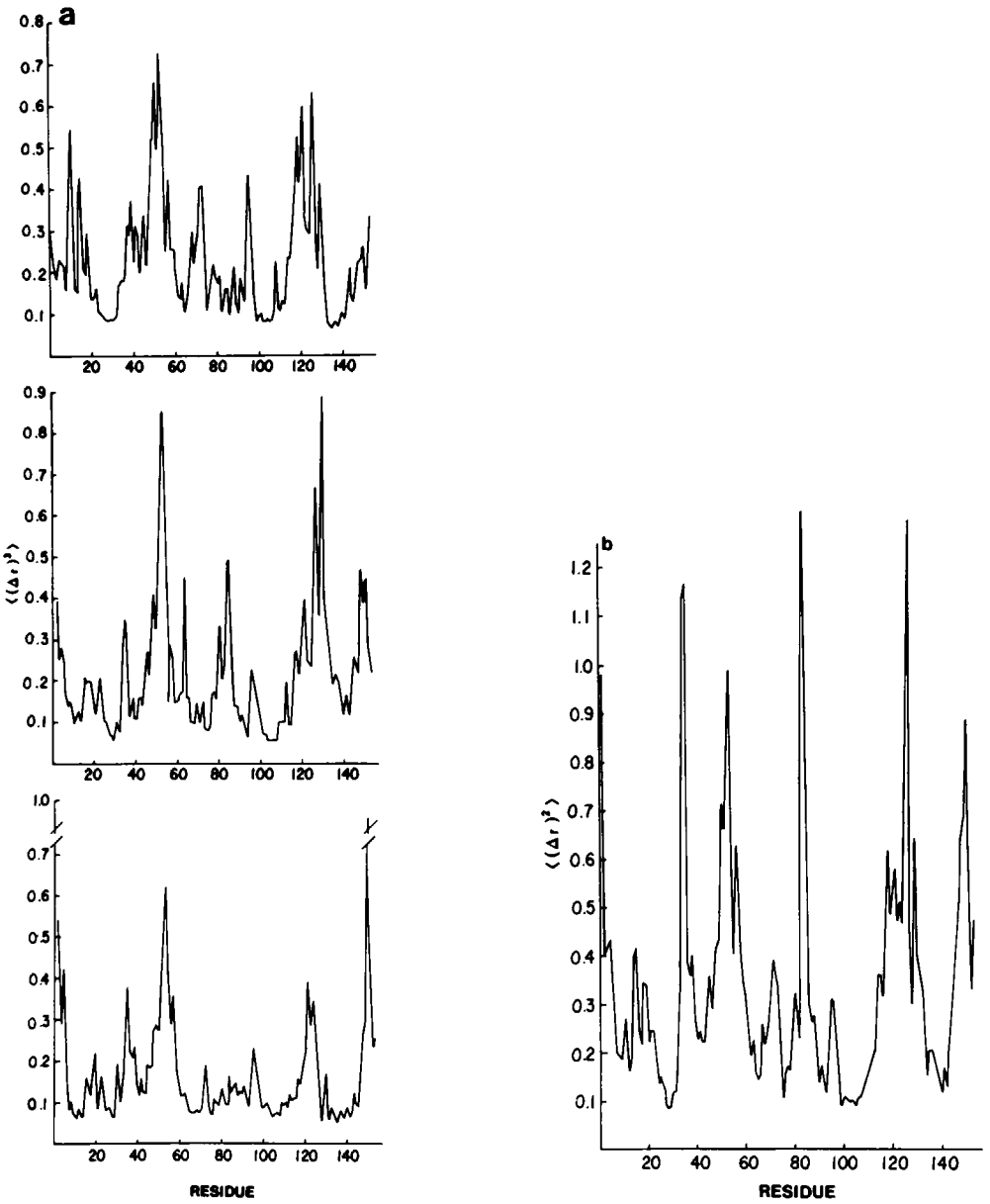

FIGURE 2 (*a*) Mean square fluctuations $\langle(\Delta r)^2\rangle = 1/3\ [\langle(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2\rangle]$ of the main chain (C^α, N, C) atoms averaged per residue: *top*, fluctuations calculated using the 0–100 ps portion of the trajectory; *middle*, fluctuations calculated using the 100–200 ps portion of the trajectory; *bottom*, fluctuations calculated using the 200–300 ps portion of the trajectory. (*b*) Same as *a* except fluctuations calculated using the entire 300-ps trajectory.

intervals and actually decrease for the final 100 ps of the trajectory (0.71 Å). When the rms displacements are calculated as a single average over the 300-ps simulation, the displacements are 20% greater than those obtained by averaging over the individual 100-ps intervals. The displacements calculated for the three 100-ps intervals show significant differences. However, they display several common features; including most notably the larger displacements in the regions of residues 40–60 and 120–130. Experimental temperature factors contain contributions from both fluctuations of atoms about local minima and from any more extensive structural oscillations that occur. The larger motions are not easily incorporated in algorithms used for protein refinement so that modeling the effect of these larger structural changes on derived temperature factors is difficult. To approximately separate the contributions to temperature factors of local atomic fluctuations and larger structural changes, we have calculated the rms displacements of the atoms during successive 25-ps intervals and then averaged the values obtained from the twelve 25-ps intervals. The (25 ps) block averaged mean square displacements are compared with the experimental estimates extracted from the temperature factors in Fig. 3.

The block averaged displacements are in better agreement with the values calculated from the temperature

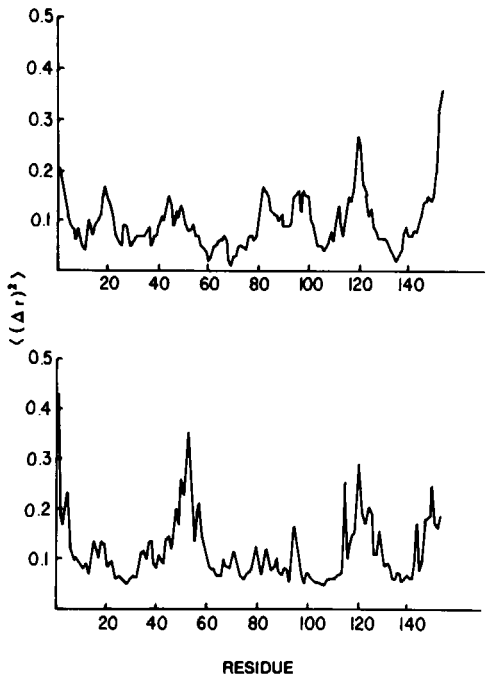

FIGURE 3 *Top*, experimental mean square fluctuations obtained from temperature factors at 300°K (reference 22). *Bottom*, average of the mean square fluctuations calculated by the results from individual 25-ps portions of the stimulation (e.g., the fluctuations obtained for 0–25 ps, 25–50 ps, . . . 275–300 ps were averaged).

factors than are values calculated as a single average over the entire 300 ps both with respect to the magnitudes and the more detailed shapes of the plots. The largest displacements occur at the amino and carboxyl termini of myoglobin and in the loop region between helices; the large values between residues 45–55 and 115–125 correspond to the CD and GH corners. Some regions of the protein that appear to be very mobile when the displacements are calculated with respect to the average coordinates for the entire 300-ps simulation have much smaller displacements when the results are block averaged over the twelve 25-ps intervals; examples are residues 34–36 and 83–86.

The effect of the block averaging is to separate the fluctuations obtained from the full dynamics into two terms $[\langle(\Delta R)^2\rangle = \langle(\Delta R)^2\rangle_1 + \langle(\Delta R)^2\rangle_2]$ in which $\langle(\Delta R)^2\rangle_1$ is the result obtained by calculating the mean square fluctuations with respect to mean coordinates in 25-ps blocks and then averaging, and $\langle(\Delta R)^2\rangle_2$ is the mean square fluctuation of the set of twelve mean coordinates corresponding to each of the 25-ps blocks. Thus, $\langle(\Delta R)^2\rangle_1$ represents an average of higher frequency fluctuations about a mean structure and $\langle(\Delta R)^2\rangle_2$ corresponds to fluctuations among mean structures. As an example, the backbone mean square fluctuations for the residues of the hydrophobic pocket of myoglobin have been decomposed in this way in Table IV. These residues do not undergo particularly large mean square displacements. Nevertheless, it is apparent that the range of mean structures, corresponding to $\langle(\Delta R)^2\rangle_2$ that are populated during the 300-ps trajectory make a contribution equal to or greater than that of the higher frequency local fluctuations $\langle(\Delta R)^2\rangle_1$ to the total mean square displacements. A similar result was observed at 300°K in the analysis by molecular dynamics of the temperature dependence of a model α-helical polypeptide (27).

TABLE III
ROOT MEAN SQUARE ATOMIC DISPLACEMENTS*‡ AVERAGED OVER DIFFERENT PORTIONS OF THE 300 ps MYOGLOBIN SIMULATION

Time interval	$\langle(\Delta R)^2\rangle^{1/2}$
ps	
0–100	0.829
100–200	0.829
200–300	0.707
0–100	0.829
0–200	0.960
0–300	1.071
0–300§	0.590
Experiment	0.650

*Average of rms displacements of backbone N, C^α, C' atoms, Å.
‡$\langle(\Delta R)^2\rangle^{1/2} = \{(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2\}^{1/2}$.
§Average obtained by calculating the rms displacements in intervals of 25 ps and then averaging the values for the 12 intervals.

NMR Order Parameters

Since NMR relaxation in proteins is determined by dynamics on the picosecond to nanosecond time scale, experimental NMR relaxation parameters can provide important information concerning picosecond motions. Time correlation functions required for determining NMR parameters may be calculated directly from molecular dynamics trajectories. Protein trajectories have been used to study motional models used in the analysis of ^{13}C NMR T_1, T_2, and Nuclear Overhauser Enhancements (NOEs) and ^1H NOEs, and as an aid in the interpretation of experiments (4–6, 8). A direct comparison between the results of a 96-ps molecular dynamics simulation (17) of PTI and an experimental ^{13}C NMR relaxation study (28) of this protein has been reported (6). Order parameters, which are measures of the extent of angular motion of the

TABLE IV
DECOMPOSITION OF BACKBONE MEAN SQUARE FLUCTUATIONS* FOR RESIDUES OF THE HYDROPHOBIC POCKET

Residue	$\langle(\Delta R)^2\rangle_1$	$\langle(\Delta R)^2\rangle_2$
Leu B13	0.19	0.25
Phe CD1	0.27	0.40
His E7	0.20	0.29
Val E11	0.26	0.40
Ile G8	0.15	0.18

*All values in angstroms squared.

bonds, were calculated from a 96-ps trajectory and compared with values extracted from the relaxation data. Although the relative flexibility of the residues studied was reasonably well described by the simulation, the theoretical order parameters were systematically larger than the experimental ones, indicating that there is less motional averaging in the 96-ps simulation than detected in experiment. It was suggested that this behavior occurred because the length of the trajectory was too short to statistically sample all accessible orientations. We have used the present 300-ps myoglobin trajectory to reexamine this question.

For ^{13}C NMR of protonated carbons with fixed bond lengths, the contribution of internal protein motions to NMR relaxation is determined by the angular correlation function $C(t)$:

$$C(t) = \sum_{a=-2}^{+2} \left\langle Y_a^2[\theta(t)\phi(t)] \cdot Y_a^2[\theta(0)\phi(0)] \right\rangle, \qquad (1)$$

where Y_a^2 are spherical harmonics and (θ, ϕ) specifies the orientation of a C^{13}-H bond in a macromolecular fixed frame. Because of the restricted nature of the motion in the protein interior, the internal correlation function (Eq. 1) usually does not decay to zero. Instead a plateau value is reached for which the internal correlation function is equal to the equilibrium orientation distribution obtained from the entire run (4–6, 29).

$$\frac{4\pi}{5} \sum_a |\left\langle Y_a^2(\Omega) \right\rangle|^2 = S^2. \qquad (2)$$

The quantity S defined by Eq. 2 is the order parameter describing the restricted motion of the ^{13}C-H vector; for a rigid system $S^2 = 1$.

The most detailed ^{13}C NMR relaxation studies in proteins including PTI and myoglobin have been reported for the ^{13}C-H bonds in methyl groups (28, 30–32). Since in previous simulations the hydrogens of all methyl groups were incorporated into the heavy atoms to which they are bound, it was not possible to calculate S^2 for the C-H vectors directly. Instead, the product approximation was made for a methyl ^{13}C-H vector (6, 29).

$$S^2 = S_{\text{axis}}^2 \cdot S_{\text{rotation}}^2 = S_{\text{axis}}^2 (1/2\{3\cos^2\theta - 1\})^2, \qquad (3)$$

where S_{axis} is the order parameter for the bond joining a methyl carbon and its nearest neighbor heavy atom (for example, the C^γ—$C^{\delta 1}$ bond of leucine, Leu), S_{rotation}^2 is the order parameter for the methyl motion about the methyl rotation axis, and θ is the methyl carbon C—C—H bond angle. Eq. 3 is valid if the motions of the methyl symmetry axis are uncoupled from the motion of the methyl protons about this axis. Since the hydrogens on the methyls of all the Leu and isoLeu residues were explicitly included in the myoglobin simulation, we can check the validity of Eq. 3. Table V lists the order parameters for selected methyl C—H bonds of the 18 Leus in myoglobin calculated explicitly from the trajectory and in the product approximation. It can be seen that the product approximation for the order parameters of methyl C-H bond vectors is excellent.

TABLE V
ORDER PARAMETERS* FOR SELECTED LEU C^1-H BONDS CALCULATED EXACTLY AND IN THE PRODUCT APPROXIMATION

Residue	S_{exact}^2	$S_{\text{axis}}^2 \cdot S_{\text{rotation}}^2$
Leu 2	0.0362	0.0355
Leu 9	0.0503	0.0492
Leu 11	0.0097	0.0094
Leu 29	0.0833	0.0813
Leu 32	0.0600	0.0587
Leu 40	0.0111	0.0111
Leu 49	0.0013	0.0013
Leu 61	0.0310	0.0303
Leu 69	0.0183	0.0177
Leu 72	0.0206	0.0202
Leu 76	0.0157	0.0155
Leu 86	0.0315	0.0302
Leu 89	0.0162	0.0155
Leu 104	0.0312	0.0311
Leu 115	0.0253	0.0250
Leu 133	0.0882	0.0859
Leu 137	0.0197	0.0190
Leu 149	0.0272	0.0263

*The entire 300 ps trajectory was used to calculate the order parameter by use of Eqs. 2 and 3.

To illustrate the convergence characteristics of NMR order parameters for this long trajectory, selected order parameters for Leu residues calculated using the complete 300-ps simulation are compared in Fig. 4 with the values calculated using the first 100-ps portion of the trajectory. The results shown in the figure are representative of the effects observed for all the residues studied. As expected order parameters for bond vectors (C^α—C^β) closer to the backbone are in general larger than those (C^γ—$C^{\delta 1}$) further out along the side chain. For the C^α—C^β bonds, the order parameters calculated from the different portions of the trajectory agree well; for the Leu methyl axis order parameters however, there are large discrepancies between the values calculated over the first 100 ps of the trajectory as compared with the complete simulation. For 14 of the 18 leucines studied, the order parameters are smaller when the entire trajectory is used to evaluate the correlation functions as compared with the first 100 ps; for seven of the eighteen residues the order parameters decrease by more than 50% when the entire 300-ps distribution of bond vector orientations is used to evaluate the order parameters. It is clear from these results that the order parameter calculations have not converged for this 300-ps trajectory. The order parameter depends on the longtime behavior of the NMR correlation function, which is determined by both the higher frequency local motions and more extensive conformational changes. We have also examined the

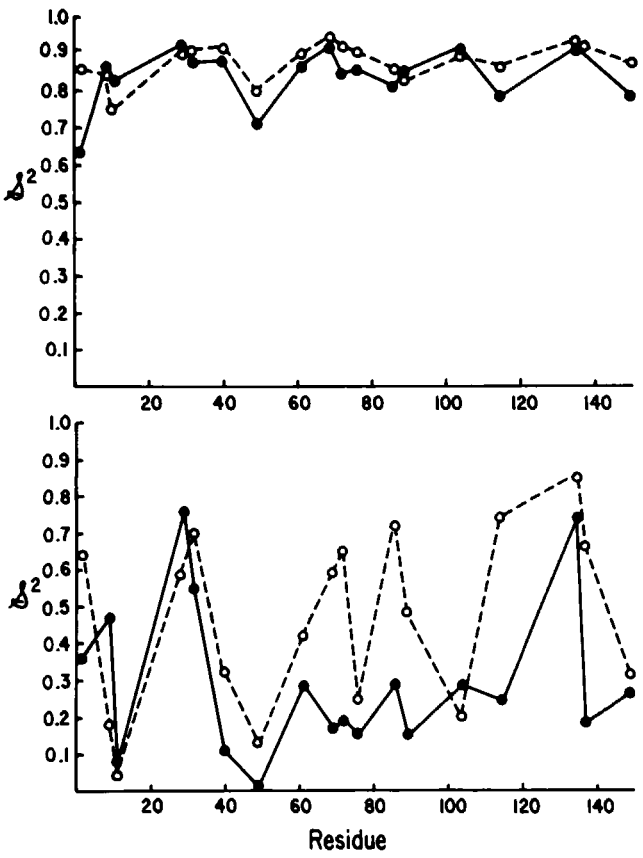

FIGURE 4 *Top*, order parameters for each of the Leu C^{α}—C^{β} bond vectors. ---, order parameters calculated using the 0–100 ps portion of the trajectory; ——, order parameters calculated using the entire 300-ps simulation. *Bottom*, same as Fig. 4, *top*, except calculations are for the Leu C^{γ}—$C^{\delta 1}$ bond vectors.

extent to which the decay of the angular correlation function at short times varies for different portions of the simulation. In Table IV, the values of selected NMR correlation functions (Eq. 1) at short times are listed; the calculations were repeated using successive 100-ps intervals as well as the entire trajectory to construct the correlation functions. The values in the table are determined by motions with relaxation times up to 32 ps. Several points are of interest. It can be seen that the extent of angular averaging increases progressively, moving out along the side chain. There is considerable variation in the short time values of the correlation function over successive 100-ps portions of the trajectory. However, as expected, the short time values of the correlation functions calculated using the entire 300-ps trajectory are close to the average of the values obtained using the three successive 100-ps segments of the trajectory. The variations in the correlation function for the bond vectors in Table VI when calculated for successive 100-ps intervals, reflect the occurrence of infrequent events; e.g., in Leu 9 there is a dihedral angle transition about C^{α}—C^{β} during the first 100-ps interval. These results are consistent with a model of protein motion in which groups of atoms (e.g., Leu side chains) oscillate about a mean conformation for tens of picoseconds and then move rapidly to a new conformation. The conformational change has a large effect on the order parameter. In addition, the effective potential in which the atoms move are somewhat different in the different conformations. We return to this point below.

DISCUSSION

The use of computer simulations to study the internal dynamics of globular proteins has attracted a great deal of

TABLE VI
VALUES OF ANGULAR CORRELATION FUNCTIONS FOR SELECTED BOND VECTORS OBTAINED USING DIFFERENT PORTIONS OF THE TRAJECTORY

Bond vector		a	b	c	d	e
		0–100 ps	100–200 ps	200-300 ps	0–300 ps	a + b + c
Leu 9	C^{α}-C^{β}	0.85	0.88	0.93	0.90	0.89
	C^{β}-C^{γ}	0.29	0.77	0.88	0.67	0.65
	C^{γ}-$C^{\delta 1}$	0.24	0.78	0.82	0.66	0.61
	C^{γ}-$C^{\delta 2}$	0.20	0.66	0.76	0.56	0.54
Leu 32	C^{α}-C^{β}	0.93	0.93	0.88	0.91	0.91
	C^{β}-C^{γ}	0.79	0.83	0.44	0.72	0.69
	C^{γ}-$C^{\delta 1}$	0.69	0.79	0.35	0.65	0.61
	C^{γ}-$C^{\delta 2}$	0.78	0.81	0.43	0.69	0.67
Leu 69	C^{α}-C^{β}	0.87	0.94	0.95	0.92	0.92
	C^{β}-C^{γ}	0.73	0.51	0.85	0.71	0.70
	C^{γ}-$C^{\delta 1}$	0.62	0.33	0.72	0.57	0.56
	C^{γ}-$C^{\delta 2}$	0.70	0.42	0.77	0.61	0.63
Leu 104	C^{α}-C^{β}	0.92	0.94	0.91	0.92	0.92
	C^{β}-C^{γ}	0.65	0.82	0.54	0.68	0.67
	C^{γ}-$C^{\delta 1}$	0.12	0.76	0.08	0.30	0.32
	C^{γ}-$C^{\delta 2}$	0.10	0.85	0.08	0.35	0.34

attention in recent years. The molecular dynamics simulation method, originally applied to simple liquids is a very powerful technique for studying the equilibrium and dynamic properties of systems composed of large numbers of atoms. Ultimately, the extent to which these simulations can quantitatively reproduce the properties of the real molecular systems for which they are models, depends on the accuracy of the potential functions used to describe the atomic and molecular interactions. For unassociated liquids it is known that the use of very simple potentials leads to quantitative results. Even for a complex liquid like water, simple potentials have been parameterized for use in computer simulations that are accurate with regard to both structural and thermodynamic properties over a range of temperatures. Packing considerations play a major role in determining the structural and dynamical properties of both liquids and proteins; because packing can be described by simple and reliable potential functions there is a strong foundation for simulation studies of protein dynamics. However, many factors distinguish molecular dynamics simulations of proteins from liquid state simulations so that it is difficult to use experience gained from molecular dynamics simulations of liquids to estimate the accuracy of the protein simulations. For liquid simulations the basic unit cell contains at least 100 molecules so that it is possible to take advantage of considerable statistical averaging in the calculation of quantities such as structure factors for comparison with experiment. For protein molecular dynamics simulations, in contrast, the computational effort required to evaluate the large number of interatomic interactions within a single protein molecule limits the simulated system to one or at most a very small number of (macro) molecules. The highly anisotropic nature of the protein interior and intrinsic interest in extracting single residue information further complicates the computational problem.

We have been concerned in several recent papers (3–11) with strengthening the relationship between molecular dynamics simulations of proteins and experimental measurements on the corresponding systems. Comparisons have been made between molecular dynamics averaged and x-ray structures, mean square atomic fluctuations and values derived from experimental temperature factors, and experimental and calculated NMR and fluorescence depolarization order parameters (1–11). For comparisons with experiment to be meaningful, it is clearly important to estimate the extent to which the calculated quantities have converged. The present results demonstrate that the structure of myoglobin changed very slowly during this long simulation. For both the mean square atomic fluctuations and NMR order parameters, there are significant quantitative differences when values calculated using shorter portions of the trajectory are compared with the results obtained for the entire 300-ps simulation. As expected, there is more atomic mobility when the entire 300-ps trajectory is used than when the shorter intervals are analyzed. For the backbone mean square displacements, the values calculated from the entire 300-ps trajectory are larger than the values estimated from the experimental temperature factors. The mean square displacements calculated by averaging the results over short (25 ps) portions of the trajectory are in better agreement with the experimental values (Fig. 3). The results suggest that the relative mobilities of the different regions of myoglobin are well described by the simulation, but that the quantitative agreement between simulation and experiment is not yet satisfactory.

To resolve differences between the simulation and experimental values, it is necessary to evaluate possible sources of error in both. With regard to x-ray experiment, the quantitative analysis of atomic displacements using temperature factors for proteins is not straightforward. It is only recently that high enough resolution x-ray data have become available from which it is possible to attempt to analyze temperature factors in terms of protein motions (22–25). A number of assumptions concerning atomic fluctuations are incorporated in the macromolecular crystallographic refinement programs that affect the subsequent refinement. We are using molecular dynamics simulations of proteins and nucleic acids to evaluate the effect of anharmonic and anisotropic motions on x-ray refinement (Kuriyan, S., M. Karplus, R.M. Levy, and G. Petsko, manuscript submitted per publication; Westhof, E., and R.M. Levy, manuscript submitted for publication). With regard to the myoglobin simulation a set of structure factors has been generated from coordinate sets sampled during the 300-ps trajectory. The simulated structure factors were then refined. By comparing B values extracted from this refinement with those calculated directly from the simulation it has been possible to evaluate the extent to which Debye Waller factors provide a quantitative measure of atomic fluctuations for macromolecules (Kuriyan, S., M. Karplus, R.M. Levy, and G. Petsko, manuscript submitted for publication). The results clearly demonstrate that for atoms with large fluctuations [$\langle(\Delta r)^2\rangle \gtrsim 0.6$ Å^2], the magnitudes are significantly underestimated by values derived from refined temperature factors. This finding has important consequences for the continued development of methodology for comparing molecular dynamics simulations of macromolecules with experimental results and for the functional interpretation of experimental B values; e.g., the postulated correlations between segmental mobility and the location of antigenic determinants on proteins (33, 34). Furthermore, in general, the mean square atomic fluctuations estimated from the x-ray refinement of the structure factors generated from the 300-ps myoglobin simulations are in better accord with the experimental results than the values calculated directly from the simulation. The results of this detailed analysis of structure factors generated from the myoglobin simulation is

reported elsewhere (Kuriyan, S., M. Karplus, R.M. Levy, and G. Petsko, manuscript submitted for publication).

With regard to improvements in the molecular dynamics methodology, the inclusion of the full unit cell in protein simulations of crystals represents a more realistic model of the system, and improved accuracy is expected. However, the required computational effort is much greater. Since the atomic motions, particularly of atoms close to the protein surface, will be more highly damped in the presence of solvent, drift of the coordinates from the x-ray structure is expected to occur more slowly for the full crystal simulations than those in vacuum. Much longer crystal simulations than those reported to date (20 ps) will be required to determine whether the potentials are sufficiently accurate for the simulations to achieve an ergodic average.

A qualitative picture is emerging from both theoretical and experimental studies of protein dynamics that suggest that the longer time motions of proteins involve multiple minima. In the present simulations these may be related to the distinct conformations obtained from subaverages of the trajectory. The variation in the mean square displacements and order parameters over different portions of this 300-ps trajectory are consistent with this picture; detailed structural studies of these conformations including an analysis of α-helix reorientations and packing changes are underway. The presence of many local potential minima near the x-ray structure severely complicates the problem of achieving an ergodic simulation. The availability of very high speed computing facilities will be required to solve the convergence problem in protein simulations. With increased computing time on the more powerful computers it will be possible to refine potential functions, include more detailed representations of the crystal, and solvent and hopefully develop more efficient simulation methods.

We thank Professor Greg Petsko for providing the myoglobin coordinates.

This work has been supported by grants from The National Institutes of Health (NIH) to R.M. Levy and M. Karplus and by a Biomedical Research Grant from Rutgers University. R.M. Levy is an Alfred P. Sloan Fellow and NIH Research Career Development Award Recipient.

Received for publication 26 November 1984 and in final form 8 May 1985.

REFERENCES

1. Karplus, M., and J. A. McCammon. 1981. The internal dynamics of globular proteins. *CRC Crit. Rev. Biochem.* 9:293–349.
2. McCammon, J. A. 1984. Protein dynamics. *Rep. Progr. Phys.* 47:1–46.
3. Levy, R. M., M. Karplus, and J. A. McCammon. 1980. Molecular dynamics studies of NMR relaxation in proteins. *Biophys. J.* 32:628–630.
4. Levy, R. M., M. Karplus, and J. A. McCammon. 1981. Increase of ^{13}C NMR relaxation times in proteins due to picosecond motional averaging. *J. Am. Chem. Soc.* 103:994–996.
5. Levy, R. M., C. M. Dobson, and M. Karplus. 1982. Dipolar NMR relaxation of nonprotonated aromatic carbons in proteins, structural and dynamical effects. *Biophys. J.* 39:107–113.
6. Lipari, G., A. Szabo, and R. M. Levy. 1982. Protein dynamics and NMR relaxation. Comparison of simulations with experiment. *Nature (Lond.)* 300:197–198.
7. Hoch, J. C., C. M. Dobson, and M. K. Karplus. 1982. Fluctuations and averaging of proton chemical shifts in the bovine pancreatic trypsin inhibitor. *Biochemistry.* 21:1118–1175.
8. Olejniczak, E. T., C. M. Dobson, M. Karplus, and R. M. Levy. 1984. Motional averaging of protein nuclear Overhauser effects in proteins. Predictions from a molecular dynamics simulation of lysozyme. *J. Am. Chem. Soc.* 106:1913–1930.
9. Levy, R. M., and A. Szabo. 1982. Initial fluorescence depolarization of tyrosines in proteins. *J. Am. Chem. Soc.* 104:2073–2076.
10. Ichiye, T., and M. Karplus. 1983. Fluorescence depolarization of tryptophan residues in proteins. A molecular dynamics study. *J. Am. Chem. Soc.* 22:2284–2293.
11. Northrup, S. H., M. R. Pear, J. D. Morgan, J. A. McCammon, and M. Karplus. 1981. Molecular dynamics of ferrocytochrome C: Equilibrium results. *J. Mol. Biol.* 153:1087–1105.
12. Mao, B., M. R. Pear, J. A. McCammon, and S. H. Northrup. 1982. Repulsive interactions between polar and apolar atoms in globular proteins. *Biopolymers.* 21:1979–1993.
13. Levy, R. M., O. Rojas, and R. A. Friesner. 1984. Quasi-harmonic method for calculating vibrational spectra from classical simulations on multi-dimensional anharmonic potential surfaces. *J. Phys. Chem.* 88:4233–4238.
14. Brooks, B., R. Bruccoleri, B. Olafson, D. J. States, S. Swaminathan, and M. Karplus. 1983. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.* 4:187–217.
15. van Gunsteren, W. F., and M. Karplus. 1982. Protein dynamics in solution and in a crystalline environment. *Biochemistry.* 21:2259–2274.
16. McCammon, J. A., B. R. Gelin, and M. Karplus. 1977. Dynamics of folded proteins. *Nature (Lond.)* 267:585–590.
17. Karplus, M., and J. A. McCammon. 1979. Protein structural fluctuations during a period of 100 ps. *Nature (Lond.)* 277:578–579.
18. Levitt, M. 1983. Molecular dynamics of native protein: (1) Computer simulation of trajectories. *J. Mol. Biol.* 168:595–620.
19. Levitt, M. 1983. Molecular dynamics of native protein: (2) Analysis and nature of motion. *J. Mol. Biol.* 168:621–657.
20. van Gunsteren, W. F., H. J. C. Berendsen, J. Hermans, W. G. J. Hol, and J. P. M. Postma. 1983. Computer simulation of the dynamics of hydrated protein crystals and its comparison with x-ray data. *Proc. Natl. Acad. Sci. USA.* 80:4315–4319.
21. Willis, B. T. M., and A. W. Pryor. 1975. Thermal Vibrations in Crystallography. Cambridge University Press, London.
22. Frauenfelder, H., G. Petsko, and D. Tsernaglou. 1979. Temperature-dependent x-ray diffraction as a probe of protein structural dynamics. *Nature (Lond.).* 280:558–562.
23. Artymiuk, P. J., C. C. F. Blake, D. E. P. Grace, S. J. Oatley, D. C. Philips, and M. J. E. Steinberg. 1979. Crystallographic studies of the dynamic properties of lysozyme. *Nature (Lond.)* 280:563–568.
24. Hartmann, H., F. Parak, W. Steigmann, G. A. Petsko, and D. Ringe Ponzi. 1982. Conformational substrates in a protein: Structure and dynamics of metmyoglobin at 80°K. *Proc. Natl. Acad. Sci. USA.* 79:4967–4971.
25. Petsko, G., and D. Ringe. 1984. Fluctuations in protein structure from x-ray diffraction. *Annu. Rev. Biophys. Bioeng.* 13:331–371.
26. Yu, H., M. Karplus, and W. A. Hendrickson. 1985. Restraints in temperature factor refinement for macromolecules: An evaluation by molecular dynamics. *Acta Cryst.* In press.
27. Levy, R. M., D. Perahia, and M. Karplus. 1982. Molecular dynamics of an α-helical polypeptide: Temperature dependence and devia-

tion from harmonic behavior. *Proc. Natl. Acad. Sci. USA.* 179:1346–1350.

28. Richarz, R., K. Nagayama, and K. Wuthrich. 1980. Carbon—13 nuclear magnetic resonance relaxation studies of internal mobility of the polypeptide chains in base pancreatic trypsin inhibitor and a selectively reduced analogue. *Biochemistry.* 19:5189–5196.
29. Lipari, G., and A. Szabo. 1982. Model free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules: Analysis of experimental results. *J. Am. Chem. Soc.* 104:4559–4570.
30. Wittebort, R. J., T. M. Rothgeb, A. Szabo, and F. R. N. Gurd. 1979. The aliphatic groups of sperm whale myoglobin. A ^{13}C NMR study. *Proc. Natl. Acad. Sci. USA.* 76:1059–1066.
31. Ribeiro, A., R. King, C. Restivo, and O. Jardetsky. 1980. An approach to the mapping of internal motions in proteins. Analysis of ^{13}C NMR relaxation in the bovine pancreatic trypsin inhibitor. *J. Am. Chem. Soc.* 102:4040–4051.
32. Gurd, F. R. N., T. M. Rothgeb, and G. Neireiter. 1982. Motions of aliphatic residues in myoglobins. *In* Biochemical Structure Determination by NMR. A. Bothner, J. Glickson and B. Sykes, editors. Marcel Dekker, Inc., New York.
33. Westhof, E., D. Altschuh, D. Moras, A. C. Bloomer, A. Mondragon, A. Klug, and M. H. V. Van Regenmortel. 1984. Correlation between segmental mobility and the location of antigenic determinants in proteins. *Nature. (Lond.).* 311:123–126.
34. Tainer, J. A., E. Getzoff, J. Alexander, R. A. Houghten, A. J. Olson, R. A. Lerner, and W. A. Hendrickson. 1984. The reactivity of anti-peptide antibodies is a function of the atomic mobility of sites in a protein. *Nature (Lond.).* 312:127–134.
35. Richards, F. M. 1977. Areas, volumes, packing, and protein structure. *Ann. Rev. Biophys. Bioeng.* 6:151–176.