Molecular dynamics simulations of heme reorientational motions in myoglobin

Eric R. Henry

Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892 USA

ABSTRACT Molecular dynamics simulations of 2-ns duration were performed on carbonmonoxymyoglobin and deoxymyoglobin *in vacuo* to study the reorientational dynamics of the heme group. The heme in both simulations undergoes reorientations of ~5° amplitude on a subpicosecond time scale, which produce a rapid initial decay in the reorientational correlation function to about 0.99. The heme also experiences infrequent changes in average orientation of ~10° amplitude, which lead to a larger slow decay of the reorientational correlation function over a period of hundreds of picoseconds. The simulations have not converged with respect to these infrequent transitions. However, an estimate of the order parameter for rapid internal motions of the heme from those orientations which are sampled by the simulations suggests that the subnanosecond orientational dynamics of the heme accounts for at least 30% of the unresolved initial anisotropy decay observed in the nanosecond time-resolved optical absorption experiments on myoglobin reported by Ansari et al. in a companion paper (Ansari, A., C. M. Jones, E. R. Henry, J. Hofrichter, and W. A. Eaton. 1992. *Biophys. J.* 64:852–868.). A more complete sampling of the accessible heme orientations would most likely increase this fraction further. The simulation of the liganded molecule also suggests that the conformational dynamics of the CO ligand may contribute significantly to discrepancies between the ligand conformation as probed by x-ray diffraction and by infrared-optical photoselection experiments. The protein backbone explores multiple conformations during the simulations, with the largest structural changes appearing in the E and F helices, which are in contact with the heme. The variations in the heme orientation correlate with the conformational dynamics of the protein on a time scale of hundreds of picoseconds, suggesting that the heme orientation may provide a useful probe of dynamical processes in the protein.

INTRODUCTION

Heme proteins have long been a major focus of studies on protein dynamics. Photodissociation experiments probe the dynamical properties of heme proteins by dissociating heme-bound ligands with intense laser pulses and spectroscopically monitoring the subsequent evolution of the protein-ligand system (for review see (1) for references). These experiments most frequently use linearly polarized laser pulses to photodissociate the ligand and to subsequently measure the absorption spectrum. To a good approximation, the heme is a circular absorber of linearly polarized light at wavelengths typically used to excite and probe the system—that is, the probability of absorption of linearly polarized light (and therefore of photodissociation) is the same when the electric vector is parallel to any direction lying in the plane of the heme and is zero for polarization perpendicular to the plane. As a result, incomplete photodissociation by linearly polarized light will produce an initially anisotropic population of photolyzed hemes whose planes are oriented preferentially parallel to the electric vector of the incident light. The sample will therefore initially exhibit linear dichroism, which will decay to zero as the distribution of orientations of the photolyzed hemes randomizes through rotational diffusion. This so-called photoselection effect complicates the analysis of spectroscopic experiments on heme proteins that make use of partial photolysis. For example, the decay of the linear dichroism will lead to an apparent increase or decrease in the fraction of liganded hemes, depending on the polarization of the probe light, thereby distorting the apparent ligand rebinding profile. Thus, the accurate measurement of ligand rebinding on time scales comparable to and shorter than the rotational diffusion time of the protein molecule requires careful consideration of the effects of photoselection.

The two companion papers in this volume describe theoretical and experimental studies of the effects of photoselection and rotational diffusion on time-resolved optical spectra of heme proteins. In their paper on the theory of photoselection, Ansari and Szabo (2) have presented a general theory of optical density and optical anisotropy measurements in polarized photolysis experiments using linearly polarized light, including the effects of rotational diffusion of the entire molecule and rapid internal motions of the heme, as well as non-ideal absorption properties of the heme and finite excitation pulse widths. In the paper on experimentation, Ansari et al. (1) have applied this theory to a series of nanosecond time-resolved spectroscopic measurements on myoglobin in viscous solvents using linearly polarized excitation and probe pulses. They found that the values of the optical anisotropy extrapolated to zero time measured as a function of excitation intensity are systematically smaller by about 10% than the values predicted by the theory. They explored three possible origins for this reduced anisotropy: (*a*) deviation from perfect circular absorption by the heme (i.e., a *z*-polarized component of

Address correspondence to Dr. Eric R. Henry, Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, Building 2, Room B1-04, National Institutes of Health, 9000 Rockville Pike 2/122, Bethesda, MD 20892, USA.

the absorption), (*b*) a tilt of the heme in response to ligand dissociation, and (*c*) rapid internal motions of the heme on subnanosecond time scales. Based on a comparison of x-ray crystallographic structures of Mb and MbCO, which produces a heme tilt of only about 3.5°, the likely contribution of a possible heme tilt to the reduced anisotropy was judged to be very small. Furthermore, an analysis of the crystal absorption spectra of liganded and unliganded myoglobin and hemoglobin indicated that the magnitude of any *z*-polarized component of the absorption is unlikely to be large enough to produce the observed reduced anisotropy. However, the theory predicts that if the excitation pulse is sufficiently long that an equilibrium is established with respect to the fast internal motions, then the effect of these fast motions on the measured anisotropy is as if the absorber had an effective out-of-plane component (2). Therefore, the effect of subnanosecond internal motions cannot be distinguished from the effect of an out-of-plane polarized component of the absorption without resolving the initial decay of the anisotropy.

An independent estimate of the contribution of rapid internal motions to the reduced anisotropy may be obtained by explicitly simulating the heme motions using molecular dynamics. In such simulations the photodissociation is most easily represented as an instantaneous event involving breakage of the iron-ligand bond and changes in the heme potential function (3). The starting point for calculating the anisotropy decay would therefore be the theoretical expression for the anisotropy for a circular absorber at both excitation and probe wavelengths in the limit of short pulses from (2) (Eq. 3.20):

$$r(t) = \frac{1}{10}\frac{a_2(\lambda_e)}{1 - a_0(\lambda_e)} \times \left[\frac{\epsilon_B^0(\lambda_p)\langle\langle P_2(\hat{n}_B(t)\cdot\hat{n}_A(0^+))\rangle\rangle - \epsilon_A^0(\lambda_p)\langle P_2(\hat{n}_A(t)\cdot\hat{n}_A(0^+))\rangle}{\epsilon_B^0(\lambda_p) - \epsilon_A^0(\lambda_p)}\right] \quad (1)$$

where the subscripts A and B refer to the pre- and post-photolysis states of the heme, respectively, and $P_2(x) = (3x^2 - 1)/2$ is the second Legendre polynomial. In this equation, $1 - a_0(\lambda_e)$ is the fraction of molecules that undergo the transition A → B and $a_2(\lambda_e)$ is proportional to the linear dichroism induced in the isotropic system by excitation with linearly polarized light; both quantities are functions of the excitation intensity (2). The quantities $\epsilon_{A,B}^0(\lambda_p)$ are the isotropic extinction coefficients for absorption at the probe wavelength in states A and B. For present purposes the interesting components of this equation are the correlation functions between angled brackets, which embody the time dependence of the anisotropy. Here the unit vectors $\hat{n}_A(t)$ and $\hat{n}_B(t)$ represent the directions at time t of the normal to the heme plane in state A (liganded) and in state B (photolyzed), respectively; $t = 0^+$ refers to the time at the end of the short intense photolysis pulse. Thus, the second expression between angled brackets is the correlation between the heme normal immediately after photolysis and the normal at a time t units later, all evaluated for the state A. This correlation function is an equilibrium property of the system in state A and as such could be estimated from a single dynamics trajectory. On the other hand, the first expression between double angled brackets is a nonequilibrium correlation function (2), which requires knowledge of the dynamics of the heme normal in state B for all times after every possible photodissociation event. A complete calculation of this correlation would require a distinct photodissociation trajectory initiated at every time point of a simulation of the system in state A, which, with current simulation capabilities would not be feasible. Therefore, for purposes of estimating from molecular dynamics simulations the anisotropy decay due to rapid internal motions of the heme, it is necessary to make the simplifying assumption that the reorientational dynamics of the heme are the same in the liganded and unliganded states. In this case, the first and second correlation functions are the same, and Eq. 1 takes the form of Eq. 3.24 of (2), which we re-write schematically in the form

$$r(t) = \frac{1}{10}\Gamma_{AB}(\lambda_e, \lambda_p)\langle P_2(\hat{n}(t)\cdot\hat{n}(0^+))\rangle. \quad (2)$$

Here, $\Gamma_{AB}(\lambda_e, \lambda_p)$ incorporates all of the absorption properties of states A and B at the excitation and probe wavelengths. The reorientational dynamics are now contained in a single correlation function, which may be estimated from an equilibrium simulation of the liganded or unliganded state.

In this paper we report the results of very long molecular dynamics simulations of both liganded (MbCO) and unliganded myoglobin (Mb) at equilibrium. A simulation of either the liganded or the unliganded state is required to calculate the correlation function in Eq. (2). Having a simulation of the unliganded state as well as the liganded state may provide information about whether there are in fact significant differences between the heme reorientational dynamics in the two states. The simulated period in each case is 2 ns, which should be long enough to at least suggest the types of heme reorientational motions that might occur on subnanosecond time scales. From these simulations we have derived the heme reorientational dynamics and calculated reorientational correlation functions, providing an estimate of the fractional reduction in the absorption anisotropy that might occur due to rapid internal motions. We have also analyzed the motions of the CO ligand in the MbCO simulation in order to compare the averaged conformations with the results of x-ray diffraction studies and picosecond infrared-optical photoselection measurements. We also present an analysis of the overall conformational dynamics of the protein in the two simulations and show how these dynamics correlate with the heme reorientational motions.

METHODS

Molecular dynamics simulations were performed on both MbCO and Mb *in vacuo,* using a program written by the author based on a program originally provided by Michael Levitt. The potential function used to describe the interatomic interactions was of the form most frequently employed in protein molecular dynamics simulations, incorporating harmonic terms for the stretching of covalent bonds and the bending of angles between bonds, periodic terms for torsional rotations about bonds, a 6–12 Lennard–Jones potential for van der Waals/hard-sphere interactions between non-bonded atoms, and electrostatic interactions between point partial charges located on each atom. The residue topology information used to construct the molecules and enumerate the bonding interactions, as well as the actual interaction parameters used in the potential function, were derived from the CHARMM19 set (4, 5); minor changes were made in the dihedral angle list and several bonding energy parameters for the heme group in order to satisfy specific structural and energetic criteria (3). The screening of electrostatic interactions at large distances was approximated by the use of a dielectric constant ϵ equal to the interatomic distance in ångströms, making the electrostatic energy effectively an inverse-square quantity (4). Both the electrostatic and the van der Waals interactions were gradually switched off over a range of interatomic distances from 8 to 10 Å using a simple multiplicative switching function (4). The starting coordinates for the MbCO simulation included the coordinates for the 1262 heavy atoms from the 1.5 Å resolution structure of sperm whale MbCO by Kuriyan et al. (6) (Brookhaven Protein Data Bank file 1MBC) determined at 260 K, to which were added coordinates of 268 polar hydrogen atoms, for a total of 1530 atoms. The starting coordinates for the Mb simulation included the coordinates of the 1260 heavy atoms from the 1.4 Å resolution structure of sperm whale deoxyMb by Phillips (Brookhaven Protein Data Bank file 1MBD), to which were added coordinates for the same set of polar hydrogen atoms, for a total of 1528 atoms. Both coordinate files include multiple sets of coordinates, or conformations, for certain groups of atoms for which conformational heterogeneity was incorporated in the refinement; in these cases, the first ("A") set for each group was used. Both starting coordinate sets were first subjected to 300 cycles of conjugate gradient energy minimization, the resulting coordinate sets serving as initial coordinates for the dynamics simulations. The equations of motion were integrated using the Beeman algorithm (7) with a time step of 0.98 fs. (For convenience, this step size will be approximated as 1 fs in calculating all time intervals presented in this paper.) The temperature of each simulation was increased to 300 K during the first 30 ps of the simulation, using a temperature increment of 30 K every 3 ps. During each 3-ps interval the required temperature was imposed several times by random reassignment of velocities of all atoms from a Maxwellian distribution of velocities. After each velocity reassignment, all velocities were corrected to eliminate any residual total momentum and angular momentum of the molecule. At the end of this initial 30 ps heating phase, an additional 10 ps equilibration phase was calculated, during which velocities were periodically rescaled to impose an instantaneous kinetic temperature of exactly 300 K. An additional 2,000 ps of dynamics after this equilibration were then used in all subsequent analysis. During each simulation the coordinates of all the atoms were saved every 20 time steps (20 fs); the list of non-bonded atom pairs used in calculating the non-bonded energies, including all pairs of non-bonded atoms separated by less than 11 Å, was updated at the same interval.

RESULTS AND DISCUSSION

General features of the simulations

It is useful to first briefly summarize some of the overall features of the simulations. Many parts of the analysis to be discussed in this and subsequent sections require or

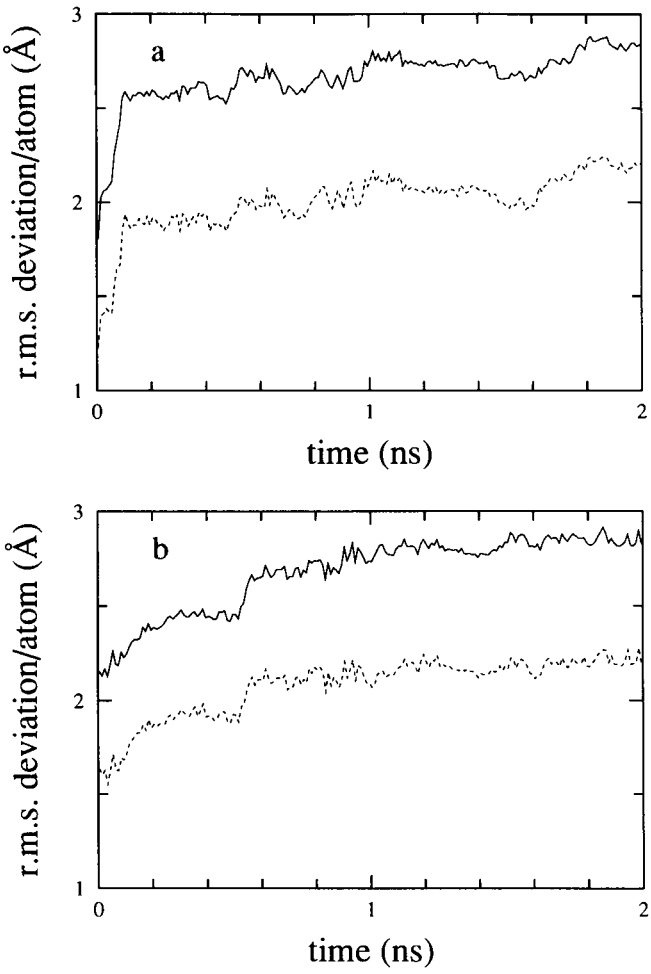

FIGURE 1 r.m.s. deviations per atom of the 200 10-ps averaged coordinate sets from the corresponding x-ray crystallographic structure used as starting point for the simulations. Deviations are shown for all atoms of the molecule (*solid line*) and for the 459 backbone atoms (*dashed line*). *a*) MbCO simulation; *b*) Mb simulation.

are greatly facilitated by a reduction of the 100,000 coordinate sets that make up one of these simulations to a much simpler form. By averaging the 500 coordinate sets in each successive 10 ps segment of the trajectory, a set of 200 averaged coordinate sets was produced, which adequately embodies the overall evolution of the simulated system. Fig. 1 shows the r.m.s. (root-mean-square) deviations per atom of these averaged structures from the starting structure as a function of time for both simulations. In both cases a significant drift from the starting structure, amounting to more than half of the ultimate deviation at the end of the trajectory, takes place during the initial energy minimization and the 40 ps of heating and equilibration (not shown). During the constant-energy analysis phase of the simulation shown, there are short periods of rapid drift (at the beginning of the MbCO simulation, and at the beginning and at 500 ps in the Mb simulation), superimposed on a general very gradual drift (~0.2 Å/ns) during the remainder of the

simulation. The final deviations at 2,000 ps (~2.2 Å for alpha carbons) are in fact comparable to deviations found in earlier, much shorter simulations of myoglobin and other heme proteins of comparable size (3, 8, 9, 10). Inspection of the principal moments of inertia of the molecule as a function of time reveals that this structural drift is associated with a small shrinkage of the molecule during the preparatory phases of the simulation (less than 3% change in any linear dimension), with very little additional change in size and shape of the molecule during the analysis phase.

It will be shown below that the myoglobin molecules in both simulations exhibit significant conformational heterogeneity. The periods of rapid drift away from the starting structure appearing in Fig. 1 are, as expected, associated with conformational changes as determined by more general criteria (see below). However, the near-absence of progressive drift from the starting structure during large segments of both trajectories does not conversely imply that the protein conformations are fixed during these periods. For example, the 10-ps average structures at 100 ps and 2,000 ps in the MbCO simulation have r.m.s. deviations from the starting structure for the alpha carbons that differ by less than 0.3 Å (Fig. 1), whereas the corresponding r.m.s. difference between these two structures is greater than 1.3 Å. Therefore, once the simulation has "stabilized" in some sense, it is able to explore different regions of conformational space without drifting significantly farther from the starting structure.

The total energy in a dynamics simulation undergoes small rapid fluctuations as a direct result of the finite-difference integration algorithm. The total energies averaged over successive 100-step (0.1 ps) blocks are extremely stable for these simulations, typically varying by ±1 kcal/mol over the entire 2000-ps length of the simulations, with no significant upward or downward drift. In the early stages of the analysis phase of each simulation, the trajectory moves very quickly into lower-lying regions of the potential surface, once the "kinetic restraints" imposed by the periodic random velocity reassignments of the preparatory phase are removed. This movement is clearly reflected in the drift from the starting structure (Fig. 1), as well as in more detailed measures of the conformation (see below). The constancy of the total energy requires that this movement be accompanied by an increase in the kinetic energy, and therefore the temperature, of the simulation. For the MbCO simulation, the early migration of the trajectory results in an increase of the temperature from about 300 K at the beginning to about 320 K by about 200 ps; this is followed by an additional slow increase to about 325 K by the end of the simulation. For the Mb simulation, the initial temperature increase is to about 315 K, with very little additional increase during the remainder of the simulation. No attempt was made to correct the temperature back to 300 K by rescaling of velocities.

Heme reorientational motions

With the assumption that the heme group is a circular absorber, the orientation of the heme for any specified coordinate set is given by the unit vector normal to the best least-squares plane through the 24 atoms of the tetrapyrrole ring; rotations about this vector for a perfect circular absorber do not contribute to the observed optical properties. The orientation of this vector can be described by the position of the tip of the vector on the surface of a fixed unit sphere centered at the origin. As the heme reorients during a simulation, the heme normal vector traces out a path on this surface. This path is most easily viewed upon projection of the spherical surface onto a flat two-dimensional plot.

The heme normal vectors were first calculated for the 200 10-ps coordinate averages for each simulation (see above). This provides a low-resolution view of the behavior of the heme orientation over an entire simulation. Projections of the paths traced out by these vectors on the surface of a unit sphere are shown in Fig. 2. In both simulations, the heme explores several more-or-less distinct regions of the orientational space separated by 5–10°, moving between regions on a time scale of hundreds of picoseconds. In the case of the MbCO simulation (Fig. 2 *a*), there are two clearly distinct orientational regions occupied during the 100–800 ps and 900–1700 ps segments of the simulation, connected by a "transition" during the 800–900 ps segment, and the apparent beginning of a third region toward the end of the simulation. In the first 600 ps of the Mb simulation (Fig. 2 *b*) the heme normal vector traces out a wide arc from the starting orientation to a second orientation about 10° away and then back, before settling into smaller reorientations within one or very few distinct states for the remainder of the simulation.

In order to obtain a higher-resolution view of the heme reorientational motions, the heme normal vector was calculated for each of the 100,000 saved coordinate sets for each simulation. Fig. 3 shows the orientations calculated from the 500 coordinate sets for a single 10-ps segment of the MbCO simulation. (The averaged coordinates from a segment of this size would provide a single orientation plotted in Fig. 2.) The plotted points at the tips of the arrows represent the direction of the heme normal vector at 20-fs intervals. The path traced out by this vector experiences frequent changes of direction every few hundred femtoseconds, between which it appears to be smoothly varying and almost ballistic in character. Reorientation of the heme during a specified time period is presumably a cumulative effect of the collisions between atoms of the heme ring and atoms of the surrounding protein that take place during that period. The heme normal vector would not be expected to respond instantaneously to these collisions, because it represents the average plane through 24 heme atoms. The resulting averaging of the effects of collisions leads to a smooth trajectory on sufficiently short time scales.

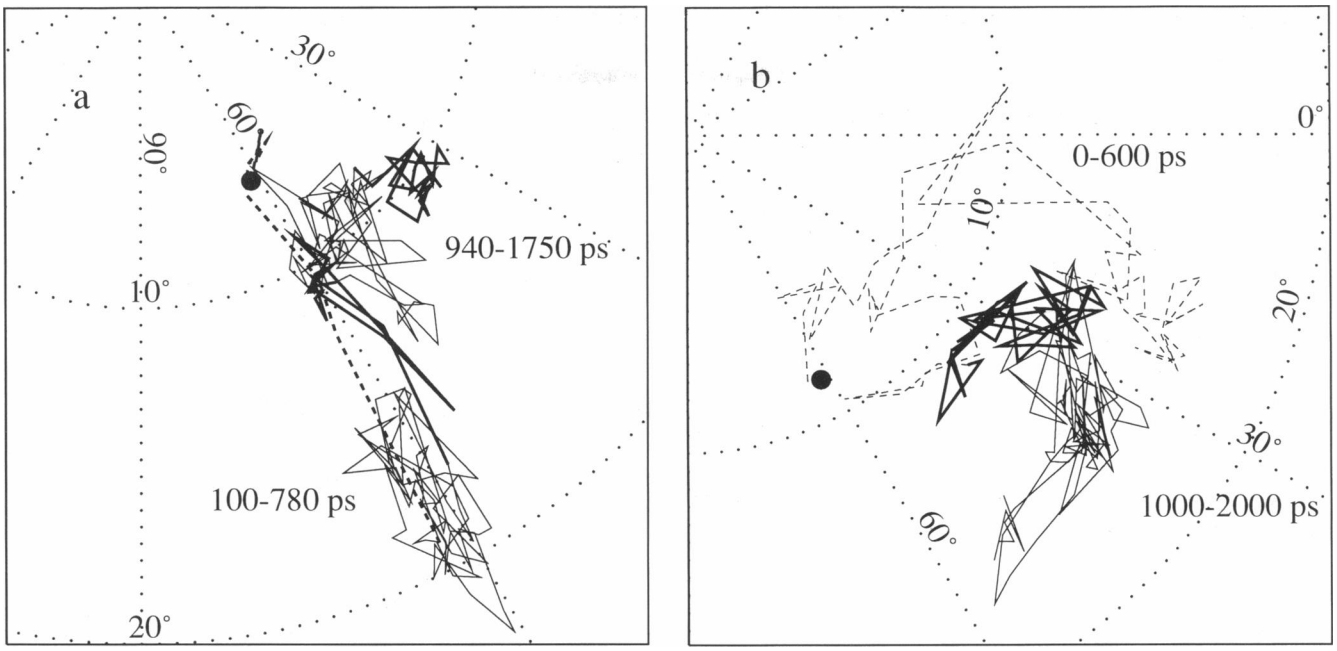

FIGURE 2 Directions of the heme normal vectors calculated from the 10-ps averaged coordinate sets. The path traced out on the surface of a unit sphere by the unit vector normal to the 24-atom heme plane in successive coordinate sets is shown in projection. In each case the "north pole" of the sphere lies in the direction of the heme normal vector in the corresponding x-ray structure, and the 0° azimuthal plane is parallel to the z coordinate direction of the Cartesian coordinate system in which the structure is defined. Segments of the path from different parts of the simulation are distinguished by the use of dashed lines and solid lines of different weights. In both panels a filled circle indicates the heme normal direction from the first coordinate set. *a*) MbCO simulation. A heavy dashed line traces the initial large motion of the heme normal during the first 100 ps. The paths during periods of relatively localized motions at 100–780 ps and 940–1,750 ps are shown with light solid lines, and the transition during 780–950 ps and the path during the final 250 ps are shown with heavy solid lines. *b*) Mb simulation. The large motions of the heme normal vector during the first 600 ps are shown as a dashed line, the path during the period 600–1,000 ps is shown as a heavy solid line, and the path during the remainder of the simulation is shown as a light solid line.

Given the amplitudes of the rapid reorientational motions shown in Fig. 3, it is expected that the finer features of the evolution of the averaged orientations in Fig. 2 are determined to some extent by the specific averaging method chosen. On the other hand, the lower-resolution view of the evolution in Fig. 2, on a time scale of hundreds of picoseconds, unequivocally shows that the heme occupies different preferred regions of orientational space at different times in the simulations. The picture that emerges is not one of the heme experiencing rapid orientational fluctuations within a small number of discrete localized orientational states, but rather one of the heme undergoing fairly large reorientations on short time scales, with the "center of gravity" of these motions itself moving among different regions of orientational space on much longer time scales. Indeed, even a "transition" such as that shown as the heavy solid line in the middle of Fig. 2 *a* is neither abrupt nor simple; under closer inspection it is resolved into a series of steps between intermediate orientations connecting the initial and the final region, with each step taking the form of a period of localized motion followed by delocalized motion of the same basic character as is shown in Fig. 3, which carries the heme to the next intermediate orientation.

The reorientational correlation function and absorption anisotropy

As outlined in the Introduction, the absorption anisotropy may be written in terms of a correlation function of the heme normal. In the limit of low excitation intensities, $\Gamma_{AB}(\lambda_e, \lambda_p) \rightarrow 1$, and Eq. 2 becomes:

$$r(t) = \frac{1}{10}\langle P_2(\hat{n}(t)\cdot\hat{n}(0^+))\rangle. \qquad (3)$$

The reorientational correlation function was calculated from a simulation by successively choosing each time point in the first half of the simulation as the "photolysis" point ($t = 0^+$), calculating the enclosed quantity as a function of time relative to this point, and averaging the resulting time-dependences over all choices of origin. The reorientational correlation functions calculated in this manner for the two entire simulations are shown in Fig. 4. These correlation functions reflect only the effects of internal reorientations of the heme with respect to the protein.[1] In both simulations there is a rapid decrease in

[1] In order to avoid disturbing the progress of these simulations, the angular momentum of the system was not reset to its initial value of zero during the analysis phase, and the consequent very small drifts in

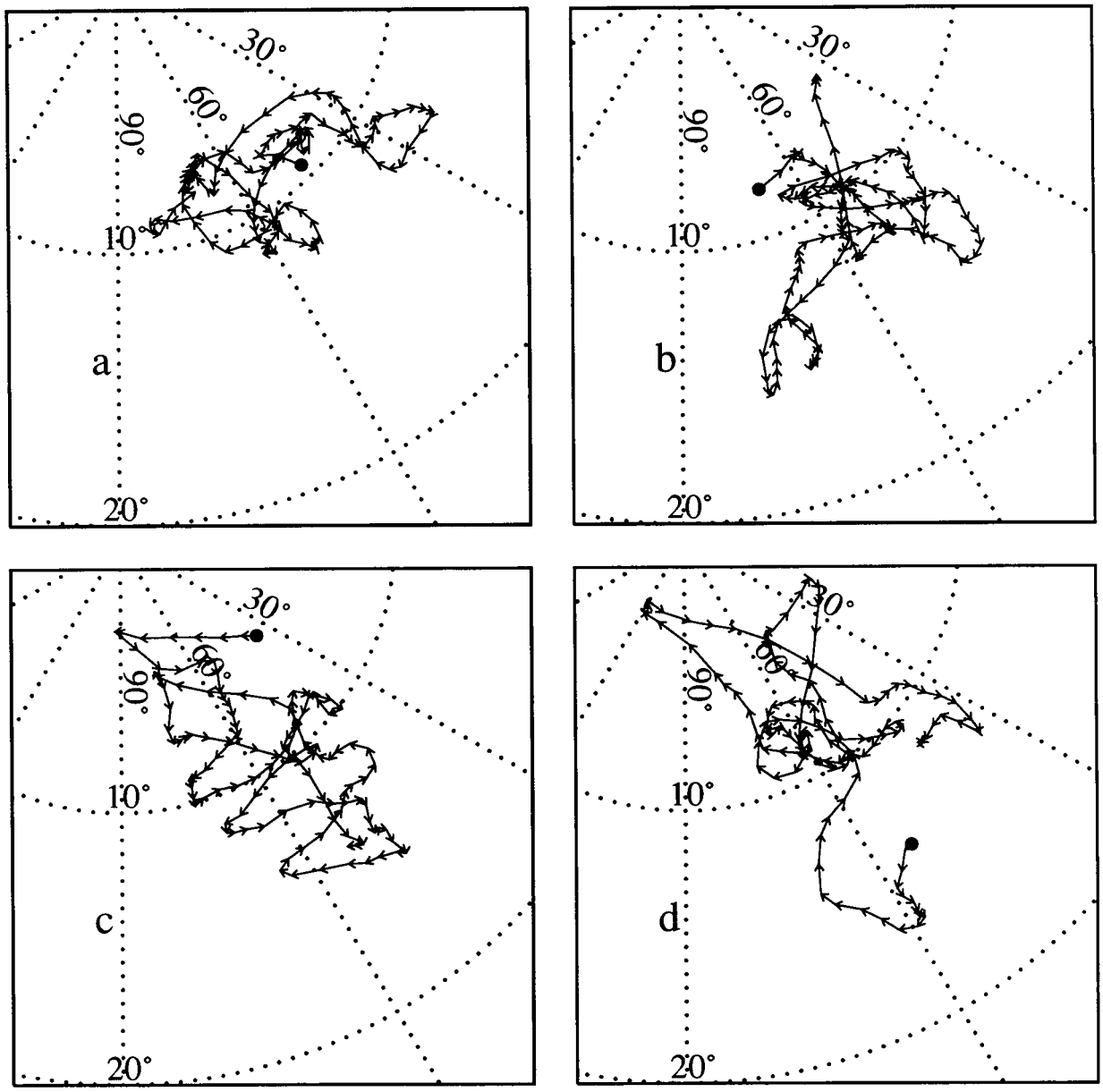

FIGURE 3 Trajectory of the heme normal vector on the surface of a unit sphere during a 10-ps period in the MbCO simulation. The coordinate system of the sphere and the orientational region enclosed in each panel are the same as in Fig. 2 *a*. The period shown is 1,200–1,210 ps, which lies within the upper localized region (*light solid lines*) of Fig. 2 *a*. The period has been divided into four 2.5-ps intervals, labeled *a–d*, for clarity. The initial orientation in each interval is indicated by a filled circle, and successive orientations at 20-fs intervals are indicated by the tips of connected arrows.

the correlation from 1 to ~0.99 in a few hundred femtoseconds, followed by a slow decay of larger amplitude over the entire 1,000 ps. The initial subpicosecond drop in the correlation arises from the types of short-time reorientational motions shown in Fig. 3, involving brief periods of smooth motion punctuated by frequent changes in direction. Larger subpicosecond decays in correlation functions have been found in simulations of the reorientational dynamics of tyrosines (11) and tryptophans (9, 12) in proteins. It was also noted that at the shortest times the motion must be largely inertial in character, so that the slope of the correlation function should approach zero as $t \rightarrow 0$ (9). The 20-fs time resolution of the curves in Fig. 4 *a, c* is not sufficient to confirm this completely, but is adequate to identify an initial period

the orientation of the protein were not periodically reset to zero. Despite this, the total drift in the orientation of the protein during the entire simulation was less than 1° in each case. Prior to computing the correlation functions reported here, this small effect was cancelled for each of the 100,000 saved coordinate sets by using least-squares superposition of the instantaneous structure of the whole protein onto a fixed standard structure to correct the heme normal and C–O bond vectors.

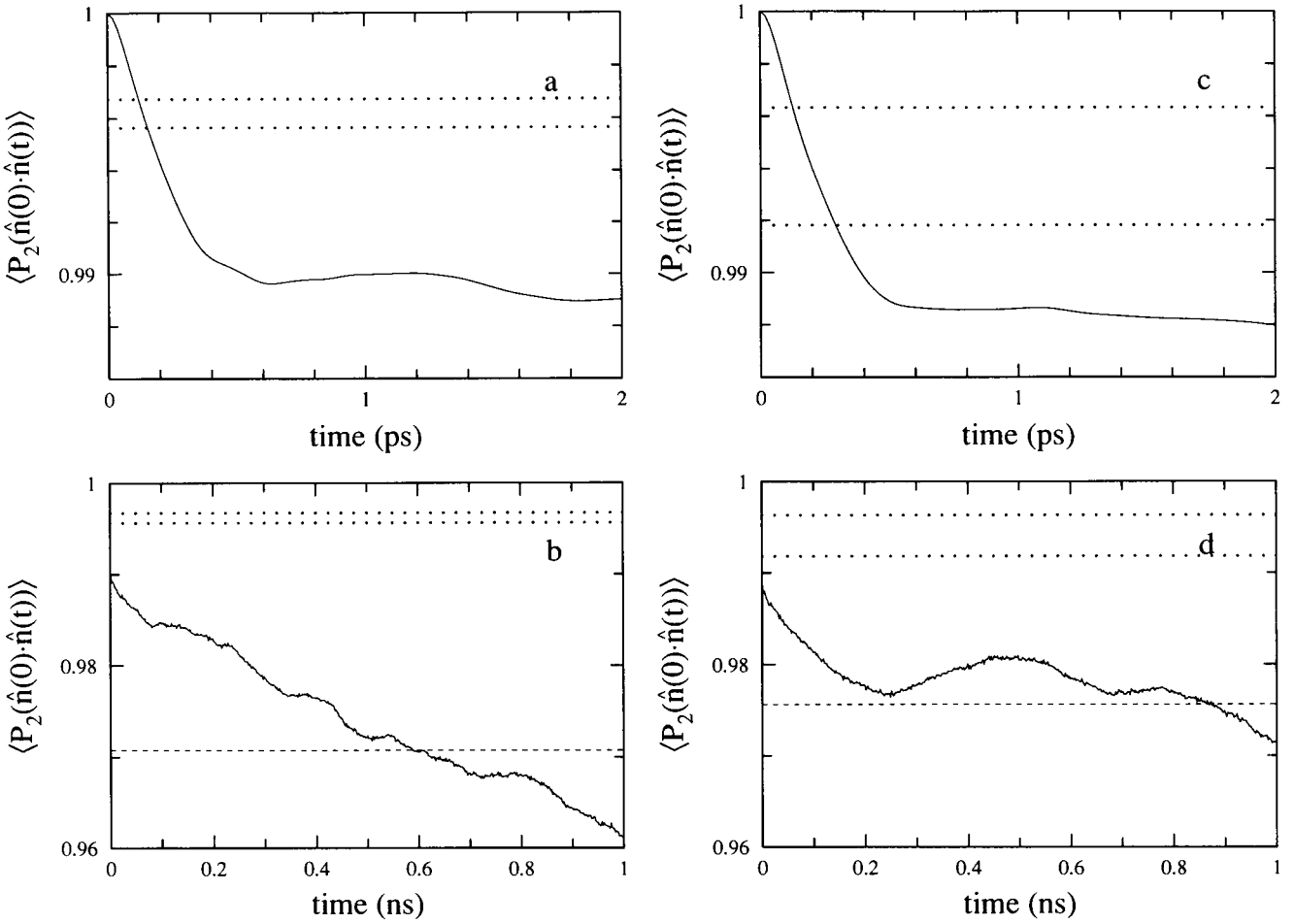

FIGURE 4 Correlation function for the orientation of the heme normal vector, calculated from complete 2-ns trajectories. *a, b*) MbCO simulation; *c, d*) Mb simulation. Panels *b* and *d* show the correlation functions to 1 ns, which is the longest interval for which a complete set of 50,000 terms contribute to the average in Eq. 3. In these panels the limiting values of the correlation functions estimated as described in the text using Eq. 5 are shown as horizontal dashed lines. Panels *a* and *c* show expanded views of the correlation functions in the first 2 ps in order to display the initial rapid decay. The dotted lines in each panel indicate the limiting values of the correlation functions calculated using Eq. 5 from "frozen globin" simulations of the corresponding system (see text).

of about 100 fs during which the second derivative of the correlation function is negative. This period corresponds well to the apparent mean period of free motion exhibited by the heme normal vector in Fig. 3. For a simple model description of the rapid reorientational motions as free diffusion within a cone of semiangle θ_0, the limiting value of the correlation function is given by $(\frac{1}{2}\cos\theta_0(1+\cos\theta_0))^2$ (13). A decay of the correlation to 0.99 then corresponds to an effective cone semiangle of ~5°, which provides a quantitative estimate of the short-time orientational mobility illustrated in Fig. 3.

The slow decay of the correlation function is caused by the tendency of the heme normal to explore different preferred orientations during the simulations, clearly shown in Fig. 2. In both simulations, but most clearly visible in the Mb simulation, this decay is non-monotonic. The roughly periodic oscillations superimposed on the decay in each case arise from a slight tendency for orientations to recur at the indicated intervals (e.g., ~0.5 and 0.8 ns for the Mb simulation). The larger overall decay in the correlation for the MbCO simulation may be explained in terms of the clear separation between the orientations in the early and late parts of the simulation, a feature not shared by the Mb simulation.

Comparison of the simulation with time-resolved optical experiments

Ansari et al. (1) have measured time-resolved absorption spectra of myoglobin in glycerol–water mixtures following photodissociation of bound carbon monoxide ligands, in which the probe beam was polarized both parallel and perpendicular to the photolysis beam. They found that the dependence of the absorption anisotropy on the degree of ligand photolysis agrees with the dependence predicted by theory (2), but the experimental values are 0.9 times the values predicted by the theory assuming that the heme is a perfect circular absorber. They attribute the differences between experimental and theo-

retical values to a combination of three possible effects: an out-of-plane component of the heme transition moment, rapid equilibrium fluctuations of the heme orientation on a subnanosecond time scale, and a subnanosecond reorientation of the heme plane in response to ligand dissociation.

One of the purposes of the simulations being reported here is to estimate the contribution of rapid equilibrium reorientations of the heme to lowering the initial anisotropy detected by nanosecond optical experiments. The decay of the absorption anisotropy due to the rotational diffusion of the protein at times that are long, compared to the time scale of internal motions of the chromophore, is given by Eq. 37 of (1), for a perfect circular absorber and in the limit of low incident light intensities:

$$r(t) = \frac{S^2}{10} \exp(-t/\tau_r). \quad (4)$$

Here τ_r is the rotational diffusion time of the molecule and S^2 is the squared generalized order parameter that incorporates the effects of rapid reorientational motions of the chromophore. S^2 is the value of the reorientational correlation function in Eq. 3 due to internal motions of the heme, evaluated at times which are short compared to the rotational diffusion time of the molecule. It would then be expected that this quantity could be estimated from these simulations as the value of the correlation function plotted in Fig. 4 extrapolated to a time of a few nanoseconds. However, such an extrapolation would not be meaningful because neither simulation has truly converged with respect to the regions of orientational space that it samples. The overall evolution in both cases is essentially unidirectional through a series of orientational regions, with little tendency to re-visit previous regions and allow the correlation function to reach a limiting value. The effects of incomplete sampling on correlation functions computed from molecular dynamics simulations have been considered recently by Chandrasekhar et al. (14).

Although the incomplete convergence of the simulations complicates a determination of the limiting value of the correlation functions by direct calculation using Eq. 3, it is possible to obtain an indirect estimate of the limiting value with certain assumptions. The major assumption is that the simulation in fact provides a complete sampling of the conformational space accessible to the system at equilibrium, with correct relative probabilities of occupying the different regions of this space. This would imply that a continuation of the simulation would simply explore the same regions in a random fashion, so that over a sufficiently long time the probability of finding the system in a specific state at a given time would be the same as the occupancy determined from the 2-ns simulation. Under these conditions the limiting value of the correlation function may be determined by extrapolating the correlation function in equation 3 to infinite time, which allows the limiting value to be written in terms of equilibrium occupancies of the accessible states (13):

$$S^2_\infty = \sum_i \sum_j p_i^{eq} p_j^{eq} P_2(\hat{n}_i \cdot \hat{n}_j). \quad (5)$$

Here both sums are over all accessible states, and p_i^{eq} and $\hat{n}_i$ are the equilibrium fractional occupancy and the heme normal vector, respectively, for state i. This expression may be evaluated from a simulation by treating each set of saved coordinates as a single state with occupancy p_i^{eq} = 1/(number of saved coordinate sets), and performing the double summation over all saved coordinate sets. The results calculated from 100,000 saved coordinate sets are 0.971 for the MbCO simulation and 0.976 for the Mb simulation.

Thus, these simulations provide an estimate of 0.97–0.98 for the squared order parameter for internal reorientational motions of the heme at equilibrium, including the effects of both the subpicosecond local reorientations and the orientational heterogeneity on a time scale of hundreds of picoseconds. It should be emphasized that this estimate required the restrictive assumption that the simulations produce all of the conformations accessible to the system. It is likely that a continuation of either simulation will produce new heme orientations, which will tend to reduce the squared order parameter estimated using Eq. 5. Therefore, it is reasonable to regard the numerical estimates from the 2-ns simulations as effective upper limits on the values that would be calculated from longer simulations produced in the same way. However, in comparing such estimates with the experimental results, it must be kept in mind that the effective temperature of these simulations is somewhat higher than the temperature at which the measurements were made (see the section above on General Features of the Simulations). The effect of the elevated temperature on the squared order parameter calculated using Eq. 5 cannot be determined without more complete knowledge of the angular distribution and relative stabilities of the accessible heme orientational states.

Analysis of ligand conformation and dynamics

The availability of a long molecular dynamics simulation of MbCO also allows questions relating to the orientation and dynamics of the CO ligand to be addressed. The specific question of the detailed configuration of the Fe–C–O triplet has been the subject of some debate. In the protein this triplet is distorted from a linear unit perpendicular to the heme plane, presumably by interactions with protein sidechains in the heme pocket, and this strain may affect the binding affinity of myoglobin for CO (6). Case and Karplus (15) used conformational energy calculations on a model system consisting of a CO ligand in the x-ray structure of metmyoglobin (in lieu of MbCO, for which a structure was not then avail-

able) to conclude that the most energetically favorable configuration of Fe–C–O in the protein is as a linear unit displaced from the heme normal by steric interactions with the protein. Hanson and Schoenborn (16) refined a neutron diffraction structure of MbCO with a linear Fe–C–O unit tilted at an angle of 24° with respect to the heme normal. Kuriyan et al. (6) refined the x-ray structure of MbCO at 260 K to 1.5 Å resolution with the oxygen of the CO ligand in two distinct, partially occupied positions. The Fe–C–O angles in the two conformations were 141° for the conformation with 78% occupancy and 120° for the conformation with 22% occupancy; the Fe–C bond was refined to a position very near the heme normal, although the position of the carbon atom could not be defined accurately. Moore et al. (17) used the technique of picosecond infrared-optical photoselection to deduce angles between the C–O bond vector and the heme normal of 20° and 35° in two distinct conformations, and combined these results with structural information from x-ray diffraction studies to produce Fe–C–O configurations in these two conformations which are both tilted and bent. (These authors also provide an extensive discussion of the various spectroscopic and structural studies performed to date that bear on the question of the Fe–C–O conformation in myoglobin and hemoglobin.)

In the current MbCO simulation, the orientational heterogeneity of the heme complicates a determination of CO positions that might be compared with results from an x-ray structure containing a stationary heme. To correct for the reorientational motions of the heme, an axis system fixed with respect to the heme was constructed, and the position of the CO ligand was calculated in this axis system, for each of the 100,000 saved coordinate sets of the simulation. The distribution of orientations of the C–O bond vector with respect to the heme normal in the heme-fixed axis system is shown in Fig. 5. The distribution is shifted to the lower left by about 3°. The distribution of orientations of the Fe–C bond vector (not shown) is shifted in the same direction by about 5°. (Both distributions are quite axially symmetric and show no evidence of multiple discrete orientations.) Therefore, a hypothetical x-ray diffraction measurement that probes the set of CO positions produced by this simulation would produce an Fe–C–O configuration which is tilted by 5° and bent from linear by about 2° (=5–3) in the opposite direction. This prediction of a nearly linear Fe–C–O almost perpendicular to the heme plane does not agree with the results of the structural studies (6, 16). It is possible that the drift from the x-ray structure that takes place during the preparatory and early analysis stages of the simulation is sufficient to relax much of the steric restraint from protein sidechains in the heme pocket that might force the ligand to assume a tilted and/or bent configuration.

The x-ray diffraction measurements probe the distribution of positions of the ligand. The infrared-optical

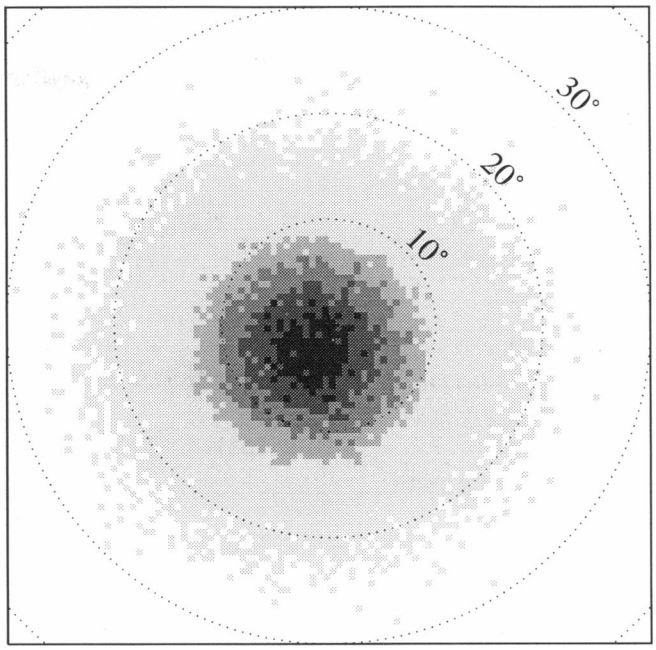

FIGURE 5 Distribution of directions of the ligand C–O bond vector in the MbCO simulation. For each of the 100,000 saved coordinate sets, a heme-fixed coordinate system was constructed with the instantaneous heme normal as the z axis. The x axis was defined as the unit vector parallel to the vector connecting the instantaneous positions of the "A" and "C" *meso* carbons, projected into the heme plane, and the y axis was defined as the outer product of the z and x axis vectors. The angles shown are the angle between the C–O bond vector transformed to the instantaneous heme-fixed frame and the direction normal to the heme plane, which is at the center of the plot. The shading of each 0.6° square in the plot reflects the number of coordinate sets in which the transformed C–O bond vector points in the corresponding direction; the lightest shading corresponds to 1–35 occurrences, and the darkest shading to greater than 125 occurrences.

photoselection measurements, on the other hand, directly probe the angle between the heme normal and the transition moment direction, which is parallel to the C–O bond vector. The infrared absorption anisotropy measured in these experiments may be written in the form (2, 18)

$$r(t) = -1/5C\langle P_2(\hat{n}_{CO}(t)\cdot\hat{n}(0^+))\rangle \qquad (6)$$

where $\hat{n}(0^+)$ is the unit vector normal to the heme plane immediately after excitation, $\hat{n}_{CO}(t)$ is the unit vector parallel to the C–O bond vector at time t, and C is a constant that incorporates absorption properties of the heme in its initial state at the excitation wavelength (cf. Eq. 2). The correlation function in this equation may be calculated from the simulation in the same way as the correlation function of the heme normal direction in Eq. 2. However, it should be noted that, unlike the correlation function of the heme normal, the computed value of this correlation function is not *de facto* invariant under reversal of the direction of time in the simulation. As discussed earlier, these simulations have apparently not

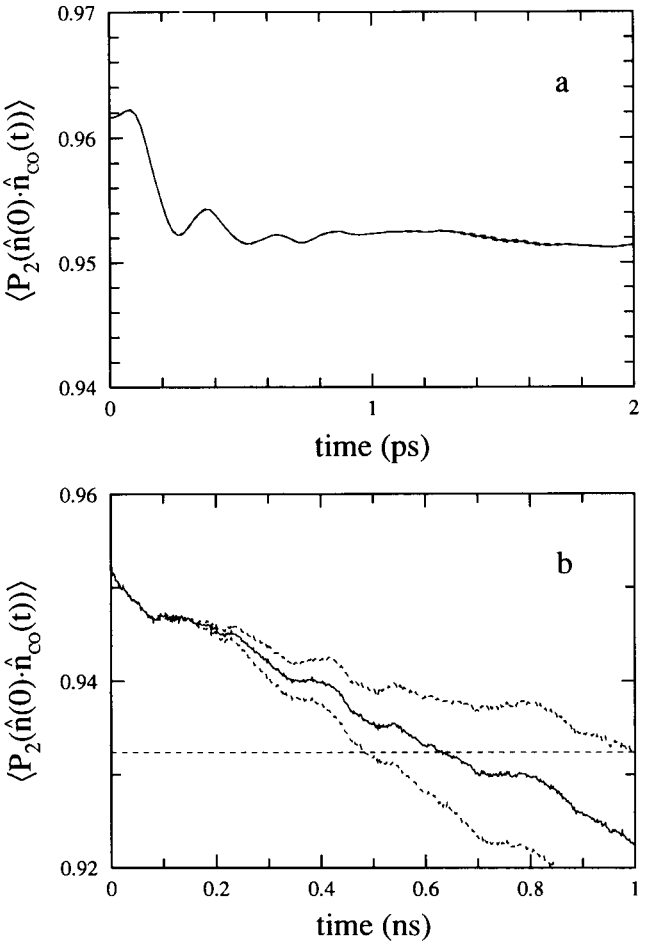

FIGURE 6 Correlation functions for the directions of the heme normal and the C–O bond vector, calculated from the complete 2-ns MbCO simulation. In each panel, the correlation functions calculated in the forward direction from the beginning of the simulation and in the reverse direction from the end of the simulation (see text) are shown as dashed lines, and the average is shown as a solid line. *a*) the 0–2 ps segment of the correlation functions, showing the initial rapid decay from ~0.962 to 0.952, on which is superimposed oscillations due to rapid relative motions of the heme normal and the C–O bond vector. *b*) the correlation functions to 1 ns. The horizontal dashed line indicates the limiting value of all three correlation functions estimated as described in the text using an expression analogous to Eq. 5.

converged with respect to the heme orientations (and most likely the C–O bond vector directions as well). Therefore, we would not *a priori* expect that correlation functions of these properties would be the same whether calculated from the simulation in the forward or in the reverse direction. This is confirmed in Fig. 6, which shows the correlation functions calculated in both directions as dashed lines. There is no physical basis for choosing between these two functions; the simplest "choice" is the average of the two, which is shown as the solid line in Fig. 6.

The first noteworthy feature of this correlation function is that it does not approach a value of 1 as t approaches 0. The value at $t = 0$ is equal to the average $\langle P_2(\cos(\alpha))\rangle$, where α is the instantaneous angle between the heme normal vector and the C–O bond vector. Because α is generally non-zero at any given time, the average $\langle P_2(\cos\alpha)\rangle$ over all times in the simulation will be less than 1. Indeed, this average represents the effective angle between the heme normal and the C–O bond vector that would be determined from an optical experiment having infinite time resolution, through the relation $P_2(\cos(\alpha_0)) = \langle P_2(\cos(\alpha))\rangle$. The value of α_0 calculated from this simulation is 9°. It is also clear that the correlation function does not approach a limiting value during the period shown, which is not surprising in light of the incomplete convergence of the simulation. It is still possible to estimate the limiting value of the correlation function—i.e., the squared order parameter of the relative motions of the heme normal and the C–O bond vector—in much the same way as was done earlier for the correlation function of the heme normal. This determination again requires the assumption that the simulation provides accurate occupancies for all accessible states, and makes use of an expression analogous to Eq. 5, with $\hat{n}_i \cdot \hat{n}_j$ replaced by $\hat{n}_i \cdot \hat{n}_{(CO)j}$. This yields a value of 0.932, which through the relation $P_2(\cos(\alpha_\infty)) = S^2_\infty$ yields an effective angle of 12° that would be determined from an optical experiment at long times.[2]

Although the detailed structural features of the simulation do not agree with the x-ray measurements, the dynamical information is helpful in interpreting the results of the optical-infrared photoselection experiments. We have seen that, because the two experiments probe the heme–CO system in different ways, the angle between the heme normal and the C–O bond vector deduced from the x-ray structure (predicted from the simulation to be at most a few degrees) will not agree with the value of the same angle deduced from an optical experiment (calculated to be ~10°). The relationship between the two measurements may be clarified for a simple case by the following analysis: In the approximation that the reorientations of the CO ligand with respect to the heme are independent of the reorientations of the heme itself (see footnote 2), we may write the correlation function in Eq. 6 as the product of separate factors incorporating the effects of the motions of the heme and the motions of the CO with respect to the heme (19), i.e.,

$$\langle P_2(\hat{n}(0)\cdot\hat{n}_{CO}(t))\rangle = \langle P_2(\hat{n}\cdot\hat{n}_{CO})\rangle_{eq}\langle P_2(\hat{n}(0)\cdot\hat{n}(t))\rangle \quad (7)$$

[2] It is interesting to note that this estimated order parameter is almost exactly equal to the estimated order parameter for the motions of the heme normal vector determined earlier, multiplied by the $t = 0$ value of the heme–CO correlation function, that is S^2_∞ (heme–CO) $\cong \langle P_2(\hat{n}\cdot\hat{n}_{CO})\rangle S^2_\infty$ (heme–heme). In addition, the heme–heme correlation function for the MbCO simulation (Fig. 4), scaled by $\langle P_2(\hat{n}\cdot\hat{n}_{CO})\rangle$, almost exactly superimposes on the average of the heme–CO correlation functions calculated in the forward and reverse directions (Fig. 6). These observations are consistent with the view that the reorientational motions of the CO ligand relative to the heme are essentially independent of the reorientational motions of the heme itself.

where $\langle \cdots \rangle_{eq}$ indicates an equilibrium average. Now consider a C–O bond vector which is undergoing reorientations that are axially symmetric about a direction vector **b** which lies at an angle β with respect to the heme normal. As the center of the orientational distribution of the C–O vectors, this direction vector represents the direction of the C–O bond that would be determined in a hypothetical x-ray diffraction measurement on this system. By the addition theorem for spherical harmonics (20), the equilibrium average in Eq. 7 may be written

$$\langle P_2(\hat{n} \cdot \hat{n}_{CO}) \rangle_{eq} = \frac{4\pi}{5} \sum_{m=-2}^{2} Y_2^{m*}(\Omega_n) \langle Y_2^m(\Omega_{CO}) \rangle_{eq} \quad (8)$$

where Ω_n is the (fixed) orientation of the heme normal vector in the coordinate frame defined by the vector **b**, and Ω_{CO} is the orientation of the C–O bond vector in the same frame. Because the reorientations of this vector are axially symmetric in this frame, only the average for $m = 0$ is non-zero, and the sum reduces to

$$\langle P_2(\hat{n} \cdot \hat{n}_{CO}) \rangle_{eq} = \frac{4\pi}{5} Y_2^{0*}(\Omega_n) \langle Y_2^0(\Omega_{CO}) \rangle_{eq}$$

$$= P_2(\cos \beta) \langle P_2(\hat{\mathbf{b}} \cdot \hat{n}_{CO}) \rangle_{eq}. \quad (9)$$

It is clear from Eqs. 7 and 9 that the effective angle between the heme normal and the C–O bond vector that would be measured in an infrared-optical photoselection experiment on this system will differ from the angle β determined by x-ray diffraction by an amount that depends on both the reorientational dynamics of the heme itself and the reorientational dynamics of the C–O bond vector with respect to the heme normal as reflected in the average on the right-hand side of Eq. 8. Both correction factors may be estimated from the present MbCO simulation. First, the contribution from the heme motions is just the correlation function plotted in Fig. 4, evaluated at a time determined by the time resolution of the experiment. Second, the equilibrium average on the left side of Eq. 9 is the value at $t = 0$ of the correlation function shown in Fig. 6, which was calculated from a nearly axially symmetric distribution of C–O bond vector directions the center of which (the direction **b**) lies about 3° from the heme normal (Fig. 5). Thus, the equilibrium average on the right-hand side of Eq. 9 calculated in the coordinate frame of this vector would be this zero-time value divided by $P_2(\cos 3°)$.

Locke et al. (19) have performed infrared-optical photoselection measurements with 200-fs time resolution on carbonmonoxyhemoglobin. They consider a number of possible reasons for the difference between their estimate of 18° for the angle between the C–O bond vector and the heme normal and the angle of 11–13° determined by x-ray diffraction. We wish to use Eqs. 7 and 9 to estimate the contributions of heme and CO reorientational dynamics to this difference, while keeping in mind that the dynamical quantities are calculated for myoglobin rather than hemoglobin. To proceed, we choose $\beta = 12°$ as the value of the angle measured by x-ray diffraction, and we assume that the reorientational dynamics of the C–O bond vector about this direction are the same as the simulated reorientational dynamics about the axial direction nearly parallel to the heme normal (Fig. 5). We can then take the equilibrium average on the right-hand side of Eq. 9 to be the zero-time value of the curve in Fig. 6 *a* (~0.962) divided by $P_2(\cos 3°)$, and a value of 0.99 for the heme–heme correlation function (Eq. 7) evaluated from Fig. 4 *a* at $t = 200$ fs. We arrive at a value of the angle α for the optical experiment given by $P_2(\cos \alpha) = 0.894$, or $\alpha = 15.4°$. Therefore, about half of the difference between the optical and x-ray determinations of the angle may be accounted for by the reorientational dynamics of the heme and CO estimated from the simulation.

Evolution of the protein conformation

It is clear from the evolution of the r.m.s. deviations of the dynamical structures from the x-ray structure (Fig. 1) that some changes in the conformation of the protein take place in at least the early parts of both simulations. However, the r.m.s. deviations of a set of structures from a single reference structure are not necessarily sensitive to pairwise differences between structures in the set, that is, two structures with similar r.m.s. deviations are not necessarily similar to each other. Considerably more information is contained in the pairwise distance matrix for a set of structures, defined by

$$D_{ij} = \sqrt{\frac{\sum_{\alpha} |(\mathbf{r}_\alpha)_i - (\mathbf{r}_\alpha)_j|^2}{n_\alpha}} \quad (10)$$

where $\mathbf{r}_\alpha$ is the coordinate vector of atom α, and the coordinate index α ranges over the set of atoms (e.g., backbone atoms or alpha carbons) chosen as the basis for the pairwise comparison of structures, and n_α is the number of atoms in this set. This description of a set of N structures taken from a dynamics simulation or any other source effectively represents each structure in the set as a $N - 1$-dimensional vector consisting of the distances of this structure from all the other structures in the set. Despite the conceptual simplicity of this description, direct visualization of the evolution of structures represented in this manner becomes prohibitively difficult for even small numbers of structures. In order to facilitate the visualization of the information contained in the distance matrix, Levitt (21) proposed a method for reducing the dimensionality of this representation by means of least-squares fitting. Specifically, an n-dimensional representation of a set of N structures ($n < N$) requires that each structure be represented as an abstract n-vector (x_1, . . . , x_n). The embedding of these structures into the n-dimensional space is accomplished by choosing the components of these vectors so that the $N \times N$ distance matrix calculated from the vectors is an optimal least-

squares approximation to the actual distance matrix, that is, so that the quantity

$$\sum_{i=1}^{N}\sum_{j=1}^{N}\left(D_{ij}-\sqrt{\sum_{k=1}^{n}((x_k)_i-(x_k)_j)^2}\right)^2 \qquad (11)$$

is a minimum with respect to all the $\{(x_k)_i: k = 1, \ldots, n; i = 1, \ldots, N\}$. The set of N optimal n-vectors is only defined in terms of distances among the vectors and therefore can only be determined to within an arbitrary uniform translation and/or rotation of all the vectors in the n-dimensional space. Clearly the representation with $n = N - 1$ is exact—i.e., the quantity (11) will be zero for the optimal set of vectors, and in general the error of the representation will increase as the dimensionality decreases. However, the choice of n is dictated primarily by the feasibility of displaying the resulting n-vectors for visualization purposes. By this criterion, only $n = 2$ and $n = 3$ are practical—$n = 2$ is the straightforward display of ordered pairs (x_1, x_2) in a plane, and $n = 3$ is the somewhat less convenient display of ordered triplets (x_1, x_2, x_3) in a volume.

A two-dimensional ($n = 2$) representation of a simulation was produced, starting from the 200 10-ps averaged coordinate sets (see above). The 200×200 distance matrix D was calculated from these sets using Eq. 10, with the r.m.s. distances between coordinate sets calculated for the set of alpha carbons which lie in helical regions of the protein. This restricted set of 121 atoms was chosen to reflect the overall structure of the backbone of the protein while suppressing the effects of the relatively rapid and uncorrelated motions of sidechain atoms and atoms in inter-helical regions. An optimal set of 200 coordinate pairs (x_1, x_2), one pair corresponding to each 10-ps average, was then determined by minimizing the error (Eq. 11) using a conjugate-gradient algorithm. In general, the final values of the 400 parameters produced by the minimization depended on the choices of initial values, with the procedure producing a number of distinct local minima. Therefore, a "best" set of optimal parameters was produced for each simulation by performing a large number (of the order of 100) of minimizations starting from randomly chosen values of all the parameters, and selecting the final set of parameters that corresponded to the smallest value of the error (Eq. 11).

The resulting best sets of coordinate pairs for the two simulations are shown in Fig. 7, with the evolution of the conformations indicated by straight lines connecting points corresponding to consecutive 10-ps averages. The distances between pairs of points in these figures approximate the actual r.m.s. distances between corresponding coordinate sets with a mean error of 0.11 Å for the MbCO simulation and 0.14 Å for the Mb simulation.[3] In both cases the protein conformation shows a tendency to "wander" throughout the simulation. In the first 200 ps of the MbCO simulation the conformation moves rapidly away from the starting point toward a new region where it remains "confined" within about a 0.3 Å radius for the next 500 ps. Between 700 and 1000 ps, the conformation again migrates fairly steadily toward a new region, in which it remains quite localized for the next 600 ps. Between 1600 and 1700 ps the conformation again moves to a new region, where it remains for the rest of the simulation. The Mb simulation also exhibits a large initial conformational change during the first 300 ps; it remains localized for the next 100 ps, then undergoes a second large change between 400 and 600 ps. For the remainder of the simulation, the conformation occupies several distinct regions, with periods of tight localization (roughly 600–850 ps, 1200–1400 ps, and 1700–2000 ps) connected by periods of relatively large drift. In both cases the early conformational changes correlate quite well with the initial increases in the r.m.s. deviation of the structure from the x-ray structure (Fig. 1), after which the changes proceed in a direction largely "perpendicular" to the initial drift, with relatively little effect on the r.m.s. deviation.

Protein conformational changes correlate with heme reorientational motions

A comparison of the two-dimensional mappings of the protein conformation in Fig. 7 with the displays of the heme normal vectors in Fig. 2 reveals that at a time resolution of 100 ps or so the features of the protein conformational evolution match quite closely the evolution of the heme orientation. In the MbCO simulation, the periods of localized reorientational motions during the 200–800 ps, 1000–1600 ps, and 1700–2000 ps segments coincide with periods of localized conformational evolution, and the initial large change in orientation during 0–200 ps as well as the transitions between orientational regions correlate quite well with significant changes in protein conformation. In the Mb simulation, the large "back-and-forth" motion of the heme normal vector during the first 600 ps coincides with a large change in the protein conformation; presumably the initial conformational drift during the 0–300 ps period allows the heme to undergo a large reorientation, and the subsequent further conformational change during the 400–600 ps period allows the heme normal vector to return close to its initial position. The gross features of the further evolution of the heme orientation in this simulation correspond to those of the evolution of the protein conformation, although the relatively small magnitude of the heme orientational changes makes it more difficult

[3] To determine the dependence of these deviations on the dimensionality of the embedding space, we also produced representations in terms of coordinate triplets (x_1, x_2, x_3) by the same method. The "best" such three-dimensional representation yielded an average distance matrix deviation of 0.065 Å for the MbCO simulation and 0.087 Å for the Mb simulation.

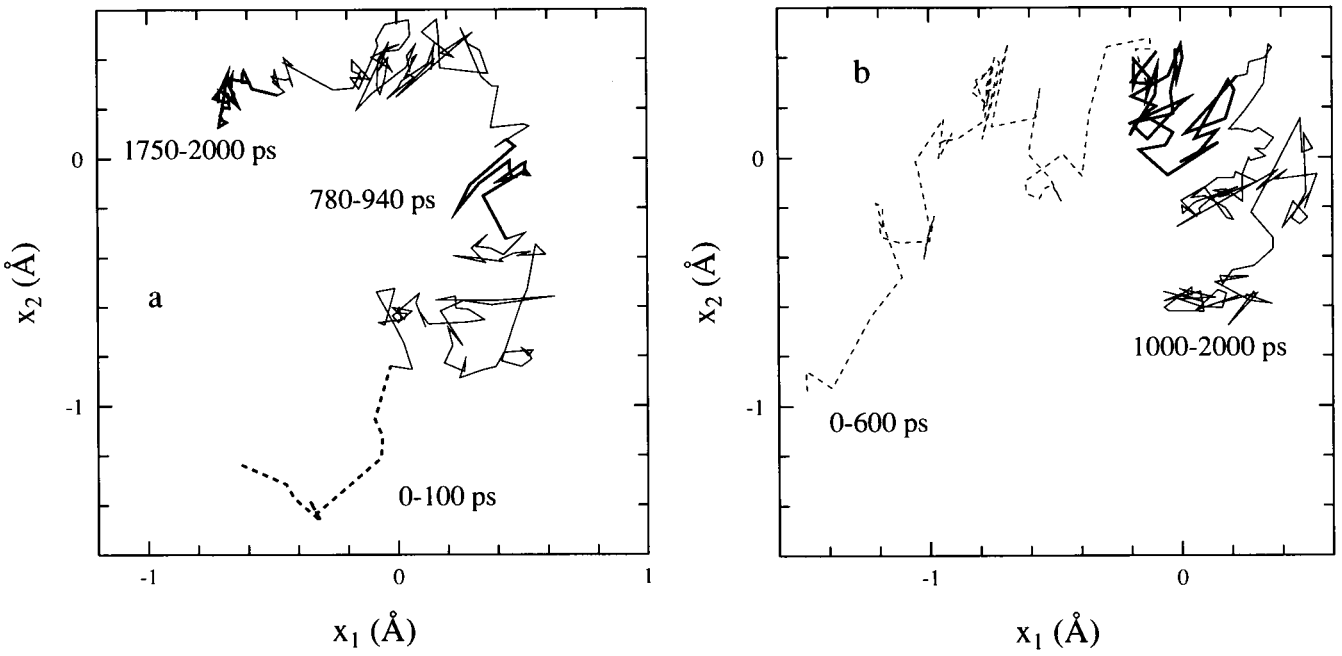

FIGURE 7 Mapping of the evolution of the protein conformation onto a 2-dimensional space. Abstract coordinate pairs (x_1, x_2) corresponding to each of the 200 10-ps averaged coordinate sets were calculated by minimizing the differences between the matrix of pairwise distances between abstract coordinate pairs and the matrix of pairwise distances between coordinate sets, calculated for the 121 alpha carbon atoms in helical regions of the protein (see text). Segments from different parts of each simulation are distinguished through the use of dashed lines and solid lines of different weights, and the correspondences between line types and times within a simulation are identical to those used in the corresponding panel of Fig. 2. *a*) MbCO simulation. The average deviation between the pairwise distances calculated from this representation and the actual pairwise distances is 0.11 Å. *b*) Mb simulation. The average deviation between the distance matrices in this representation is 0.14 Å.

to delineate regions of localized motion in this case than in the case of the MbCO simulation.

These observations suggest that the heme orientation, as monitored by the heme normal, is sensitive to the overall conformation of the protein. This is not unexpected, because at least two helices (E and F) are in van der Waals contact with the heme ring, and a conformational change involving changes in the relative positions and/or orientations of these helices might also require the heme to reorient. The degree of coupling between heme reorientations and the motions of protein atoms may be demonstrated by considering the heme dynamics in a system in which the protein atoms are not permitted to move. To this end, two coordinate sets from different parts of each simulation were used as starting structures for new 2-ns simulations, which were prepared using the procedure described in Methods and in which only the heme atoms were allowed to move. Reorientational correlation functions (not shown) calculated from these simulations show that the heme motions are significantly restricted by freezing the globin. In all but one case the correlation function decays within a few hundred femtoseconds to levels very close to its theoretical limiting value estimated using Eq. 5, and exhibits no additional decays on longer time scales. This behavior indicates that these "frozen globin" simulations are essentially converged with respect to the heme reorientations, with no significant appearance of incompletely sampled slow motions.[4] The limiting values (shown in the corresponding panels of Fig. 4 for comparison) were 0.996 and 0.997 for the liganded, and 0.992 and 0.996 for the unliganded simulations, reflecting overall decays that are smaller than the rapid decays seen in the full simulations. Therefore, the protein motions in these simulations play a significant role in even the short-time orientational dynamics of the heme.

The nature of the protein conformational change

The analysis in Fig. 7 represents the differences between any two structures using a single r.m.s. distance and therefore does not resolve contributions to this difference from different parts of the molecule. Additional analysis is required in order to determine which parts of the molecule typically undergo the largest displacements during the conformational changes shown in Fig. 7. We proceed by selecting for each simulation two periods, well separated in time such that the conformations do not wander significantly during each period, and calcu-

[4] During one of the unliganded simulations there was a single apparent change in average orientation of about 5° amplitude which produced an additional decay in the reorientational correlation function. The value of the correlation function at 1,000 ps was 0.99, reflecting an overall decay which is still less than half that calculated from the full liganded simulation (Fig. 4 *b*).

lating the average structure and the atom-by-atom r.m.s. fluctuations about the average for each period. The net atomic displacements that occur between the two periods are calculated from the atom-by-atom scalar differences between the average structures. Those atoms whose displacements are large compared to their r.m.s. fluctuations in position during either period are considered to have made a significant contribution to the conformational change that takes place between the two periods. The results of this analysis are shown in Fig. 8. In both cases there are clearly distinct parts of the protein sequence that make major contributions to the conformational differences. In the MbCO simulation, the backbone atoms in the D and E helices on the distal side of the heme undergo the largest displacements. On the other hand, in the Mb simulation these helices are relatively fixed, and parts of the F and G helices and E–F and G–H interhelical regions experience the largest shifts.

It was suggested above that the heme orientation is coupled somehow with the protein conformation in these simulations, based on similarities in the temporal evolutions of the heme orientation and the protein conformation. A possible corollary of such a coupling would be a tendency for parts of the protein close to the heme to experience larger shifts during a conformational change than more distant parts of the protein—i.e., there would be an inverse relationship between the displacements of atoms shown in Fig. 8 and the distance of these atoms from the heme. In order to test this possibility, the distance of a backbone atom from the heme was defined as simply the distance between this atom and the heme iron atom in the x-ray structure of the molecule in question, and linear correlation coefficients were evaluated between the reciprocal distances for all backbone atoms and the displacements of these atoms shown in Fig. 8. For MbCO the value of this correlation coefficient is 0.19, which for $N = 459$ backbone atoms indicates a statistically significant correlation. When only the 286 backbone atoms on the distal side of the heme are included, the correlation coefficient increases to 0.41. When the averaged coordinates in the final conformation, rather than the crystallographic coordinates, are used both to calculate the distances of atoms from the heme iron and to identify distal atoms, the correlation coefficient for the resulting set of 312 distal backbone atoms is 0.55. Thus, in this simulation there is a clear relationship for the backbone atoms between the distance from the iron atom and the magnitude of the displacement during the conformational change, particularly for atoms on the distal side of the heme where the displacements are generally larger. For Mb the overall correlation coefficient has the very similar value of 0.20. In this case, however, the correlation is essentially zero when evaluated for the distal atoms identified from the crystallographic structure, which is not surprising because there is relatively little movement of these atoms in the conformational change. On the other hand, for the set of 166 backbone atoms on the *proximal* side of the heme the correlation coefficient has the value 0.41. This qualitative difference between the two simulations, with the conformational heterogeneity focused on distal atoms in the MbCO simulation and proximal atoms in the Mb simulation, may arise directly (through steric interactions in the heme pocket) or indirectly (through differences in the heme potential function) from the presence of the CO ligand in only one of the simulations, but additional simulations of each molecule would be required to demonstrate this connection unequivocally.

CONCLUDING REMARKS

A major objective in performing these calculations has been to estimate the contribution of rapid internal motions of the heme group to the reduction of the initial absorption anisotropy detected in nanosecond optical experiments on myoglobin. The simulations have evidently not converged for purposes of evaluating the heme orientational dynamics on all time scales. However, the correlation function of the heme normal during the first few picoseconds (Fig. 4 *a, c*) has apparently nearly converged, and a limiting value ($\sim$0.99) may be discerned. To the extent that molecular dynamics is able to describe the motions of the heme on the shortest time scales, this result provides a reliable *lower* limit of about 10% for the fractional contribution of rapid heme motions to the observed reduction in anisotropy, even in the absence of any additional reorientations on longer time scales. We have estimated from the set of heme orientations provided by the simulations that the infrequent large-scale motions account for an additional 20% of the anisotropy reduction. The detailed features of these motions derived from the simulations are not statistically meaningful, but their overall character suggests that the protein is able to accommodate multiple heme orientations which differ by $\sim$5–10° within a 2-ns period. The actual set of such orientations is probably larger than we have seen in these simulations, so that—subject to the caveat given above concerning the elevated temperatures of the simulations—the present estimate of the contribution of the infrequent transitions to the anisotropy reduction may reasonably be regarded as a conservative lower limit.

From this work the picture has arisen of the heme group as a major participant in, and perhaps a useful monitor of the overall dynamics of the protein. Whatever the detailed mechanism of the apparent coupling between the heme orientation and protein conformation, the suggestion from these simulations that the heme orientation may be a reporter of protein conformational changes provides a possible direct probe of structural dynamics in heme proteins. On a picosecond time scale, a time-resolved linear dichroism experiment could monitor the reorientation of a photoselected population of photolyzed deoxyhemes before the orientations

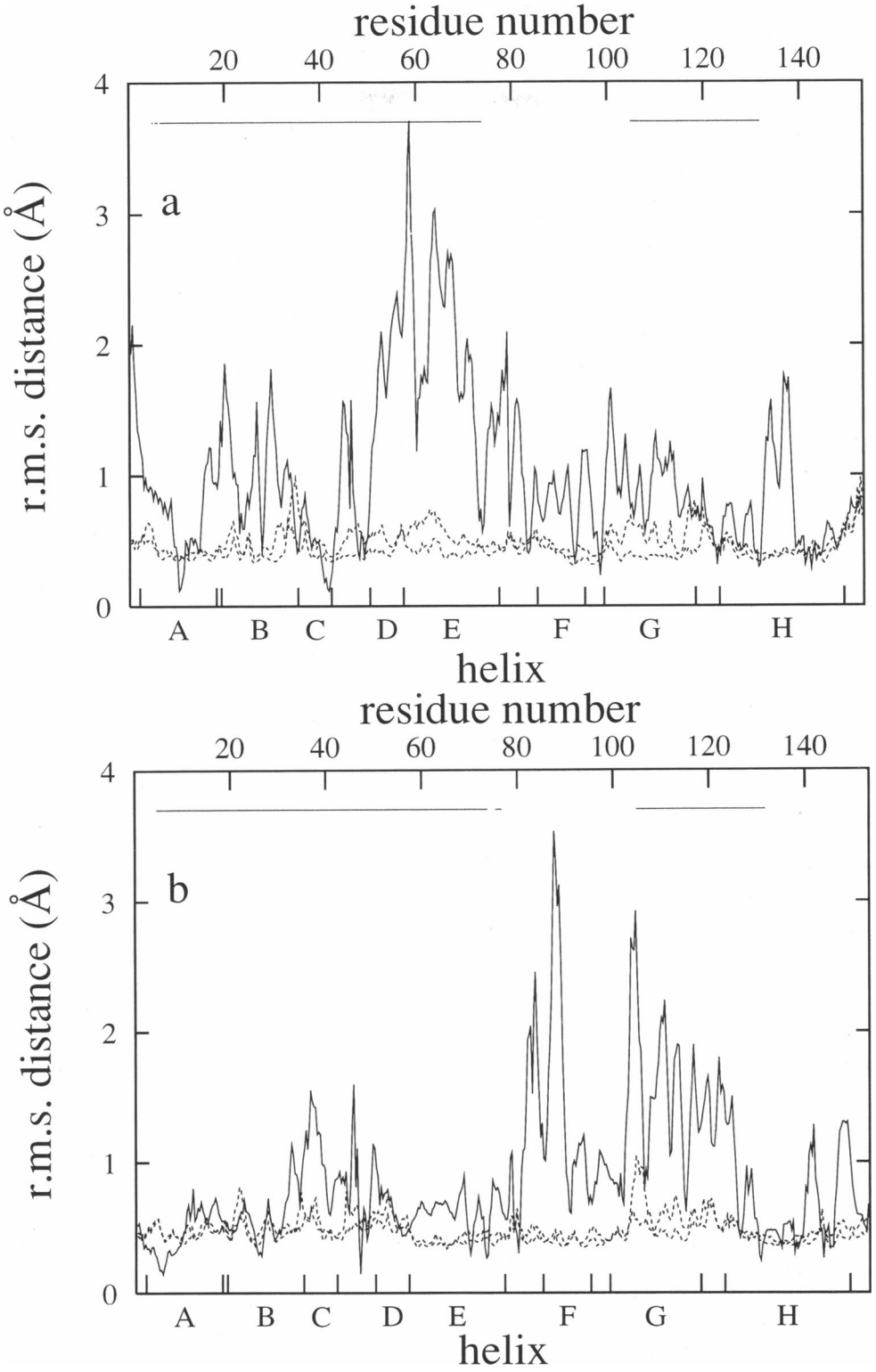

FIGURE 8 Atomic displacements contributing to conformational changes in the simulations. In each case, two 200-ps segments of the simulation were selected which are well separated in time, and the average structure and r.m.s. fluctuations about this average were calculated for each segment. The atom-by-atom distances between the average structures from the two segments are shown for all backbone atoms as a solid line, and the fluctuations during both segments are shown for comparison as dashed lines. (An additional point of comparison is obtained by performing a least-squares superposition of the backbone atoms from the x-ray structures of Mb and MbCO. All differences between the superimposed structures are less than 0.6 Å, except for residues very close to the NH_2 and COOH termini.) The horizontal line near the top of each panel indicates those parts of the backbone which are on the proximal side of the heme in the x-ray structure. (A proximal atom is defined as one whose displacement vector from the iron atom lies at an angle of less than 90° from the heme normal vector pointing in the direction of the proximal histidine residue.) *a*) MbCO simulation. The 300–500 and 1,800–2,000 ps segments of the simulation were used in the comparison. *b*) Mb simulation. The 220–420 and 1,800–2,000 ps segments of the simulation were used in the comparison.

are randomized by rotational diffusion of the protein. Such reorientations would be sensitive indicators of structural events taking place in the surrounding protein in response to ligand dissociation.

These simulations have also highlighted the question of conformational heterogeneity in proteins. The complexity of the potential energy surface for proteins, and the corollary existence of large numbers of nearly isoenergetic conformational substates, has been the subject of intensive experimental and theoretical research (for review see reference 22). This complexity is likely to be an intrinsic property of large inhomogeneous molecular systems, independent of the detailed description of interatomic interactions that is employed. Thus, it is reasonable to speculate that all molecular dynamics simulations of complex molecular systems will exhibit multiple conformational states on some length scale. For example, Levitt (21) identified four distinct regions of conformational space (on a distance scale of ~0.5 Å r.m.s. displacement) in a 132-ps simulation of bovine pancreatic trypsin inhibitor. Multiple distinct conformations were also observed by Daggett and Levitt (23) in a 200-ps simulation of the COOH-terminal fragment of the L7/L12 ribosomal protein in solution. Elber and Karplus (24) estimated that approximately 2,000 potential energy minima were sampled in the 300-ps simulation of myoglobin by Levy et al. (10). As discussed earlier and displayed in Fig. 7, the present simulations clearly exhibit multiple conformations on relatively large time and distance scales (hundreds of picoseconds and ~1 Å). It is also straightforward to demonstrate finer-grained heterogeneity: Energy minimization of 100 individual coordinate sets selected from the MbCO simulation at 20-ps intervals results in 100 distinct energy minima; the r.m.s. differences between the backbone conformations of consecutive minima vary between 0.1 and 1 Å. The energy distribution of these minimized conformations is very similar to that reported by Elber and Karplus (24). This unfolding of the complexity of the potential energy surface is likely to continue if the analysis is extended to shorter time scales (24).

The most significant shortcoming of these simulations is that they neglect the effects of solvent; with the current state of simulation technology a simulation of comparable length for myoglobin with a proper complement of water molecules is not yet practical. It is difficult to ascertain the likely effects of solvent on the dynamical properties being studied here. On the one hand, there is experimental evidence that the dynamics of protein conformational changes are influenced by solvent viscosity, even at viscosities close to that of water (25), and it is reasonable to expect the presence of water molecules to influence the degree of participation of surface residues in the conformational dynamics. On the other hand, a simulated protein in water is coupled to a thermal bath, which might enable more extensive exploration of conformational space than would be seen in a vacuum simulation (26). The present simulations have exposed possible features of the dynamics of myoglobin that would not be apparent from shorter simulations, and we expect that extending this type of calculation both to longer times and to a more complete description (i.e., including solvent) will only add to the set of possible dynamical behaviors of these systems.

The author is indebted to Dr. Attila Szabo for many enlightening discussions, and to Dr. William Eaton for motivating this work. The author would also like to thank the Computer Center Branch of the Division of Computer Research and Technology at the National Institutes of Health for generous access to their resources during the course of this study, and one of the reviewers for suggesting the "frozen globin" simulations.

Received for publication 4 August 1992 and in final form 15 October 1992.

REFERENCES

1. Ansari, A., C. M. Jones, E. R. Henry, J. Hofrichter, and W. A. Eaton. 1992. Photoselection in polarized photolysis experiments on heme proteins. *Biophys. J.* 64:852–868.
2. Ansari, A., and A. Szabo. 1992. Theory of photoselection by intense light pulses: The influence of reorientational dynamics and chemical kinetics on absorbance measurements. *Biophys. J.* 64:838–851.
3. Henry, E. R., M. Levitt, and W. A. Eaton. 1985. Molecular dynamics simulation of photodissociation of carbon monoxide from hemoglobin. *Proc. Natl. Acad. Sci. USA.* 82:2034–2038.
4. Brooks, B. R., R. E. Bruccoleri, B. D. 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.
5. Reiher, W. E. 1985. Ph.D. thesis, Harvard University, Cambridge, MA. 423 pp.
6. Kuriyan, J., S. Wilz, M. Karplus, and G. A. Petsko. 1986. X-ray structure and refinement of carbon-monoxy (Fe II)-myoglobin at 1.5 Å resolution. *J. Mol. Biol.* 192:133–154.
7. Beeman, D. 1976. Some multistep methods for use in molecular dynamics calculations. *J. Comp. Phys.* 20:130–139.
8. Henry, E. R., W. A. Eaton, and R. M. Hochstrasser. 1986. Molecular dynamics simulations of cooling in laser-excited heme proteins. *Proc. Natl. Acad. Sci. USA.* 83:8982–8986.
9. Henry, E. R., and R. M. Hochstrasser. 1987. Molecular dynamics simulations of fluorescence polarization of tryptophans in myoglobin. *Proc. Natl. Acad. Sci. USA.* 84:6142–6146.
10. Levy, R. M., R. P. Sheridan, J. W. Keepers, G. S. Dubey, S. Swaminathan, and M. Karplus. 1985. Molecular dynamics of myoglobin at 298°K: Results from a 300-ps computer simulation. *Biophys. J.* 48:509–518.
11. Levy, R. M., and A. Szabo. 1982. Initial fluorescence depolarization of tyrosines in proteins. *J. Am. Chem. Soc.* 104:2073–2075.
12. Ichiye, T., and M. Karplus. 1983. Fluorescence depolarization of tryptophan residues in proteins: A molecular dynamics study. *Biochemistry.* 22:2884–2893.
13. Lipari, G., and A. Szabo. 1982. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* 104:4546–4559.
14. Chandrasekhar, I., G. M. Clore, A. Szabo, A. M. Gronenborn, and

B. R. Brooks. 1992. A 500 ps molecular dynamics simulation study of interleukin-1β in water: Correlation with nuclear magnetic resonance spectroscopy and crystallography. *J. Mol. Biol.* 226:239–250.

15. Case, D. A., and M. Karplus. 1978. Stereochemistry of carbon monoxide binding to myoglobin and hemoglobin. *J. Mol. Biol.* 123:697–701.
16. Hanson, J. C., and B. P. Schoenborn. 1981. Real space refinement of neutron diffraction data from sperm whale carbonmonoxymyoglobin. *J. Mol. Biol.* 153:117–146.
17. Moore, J. N., P. A. Hansen, and R. M. Hochstrasser. 1988. Iron-carbonyl bond geometries of carboxymyoglobin and carboxyhemoglobin in solution determined by picosecond time-resolved infrared spectroscopy. *Proc. Natl. Acad. Sci. USA.* 85:5062–5066.
18. Hansen, P. A., J. N. Moore, and R. M. Hochstrasser. 1989. Determination of the iron–carbonyl bond geometry of carboxyprotoheme in solution using picosecond infrared-optical photoselection. *Chem. Phys.* 131:49–62.
19. Locke, B., T. Lian, and R. M. Hochstrasser. 1991. Determination of Fe–CO geometry and heme rigidity in carbonmonoxyhemoglobin using femtosecond IR spectroscopy. *Chem. Phys.* 158:409–419.
20. Brink, D. M., and G. R. Satchler. 1961. Angular momentum. 2nd ed. Oxford University Press, Oxford, UK. 160 pp.
21. Levitt, M. 1983. Molecular dynamics of native protein. II. Analysis and nature of motion. *J. Mol. Biol.* 168:621–657.
22. Frauenfelder, H., S. G. Sligar, and P. G. Wolynes. 1991. The energy landscapes and motions of proteins. *Science (Wash. DC).* 254:1598–1603.
23. Daggett, V., and M. Levitt. 1991. A molecular dynamics simulation of the C-terminal fragment of the L7/L12 ribosomal protein in solution. *Chem. Phys.* 158:501–512.
24. Elber, R., and M. Karplus. 1987. Multiple conformational states of proteins: A molecular dynamics analysis of myoglobin. *Science (Wash. DC).* 235:318–321.
25. Ansari, A., C. M. Jones, E. R. Henry, J. Hofrichter, and W. A. Eaton. 1992. The role of solvent viscosity in the dynamics of protein conformational changes. *Science (Wash. DC).* 256:1796–1798.
26. Cooper, A. 1984. Protein fluctuations and the thermodynamic uncertainty principle. *Prog. Biophys. Mol. Biol.* 44:181–214.