Reprinted from the Journal of the American Chemical Society, 1990, *112*.

Enhanced Sampling in Molecular Dynamics: Use of the Time-Dependent Hartree Approximation for a Simulation of Carbon Monoxide Diffusion through Myoglobin†

R. Elber‡,§ and M. Karplus*,‡

Contribution from the Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138, and Department of Chemistry, University of Illinois at Chicago, Chicago, Illinois 60680. *Received April 30, 1990*

Abstract: To obtain enhanced sampling in the study of carbon monoxide motion through myoglobin, a classical version of the time-dependent Hartree approximation is introduced. The method is derived from the Liouville equation by separating the system of interest into two parts, each of which moves in the average field of the other. In the application to myoglobin, the method makes it possible to treat a swarm of ligand molecules in the presence of a *single* trajectory for the protein. This results in a calculation that is approximately a factor of N faster than the N separate protein-ligand trajectories that would have to be used in standard simulations. Corresponding savings in time can be expected in other applications of the method to appropriate condensed phase problems. The enhanced sampling (60 CO molecules with initial positions in the heme pocket) make it possible to find widely different pathways for the escape of CO from the protein. Most of the pathways involve internal cavities that have been observed in an X-ray structure of Xe saturated myoglobin. The individual trajectories spend most of the time in the cavities; the transitions between cavities are rare events that are rapid and involve the crossing of barriers. By a comparison with the results obtained with 60 high-temperature ligands in a room-temperature protein and in a rigid protein, it is shown that even high-temperature ligands are prevented from escaping in the latter. Thus, the present results confirm the conclusion from earlier work that protein fluctuations are essential for the escape of ligands. The most important exit routes are concentrated in the region between the A, B, and E helices. Others involve the CD corner, on the proximal side of the heme and between B and G helices. A short exit path near the distal histidine found in previous simulations and supported by mutation studies is important only when the fluctuations of side chains are enhanced by increasing their temperature. This suggests that the dominant ligand pathway through the protein may depend on the system temperature.

I. Introduction

The oxygen storage and transport proteins myoglobin and hemoglobin have a binding pocket at the heme that protects the iron from the solvent in the absence of ligand.[1,2] This fact is of biological importance since water can oxidize the ferrous heme iron into the inactive ferric state.[1-3] However, as was already noted in the original X-ray structure of myoglobin,[4] the protein environment of the heme, if rigid, would prevent the entrance and the exit of ligands (e.g., oxygen, carbon monoxide) that are of approximately the same size as a water molecule. More recent high-resolution X-ray structures have confirmed that there is no obvious pathway for ligand escape in both liganded and unliganded myoglobin.[5-7] Further, empirical energy function calculations based on the crystal structures have shown that possible pathways in the neighborhood of the heme pocket have barriers on the order of 100 kcal/mol in the rigid protein;[8] such barriers would lead to escape times on the order of 10^{60} s at room temperature, infinitely long on the biological time scale. Thus, for myoglobin protein fluctuations are involved in the biological function, which requires that oxygen be able to enter and leave the heme pocket at a relatively rapid rate.

†Supported in part by a grant from the National Science Foundation (M.K.) and from the National Institutes of Health (R.E.).
‡Harvard University.
§University of Illinois at Chicago.

(1) Antonini, E.; Brunori, M. *Hemoglobin and Myoglobin and Their Reactions with Ligands*; North-Holland: Amsterdam, 1971.
(2) Bunn, H. F.; Forget, B. G. *Hemoglobin: Molecular, Genetic and Clinical Aspects*; Saunders: New York, 1980; W. B. Saunders: Philadelphia, 1986.
(3) Stryer, L. *Biochemistry*, 2nd ed.; W. H. Freeman: 1981; p 53.
(4) Perutz, M. F.; Mathews, F. S. *J. Mol. Biol. 1965, 21*, 199.
(5) Takano, T. *J. Mol. Biol.* **1977**, *110*, 537, 569.
(6) Phillips, S. E. V. *J. Mol. Biol.* **1980**, *142*, 531.
(7) Kuriyan, J.; Wilz, S.; Karplus, M.; Petsko, G. A. *J. Mol. Biol.* **1986**, *192*, 133.
(8) Case, D. A.; Karplus, M. *J. Mol. Biol.* **1979**, *132*, 343.

0002-7863/90/1512-9161$02.50/0

Molecular dynamics simulations of proteins and experimental data[9-11] have demonstrated that a rigid picture of proteins is inappropriate. It is now recognized that the atoms of which a protein is composed are in a state of constant motion. Crystallographic studies provide the average structure of the protein, but the atoms undergo fluctuations of sizeable amplitudes about these average positions at ordinary temperatures. Simulation can be used to determine whether the transient fluctuations of atomic positions (e.g., side-chain motions) are sufficient to permit ligands to pass through the protein matrix at the experimentally observed rates.

Phenomenologically, the ligand-binding reaction in myoglobin can be written as

$$\mathrm{MbCO} \underset{k_{-a}}{\overset{k_a}{\rightleftharpoons}} [\mathrm{Mb}][\mathrm{CO}]_{\mathrm{pocket}} \underset{k_{-b}}{\overset{k_b}{\rightleftharpoons}} [\mathrm{Mb}\cdots\mathrm{CO}]_{\mathrm{protein}} \underset{k_{-c}}{\overset{k_c}{\rightleftharpoons}} [\mathrm{Mb}\cdots\mathrm{CO}]_{\mathrm{solution}}$$

Our concern in this paper is with steps b and c after photodissociation: the escape of the ligand from the heme pocket into the protein matrix and the passage of the ligand through the protein matrix to the outside. The solvent barrier, if any, is not included because the simulations are done in vacuum.

In an earlier simulation of ligand motion through the protein matrix,[8] the protein structure was treated as rigid and the O_2 ligand was modelled by a high-energy sphere (a mass of 32 amu, kinetic energy of 15 kcal/mol) of reduced diameter (2.8 Å versus the normal diatomic value of 4.2–4.9 Å). One hundred fifty ligand trajectories were examined. Most of the molecules that escaped from the protein matrix (31 out of 150) did so between the distal histidine 64 E7 and the side chains of Thr 67 E10 and Val 68 E11 (see Figure 5 of ref 8). A second significant pathway was also found; this involves a more complicated motion along an extension of the heme pocket into a space surrounded by Leu 29 B10, Leu 61 E4, and Phe 33 B14 and escape from the protein between Leu 61 E4 and Phe 33 B14 (see Figure 6 of ref 8). It was suggested, based on the range of motions found in the different trials, that additional, more complicated pathways may also exist; e.g., a number of ligands left the heme pocket but remained in the protein at the end of the simulation. The length of the trajectories, which were limited by the available computer time, was too short (3.75 ps) to obtain a complete picture of the possible paths.

The two major pathways just described have very high barriers (on the order of 100 kcal/mol) for a ligand of normal size in the rigid X-ray structure. An examination of the energetics of barrier relaxation was made to determine whether either of the pathways had acceptable activation enthalpies. Dihedral rotations of key side chains, analogous to the tyrosine side-chain oscillations explored by adiabatic mapping[12] and molecular dynamics,[13] indicated that barriers on the order of 10 kcal/mol would be present in a fluctuating protein.[8] These barriers are comparable to those estimated from the photolysis, rebinding studies for CO myoglobin by Austin et al.[14] and Ausari et al.[15] Comparable results have been obtained recently by Chatfield.[40]

The energy minimization study has been complemented by activated dynamics calculations[16,17] that investigated the rate for crossing the barrier between His E7 and Val E11 along the primary path. The computed value of the rate constant is $k \approx 4 \times 10^7\ s^{-1}$, a value in reasonable agreement with the overall kinetics of ligand binding,[14,15,18,19] although such a comparison on the basis of an analysis of only one of the possible paths is not necessarily appropriate. A surprising result of the calculation is that the barrier is entropy dominated, in contrast to the mapping studies[8] and experimental estimates,[14,15,49] in which the activation energy plays the dominant role.

One of the paths suggested by the energy calculations corresponds to a high mobility region in the protein as determined by X-ray temperature factors.[20] Also, a crystal structure of myoglobin with a CO ligand[7] has shown that a key residue (Arg 45 CD3), which is in contact with His 64 E7, has two alternative conformations. In an X-ray structure of myoglobin with a phenyl group bound to the iron,[21] the large size of the ligand leads to a distortion of the heme pocket; the distal His 64 E7 has moved outward from the pocket with Arg 45 CD3 in the alternative conformation that permits this displacement. It was suggested that the phenyl ligand acts as a "door stop" and that the observed structural changes might mimic those occurring when ligands enter and exit from the heme pocket. Similar results have been obtained for imidazole metmyoglobin,[22] methemoglobin,[23] and ethyl isocyanide.[50] The crystallographic studies are of considerable interest in demonstrating possible protein distortions. However, since the ligands are bound to the iron, the van der Waals repulsions resulting from their large size tend to induce significant local distortions in the protein matrix. This does not necessarily duplicate what happens for small ligands that are able to move through the protein.

More information related to the possible escape routes of the ligand has been obtained from the study of mutants.[24] A detailed analysis of photolysis experiments[51] with O_2 and CO for several His 64 E7 mutants indicated that polarity is more important than size in its contribution to the barrier between the heme pocket and the outside of the protein; e.g., for Val 64 E7 the barrier is slightly lower than that for Gly 64 E7. The mutation of Arg 45 CD3 to Gly (the arginine is believed to inhibit the distal histidine motion [ref 21, see above]) increases the rate for O_2 by a very small amount.[52] This suggest that the distal histidine has a role in blocking ligand motion from the heme pocket to the solvent. However, the results do not demonstrate that in native myoglobin the major pathway goes by the distal histidine whose fluctuations make ligand escape possible. A more definitive mutant would be one which *slowed* the rate.

All of the structural and theoretical studies described so far focused on paths that lead from the heme pocket through the protein to the surface in the neighborhood of the heme. Tilton et al.[26] have calculated trajectories for the escape of xenon atoms from myoglobin; they examined four experimentally observed xenon bending sites, plus a site in the heme pocket.[27] One hundred 5-ps trajectories were run with a rigid protein by adapting the procedure described in ref 8. The results of this simulation suggested that xenon might exit via longer paths through the protein. A molecular dynamics simulation performed with all the protein atoms allowed to move[26] confirmed that the xenon atoms can undergo rather lengthy excursions through the protein matrix

(9) McCammon, J. A.; Gelin, B.; Karplus, M. *Nature* **1977**, *267*, 585–590.
(10) Brooks, C. L., III; Karplus, M.; Pettitt, B. M. *Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics*; Adv. Chem. Phys. LXXI John Wiley & Sons, 1988.
(11) McCammon, J. A.; Harvey, S. *Dynamics of Proteins and Nucleic Acids*; Cambridge University Press: Cambridge, 1987.
(12) Gelin, B. R.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72*, 2002–2006.
(13) McCammon, J. A.; Wolynes, P.; Karplus, M. *Biochemistry* **1979**, *18*, 927–942.
(14) Austin, R. H.; Beeson, W. K.; Eisenstein, L.; Frauenfelder, H.; Gunsalus, I. C. *Biochemistry* **1975**, *14*, 5355.
(15) Ansari, A.; Berendzen, J.; Bowne, S. F.; Frauenfelder, H.; Iben, I. E. T.; Sauke, T. B.; Shyamsunder, E.; Young, R. D. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 5000.
(16) Case, D. A.; McCammon, J. A. *Ann. N.Y. Acad. Sci.* **1986**, *482*, 222.
(17) Kottalam, J.; Case, D. A. *J. Am. Chem. Soc.* **1988**, *110*, 7690.
(18) Henry, E. R.; Sommer, J. H.; Hofrichter, J.; Eaton, W. A. *J. Mol. Biol.* **1983**, *166*, 443.
(19) Olson, J. S.; Rohlfs, R. J.; Gibson, Q. H. *J. Biol. Chem.* **1987**, *262*, 2930.
(20) Frauenfelder, H.; Petsko, G. A.; Tsernoglou, D. *Nature (London)* **1979**, *280*, 558–563.
(21) Ringe, D.; Petsko, G. A.; Kerr, D. E.; Ortiz de Montellano, P. R. *Biochemistry* **1984**, *23*, 2.
(22) Bolognesi, M.; Cannillo, E.; Ascenzi, P.; Giacometti, G. M.; Merli, A.; Brunori, M. *J. Mol. Biol.* **1982**, *158*, 305–315.
(23) Bell, J. A.; Korzun, Z. R.; Moffat, K. *J. Mol. Biol.* **1981**, *147*, 325.
(24) Olson, J. S.; Mathews, A. J.; Rohlfs, R. J.; Springer, B. A.; Egeberg, K. A.; Sligar, S. G.; Tame, J.; Renaud, J. P.; Nagai, K. *Nature* **1988**, *336*, 265–266.
(25) Rohlfs, R. J.; Mathews, A. J.; Olson, J. S.; Springer, B. A.; Springer, K. A.; Egeberg, K. D.; Sligar, S. G. *J. Mol. Biol.* In press.
(26) Tilton, R. F., Jr.; Singh, U. C.; Weiner, S. J.; Connolly, M. L.; Kuntz, I. D., Jr.; Kollman, P. A.; Max, N.; Case, D. A. *J. Mol. Biol.* **1986**, *192*, 443.
(27) Tilton, R. F.; Kuntz, I. D., Jr.; Petsko, G. A. *Biochemistry* **1984**, *23*, 2849.

before escaping. Although only five trajectories were calculated, they do suggest that the earlier work on myoglobin ligands may have been too restrictive to obtain a full picture of the dynamics. This is in accord with fluorescence quenching experiments which indicate that a rather large portion of the globin can be penetrated by oxygen and other ligands.[28-30]

To overcome the two major limitations of the earlier myoglobin studies (i.e., when a large number of ligand trajectories were determined, the protein was kept rigid, while when the protein was allowed to move, only a small number of trajectories were calculated), we propose a new approach for enhanced sampling in ligand binding and related processes in macromolecular dynamics. The method is based on the time-dependent Hartree (TDH) approximation,[31] which has been used extensively for electronic structure calculations.[32] Recently the TDH approach has been extended to classical dynamics by Gerber et al.[33] and applied to a range of molecular dynamics problems.[33-36] Gerber et al. introduced an efficient new method to calculate ensemble averages by means of "trajectory bundles"[33,34] and applied it to energy redistribution and predissociation of van der Waals molecules.

The essence of the TDH method, in trajectory calculations as in electronic structure studies, is the separation of the system into two or more parts, each of which moves in the coordinate averaged (in general, time dependent) potential due to the other parts. In the full TDH method, this reduces the many-body problem (say for N particles) into N one-body problems that are solved simultaneously. Due to the time dependence of the interaction potential, coupling between the different parts of the system is present, and energy transfer can occur. In the present application we separate the system into only two subsystems: the ligand and the protein molecule.

The implementation of the TDH approximation that we use here makes it possible to treat simultaneously an ensemble (say 60) of photodissociated CO molecules moving through a protein matrix which is being simulated by molecular dynamics. The required computer time is little more than that used in a single standard protein trajectory since it is the dynamics of the protein that is the most time intensive part of any such simulation; e.g., the number of degrees of freedom of a polar hydrogen model for myoglobin is 4602, so that addition of 100 diatomic ligands, each with six degrees of freedom, increases the computational efforts by less than a factor of 2. This is much less than the factor of 100 which is required by the exact approach (i.e., separate calculations of 100 trajectories with one protein and one ligand with different initial conditions). Thus, the TDH method permits a much more complete search of exit pathways than was possible previously. By choosing the initial velocities with random directions, a large portion of the accessible conformational space in the protein interior can be examined without the need for excessive amounts of computer time. As in any Hartree-type formulation, correlation effects (here those corresponding to artificial correlations introduced by the interactions of several CO molecules with the protein) are assumed not to be important.

Although the TDH approximation makes possible a more efficient search of the phase space available to the CO molecule at the chosen temperature, there remains the problem of the energy barriers that make the mean escape time for a single CO molecule from the heme pocket much too long for a standard molecular dynamics simulation (i.e., the time scale is on the order of 100 ns rather than picoseconds[37]). As indicated by the earlier studies,[8] escape may require rare fluctuations of the protein structure that are not sampled efficiently by the TDH (or straightforward molecular) dynamics at the experimental temperature. To overcome this problem, we increased the ligand kinetic energy by heating the ligand (i.e., scaling the velocity) at fixed time intervals for short periods and then gradually cooling it to room temperature; the protein temperature is kept near 300 K by velocity scaling. Thus, as in alternative approaches (scaling of the particle mass and/or radius), the resulting trajectories are of significance primarily for determining pathways that can serve as starting points for more detailed simulations of the reaction dynamics.[16,17] Reaction paths with barriers accessible at room temperature will be sampled but so will others with higher barriers. By following the CO molecules along their trajectories, the energy involved in crossing the barriers can be estimated.

In section II we describe briefly the TDH approximation as applied to enhanced sampling in molecular dynamics and introduce the specific form that is employed for the ligand escape simulation in myoglobin; the derivation of the TDH method based on the Liouville equation is given in the Appendix. The computational procedure is outlined in section III. Section IV presents a description of the pathways determined by the TDH approach. The conclusions are given in section V.

II. Time-Dependent Hartree Method

In standard simulations of the classical dynamics of a many-particle system, trajectories for the particles are determined. Such simulations may treat the rate of a chemical reaction, where a series of trajectories leading from reactants to products are calculated, or the equilibrium dynamics of a protein, where a single trajectory is usually employed. In such simulations, the coordinates and the velocities of the individual particles interacting through a specified potential function are propagated in time by the use of numerical solution of Newton's equation of motions. For the purpose of introducing the TDH approximation, it is convenient to utilize an alternative (equivalent) formulation of classical dynamics in terms of the Liouville equation. We consider the phase space probability density, $\rho(\mathbf{P},\mathbf{Q},t)$. Here, $\mathbf{P}$ and $\mathbf{Q}$ are, respectively, canonical momenta and coordinates representing the state of the N degrees of freedom of the entire system, and $\rho(\mathbf{P},\mathbf{Q},t)\, d\mathbf{P}\, d\mathbf{Q}$ is the probability of finding the system in the region between $(\mathbf{P},\mathbf{Q})$ and $(\mathbf{P}+d\mathbf{P},\mathbf{Q}+d\mathbf{Q})$ at time t. The quantity $\rho(\mathbf{P},\mathbf{Q},t)$ satisfies the Liouville equation[38]

$$\frac{\partial \rho}{\partial t} - \sum_i \left(\frac{\partial \mathcal{H}}{\partial Q_i} \frac{\partial \rho}{\partial P_i} - \frac{\partial \mathcal{H}}{\partial P_i} \frac{\partial \rho}{\partial Q_i} \right) \tag{1}$$

where $\mathcal{H}$ is the classical Hamiltonian of the system, and $\rho(\mathbf{Q},\mathbf{P},t)$ is normalized according to

$$\int_{-\infty}^{\infty} \rho(\mathbf{P},\mathbf{Q},t)\, d\mathbf{P}\, d\mathbf{Q} = 1 \tag{2}$$

The ensemble average for any dynamical variable $A(\mathbf{P},\mathbf{Q})$ is then given by

$$\langle A \rangle = \int_{-\infty}^{\infty} A(\mathbf{P},\mathbf{Q})\, \rho(\mathbf{P},\mathbf{Q},t)\, d\mathbf{P}\, d\mathbf{Q} \tag{3}$$

where $\langle A \rangle$ is, in general, a function of time.

We now wish to derive the equations of motions for the ensemble averages of the coordinates and momenta. With eq 3, we have

$$\frac{\partial}{\partial t}\langle Q_j \rangle = \int Q_j \frac{\partial}{\partial t} \rho(\mathbf{P},\mathbf{Q},t)\, d\mathbf{P}\, d\mathbf{Q} \tag{4}$$

$$\frac{\partial}{\partial t}\langle P_j \rangle = \text{int} P_j \frac{\partial}{\partial t} \rho(\mathbf{P},\mathbf{Q},t)\, d\mathbf{P}\, d\mathbf{Q}$$

Introducing the Liouville equation for the time derivative of $\rho(\mathbf{Q},\mathbf{P},t)$ from eq 1 into eq 4, we obtain

(28) Lakowitz, J. R.; Weber, G. *Biochemistry* **1983**, *12*, 4171.
(29) Eftink, M. R.; Ghiron, C. A. *Anat. Biochem.* **1987**, *114*, 199.
(30) Englander, S. W.; Kallenbach, N. R. *Quart. Rev. Biophys.* **1983**, *16*, 521.
(31) Dirac, P. H. M. *Proc. Cam. Phil. Soc.* **1930**, *26*, 376.
(32) Kolker, H. J.; Karplus, M. *J. Chem. Phys.* **1961**, *35*, 2235-2236.
(33) Gerber, R. B.; Buch, V.; Ratner, M. A. *J. Chem. Phys.* **1982**, *77*, 3022.
(34) Schatz, G. C.; Buch, V.; Ratner, M. A.; Gerber, R. B. *J. Chem. Phys.* **1983**, *79*, 1808.
(35) Eslava, L. A.; Gerber, R. B.; Ratner, M. A. *Mol. Phys.* **1985**, *56*, 47.
(36) Marki, N.; Miller, W. H. *J. Chem. Phys.* **1989**, *87*, 5781.
(37) Anfinrud, P. A.; Han, C.; Hochstrasser, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 8387.

$$\frac{\partial}{\partial t}\langle Q_j\rangle = \sum_i \int Q_j \left[\frac{\partial \mathcal{H}}{\partial Q_i}\frac{\partial \rho}{\partial P_i} - \frac{\partial \mathcal{H}}{\partial P_i}\frac{\partial \rho}{\partial Q_i}\right] \mathrm{d}\mathbf{P}\,\mathrm{d}\mathbf{Q} \quad (5)$$

$$\frac{\partial}{\partial t}\langle P_j\rangle = \sum_i \int P_j \left[\frac{\partial \mathcal{H}}{\partial Q_i}\frac{\partial \rho}{\partial P_i} - \frac{\partial \mathcal{H}}{\partial P_i}\frac{\partial \rho}{\partial Q_i}\right] \mathrm{d}\mathbf{P}\,\mathrm{d}\mathbf{Q}$$

Integrating by parts and making use of the fact that ρ must vanish sufficiently rapidly for large $\mathbf{Q}$ and $\mathbf{P}$ so that the normalization integral (eq 2) exists, we find after some manipulation

$$\frac{\partial\langle Q_j\rangle}{\partial t} = \left\langle \frac{\partial \mathcal{H}}{\partial P_j}\right\rangle \quad (6a)$$

$$\frac{\partial\langle P_j\rangle}{\partial t} = -\left\langle \frac{\partial \mathcal{H}}{\partial Q_j}\right\rangle \quad (6b)$$

$$j = 1, 2, \ldots, N$$

Equations 6a and 6b demonstrate that Hamiltonian's equations are valid for the ensemble averages of the individual particle coordinates and momenta.

To express the above results in a form that is useful for calculations, we first reduce eq 6 for the phase-space averages to those corresponding to a single trajectory. This is done by assuming that the distribution $\rho(\mathbf{P},\mathbf{Q},t)$ is a δ-function; i.e.

$$\rho(\mathbf{P},\mathbf{Q},t) = \delta(\mathbf{P}-\mathbf{P}^\circ(t),\mathbf{Q}-\mathbf{Q}^\circ(t)) \quad (7)$$

By the Gibbs principle of the conservation of probability for canonical variables in phase space,[38] if $\rho(\mathbf{Q},\mathbf{P},t)$ is a δ function at time t it remains a δ function at all times. Substituting into eq 6, we obtain

$$\frac{\partial Q_j^\circ}{\partial t} = \frac{\partial \mathcal{H}}{\partial P_j^\circ}$$

$$\frac{\partial P_j^\circ}{\partial t} = -\frac{\partial \mathcal{H}}{\partial Q_j^\circ} \quad (8)$$

$$(j = 1, 2, \ldots, N)$$

Equations 8 are the standard Hamilton's equations of motion for the coordinates and momenta which are associated with the N degrees of freedom of the system.

To extend this to a more general distribution in phase space, we make use of the fact that δ-functions form a complete orthonormal set.[39] Thus, we can expand $\rho(\mathbf{P},\mathbf{Q},t)$ in terms of δ-functions, as

$$\rho(\mathbf{P},\mathbf{Q},t) = \sum_w w\delta_w(\mathbf{P}-\mathbf{P}_w^\circ(t),\mathbf{Q}-\mathbf{Q}_w^\circ(t)) \quad (9)$$

where w represents the weight factors determined from the initial form of the distribution; i.e.,

$$w = \int \rho(\mathbf{P},\mathbf{Q},0)\ \delta_w(\mathbf{P}-\mathbf{P}_w^\circ(0),\mathbf{Q}-\mathbf{Q}_w^\circ(0))\ \mathrm{d}\mathbf{P}\,\mathrm{d}\mathbf{Q} \quad (10)$$

From eqs 9 and 10, the time dependence of $\rho(\mathbf{P},\mathbf{Q},t)$ is in the "parameters" $\mathbf{P}_w^\circ(t)$ and $\mathbf{Q}_w^\circ(t)$. From eqs 3 and 9, the phase space averages of $\mathbf{P}_j$ and $\mathbf{Q}_j$ can be written

$$\langle P_j\rangle = \sum_w wP_{wj}^\circ;\ \langle Q_j\rangle = \sum_w wQ_{wj}^\circ \quad (11)$$

The equations of motion for the phase-space average of the parameters Q_{wj}°, P_{wj}° are (see eq 6)

$$\frac{\partial}{\partial t}\sum_w wQ_{wj}^\circ = \sum_w w\frac{\partial \mathcal{H}}{\partial P_{wj}^\circ}$$

$$\frac{\partial}{\partial t}\sum_w wP_{wj}^\circ = -\sum_w w\frac{\partial \mathcal{H}}{\partial Q_{wj}^\circ} \quad (12a)$$

$$(j = 1, 2, \ldots, N)$$

(38) Tolman, R. C. *The Principles of Statistical Mechanics*; Oxford, 1938.
(39) Messiah, A. *Quantum Mechanics*; North Holland: Amsterdam, 1964.

For these first-order differential equations with the initial conditions corresponding to $\rho(\mathbf{Q},\mathbf{P},t = 0)$, there exists one and only one solution. Equation 12a is satisfied when each $Q_{wj}^\circ,P_{wj}^\circ$ individually satisfies the equations of motion; i.e.

$$\frac{\partial Q_{wj}^\circ}{\partial t} = \frac{\partial \mathcal{H}}{\partial P_{wj}^\circ}$$

$$\frac{\partial P_{wj}^\circ}{\partial t} = -\frac{\partial \mathcal{H}}{\partial Q_{wj}^\circ} \quad (12b)$$

$$(j = 1, 2, \ldots, N)$$

where the Hamilton is

$$\mathcal{H} = \mathcal{H}(\mathbf{P}_w^\circ(t),\mathbf{Q}_w^\circ(t)) \quad (13)$$

Equations 10–12 represent the exact connection between the Liouville formulation and the usual molecular dynamics. Since the $Q_{wj}^\circ,P_{wj}^\circ$ that obey eq 12b provides the solution for eq 12a, the desired phase-space averages (eq 11) can be evaluated from a series of standard trajectory calculations (eq 12b) with the appropriate weights (w) determined by the initial conditions (eq 10) for the system of interest.

TDH Approximation. In the TDH approximation applied to the classical dynamics, it is *assumed* that ρ can be approximated as a product of probability densities of different subsets of coordinates; that is,

$$\rho(\mathbf{P},\mathbf{Q},t) \sim \prod_s^N \rho_s(\mathbf{P}_s,\mathbf{Q}_s,t) \quad (14)$$

where the index s denotes a subset of the complete set of coordinates. This is similar to what is done in electronic structure calculations in which the electronic wave function is written as a product (or determinant) of one-electron wave functions. In the present version of the TDH approximation we do not employ a separation into single coordinate probability densities but instead use a partitioning in which the phase space distribution is factored into a product of two probability densities, each of which may include more than one coordinate. This is especially useful if we are interested primarily in the details of a small part of the system, and the rest of the system, which is not of direct interest, can be assumed to act as a "bath" that is sufficiently large to be "self-averaged"; i.e., a direct phase-space average (average over coordinates and momenta rather than over time) is required only for the few degrees of freedom making up the "small part" of the system. In the present application, the ligand is the "small part" of the system.

We assume the phase-space distribution can be written as

$$\rho(\mathbf{P},\mathbf{Q},t) \approx \rho_s(\mathbf{P}_s,\mathbf{Q}_s,t)\rho_{N-s}(\mathbf{P}_{N-s},\mathbf{Q}_{N-s},t) \quad (15)$$

where s designates the coordinates of the subsystem of interest, and $N - s$ denotes the rest of the coordinates. To derive the equations of motion for the coordinates s and $N - s$, we expand ρ_s in a set of δ-functions (as in eq 9)

$$\rho_s(\mathbf{P}_s,\mathbf{Q}_s,t) = \sum_w w\delta_w[\mathbf{P}_s-\mathbf{P}_{s,w}^\circ(t),\mathbf{Q}_s-\mathbf{Q}_{s,w}^\circ(t)] \quad (16)$$

and for the $N - s$ bath subsystem, we assume that it can be represented by a single δ-function

$$\rho_{N-s}(\mathbf{p}_{N-s},\mathbf{Q}_{N-s},t) = \delta(\mathbf{P}_{N-s}-\mathbf{P}_{N-s}^\circ(t),\mathbf{Q}_{N-s}-\mathbf{Q}_{N-s}^\circ(t)) \quad (17)$$

With the phase-space distributions for eqs 15–17, the equations of motion are obtained by a procedure corresponding to that used for deriving eq 12; the details are given in the Appendix. The resulting equations are

$$\frac{\partial}{\partial t}\mathbf{Q}_{s,w}^\circ = \frac{\partial \mathcal{H}(P_{s,w}^\circ,\mathbf{Q}_{s,w}^\circ,\mathbf{P}_{N-s}^\circ,\mathbf{Q}_{N-s}^\circ)}{\partial \mathbf{P}_{s,w}^\circ} \quad (17a)$$

$$\frac{\partial}{\partial t}\mathbf{P}_{s,w}^\circ = -\frac{\partial \mathcal{H}(\mathbf{P}_{s,w}^\circ,\mathbf{Q}_{s,w}^\circ,\mathbf{P}_{N-s}^\circ,\mathbf{Q}_{N-s}^\circ)}{\partial \mathbf{Q}_{s,w}^\circ} \quad (17b)$$

and

$$\frac{\partial}{\partial t}\mathbf{Q}^{\circ}_{N-s} = \sum_{w} w \frac{\partial \mathcal{H}(\mathbf{P}^{\circ}_{s,w},\mathbf{Q}^{\circ}_{s,w},\mathbf{P}^{\circ}_{N-s},\mathbf{Q}^{\circ}_{N-s})}{\partial \mathbf{P}^{\circ}_{N-s}} \quad (18a)$$

$$\frac{\partial}{\partial t}\mathbf{P}^{\circ}_{N-s} = -\sum_{w} w \frac{\partial \mathcal{H}(\mathbf{P}^{\circ}_{s,w},\mathbf{Q}^{\circ}_{s,w},\mathbf{P}^{\circ}_{N-s},\mathbf{Q}^{\circ}_{N-s})}{\partial \mathbf{Q}^{\circ}_{N-s}} \quad (18b)$$

Equations 17 and 18 are the essential results of the present development. They are derived with the TDH approximation from a Liouville formulation of classical mechanics. Equations 17a and 17b are the equations of motion for an ensemble of trajectories corresponding to the s subsystem, each one of which moves in the potential determined by the coordinates of the $N - s$ subsystem. The equations of motion for the $N - s$ subsystem (eqs 18a and 18b) are solved simultaneously with a single set of initial conditions and the effective potential obtained by averaging over the ensemble of s subsystem trajectories. This contrasts with the exact formulation where a series of trajectories with appropriately chosen initial phase-space distribution for the s and $N - s$ ensembles would be determined, and the results would be obtained by averaging over these distributions.

The utility of the TDH approximation in the present formulation is to enhance the sampling of the subsystem of interest (the s subsystem). This contrasts with the approach of Gerber et al.[33,34] where full approximate phase-space sampling of the entire N degree of freedom system is of interest and the subsystems s and $N - s$ are essentially of equal size. The present approach is primarily of interest if $s \ll (N - s)$, and one can assume (as we do) that the $N - s$ system can be treated in terms of a single δ-function distribution. It is in this way that the essential saving in time is achieved; i.e., we are able to use a realistic phase-space distribution of initial conditions for the small s subsystem and calculate a bundle of trajectories, while doing only a single large $N - s$ subsystem trajectory. This type of approach is likely to be useful not only for the present problem involving a protein and ligand but also in other condensed matter simulations where a small subsystem is of primary interest (e.g., a reactive system) and self-averaging of the solvent or bath (protein) coordinates can be assumed.

Application to Myoglobin. To illustrate the TDH method and prepare for the following sections, we consider the application to ligand motion through myoglobin. We divide the system into two parts (the CO ligand and the protein) and introduce an initial phase space probability distribution for each. The initial phase space probability density for the protein (the $N - s$ system) is assumed to be a δ-function (eq 17) of the form

$$\rho_{myo}(\mathbf{P}_{N-s},\mathbf{Q}_{N-s},t=0) = \delta(\mathbf{P}_{N-s}-\mathbf{P}^{\circ}_{N-s}(0))\delta(\mathbf{Q}_{N-s}-\mathbf{Q}^{\circ}_{N-s}(0)) \quad (19)$$

where $\mathbf{Q}^{\circ}_{N-s}(0)$ are the protein coordinates obtained from the X-ray structure of carboxymyoglobin after a minimization with parameters for an unliganded heme and CO (see below for computational details). The momentum (or velocity) vectors, $P^{\circ}_{N-s}(0)$, for the protein atoms are selected from a Boltzmann distribution at a temperature T in the standard way. Since we wish to achieve enhanced sampling for the ligand, we introduce a "swarm" of L CO molecules and use an initial distribution corresponding to the sum over δ-functions in eq 16. We have

$$\rho_{CO}(\mathbf{Q}_{CO},\mathbf{P}_{CO},t=0) = \delta(\mathbf{Q}_{CC}-\mathbf{Q}^{\circ}_{CO}(0))\sum_{w} w\delta(\mathbf{P}_{CO}-\mathbf{P}^{\circ}_{CO,w}(0)) \quad (20)$$

In eq 20, $\mathbf{Q}^{\circ}_{CO}(0)$ is the initial position of the ligand obtained by minimizing the carboxymyoglobin structure, while the momentum values for the C and O atoms of each CO in the δ-function representation are chosen via a random number from a Boltzmann distribution at the temperature T; that is, the weights w correspond to a distribution

$$w = \frac{\exp[-\mathbf{P}^2_{C,w}/2M_C kT]\exp[-\mathbf{P}^2_{O,w}/2M_O kT]}{\sum \exp[-\mathbf{P}^2_{C,w}/2M_C kT]\exp[-\mathbf{P}^2_{O,w}/2M_O kT]} \quad (21)$$

where M_C and M_O are the atomic masses and k is the Boltzmann constant.

The initial conditions described by eqs 19–21 make it possible to calculate many CO trajectories in the presence of a single protein trajectory. The differential equations of motion for the CO's are exact; i.e., in the TDH approximation the CO are moving in the "average" field of the protein but since there is only a single set of protein coordinates, this average coincides with the exact result. The protein trajectory is approximate since the atoms are moving in the average field of a swarm of L CO molecules; that is, eq 18b has the form

$$\frac{\partial}{\partial t}\mathbf{P}^{\circ}_{N-s} = -\frac{1}{L}\sum_{w}\frac{\partial V(\mathbf{Q}^{\circ}_{N-s},\mathbf{Q}^{\circ}_{CO,w})}{\partial \mathbf{Q}^{\circ}_{N-s}} \quad (22)$$

where each member of the CO swarm has the weight $(1/L)$. Clearly this procedure is best when all the CO molecules are close together, although the imbalance of forces between the protein atoms and the ligand replica leads to non-Newtonian dynamics.[40] This causes no problem in the present search for CO pathways but would effect the details of the dynamics, which are not considered here; a modified procedure that corrects this imbalance has been suggested (Ulitsky, A.; Elber, R., work in progress). When a CO molecule escapes, based on a distance criterion of 8 Å from the closest protein atom, the factor L in eq 22 was reduced by 1 to renormalize the interactions of the protein with the CO molecules and the escaping CO was excluded from the CO list.

The use of a single set of protein coordinates and momenta is expected to be a satisfactory approximation for the present case. The errors in the CO trajectories, which are of primary interest, are of second order; i.e., since the differential equations of motion for the CO are exact and it is the differential equation of motions for the protein atoms that are approximate, the errors in the forces on the CO molecules are second order. The possible effect of this in the present problem would be to introduce artificial correlations in the CO motions; e.g., if one ligand "opened" a hole, a second ligand might be able to pass through and bias the results. It is expected that the protein fluctuations are dominated by interactions within the protein and that the ligands do not introduce a major perturbation. As we show in section IV, this expectation is borne out by a comparison of the average properties of the protein in the present simulation with those obtained in standard molecular dynamics simulations.

III. Computational Procedure

The TDH methodology has been incorporated into the program CHARMM.[41] In the generation of the PSF (Protein Structure File), the data structure which defines the nature of the atoms and the covalent connectivities of the system under consideration, special TDH residues (here the CO molecules) are defined, and the bookkeeping is done to introduce the force renormalization in eq 22. Also, the generation of the nonbonded list was altered to exclude interaction between the CO molecules. Finally, the "fast" energy routine used in the dynamics was modified to take account of the TDH residues and to correctly calculate the normalized force due to them. Except for these modifications the standard CHARMM program was employed, including the usual parameters (PARAM19, polar hydrogen set) for the protein, the heme group, and the CO molecules. A constant dielectric of unity was used, and both the electrostatics and van der Waals interactions were smoothly truncated with a switching function at 8 Å over a distance of 1.0 Å. In all of the dynamics, a 1-fs time step was used.

The initial CO coordinates, $\mathbf{Q}^{\circ}_{CO}(0)$, were obtained by using the X-ray coordinates of the most populated bound conformation of carboxymyoglobin,[7] switching to the deoxy heme parameters[42] and minimizing for 200 steps with the Powell minimization al-

(40) Straub, J.; Karplus, M. Submitted for publication.
(41) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(42) Kuczera, K.; Kuriyan, J.; Arata, Y.; Karplus, M. Unpublished.

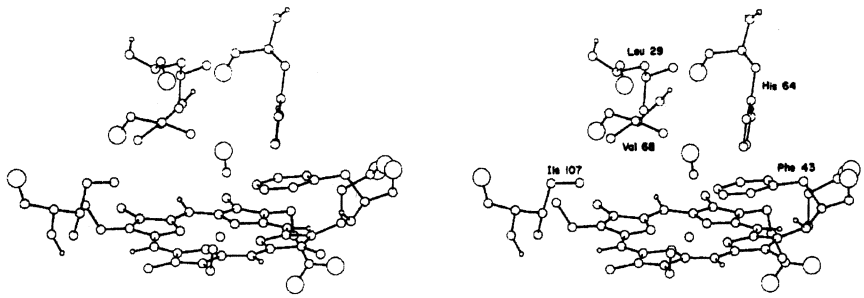

Figure 1. The heme pocket region: the heme, selected myoglobin residues forming the heme pocket, and the dissociated CO ligand are shown. The structure was obtained from the carboxymyoglobin X-ray coordinates,[7] after 200 Powell minimization steps with parameters appropriate for deoxymyoglobin. This is the initial structure used for the TDH simulations.

gorithm.[43] Only nonbonded van der Waals and electrostatic interaction between the heme and CO were included. During the minimization, the iron moves out of the heme plane into the unliganded position, and the CO moves about 1 Å from its liganded position. It remains in the heme pocket with the C atom at a distance of 3.5 Å from the average of the heme nitrogen positions, in approximate van der Waals contact with residues Phe 43 CD1, Leu 29 D10, His 64 E7, Val 68 E11, and Ile 107 G8 (see Figure 1). The CO position is similar to that found in a minimization study.[44] However, in contrast to that result, the CO remains nearly perpendicular to the heme (76° with respect to the plane of the four heme nitrogens). This position, with no force on the CO, is appropriate for initiating the simulation of CO motion in the heme pocket and through the protein matrix. Since the CO molecules spend a relatively long time exploring the heme pocket, the exact choice of initial coordinates makes little difference.

Simulation Protocols. Three different simulations were performed to permit analysis of the role of protein fluctuations in the escape of the CO molecules from the heme pocket and their penetration through the protein matrix. Sixty CO molecules were used in each of the simulations, and they initially occupied the same point in coordinate space (see above). Where the simulations differed is in the treatment of the dynamics. In simulation (a), all protein atoms, the heme, and the ligands were included in the dynamics. In the second (b), the protein atoms were fixed (rigid protein), and only the ligands were allowed to move. In the third simulation (c), the heme and the polypeptide chain were kept rigid, but the protein side chains and the ligands were allowed to move.

Simulation (a). In the main simulation, all protein and heme atoms as well as the CO molecules were allowed to move. To accelerate the barrier crossings, the CO molecules were given excess energy, while the protein and heme atoms were maintained near room temperature at all times. This was achieved by preparing a file of random velocities that are selected from a 300 K Boltzmann distributioin for the protein and heme atoms and 10000 K Boltzmann distribution for the CO atoms. This corresponds for a mean value for the initial kinetic energy of 30 kcal/mol in center-of-mass translation, 20 kcal/mol in rotation, and 20 kcal/mol in vibration; the initial CO potential energy is small, since a minimized structure was used. After the velocities are assigned, 0.1 ps of standard dynamics are calculated. Then, dynamics are run for 0.9 ps, with the temperature of the *entire* system (protein-heme plus CO) adjusted each 0.1 ps to 300 K by scaling the velocities. This cycle is repeated 100 times (choice of new set random velocities for 300 K protein heme and 10000 K CO, 0.1 ps free dynamics, 0.9 ps scaled dynamics) to obtain the full 100-ps trajectory. At the end of 0.1 ps in each cycle, the translational kinetic energy of CO is essentially equal to 300 K, but the molecule is still vibrationally hot. By doing the velocity scaling on the protein and heme as well as the CO the protein also is prevented from being heated locally; if the protein were allowed to heat up, it could undergo significant structural changes.

Simulation (b). In the 100-ps rigid protein calculation, which is most similar to earlier studies[8,26] the initial kinetic energy of the ligand swarm was chosen from a Boltzmann distribution at a temperature of 10000 K. Since the protein was treated as rigid, no energy transfer takes place, and the CO molecules move with fixed total energy. The standard radii and masses were used for the C and O. This simulation permits us to examine systematically the possibility of open pathways in a rigid protein structure and, by comparison with the other simulations, to determine the role of protein fluctuations. Within the rigid protein approximation the TDH is exact, and calculating N separate ligand trajectories is essentially equivalent in computer time on a scalar machine to calculating one trajectory with N ligands. However, with a vector machine in which the forces on the ligands can be calculated in one operation, the TDH method still saves time.

Simulation (c). The third simulation involved a rigid polypeptide chain and heme, but the side chains (starting with C_β) and the CO molecules were allowed to move. The same minimized structure, as in the rigid protein calculation, was used for the initial coordinates. The initial velocities for the CO and side chains were assigned from a Boltzmann distribution at 300 K. Then, 1-ps heating and cooling cycles were used to generate a simulation 20 ps in length. By linear scaling of the velocities at each step for 0.5 ps, the CO molecules and side chains were heated up to 900 K; by corresponding linear scaling the temperature was reduced to 300 K over another 0.5 ps. This calculation addresses the role of side chain motions in the ligand escape. A lower maximum energy was used than for simulations (a) and (b), since here a larger number of atoms are raised in temperature, while in the other cases only the swarm of CO was heated and the protein was at room temperature (a) or kept rigid (b). In simulation (c), the maximum total energy is about three times that in simulation (a).

IV. Results

In this section we present the simulation results and their analysis. We consider first the simulation in which both the CO and the myoglobin are allowed to move with the CO pulsed to high temperature and the protein plus heme at approximately room temperature (see section III). To obtain increased insight into the significance of the results, we compare them to those from simulations in which both the CO and the protein side chains are heated to a relatively high temperature and in which the CO is at a high temperature but the protein is not allowed to move.

Simulation with High-Temperature CO and Room-Temperature Protein. To verify that the TDH simulation with a large number of CO molecules does not introduce artificial behavior as far as the structure or dynamics of the protein are concerned, we first examine some of the protein simulation results and compare them with standard equilibrium simulations and with the X-ray structure. We then describe the dynamics of the CO molecules, which are of primary interest for understanding the effect of the protein matrix.

Analysis of Protein Trajectory. Figure 2a shows the time dependence of the rms deviations of the simulation coordinate sets from the initial liganded X-ray structure.[7] One picosecond av-

(43) Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. *Numerical Recipes*; Cambridge University Press: 1986.
(44) Sassaroli, M.; Rousseau, D. L. *J. Biol. Chem.* **1986**, *261*, 16292.

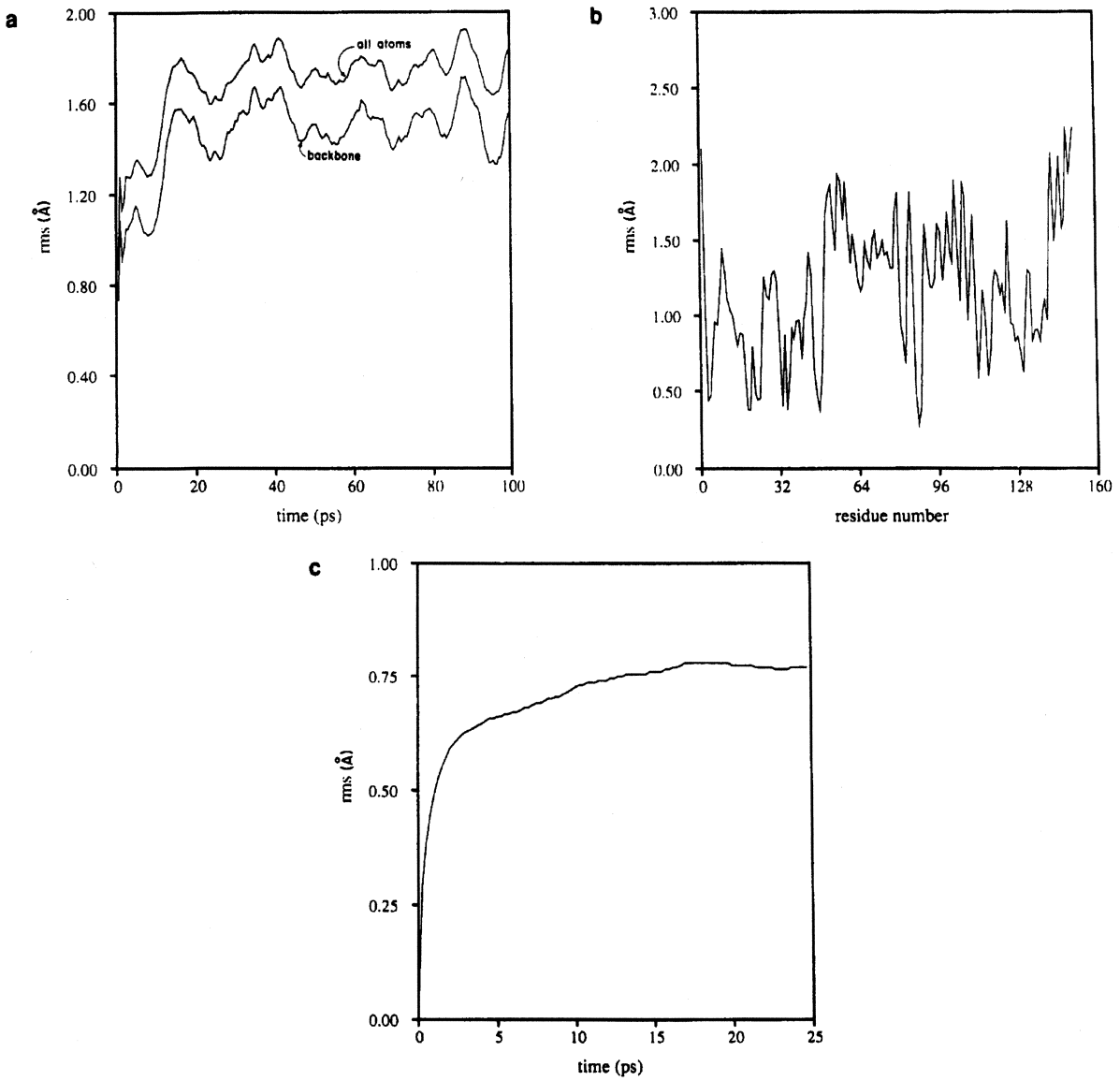

Figure 2. Characterization of the protein in simulation (a) (see text). (a) The root mean square difference (rms) between the X-ray coordinates and the coordinate sets of the protein trajectory as a function of time. (b) The rms difference for each residue between the average coordinates from the simulation and the X-ray structure as a function of the residue number. (c) The average rms difference between coordinate sets as a function of the time between the coordinate sets. (The average is over the complete set employing cyclic boundary conditions in time; see text for details.)

erages are shown and separate results are given for the backbone and all atoms; mass-weighted coordinates are used, and hydrogen atoms are not included. As can be seen, the trajectory converges after approximately 20 ps to an equilibrium structure which deviates from the X-ray results by about 1.45 Å for the main chain and 1.9 Å for the side chains. The fact that the unliganded heme structure has been introduced makes only a small difference since the rms difference between the Mb and MbCO X-ray structures is small; i.e., it is 0.29 Å for the main-chain atoms. The deviation from the X-ray structure is slightly larger than that obtained in other recent myoglobin simulations (e.g., the 100 ps Xe simulation[26] yielded 1.25 and 1.71 Å for the main chain and all atoms, respectively) and somewhat smaller than that in an earlier simulation (i.e., 1.81 and 2.12 Å for the main chain and all atoms over 100 ps.[45]

Figure 2b shows the average backbone rms deviation for the period 50–100 ps from the X-ray as a function of residue number; all heavy atoms are included. The deviations from the crystal structure are distributed throughout the entire protein molecule. The largest deviations of the individual rms of residues are comparable to those of Levy et al.[45] The structural differences at the heme pocket are of special interest. In the average structure the E helix is bent away from the heme group, so that the heme pocket is somewhat larger in the simulation than in the liganded X-ray structure; similar effects are found in the standard simulations. The changes are on the order of 1 Å; e.g., the distance of the C_α of Thr 67 E10 from the heme iron increases from 9.52 in the X-ray structure to 10.5 Å in the average structure, the distance of the C_α of His 64 E7 changes from 8.9 to 9.6 Å, and Phe 43 CD1 is shifted by a rigid body motion of 1.6 Å from the heme iron in the direction opposite to the E helix displacement. The B helix contact residue Leu 29 is closer to the iron by 1 Å; its distance changes from 7.6 to 6.6 Å.

Figure 2c presents the average rms difference between coordinate sets as a function of the time difference between them. For calculating the averages, cyclic boundary conditions (in time) are employed. Thus, if T is the last structure in the trajectory, structure T+1 is chosen to be the first coordinate set. In this way

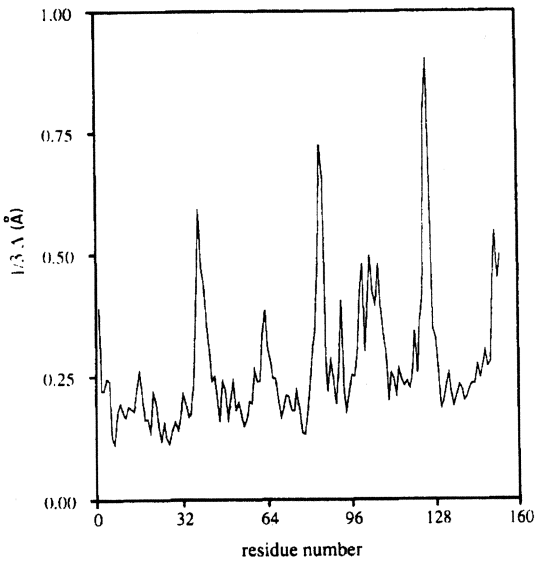

Figure 3. The rms fluctuations of the main-chain atoms in the trajectory as a function of residue number; the calculations are an average over the final 50 ps of the trajectory. Δ is the length of the displacement vector; i.e., $\Delta = [\langle \Delta x^2 \rangle + \langle \Delta y^2 \rangle + \langle \Delta z^2 \rangle]^{1/2}$.

each of the averages has a sample of the same size, making the analysis of the average rms more meaningful. The use of cyclic boundary conditions is appropriate if the simulation is converged, and the difference between the last and the first points is not too large. The plot starts at 50 ps and shows only the main-chain atom. In agreement with standard molecular dynamics simulations, the rms differences are significantly smaller than the deviations from X-ray structure and reach a limiting value of about 0.75 Å. The convergence suggests that equilibrium has been achieved, and the ligand heating and cooling does not have any "global" effect on the protein structure.

Figure 3 shows the rms fluctuations of the main-chain atoms as a function of the residue number. The fluctuations were calculated over the period 50–100 ps. Similar to Levy et al.,[45] the fluctuations tend to be localized in specific protein segments, and there are sharp spikes on a uniform background of about 0.2 Å. The largest fluctuations are in the GH loop in both simulations. The residues at other peak positions are similar in the two simulations and correspond to the GH, EF, CD loops, the F helix, and the C terminus. A small peak exists in the current simulation at the E helix position which is missing in the earlier simulation. This change at the E helix may result from the special dynamics protocol since it is the most likely main-chain region to be affected by the large number of hot CO molecular replicas.

Since side-chain motions are believed to play an important role in determining the rate of ligand diffusion through the protein,[8,16] we have examined the residues which undergo dihedral angle transitions during the trajectory. Although there are significant dihedral angle fluctuations, only 10% of the residues (16 out of 153) undergo χ_1 dihedral angle transitions and 13 undergo χ_2 dihedral angle transitions; the residues and dihedral angles are listed in Table I. These values are comparable to those obtained in standard 100-ps room-temperature simulations.[45,46] Residues undergoing transitions are distributed throughout the protein. Three of the χ transition occur in loop residues and one at the C terminus, and 12 transitions are associated with residues in helices. For χ_2, one transition occurs at the C terminus and one in a loop; the other 11 are associated with helices. Thus, considering the fraction of nonhelical residues (0.144), the transitions

Table I

residue	χ_1	χ_2	residue	χ_1	χ_2
a. Myoglobin Side-Chain Transition in Trajectory[a]					
Val 1 NA1	+		Ser 92 F7	+	+
Leu 9 A7		+	Lys 96 FG3	+	
Leu 11 A9		+	Ile 99 FG5	+	
Val 13 A11	+		Phe 106 G7		+
Leu 32 B13	+		Ile 107 G8	+	
Phe 33 B14	+		Ser 108 G9	+	
Glu 38 C3		+	Val 114 G15	+	
Lys 47 CD5	+		Phe 123 GH5		+
Thr 51 D1	+	+	Gln 128 H6	+	+
Leu 61 E4		+	Met 131 H8		+
Lys 62 E5	+		Leu 137 H14		+
Leu 72 E15	+		Lys 145 H22	+	+
			Tyr 151 HC2		+
b. Myoglobin Side-Chain Transitions in Simulation (c)[a]					
Ser 3		+	Leu 61	+	+
Leu 9		+	Lys 63	+	
Leu 11	+	+	His 64	+	+
Val 13	+		Thr 67		+
Lys 16		+	Thr 70		+
Val 21	+		Leu 72	+	
Leu 29		+	Lys 87	+	+
Ile 30	+		Leu 89	+	+
Lys 34		+	Ser 92	+	+
Gln 38		+	Thr 95		+
Leu 40	+		Lys 96	+	
Arg 45		+	Ile 101	+	
Leu 49		+	Ser 117		+
Lys 50		+	Tyr 146		+
Glu 52		+	Leu 149		+
Lys 56		+	Gln 128	+	

[a] A side-chain transition is defined by a shift of the χ angle by 100° relative to the initial position.

are relatively uniformly distributed. As noted previously,[47] transitions are expected in flexible loops and in the contacts between helices as the protein makes a transition from one minimum to another. There appears to be little or no correlation among the χ_1 and χ_2 transitions; only three χ_1 and χ_2 transitions occur in the same residue.

Some of the residues that underwent transitions are in regions that lie along the path of the escaping CO molecules. Three χ_1 transitions of important colliding residues (at least 500 collisions with the CO in the trajectory, see below) were noted; they are Leu 72 E15, Ile 107 G8, and Ser 108 G9. Figure 4a–c gives the time history of the three dihedral angles; for comparison, we also show χ_1 of Val 68 E11 (Figure 4d), which did not undergo any transition. The transition of Ser 108 G9 occurs at around 33 ps and is followed by the transition of Ile 107 G8 at 36 ps. At 49 ps (13 ps later) Leu 72 E15 underwent a transition, which is clearly uncorrelated. Leu 72 E15 and Ile 107 G8 are in the neighborhood of the heme pocket, and their transitions together with the bending of the E helix tend to increase the volume of the heme pocket. However, they do not create new exit passages, though they do appear to change the existing openings somewhat. His 64 E7, Val 68 E11, and Thr 67 E10 do not undergo transitions and maintained average conformations that are close to the X-ray structure.

Summarizing the above analysis, we conclude that the protein trajectory is well behaved in the presence of the hot CO molecules. The most significant difference from the X-ray structures[5-7] is the somewhat larger heme pocket caused mainly by the bending of the E helix; such behavior was also observed in the trajectory of Levy et al.[45]

Analysis of Ligand Trajectories. We consider first the motion of the ligands in the heme pocket and then describe the exit from the pocket and the escape from the protein by several different paths. Of the 60 CO molecules included in the simulation, 30 had escaped at the end of the 100-ps trajectory; the others were distributed along the various paths inside the protein.

(45) Levy, R. M.; Sheridan, R. P.; Keepers, J. W.; Dubey, G. S.; Swaminathan, S.; Karplus, M. *Biophys. J.* **1985**, *48*, 509–518.
(46) van Gunsteren, W. F.; Karplus, M. *Biochemistry* **1982**, *21*, 2259–2274.
(47) Elber, R.; Karplus, M. *Science* **1987**, *235*, 318–321.

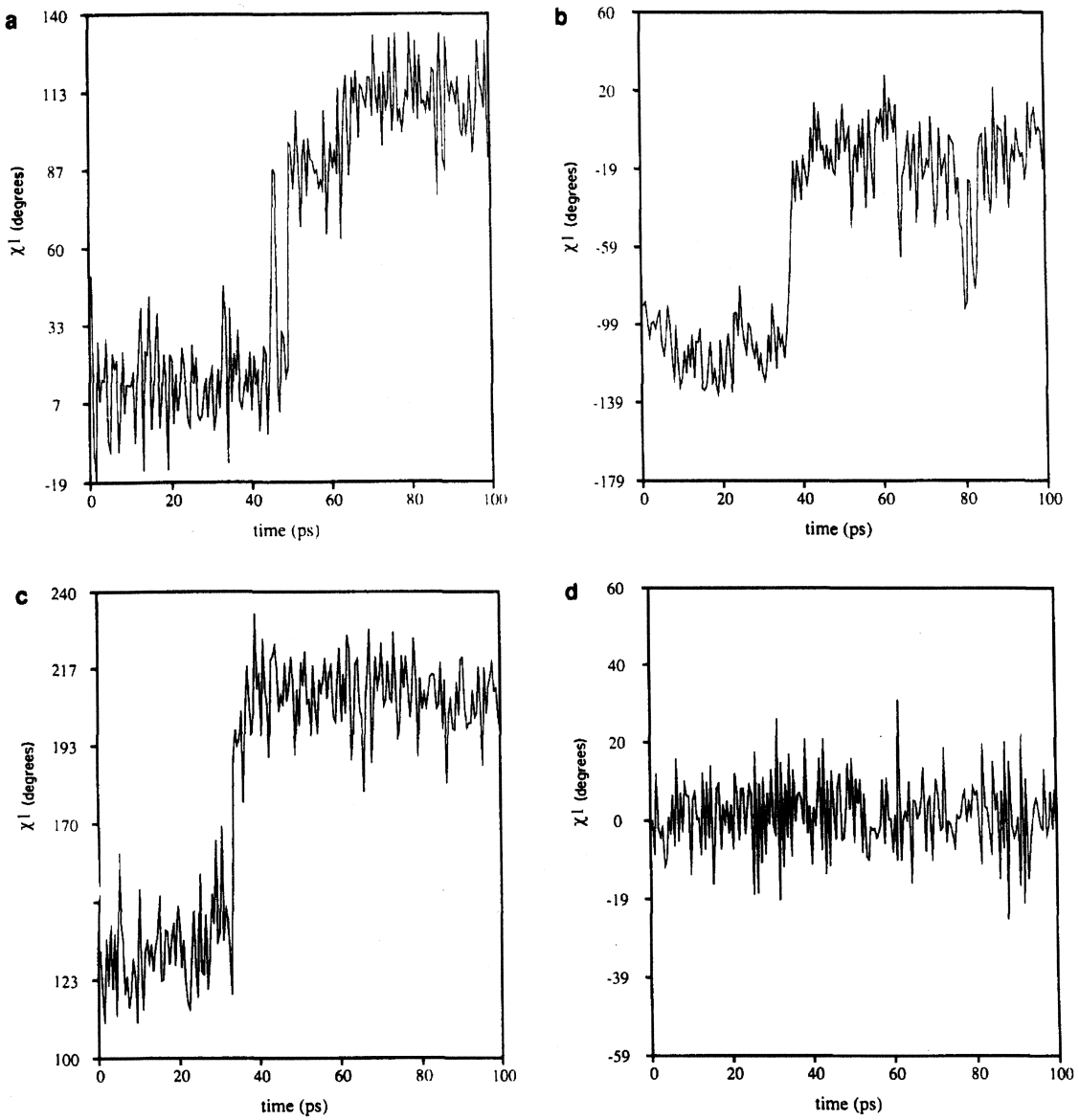

Figure 4. Time history of certain side-chain dihedral angles (χ_1) in simulation (a): (a) Leu 72 E15, (b) Ile 107 G8, (c) Ser 108 G9, and (d) Val 66 E9.

(a) Motion in the Heme Pocket. The initial velocity directions of the center-of-mass of the CO molecules were selected from an isotropic Boltzmann distribution at 10000 K. Figure 5 shows the probability density for the center-of-mass kinetic energy.

The heme pocket is delineated by the following residues: Leu 29 B10, His 64 E7, which form a "wall" in the direction of the CD corner, His 64 E7 and Val 68 E11 which prevent the direct escape near the E helix, and Ile 107 G8 which is the dominant residue that blocks motion from the B helix in the direction of the average of helices G and H (see Figure 1). Figure 6a shows a histogram of the distribution of times that the CO molecules spent in the heme pocket. The heme pocket is here defined operationally by use of Phe 43 C^{ζ}; i.e., the CO is in the heme pocket if either the C or O atom is less than 6.68 Å from Phe 43 C^{ζ}. The times range from 9 to 79 ps with five CO's still there at the end of the trajectory. Only four CO's returned to the heme pocket after escaping; in two trajectories the motion was heme pocket → EF/heme → B/G → heme pocket, and in the other two it was heme pocket → B/G → heme pocket. Thus, the CO molecules appear to diffuse back into the heme pocket only from the B/G cavity.

Most of the CO molecules undergo many collisions with the residues forming the walls of the heme pocket. In Figure 7, the CO behavior in the heme pocket is illustrated by a trajectory that remains in the pocket for 19 ps and then escapes to the B/G cavity. It is clear that a significant fraction of the pocket is sampled and that the ligand collides many times with the walls of the heme cavity before escaping between Ile 107 68 and Val 68 E11. Most CO's escape only after having explored much of the pocket.

To further analyze the motion in the heme pocket, coordinate sets for all CO trajectories were examined every 0.5 ps, and a "collision" between a protein atom and a CO molecule was assumed to have occurred if the distance between either atom of a CO and any protein heavy atom was less than 4 Å; this collision radius is a mean value based on the van der Waals radii of the heavy atoms. The results for all protein atoms summed over all the CO molecule trajectories are given in Table II. The collisions with the heme group are not included; there are 8554 of them. The largest number of collisions with protein residues are with three hydrophobic amino acids, Leu 29 B10, Val 68 E11, and Ile 107 G8, which form part of the walls of the pocket (see Figure 1 and Figure 1 of ref 8). The CO molecules collide many times

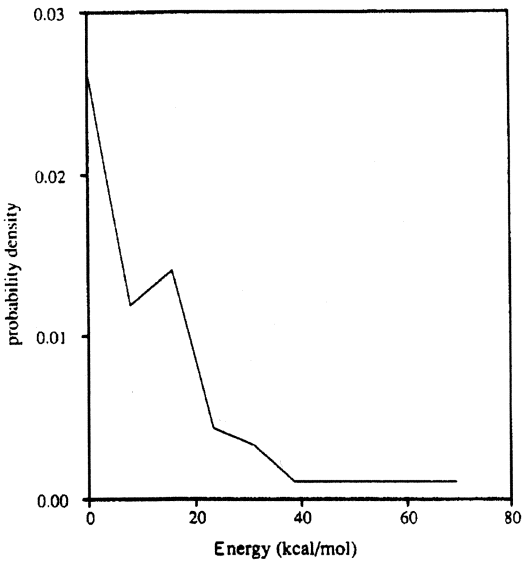

Figure 5. The average center-of-mass energy for the 60 CO replicas used in simulation (a).

Table II. Collisions of the 60 CO with Protein Residues in Simulation (a)

residue	no. of collisions	residue	no. of collisions
Leu 2 NA2	3[a]	Lys 79 EF2	9
Trp 7 A5	102	His 82 EF5	11
Val 10 A8	124[a]	Leu 89 F4	348[a]
Val 13 A11	1	Ala 90 F5	127
Trp 14 A12	73[a]	Gln 91 F6	2
Val 17 A15	19[a]	Ser 92 F7	22
Glu 18 A16	4	His 93 F8	872
Asp 20 B1	7	Ala 94 F9	31
Val 21 B2	7	His 97 FG3	20
Ala 22 B3	21	Lys 98 FG4	4
His 24 B5	66	Ile 99 FG5	46
Gly 25 B6	713[a]	Tyr 103 G4	29
Gln 26 B7	29	Leu 104 G5	376
Ile 28 B9	1024	Glu 105 G6	2
Leu 29 B10	2356[a]	Phe 106 G7	10
Ile 30 B11	4	Ile 107 G8	2036
Leu 32 B12	179[a]	Ser 108 G9	450
Phe 33 B13	25	Glu 109 G10	1
Thr 39 C4	14[a]	Ala 110 G11	17
Leu 40 C5	20	Ile 111 G12	840
Lys 42 C7	32	Ile 112 G13	6
Phe 43 CD1	435[a]	Val 114 G15	8
Phe 46 CD4	8[a]	Leu 115 G16	43[a]
Lys 47 CD5	14	Arg 118 G19	9
Leu 49 CD7	8[a]	Phe 123 GH5	3[a]
Lys 50 CD8	6	Gly 124 H1	3
Met 55 D5	13	Ala 125 H2	2
Leu 61 E4	56	Asp 126 H3	15
Lys 62 E5	54[a]	Ala 127 H4	3
Lys 63 E6	14	Ala 130 H7	4
His 64 E7	363[a]	Met 131 H8	183
Gly 65 E8	582[a]	Asn 132 H9	1
Val 66 E9	34	Lys 133 H10	1[a]
Thr 67 E10	1	Ala 134 H11	422
Val 68 E11	3426[a]	Leu 135 H12	808
Leu 69 E12	639	Glu 136 H13	2
Thr 70 E13	2	Leu 137 H14	73
Leu 72 E15	514	Phe 138 H15	1171
Gly 73 E16	4	Arg 139 H16	48
Ile 75 E18	1	Ile 142 H19	31
Leu 76 E19	377	Tyr 146 H23	51[a]
Lys 77 E20	6	heme 154	8554
Lys 78 EF1	4		

[a] Indicates conserved residues present in α and β chains of horse and human hemoglobin.

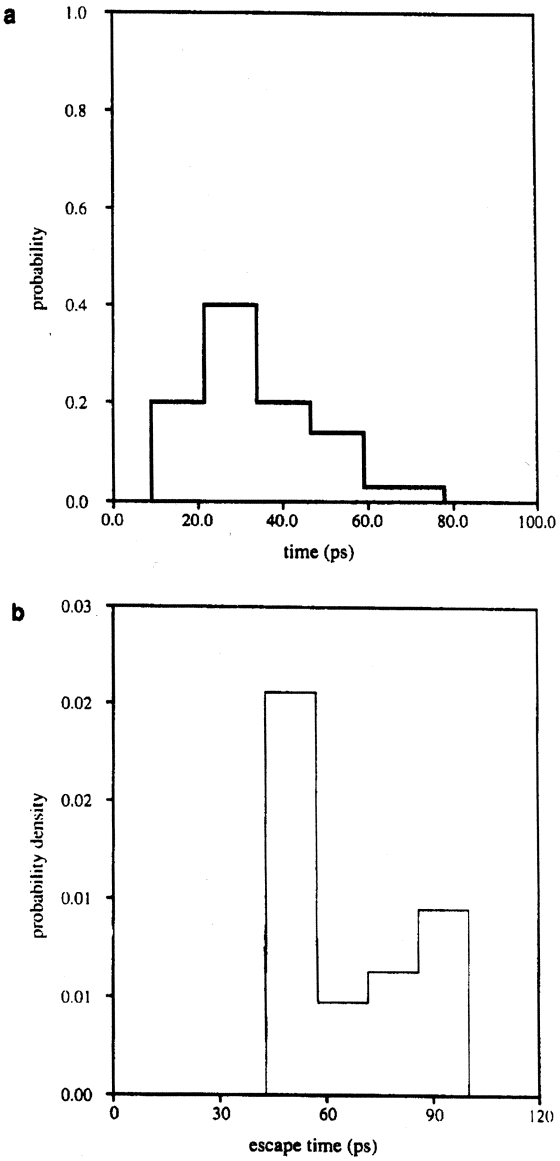

Figure 6. Time distributions: (a) time spent by ligands in heme pocket and (b) time spent by ligands in protein. The histograms are constructed by dividing the difference between the maximum and minimum values of the time into five equal intervals.

with each of these. They collide often but not as much with His 64 E7 and Phe 43 CD1 because they tend to drift from the minimum energy position in the direction of the cavity between the AB/G helices and the cavity between the EF/G helices (see Figure 8). The transition to the AB/G cavity is between Leu 29 B10, Val 68 E11, and Ile 107 G8, in the N-terminal direction at approximately 70° to the heme plane; 42 CO molecules escape by this route. The transition to the EF/G cavity is also between Val 68 E11 and Ile 107 G8 in the N-terminal direction, but the motion is almost parallel (within 10°) to the heme plane; 10 CO molecules escape by this route. Both paths have many collisions with Val 68 E11 and Ile 107 G8, but the former path also involves collisions with Leu 29 B10 while in the latter there are many contacts with the heme. Even with the protein motions, the barriers for escape from the heme pocket are high except in these directions. The dominant barriers along these two pathways are between the residues cited above.

Motion through Protein. Of the ligands that escaped from the heme pocket, 26 remained within the protein and 30 reached the

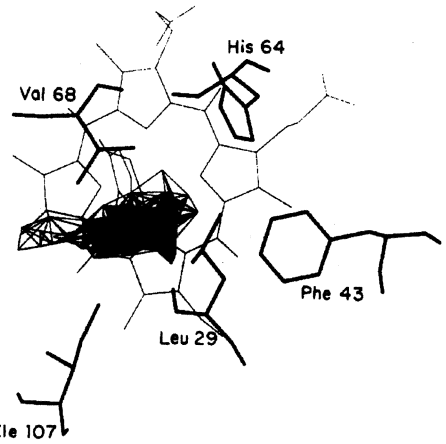

Figure 7. Trajectory of a single carbon monoxide molecule in the heme cavity. The continuous line corresponds to the motion of CO center of mass.

outside of the protein during the 100-ps simulation; the distribution of times required to escape is shown in Figure 6b.

There are five major pathways given in decreasing order of importance; "major" implies that small local variations may occur, which make it impossible to identify the escape with motion of a single residue but rather suggest a "hot spot" in the protein structure. Figure 9 illustrates these pathways by trajectories that follow them. They can be classified as (a) EF and N terminal, (b) A/E, (c) AB/G, (d) proximal histidine, and (e) CD loop. All of these pathways sample certain cavities of myoglobin. There are four significant cavities (including the heme cavity), plus a "semicavity", in which the CO's are found during the trajectories, though not all the cavities are sampled by every trajectory. The cavities are shown in Figure 8 and can be defined in terms of certain residues: heme (His 64 E7, Phe 43 CD1, Val 68 E11, Leu 29 B10, Ile 69 E12, Ile 107 G8); AB/G (Ile 111 G12, Trp 12 A14, Val 15 A17, His B6 24, Leu 115 G16); proximal (heme, His 93 F8, Leu 89 F4, Ile 142 H19, Leu 104 G5, Ile 99 FG5); EF/G (Trp 7 A5, Leu 76 E19, Val 10 A8, Met 131 H8, Leu 135 H12, Ala 134 H11); semicavity (Trp 7 A5, Leu 76 E19, His 82 EFG, Leu 132 H9, Ala 134 H11). They are similar to the ones observed in the MbCO, Xe X-ray structure and the Xe molecular dynamics trajectory. In the paper of Tilton et al., five cavities are described and numbered; 1 is the proximal histidine cavity, 2 is the EF/G cavity, 4 is the AB/G cavity, and 5 is the heme pocket. Number 3 does not correspond to a cavity in the current simulation. Since it is quite close to the EF/G cavity, the difference between it and 3 is smeared out in the trajectory; i.e., 2 plus 3 behave like one large cavity in the present trajectory. Thus, all of the important

Table III. Times in Cavities from the 60 CO Trajectories[a]

	heme	AB/G	EF/G	proximal
Escaped from Protein				
no. of trajectories	30	26	22	7
av time	23.4	34.0	29.2	13.4
max. time	41.0	78.0	88.0	40.0
min. time	9.0	9.0	2.0	3.0
Trapped in Protein				
no. of trajectories	30	30	10	0
av time	41.5	40.5	18.0	
max. time	75.3	51.1	22.1	
min. time	20.7	10.0	3.1	

[a] Results from simulation (a); times in picoseconds.

cavities are present in the X-ray structure. They are not a consequence of dynamic fluctuations, although they are somewhat distorted by the protein motions.

The general features of the ligand diffusion process are the same for all of the paths. Overall, the motion can be described as a few-site random hopping picture, in which the ligand is trapped for a significant time in individual cavities and then hops to another cavity or finally to outside the protein. We list the cavities in Table III and indicate the number of the trajectories that pass through each one. The mean time spent in each and the range of times involved are also given; although the absolute values of the times are not meaningful because of the simulation method, the relative times spent in various regions is indicative. Also, we give the results for the ligands that did not manage to escape during the trajectory. In general, their behavior was similar to that for those that did escape, but they were still trapped in one or another of the cavities at the end of the trajectory. This makes it likely that if the trajectory were extended, they would also escape.

Transitions back and forth between the EF/G and AB/G cavities occur in many trajectories. Jumps between the semicavity close to the N-terminus and the EF/G cavity occur in about half of the trajectories which reach that point. When the proximal histidine cavity is populated it communicates with the EF/G cavity.

To determine the details of the motion outside of the heme pocket we again consider the collisions of the ligand with the residues of the protein (Table II). Figure 10 shows the protein residues which are involved in collisions with the CO's. Some of the most frequently colliding residues are along a sort of "channel" for the escaping CO; i.e., there is a long path from the heme pocket to the region of the N terminus near Trp 7. Its general direction follows that of the E helix. Most of the CO which escape (28 out of the 30) follow this path at least partially, with individual molecules escaping from the protein at different points along the path. These involve the contact regions between the A, B, and E helices. More specifically the CO's spend most of their time

Figure 8. Internal cavities sampled by the ligands. A single carbon monoxide trajectory which passes through all these cavities is used to demonstrate their locations. The trajectory starts at the heme cavity, continues to the EF/G cavity, and then goes to the "semi" cavity, back to EF/G, to the proximal cavity, and back to the EF/G cavity again. It finally escapes between the A, B, and G helices from the AB/G cavity.

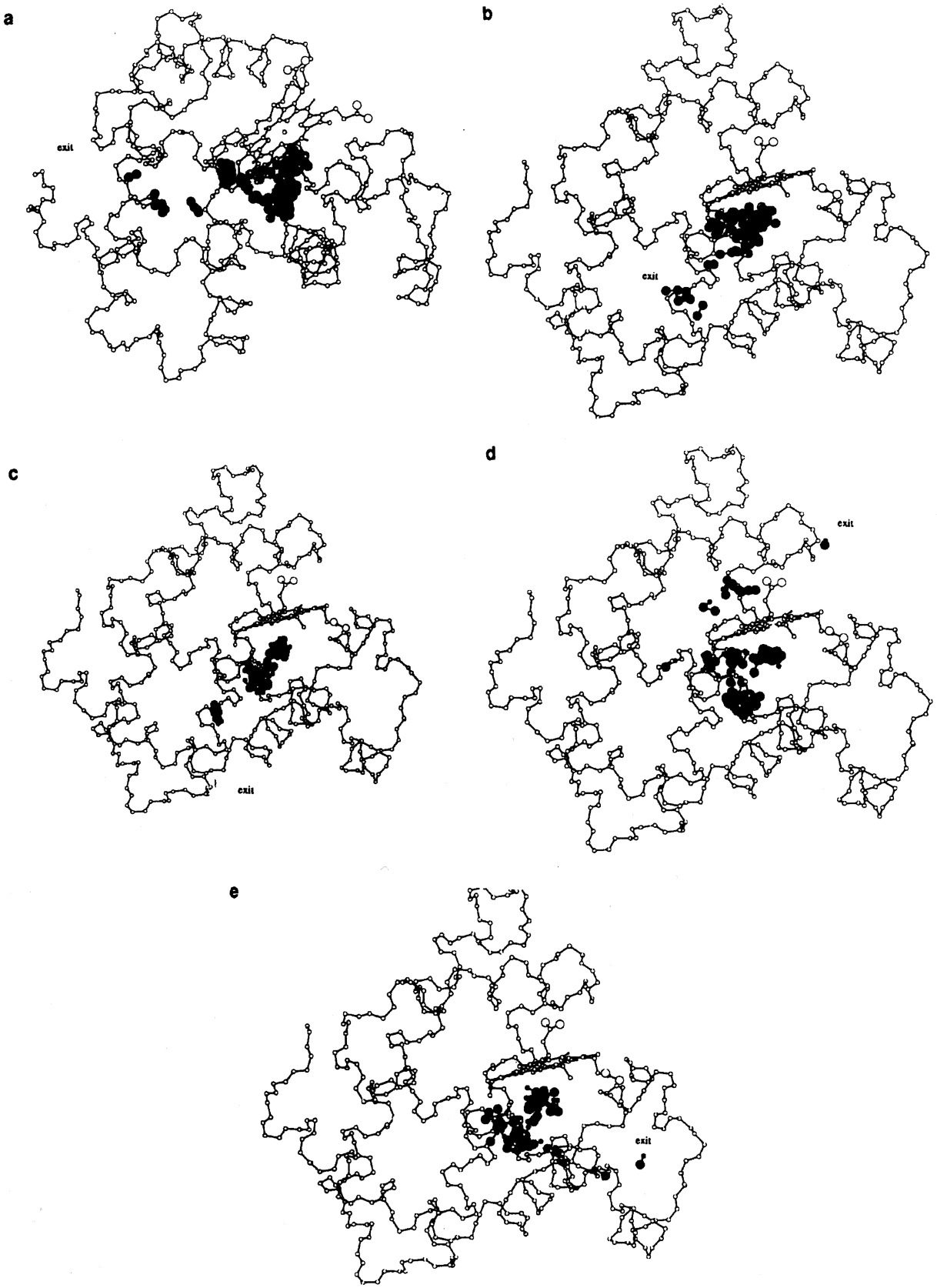

Figure 9. The different exits from the protein are illustrated by ligand trajectories: (a) EF and N terminal exit; (b) A/E exit; (c) AB/G exit; (d) proximal histidine exit; (e) CD exit.

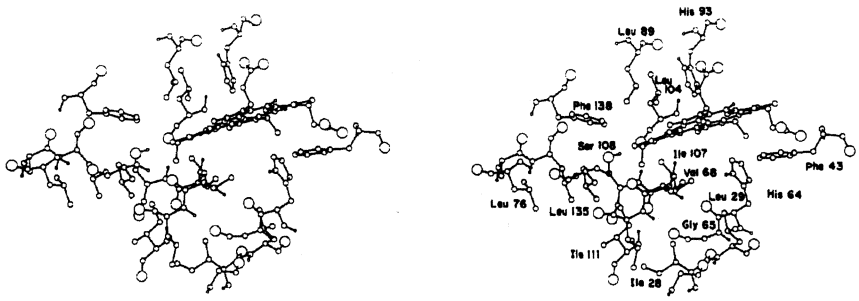

Figure 10. Residues which have more than 350 collisions with the ligand. (See text and Table II for definition of a collision.)

in the B/G and EF/G cavities and to a smaller extent in the proximal histidine cavity (see Figure 8), in addition to the time spent in the heme pocket. The probability of returning to the heme pocket after escape is small (only four CO do so), while transitions between B/G and EF/G occur in almost all trajectories. Transitions back and forth between the AB/G and EF/G cavities occur in most of the trajectories. They take place near the heme and go over a barrier formed by residues Val 68 E11 and Ile 107 G8; there are a total of 105 jumps between these two. All of the trajectories pass near the heme and Ile 107 G8, while 60% also collide with Val 68 E11. After hopping one to three times between the AB/G and EF/G cavities the ligand escapes by one of several routes: (1) From the AB/G cavity between the A and E helices and it comes in contact with residues Trp 14 A12, Leu 72 E15, Gly 73 E16, Leu 76 E19, and Lys 77 E20; (2) From the EF/G cavity toward the N terminus undergoing contacts with Phe 138 H15, Leu 135 H12, and His 82 EF5 to the "semi"-cavity (six trajectories stayed a few picoseconds there before finally escaping). From the semicavity, the escape is direct.

Five CO's do not follow this general pathway. Two escape near the CD corner, and three exit from the proximal histidine cavity; this cavity was found to be populated in the X-ray studied of Xe myoglobin.[27] The CD escape occurs after the ligand is trapped in the B/G cavity, in one case for the period of 20 ps and in the other for 3 ps. The first path came close to the B/E contact (near the backbone of residues Gly 25 B6, Gln 26 B7, Leu 29 B10, His 64 E7, and Gly 65 E8); it then continued in the CD loop, D helix direction (residues Ile 30 B11, Phe 33 B14, and Met 55 D5). The exit from the protein occurred between Phe 46 CD4, Leu 49 CD7, and Met 55 D5. All the residues mentioned were involved in collisions for this specific trajectory (see Table II). The second CD trajectory also starts at the B/G cavity, and it follows a similar route but comes somewhat closer to the E helix; from the B/G the ligand jumps directly to interact with Phe 33 B14, Leu 40 C5, Met 55 D5, and Leu 61 E4, from there it continues to the CD loop (residues Phe 46 CD4, Lys 47, CD5, His 48 CD6, Leu 49 CD7, and Lys 50 CD8. It finally escapes between Lys 50 CD8 and Leu 40 C5. Both trajectories are relatively short, with rapid escape from the protein after leaving the heme pocket.

The proximal cavity path is considerably longer than those involving the CD loop. The two ligands that followed it required 63 and 100 ps for escape. There is basically a single path which involves entrance to the proximal cavity from the EF/G cavity. There are collisions with the heme, Leu 89 F4, Leu 104 G5, Ile 107 G8, Ser 108 G9, Leu 135 H12, Phe 138 H15, and Ile 142 H20. Only seven trajectories visited the proximal cavity. Of these, three escaped directly, and the others returned to the EF/G cavity. One ligand passes between the C_α of His 93 FG and the backbone nitrogens of Lys 98 FG4 and Ile 99 FG5 to the other side of the histidine. It then continued to the propionic groups of the heme and the C and the G helices (residues Thr 39 C4, Lys 42 C7, and Tyr 103 G4 are the last "colliders"). The second ligand followed a direct path to escape between the F and H helices (the last colliders are Leu 89 F4, Phe 138 H15, and Ile 142 H20). The third ligand, like the first, passed by His 93. However, it escaped between the N_δ of His 93, O_γ of Ser 92 F7, and the main chain of Ala 90 F5.

None of the residues involved in frequent collisions with the CO (Table II) underwent dihedral angle transitions. The maximal change observed in χ_1 or χ_2 was 42°, which indicates that some of the side chains underwent significant fluctuations. The importance of motion, per se, is confirmed by comparison with the rigid protein trajectory results (trajectory (b)).

To make clearer the nature of the dynamics, we describe some individual trajectories in which the ligand escapes from the protein. This permits us to illustrate general as well as special features of each trajectory.

Trajectory I (Figure 9a). The CO molecule spends considerable time in the heme pocket. Only one direction is approximately open in the heme pocket, and this is where the CO goes. It is parallel to the E helix and passes just below the heme. Since a barrier exists even in that direction, the CO cannot exit immediately. The side chains of Ile 107 G and Val 68 E contribute significantly to this barrier which inhibits a jump to the EF/G cavity. The last step in the ligand escape is an almost *one-dimensional* diffusive motion of 10 Å to the exit near the beginning of the A helix. The direction follows the E helix. The residues which the carbon monoxide passes in order along the trajectory are Leu 72 E15, Trp 14 A12, Leu 76 E19, Leu 135 H12, Met 131 H8, Val 10 A8, and Trp 7 A5. The exit is located in the neighborhood of Trp 7 A5.

Trajectory II (Figure 9b). The CO is trapped in the heme pocket for 90% of the simulation time. Since the collisions in the heme pocket were described already, we discuss only the remaining 10% of the trajectory. The ligand escapes in a direction approximately perpendicular to the heme to the AB/G cavity between the E helix and the AB corner. It passes the barrier formed by the residues Leu 69 E12, Gly 25 B6, and Ile 28 B8 from the heme pocket to the AB/G cavity. From this cavity the CO follows the shortest possible path to exit between the A and E helices. The exit is between Val 17 A15 and Thr 70 E13 and it is of major importance, since many trajectories escape in this way. However, not all trajectories which escape from the protein by this exit follow the same diffusive path to reach it.

Trajectory III (Figure 9c). This trajectory is similar to the previous and includes the same elementary diffusion steps in the protein interior. A considerable period of the simulation is spent in the heme pocket (90%), and then the ligand hops to the AB/G cavity. It stays in this cavity only 10% of the simulation period. The transition from the heme pocket to the AB/G cavity was between the residues Ile 28 B9, Leu 29 B10, and Val E11 68. After hopping to the AB/G cavity the ligand remained (until escaping) in the proximity of the residues Trp 14 A12, Val 17 A15, His 24 B5, Ile 28 B9, and Leu 29 B10.

(b) Simulation with High-Temperature CO and Rigid Protein. The hot CO, rigid protein simulation was performed to obtain a comparison with the results for the case where the protein underwent normal room-temperature fluctuations (simulation (a)). Most of the CO molecules (45 out of 60) remained in the heme pocket for the entire trajectory. Those that did not, escaped more slowly from the heme pocket than in the fluctuating protein trajectory; i.e., the average "time" in the heme pocket for the ligands that did escape (15 CO) was 70 ps. The CO's explored several of the cavities that play a role in the full simulation. The

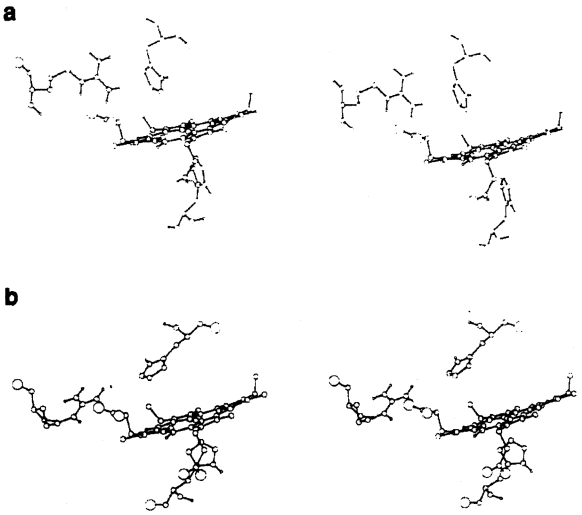

Figure 11. Closed and open heme pocket structure resulting from a transition of the distal histidine side chain in simulation (c): (a) closed (initial) structure and (b) open structure after 10 ps of simulation.

cavities include the one in the AB/G corner, which was reached by 10 CO's, and the cavity EF/G, which was reached by three CO's directly from the heme pocket.

Only two CO molecules escaped from the proteins during the 100-ps trajectory. They both followed the most common pathway of the fluctuating protein trajectory. From the heme pocket, they went to the AB/G cavity, made a second jump to the EF/G cavity, and escaped near Trp 14 A12. The first CO stayed 85% of the simulation in the heme pocket and 15% in the AB/G cavity and then escaped; the second CO stayed for 50% of the simulations in the heme pocket, 22% in the AB/G cavity, and 11% in the EF/G cavity before escaping by the same exit.

If the time-scale difference of the fluctuating and rigid protein simulations is ignored, there are similarities in the overall behavior. This would suggest that the possible pathways are determined by the average structure of the protein but that the barriers (which determine the rates) are lowered significantly by protein fluctuations.

(c) Simulation with High-Temperature CO and Side Chains and Rigid Protein Main Chain and Heme. In this simulation, a temperature of 900 K was employed for the protein side chains and the 60 CO molecules; the heme and the polypeptide chain were kept fixed. As a result of their higher temperature, the side chains have much larger fluctuations than in the room-temperature protein simulation. In the present 20-ps trajectory, 15 χ_1 transitions occur, as compared with 16 in the entire 100-ps room-temperature trajectory. Thus, the rate of transition was increased by approximately a factor of 4. Also, the mean amplitude of the dihedral angle fluctuations is significantly larger even when there are no transitions. The most critical transition occurs at 5 ps when His 64 E7 swings out from the heme pocket and creates a pathway to exterior. It remains in this position for the remainder of the trajectory. This orientation is essentially that observed in the X-ray structures with a large ligand bound to heme. As found in the "door stop" structure,[21] Arg 45 moves away from the His in the CD direction (N_ϵ moves by 3.1 Å); the simulation structure has $\chi_1 = -61$; $\chi_2 = 174$; $\chi_3 = -64$ while the X-ray $\chi_1 = -160$; $\chi_2 = -151$; $\chi_3 = -61$. There is also a transition of Phe 43 CD1 whose χ_2 changes by 90°. No other transition of heme pocket residues are detected in the simulation. The behavior is in accord with the mutant studies in that *removal* of the barrier due to the presence of His 64 E7 (either by mutation or by reorientation) makes this path the dominant one. This does not show, however, that when the His is present in the normal orientation the same path is most important (Figure 11). During the trajectory, the 10 ligands that escaped followed a path directly through the hole opened by the His between heme and the E helix. The general direction is close to the minimum distance path found in the original simulation,[8] but because the His has moved out of the way the barriers due to Val 68 E11 and Thr 44 E10 do not play a role (see Figure 11, ref 8). If one examines all 60 ligands, one finds that some leave the heme pocket and enter the cavities occupied in the other trajectories.

Overall the CO motions in this trajectory are more localized than in the room-temperature protein simulation. One reason is that the trajectory is shorter. Thus, most of the ligands (40) are still in the heme pocket at the end of the trajectory, 10 diffused to the AB/G cavity, and 10 escaped. None of the ligand reached the other mentioned cavities. In comparison with the first 20 ps of the room-temperature trajectory, the two simulations are rather similar with respect to the diffusion through the protein. The major difference is that no escapes happen in that time in the room-temperature simulation.

V. Conclusion

A new method for enhanced sampling of ligand dynamics in a protein has been presented and applied to the exit pathways for carbon monoxide from myoglobin after photodissociation. This has permitted us to examine a large number of CO trajectories with a relatively small amount of computer time. It is clear that the ligand motion is more complex than was thought from previous studies. The CO pathways found by high-temperature ligands moving through a room-temperature protein involve mainly the cavities which were observed in the X-ray structure. The general behavior of most CO molecules can be described as a diffusive hopping from one cavity to another. Inspection of the trajectories show that barriers between the cavities are reduced transiently by motions of the protein backbone coupled with fluctuations in side-chain orientations. These are of the type observed in transitions from one substate to another in a molecular dynamics analysis.[47] Even with high-temperature ligands, few escape in the simulation with a rigid protein. Those that do follow pathways that are sampled in the fluctuating protein trajectories.

The ligand diffusion process is of considerable complexity so that a simple, single reaction coordinate approach is not likely to yield a full description of the dynamics. Instead an average over reaction paths, including multiple barrier crossings for each one, must be performed. Although the relative importance of the various paths is not determined by the present simulations, the results may suggest the possibility of a transition in the dynamics of ligand escape. At low temperatures, side-chain transitions may be sufficiently rare so that the "rigid" protein paths are dominant; e.g., for a 80 K myoglobin simulation there are no dihedral angle transitions.[48] At higher temperature, the side-chain fluctuations become more important, and a new channel (the "direct" His 64 E7 gate) for escape may be opened.

The present study, although an improvement over earlier work, still suffers from several limitations. The CO energy is four times larger than the typical size of barriers which were observed experimentally.[14,49] The measured barriers are on the order of a few kcal/mol in myoglobin. Thus, it is possible that some of the paths which we observed are inaccessible at lower energies. Also, solvent was not included in the simulation. Though the direct influence of the solvent on the internal diffusion of the ligand is expected to be small, it may affect the average structure of the protein and hence the nature of the cavities and pathways between them. The most significant deviation from the X-ray structure is the change in the heme pocket, which pushes the ligand away from the E helix and from the path near His 64E7 suggested in the earlier simulation.[8] Also, it is possible that there are alternative

(48) Kuczera, K.; Kuriyan, J.; Karplus, M. *J. Mol. Biol.* **1990**, *213*, 351.
(49) Chatfield, M. D.; Walda, K. N.; Magde, D. *J. Am. Chem. Soc.* **1990**, *112*, 4680.
(50) Johnson, K. A.; Olson, J. S.; Phillips, G. N. *J. Mol. Biol.* **1989**, *207*, 459.
(51) Carver, T. E.; Rohlfs, R. J.; Olson, J. S.; Gibson, Q. H.; Blackmore, R. S.; Springer, B. A.; Sligar, S. G. *J. Biol. Chem.* In press.
(52) Carver, T. E.; Olson, J. S.; Swerdon, S. J.; Krzywda, S.; Wilkinson, A. J.; Gibson, Q. H.; Blackmore, R. J.; Ropp, J. D.; Sligar, S. G. *Biochemistry*. Submitted for publication.

paths which involve rare protein fluctuations (e.g., side-chain motions of certain residues that are not sampled properly in the simulation.)

In spite of these difficulties, the present study provides new insights into the process of ligand motion through the protein matrix. Simulations with solvent as well as activated dynamics of some of the barrier crossing events are in progress. Of particular interest also is the possibility of testing the results by kinetic studies of appropriately constructed myoglobin mutants.

Appendix

To obtain eqs 17a and 17b we assume eqs 15 and 16. With eq 6a applied to the coordinates s

$$\frac{\partial}{\partial t}\langle \mathbf{Q}_s\rangle = \left\langle \frac{\partial \mathcal{H}}{\partial \mathbf{P}_s}\right\rangle \tag{A1}$$

we can obtain an explicit expression for the left-hand side by introducing eqs 15 and 16 for the form of the probability density; i.e.

$$\frac{\partial}{\partial t}\langle \mathbf{Q}_s\rangle = \frac{\partial}{\partial t}\int \rho_{N-s}(\mathbf{P}_{N-s},\mathbf{Q}_{N-s},t)\, \mathrm{d}+\mathbf{P}_{N-s}\, \mathrm{d}\mathbf{Q}_{N-s} \tag{A2}$$

$$x\sum_w w \int \mathbf{Q}_s\delta_w(\mathbf{P}_s-\mathbf{P}^\circ_{s,w}(t),\mathbf{Q}_s-\mathbf{Q}^\circ_{s,w}(t))\, \mathrm{d}\mathbf{P}_s\, \mathrm{d}\mathbf{Q}_s \tag{A3}$$

Since the (normalization) integral over the $N-s$ subsystem is equal to unity, we have

$$\frac{\partial}{\partial t}\langle \mathbf{Q}_s\rangle = \sum_w w\frac{\partial}{\partial t}\int \mathbf{Q}_s\delta_w(\mathbf{P}_s-\mathbf{P}^\circ_{s,w}(t),\mathbf{Q}_s-\mathbf{Q}^\circ_{s,w}(t))\, \mathrm{d}\mathbf{Q}_s\, \mathrm{d}\mathbf{P}_s \tag{A4}$$

or

$$\frac{\partial}{\partial t}\langle \mathbf{Q}_s\rangle = \sum_w w\frac{\partial}{\partial t}\mathbf{Q}^\circ_{s,w}(t) \tag{A5}$$

To evaluate the right-hand side of eq A1, we write out the average explicitly and find

$$\left\langle \frac{\partial \mathcal{H}}{\partial \mathbf{P}_s}\right\rangle = \int \rho_{N-s}(\mathbf{P}_{N-s},\mathbf{Q}_{N-s},t)\sum_w w\delta(\mathbf{P}_s-\mathbf{P}^\circ_{s,w}(t),\mathbf{Q}_s-\mathbf{Q}^\circ_{s,w}(t))\frac{\partial \mathcal{H}}{\partial \mathbf{P}_s}\, \mathrm{d}\mathbf{P}_{N-s}\, \mathrm{d}\mathbf{Q}_{N-s}\, \mathrm{d}\mathbf{P}_s\, \mathrm{d}\mathbf{Q}_s \tag{A6}$$

Integrating over the δ-functions, we obtain

$$\left\langle \frac{\partial \mathcal{H}}{\partial \mathbf{P}_s}\right\rangle = \sum_w w\int \rho_{N-s}(\mathbf{P}_{N-s},\mathbf{Q}_{N-s},t)\frac{\partial \mathcal{H}}{\partial \mathbf{P}^\circ_{s,w}}\, \mathrm{d}\mathbf{P}_{N-s}\, \mathrm{d}\mathbf{Q}_{N-s} \tag{A7}$$

where the integral does not separate into integrals over s and $N-s$ because H is a function of both sets of variables; i.e.

$$\frac{\partial \mathcal{H}}{\partial \mathbf{P}^\circ_{s,w}} = \frac{\partial \mathcal{H}(\mathbf{P}_{N-s},\mathbf{Q}_{N-s},\mathbf{P}^\circ_{s,w}(t)\mathbf{Q}^\circ_{s,w}(t))}{\partial \mathbf{P}^\circ_{s,w}} \tag{A8}$$

The remaining integral in eq A7 gives the average of $\partial\mathcal{H}/\partial\mathbf{P}^\circ_{s,w}$ over the distribution $\rho_{N-s}(\mathbf{P}_{N-s},\mathbf{Q}_{N-s},t)$

$$\left\langle \frac{\partial \mathcal{H}}{\partial \mathbf{P}_s}\right\rangle = \sum_w w\left\langle \frac{\partial \mathcal{H}}{\partial \mathbf{P}^\circ_{s,w}}\right\rangle_{N-s} \tag{A9}$$

Finally, from eqs A1, A5, and A9, we have

$$\sum_w w\frac{\partial}{\partial t}(\mathbf{Q}^\circ_{s,w}) = \sum_w w\left\langle \frac{\partial \mathcal{H}}{\partial \mathbf{P}^\circ_{s,w}}\right\rangle_{N-s} \tag{A10}$$

and, correspondingly, for the momentum equation

$$\sum_w w\frac{\partial}{\partial t}(\mathbf{P}^\circ_{s,w}) = -\sum_w w\left\langle \frac{\partial \mathcal{H}}{\partial \mathbf{Q}^\circ_{s,w}}\right\rangle_{N-s} \tag{A11}$$

where the "derivatives" with respect to a vector denote gradients to take account of the fact that the s subsystem can have more than one degree of freedom; eqs 23 and 24 are first-order differential equations for the parameters $\mathbf{Q}^\circ_{s,w}$. As for eq 12, there is one and only one solution, that involving separable trajectories that determine the parameters $\mathbf{Q}^\circ_{s,w}(t)$ and $\mathbf{P}^\circ_{s,w}(t)$, so that

$$\frac{\partial}{\partial t}\mathbf{Q}^\circ_{s,w} = \left\langle \frac{\partial \mathcal{H}}{\partial \mathbf{P}^\circ_{s,w}}\right\rangle_{N-s} \tag{A12a}$$

$$\frac{\partial}{\partial t}\mathbf{P}^\circ_{s,w} = -\left\langle \frac{\partial \mathcal{H}}{\partial \mathbf{Q}^\circ_{s,w}}\right\rangle_{N-s} \tag{A12b}$$

Equations A12a and A12b differ from the exact equations (eqs 12a and 12b) in that coordinates of subsystem s are determined in the average field of the remaining $N-s$ coordinates. To be able to evaluate these averages as a function of time we need to know $\rho_{N-s}(\mathbf{P}_{N-s},\mathbf{Q}_{N-s},t)$; we make the simplification, already mentioned, that we can represent the $N-s$ bath subsystem, by a single δ-function (eq 17), and the average over the $N-s$ coordinates in eq 25a becomes

$$\left\langle \frac{\partial \mathcal{H}}{\partial \mathbf{P}^\circ_{s,w}}\right\rangle_{N-s} = \int \delta(\mathbf{P}_{N-s}-\mathbf{P}^\circ_{N-s}(t),\mathbf{Q}_{N-s}-\mathbf{Q}^\circ_{N-s}(t))\frac{\partial \mathcal{H}}{\partial \mathbf{P}^\circ_{s,w}}\mathrm{d}\mathbf{P}_{N-s}\, \mathrm{d}\mathbf{Q}_{N-s} = \frac{\partial \mathcal{H}(\mathbf{P}^\circ_{s,w},\mathbf{Q}^\circ_{s,w},\mathbf{P}^\circ_{N-s},\mathbf{Q}^\circ_{N-s})}{\partial \mathbf{P}^\circ_{s,w}} \tag{A13}$$

Substituting into eq A12a we obtain eq 17a, and by a corresponding procedure we obtain eq 17b. Finally, we need the equations of motion for the $N-s$ system coordinates and momenta. From the first of eq 6 for the $N-s$ coordinates

$$\frac{\partial}{\partial t}\langle \mathbf{Q}_{N-s}\rangle = \left\langle \frac{\partial \mathcal{H}}{\partial \mathbf{P}_{N-s}}\right\rangle \tag{A14}$$

we have for the left-hand side

$$\frac{\partial}{\partial t}\langle \mathbf{Q}_{N-s}\rangle = \int \delta(\mathbf{P}_{N-s}-\mathbf{P}^\circ_{N-s}(t),\mathbf{Q}_{N-s}-\mathbf{Q}^\circ_{N-s}(t)) \tag{A15}$$

$$x\sum_w w\delta(\mathbf{P}_{s,w}-\mathbf{P}^\circ_{s,w}(t),\mathbf{Q}_{s,w}-\mathbf{Q}^\circ_{s,w}(t))\frac{\partial}{\partial t}(\mathbf{Q}_{N-s})\mathrm{d}\mathbf{P}_{N-s}\, \mathrm{d}\mathbf{P}_s\, \mathrm{d}\mathbf{Q}_s$$

The integration over the s subsystem yields unity, and we obtain

$$\frac{\partial}{\partial t}\langle \mathbf{Q}_{N-s}\rangle = \frac{\partial}{\partial t}(\mathbf{Q}^\circ_{N-s}) \tag{A16}$$

Proceeding similarly with the right-hand side of eq A14, we obtained eqs 18a and 18b.