Application of the Locally Enhanced Sampling (LES) and a Mean Field with a Binary Collision Correction (cLES) to the Simulation of Ar Diffusion and NO Recombination in Myoglobin

Alex Ulitsky and Ron Elber*,†

Department of Chemistry, M/C 111, University of Illinois at Chicago, P.O. Box 4348, Chicago, Illinois 60680, and Department of Physical Chemistry, Fritz Haber Center and the Institute of Life Sciences, The Hebrew University, Givat Ram, Jerusalem 91904, Israel

Received: August 26, 1993®

Simulations of argon diffusion and of picosecond recombination of nitric oxide to myoglobin are presented. Three computational methods are compared, usual classical trajectories, mean field trajectories, and mean field trajectories with a binary collision correction. The ligand mobility and the spatial distribution of the ligand in the protein interior are examined. Time-dependent properties calculated with the binary collision model are in good agreement with exact simulations. This is to be contrasted with the original mean field protocol that is not working well for dynamics.

1. Introduction

Computer simulations had contributed significantly to the understanding of structure and dynamics of ligand recombination and diffusion in the globins.[1-14] Nevertheless, theoretical investigations of nonequilibrium dynamics associated with ligand photodissociation are computer intensive and therefore attempted only for very short times.[9-14] For a system that is not at thermal equilibrium, an ensemble of trajectories with different initial conditions is needed to sample configurations for the required averages. It is the purpose of this paper to present an application of a promising method for more efficient phase-space averaging in nonequilibrium problems.

To accelerate the sampling process for only a part of the system, an approximate simulation methodology based on the classical time-dependent self-consistent field (TDSCF)[15] was introduced in the past.[4] It was called locally enhanced sampling (LES). LES was employed in a number of *exact* calculations of time-independent properties. Examples include calculations of free energy differences[16] and global energy minimization.[17-19] However, LES is an approximation whose validity is hard to assess when applied to study time-dependent properties.[20,21]

Recently, a binary collision model (cLES) was proposed as a dynamic correction to the LES approximation.[21] cLES was tested in calculations of equilibrium (temperature, force virials), structural (spatial correlation function), and nonequilibrium properties (escape times of a trapped particle from rare-gas matrices). It was shown that cLES significantly improves the above properties compared to LES. The good agreement with the results of regular classical trajectories was obtained at a considerable saving of CPU time using a cLES protocol.

In this manuscript, cLES is used to investigate ligand dynamics in myoglobin. Two processes are modeled, Ar diffusion in the protein interior and NO recombination to the wild type (Leu 29) and Val 29 mutant of myoglobin. The ligand mobility and the spatial distribution were computed using regular classical trajectories, LES, and cLES.

A comprehensive set of approximate calculations of these processes is available to us.[4,11] Therefore, the application of cLES to these problems is useful as a test of the new approximation for dynamics. The calculations of long-time diffusion make it possible to critically review some old calculations that employed the LES approximation.[4] Methodologically, the cLES formulation is extended in this manuscript to include collisions between particles with internal (covalent) structure. Only collisions between rare-gas atoms were considered in the past.[21]

In section 2, we discuss the extensions of the binary collision model to make it possible to study proteins with monoatomic and diatomic ligands. In the third section, the computational protocol is presented. The results are summarized in the fourth section, which is followed by conclusions.

2. LES and a Binary Collision Model

Locally enhanced sampling (LES) is based on a classical version of the time-dependent hartree (TDH) approximation.[15] In LES, the ligand is represented by N copies that are propagated simultaneously in a single-protein matrix (eqs A1 and A2 in the Appendix). The intramolecular forces of the ligand and of the protein are treated exactly. The approximation is in the calculation of ligand–protein interactions. The mean field approximation is approaching the exact result when the subsystems are interacting weakly, i.e., when the average interactions between the ligand and the protein do not contribute considerably to their dynamics. When the ligand and a protein atom collide, the weak interaction limit is not necessarily valid and a better partitioning of the mean field coordinates (instead of protein–ligand) is desirable. A sequence of different mean field approximations is used in the binary collision correction to LES (cLES) in which the coordinate representation of the mean field is reselected each time step.[21] This makes it possible to treat more accurately the interactions between the protein atoms and the ligand.

The different mean field representations are defined to account for collisions between the ligand and various protein atoms that occur during the time evolution of the system. While it is intuitively clear that two particles collide if the distance between them is less than a preset value, it is less clear how to define a collision between a protein atom and a LES particle. This is because the LES particle is "distributed" in space and represented by several copies. In the binary collision model, we assume that a collision between these particles takes place if at least one of the copies is within a collision cutoff (σ_c) from the protein atom. We further assume that at most, one (hard) collision occurs at a time and that this collision is binary (i.e., a ligand copy and a single protein atom are the only colliding particles).[21]

In the binary collision model, the LES Lagrangian L°_{LES} (eq A3) is used at times when a collision is not detected. A different mean field approximation is employed to describe hard collisions. It is based on the relative ($q_L = x_{pn} - x_L$) and the center of mass ($q_{cm} = (M_n x_{pn} + M_L x_L)/(M_n + M_L)$) coordinates for the colliding

† The Hebrew University.

® Abstract published in *Advance ACS Abstracts*, December 15, 1993.

0022-3654/94/2098-1034$04.50/0

particles (eqs A4 and A5).[21] $M_{n,L}$ are the masses of the monoatomic ligand and a colliding protein atom—n; $x_{pn,L}$ are the coordinates of these particles. The enhanced sampling is applied to the relative coordinate (which is "multiplied") (eq A4) rather than to the Cartesian coordinate of the ligand as in LES. For a diatomic ligand, the Jacobi coordinates

$$q_{L1} = x_{L1} - x_{L2}, \quad q_{L2} = x_{pn} - \frac{M_{L1}x_{L1} + M_{L2}x_{L2}}{M_{L1} + M_{L2}},$$

$$q_{cm} = \frac{M_n x_{pn} + M_{L1}x_{L1} + M_{L2}x_{L2}}{M_n + M_{L1} + M_{L2}}$$

were adopted to describe the colliding particles that are the ligand and the protein atom n (eqs A7–A9). Here, $x_{L1,L2}$ and $M_{L1,L2}$ are the coordinates and the masses of the atoms in the diatomic ligand. These new coordinates (eqs A4–A10) define a collision representation. The Lagrangian for the collision representation (L^n_{coll}) is presented in eqs A11 and A12. For a diatomic ligand, L^n_{coll} was derived using the procedure proposed in our previous paper.[21] The Jacobi coordinates with enhancement of all relative positions suggest a plausible choice of the collision representation for more structured molecules.

Up to now, we considered two different mean field descriptions for the ligand–protein system (LES and collision). What is left to be defined is the switching procedure between LES and the collision representation. This is because the transformation between different mean field approximations is not unique. Assume a collision of the LES copy with the n-th protein atom starting from a regular LES. To start the dynamics in the collision representation, the initial velocities and coordinates need to be calculated from the previous mean field representation. These coordinates (and velocities) are assigned according to eqs A4–A10 using the LES variables at the time of the collision. The potential energy of the subsystems (the ligand and the protein) is the same in both representations. However, the value of the kinetic energy is not. Therefore, we used an additional procedure to refine the initial values for the velocities in the transformation from LES to the collision representation. The velocities of the protein particles exluding the colliding one are computed according to eqs A6 and A10. For the colliding pair, the velocity of the center of mass motion q_{cm} (eqs A5 and A9) and the set of velocities for the enhanced part (denoted by index "L" in the left part of eqs A4, A7, and A8) are obtained by requiring that the total energy and therefore the kinetic energy are conserved in the transformation. From the solutions of eq A13 ($dq^\circ_{L,j}/dt$, j = 1,..., N_L), we choose the set that would minimize the deviation ($\sum_j(dq_{L,j}/dt - dq^\circ_{L,j}/dt)^2$). Here, the values $dq_{L,j}/dt$, j = 1,..., N_L are assigned according to (eqs A4, A5, and A7–A9). The total momentum for the protein–ligand system is not conserved in the switching procedure. Only the relative momenta of the ligand and the protein particles are conserved. We considered the energy criteria to be more important. Simultaneous conservation of the total energy and the total momenta is not always possible.

To complete this section, we provide an algorithm-like description of how the simulation with the switched mean field is pursued.

(a) The simulation is initiated using the LES protocol with coordinates and velocities provided by the user.

(b) A molecular dynamic step is calculated according to the LES Lagrangian.

(c) The distances between the real particles and the copies of the enhanced part are calculated. If all the distances are larger than the "distance of collision" (σ_c) go to (b). If one of the distances is smaller than σ_c, go to (d).

(d) In this step, the switch between LES and the collision representation is performed. An effective Lagrangian (eqs A11 and A12) corresponding to this collision is evaluated. The starting velocities and coordinates are extracted using eqs A4–A10. The initial guess is refined to account for kinetic energy conservation.

(e) A molecular dynamics step in the collision representation is computed using the equations of motion according to L_{coll}.

(f) The distance between the colliding particles is calculated. If the distance is smaller than σ_c, go to (e). If the distance is larger than σ_c, continue to (g).

(g) Switch back from the collision representation to LES. Coordinates and velocities are obtained using eqs A4–A10. The deviations in the kinetic energy are corrected by a uniform scaling of the velocities (i.e., by multiplying all the velocities by the same constant factor). A typical scaling factor is between 0.97 and 0.99

This completes the description of the collision correction to the LES that we call the cLES protocol.

3. Computational Procedure

3.1. Ar Diffusion Study. The cLES algorithm was implemented in the program MOIL.[22] The X-ray coordinates of Phillips et al. were used for myoglobin (Mb).[23] To prepare a solvation shell, the protein was placed in a 40-Å × 40-Å × 40-Å water box. Overlapping water molecules and water molecules with a distance larger than 6 Å from any protein atom were eliminated. The TIP3P model was used for the water molecules.[24] The RATTLE algorithm was used to constrain bond angles and bond lengths for the water molecules.[25] The protein degrees of freedom were unconstrained. An Ar atom was positioned initially in the heme cavity, and the structure was equilibrated for 10 ps at 300 K using a velocity-scaling procedure each 100 steps. The van der Waals parameters for Ar were (ϵ = 0.238 kcal/mol, σ = 3.405 Å).[26] To initiate the LES and cLES dynamics, the Ar atom was replaced by 10 copies. The starting velocities of all the particles were sampled from Boltzmann distribution at 300 K. The equations of motion were solved by the Verlet algorithm[27] using a 2-fs time step and 9-Å cutoff. The nonbonded list was updated every 10 steps.

Four 200-ps trajectories were computed in the microcanonical ensemble for myoglobin with 10 Ar copies using LES and the cLES protocol. Four 700-ps trajectories were computed using ordinary classical trajectories.

We also calculated two 200-ps trajectories at 300 K using LES with the velocity-scaling procedure. The temperature in the last simulations was maintained by scaling velocities for each particle every 200 steps of molecular dynamics.

3.2. NO Recombination Study. In the recombination study, a solvation shell was not included. During the first picoseconds after the dissociation, the ligand remains in the heme pocket and the effect of the solvent is expected to be small. This was also demonstrated in earlier studies.[11]

The set of initial structures used in the nitric oxide recombination study by Li et al.[11] was employed. Ten starting coordinates of the ligated myoglobin were prepared by the following protocol. A bound NO molecule was placed in the heme pocket. The system was heated up to 300 K from the initially minimized X-ray structure by velocity scaling over a period of 10 ps and then equilibrated for an additional 10 ps. Ten structures of the ligated protein were sampled at random from the 10-ps equilibration run. In these structures, the NO ligand was substituted by 10 copies occupying the same position in space. These coordinate sets were employed to initiate the LES and cLES trajectories. The starting velocities were sampled from Boltzmann distribution at 300 K. Two different sets of initial velocities were generated for each initial coordinate set. A total of 20 LES and 20 cLES trajectories were therefore generated and analyzed. The trajectories were initiated by instantaneously applying the potential corresponding to the excited state of NO to all ligand copies and

the unbound potential to the heme. The time step and the cutoff distance were used in a similar way to the diffusion study. The trajectory length was 10 ps.

To model the photodissociation process, we used the computational scheme that was discussed in detail previously.[11] Analysis of the rebinding data suggests that whenever ligands in the ground state approach the heme beyond $R_{cross} = 3.4$ Å (R_{cross} is a crossing point of the ground and excited electronic states of the ligand), none of these ligands returns to the unbound state. Therefore, when the ligand copy is in the ground electronic state at a distance $R_{N-Fe} < R_{cross}$ moving *toward* the heme, that copy was considered bound and it was eliminated from the simulations. This is an additional approximation that makes it impossible to study the heme relaxation after rebinding. However, it is expected to be valid for studying the ligand motion. When the ligand is passing the crossing point, moving *away* from the heme, a transition to both electronic states is allowed. Therefore only "unbound" potential[11] was used for the heme.

The semiclassical Landau–Zener model was used to describe the NO transitions between the electronic states in the vicinity of the crossing point.[28,29] The probability P of the system to remain on the adiabatic surface when passing the crossing point is

$$P = 1 - \exp\left(-\pi\frac{\gamma_{LZ}}{2}\right) \tag{1}$$

where γ_{LZ} is

$$\gamma_{LZ} = \frac{2\pi\delta^2}{h v|F_1 - F_2|} \tag{2}$$

F_1 and F_2 are the slopes of the diabatic curves at the crossing point, h is the Plank constant, v is a classical velocity along the reaction coordinate (in this model it is the distance between the nitrogen and iron atoms), and δ is the electronic coupling between the two states. The electronic coupling constant δ was 200 cm^{-1}. The ground state of NO was modeled by a Morse function,

$$V_g(x) = D(e^{-2\alpha x} - 2e^{-\alpha x}) \tag{3}$$

where D is the binding energy (D = 30 kcal/mol), α is a range parameter ($\alpha = 2$ Å^{-1}), and x is the deviation of the nitrogen–iron distance (R_{N-Fe}) from the equilibrium value (R_o). The excited state was modeled by a single exponential function,

$$V_e(r) = Ae^{-br} - B \tag{4}$$

where r is the distance between the nitrogen and iron atoms (R_{N-Fe}). The shift of B = 4 kcal/mol was used to describe the difference between the asymptotic values of the excited and ground-state energies for the NO molecule. The range parameter b and the coefficient A were assigned on the basis of the photon energy in photolysis experiments ($b = 1$ Å^{-1}, A = 80 kcal/mol).[11]

4. Results

4.1. Ar Diffusion Study. To study the effect of the approximations in LES and cLES on the properties of the protein, average structural characteristics were computed. The time dependence of the rms deviation of the trajectory structures from the initial coordinates is shown in Figure 1a. The data were collected from four trajectories computed for each protocol and averaged over 5 ps for exact, LES, and cLES protocols. The rms converges to ~1.6 Å when all the protein atoms are included and to ~1.2 Å for main-chain atoms (Figure 1b). We specifically examined the neighborhood of the heme since this is the most likely domain to be affected by the mean field approximation. Residues which board the heme cavity—the heme, His 64, Phe 43, Val 68, Leu 29, Ile 69, and Ile 107—have rms of ~1.25 Å in all protocols (Figure 1c). The solvation shell restricts the mobility of the protein and increases the time needed for the rms to converge. Convergence is achieved after ~70 ps which is approximately twice as long as that in vacuum simulations.[4,30]

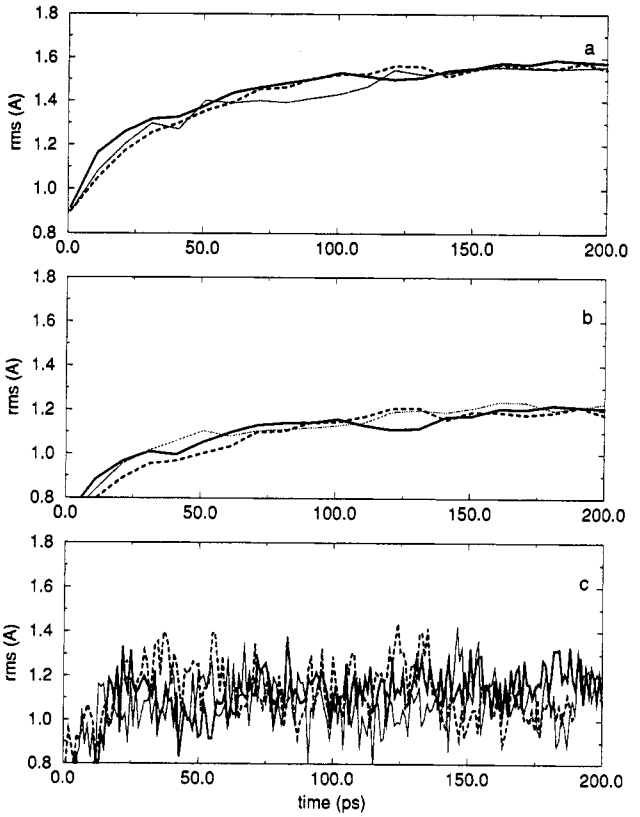

Figure 1. Root mean square difference between the initial coordinates of the protein and the trajectory coordinates as a function of time. The data is averaged over four trajectories of length 200 ps computed using the cLES (dashed line) and LES (thin line) protocols. Four 700-ps trajectories were analyzed for a single copy (thick line); (a) is the rms for all protein atoms, in (b) only the backbone atoms are included, and in (c) only atoms of the residues that boarder the heme pocket (Leu 29, Phe 43, His 64, Val 68, and Ile 107) are included.

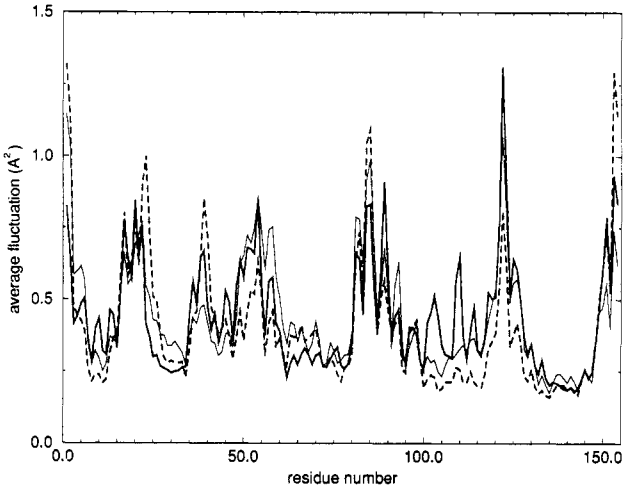

Figure 2. rms fluctuation of the myoglobin main-chain atoms as a function of the residue number. The data is averaged over four trajectories of 200-ps length computed using the cLES (dashed line) and LES (thin line) protocols. Four 700-ps trajectories were analyzed for a single copy (thick line).

The rms fluctuations of the main-chain atoms as a function of the residue number are presented in Figure 2. Four trajectories for exact, LES, and cLES protocols were analyzed. All curves show a similar pattern: localized spikes on a uniform background of about 0.3 Å. Maxima of the fluctuations correspond to the interhelicial regions (loops), the N terminus, the C terminus, and the F helix that were observed also in the previous simulations.[4]

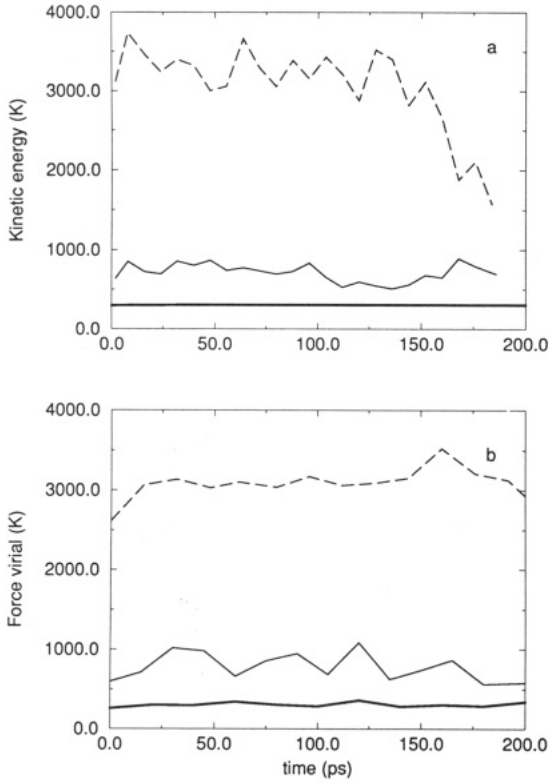

Figure 3. (a) Temperature (kinetic energy) relaxation of the Ar atom computed using the cLES (dashed line), LES (thin line), and single-copy (thick line) protocols. (b) Force virial relaxation of the Ar atom computed using the cLES (dashed line), LES (thin line), and single-copy (thick line) protocols.

From this analysis, we conclude that the cLES and LES algorithms do not affect significantly the average structural properties of the protein molecule.

The thermal equilibrium properties such as the temperature and the force virials per degree of freedom of the Ar were evaluated. The data are presented in Figure 3a,b. The temperature of the "bath" which includes the protein and the solvent is stable during the course of dynamics, with small fluctuations around 300 K. The average kinetic energy of the Ar in exact simulations relaxes to the same temperature. When LES is employed for the description of ligand motion, the kinetic energy and the force virials approach equilibrium temperature which is $\sim$10 times (10 copies were employed) higher. The deviation of the ligand temperature from 300 K at the end of the simulation reflects the escape of a few Ar copies during the 200-ps runs. The new equilibrium temperature is proportional to the number of remaining copies. The fluctuations in the values of the force virials are much larger than those in the kinetic energy. Therefore, the corresponding change in the equilibrium value of the force virial is masked by the large fluctuations of the data.

The problem with the equipartitioning of the kinetic energy and the force virials in the LES approximation was noted first by Straub and Karplus.[20] That work stimulated the introduction of the mean field with a binary collision correction (cLES).[21] The ligand described by the cLES protocol relaxes to a temperature significantly lower than in the simulations using LES. As discussed in the previous section, an analytical solution for the extent of cooling is not available. To increase the efficiency of a collision, a different value of the collision parameter (that is, the distance between the copy and the protein atom when transformation to the collision representation is performed) might be needed. However, this will also affect the frequency of the collisions.

Another equilibrium property which is related to the energy partition is the distribution of the potential energies sampled by the ligand. Figure 4 presents a normalized distribution of the potential energies during 200-ps dynamics using exact, LES, cLES, and LES with velocity-scaling protocols. The higher temperature of the ligand described by LES is reflected in a slower decay of the generated distribution (at equilibrium, $P(U) \simeq e^{-U/kT_{\text{eff}}}$). The effective temperature of the ligand was estimated by fitting the distribution of the potential energy to the Boltzmann distribution. The computed T_{eff} is about 6 times higher in LES than in the cLES simulations. When the temperature was "balanced" by velocity scaling, the potential energy distribution decays very fast as if the ligand is at a temperature lower than 300 K. LES with velocity scaling describes a system which is not at equilibrium.[21] Therefore, the analysis of the last distribution, based on the Boltzmann relation, is questionable.

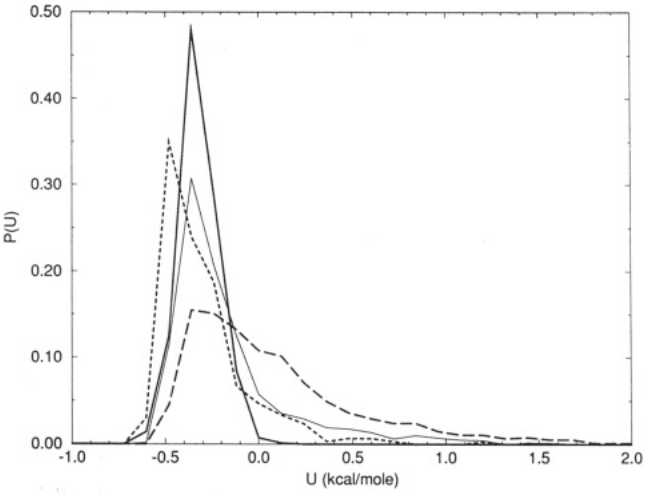

Figure 4. Distribution of potential energies for the Ar atom computed using the cLES (thin line), LES (dashed line), LES with velocity scaling (dotted line), and single-copy (thick line) protocols.

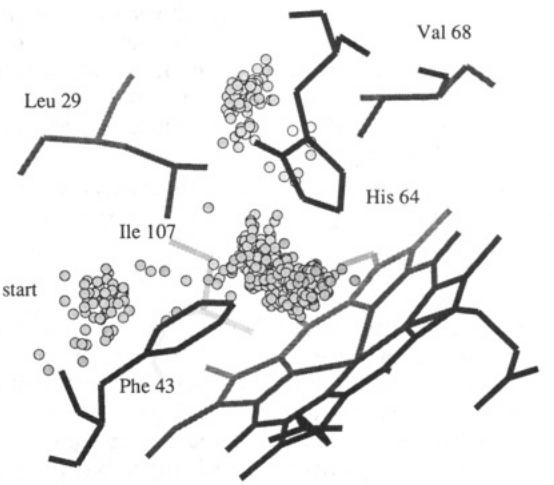

Figure 5. Distribution of the Ar position in the heme pocket computed using the single-copy protocol. A single structure for the heme and residues Leu 29, Phe 43, His 64, Val 68, and Ile 107 are plotted.

It is of interest to check if the improvement in the description of equilibrium properties is reflected in the diffusion dynamics of the Ar obtained via the cLES approximation. We therefore proceed with the analysis of the Ar motion. Initially positioned near Leu 29 and Ile 107 (Figure 5), the ligand in the exact simulations explores the heme and AB/G cavities (heme cavity, His 64, Phe 43, Val 68, Leu 29, Ile 69, and Ile 107, and AB/G cavity, Ile 111, Trp 14, Val 15, His 24, and Leu 115).[4] The distribution of the ligand coordinates mapped on the heme plane and the contour plot for that distribution are presented in Figures 6a and 7 (the data were averaged from the first 300 ps of four single-copy trajectories). The heme iron atom is located at the center of the coordinates. The first peak reflects the initial

localized position of the ligand which is the starting point. This site is near the following residues: Leu 29, Ala 110, Ile 107, and Ile 111. The second peak near the center of the coordinates is above the heme iron. The third location is of the ligand positioned in the AB/G cavity. The last cavity is populated after ~150-ps dynamics in exact simulations. In the LES and cLES simulations, several copies migrated to the AB/G and EF/G cavities after ~70 ps. For the cLES and LES trajectories, the distribution of the ligand positions above the heme plane are presented in Figure 6b,c. The data were computed during the first 75 ps and averaged over four trajectories for the cLES (Figure 6b) and LES (Figure 6c) protocols. The "hot" LES ligands explore almost uniformly the volume near the heme (the height of the peaks in LES is ~5 times smaller than those in cLES or single-copy calculations), providing considerably more delocalized distribution. The cLES protocol has a similar pattern of ligand density to that obtained from a single-copy simulation. However, the temperature of the ligand in cLES is higher than in single-copy calculations (Figure 3). This may result in the observed differences. Assuming diffusive motion in the vicinity of the heme, we proceed to estimate a diffusion constant within the heme pocket. We consider the time interval below 75 ps to analyze the motion near the heme. Making this choice, we excluded transitions between different protein cavities and we focused instead on diffusion within the heme pocket. Intercavity diffusion is likely to be activated and can not be investigated meaningfully in the length of the trajectories we run.

In this study, the diffusion tensor was calculated from the average of the ligand displacement as a function of time

$$\frac{1}{2}\langle (r - r_0) * (r - r_0) \rangle = \mathbf{D}t \qquad (5)$$

where t denotes the time and $\mathbf{D}$ is the diffusion tensor (the asterisk denotes a tensor multiplication). The square of the total displacement (the trace of diffusion tensor) as a function of time is plotted for all protocols in Figure 8. Four trajectories were analyzed and averaged for exact, LES, and cLES protocols. The linearity of the resulting curves (Figure 8) supports the suggested diffusive motion of the ligand. Estimated values for the diffusion constant $\Delta = 1 / 3(D_{xx} + D_{yy} + D_{zz})$ (from least-squares fit of the data in Figure 8) are ~0.22 ± 0.03 Å²/ps for LES, ~0.07 ± 0.02 Å²/ps for cLES, and ~0.07 ± 0.04 Å²/ps for the single-copy calculation. The variations of the diffusion constant were estimated at the 95% confidence limit for all protocols. In the work of Verkhivker et al.,[31] the diffusion constant of CO moving in the interior of leghemoglobin was estimated as $\Delta = 0.1$ Å²/ps. The larger value of this constant in leghemoglobin is consistent with the experimentally observed higher mobility of the ligand in leghemoglobin compared to myoglobin.[32,33] Of course, Ar is also a different particle.

The data from four trajectories with a single copy were not sufficient to estimate the components of the diffusion tensor. The components of the diffusion tensor for LES and cLES are presented in Table 1. The values of the diagonal components reflect a higher mobility of the ligand when LES is employed. The low value of the D_{zz} (compared to $D_{xx,yy}$) which is the direction perpendicular to the heme plane and the absence of the off-diagonal elements $D_{zx,zy}$ indicate that the Ar motion is restricted in the plane that is parallel to the heme. This concludes the analysis of the Ar motion near the heme.

In the enhanced sampling trajectories, the Ar atom explores the interior of the protein. A total of 40 ligand trajectories (four simulations employing 10 copies for each protocol) of 200 ps each were computed. Ten events of escape were found for the LES simulation and nine for the cLES. No events of escape were observed in the four exact trajectories that together lasted for 2.8 ns. Only a few (four) distinct escape pathways were obtained using the LES and cLES protocols. Two major paths were

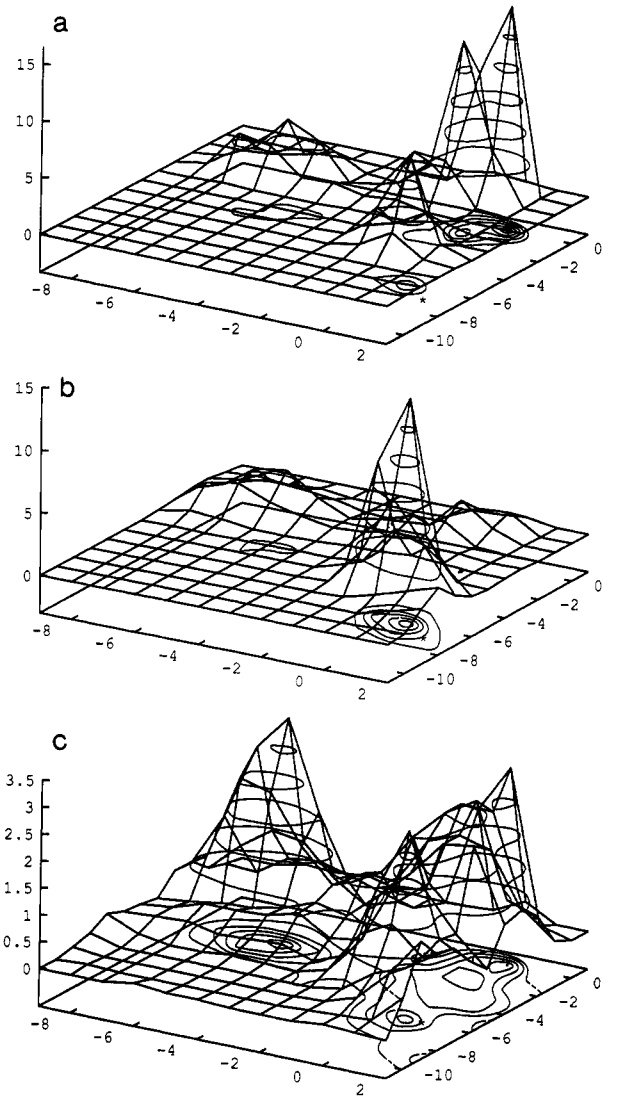

Figure 6. Distribution of the Ar atom projected on the heme plane. The asterisk denotes the initial position of the Ar. The heme iron is located at (0,0): (a) single copy, (b) cLES, and (c) LES.

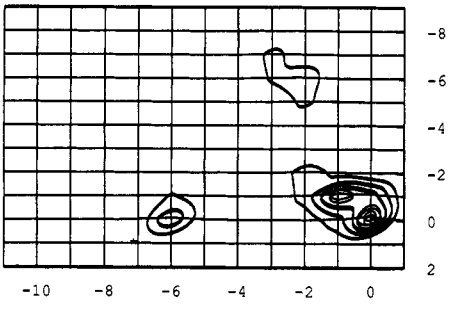

Figure 7. Contour plot of the Ar atom distribution mapped on the heme plane (computed using single-copy simulations). The asterisk denotes the initial position of the Ar. The heme iron is located at (0,0).

detected when LES was used (Figure 9). Both of them have a similar part that is a passage from the heme cavity through the AB/G to EF/G cavity (Trp 7, Leu 76, Val 10, Met 131, Leu 132, and Ala 134). This route involves the "semicavity" (Trp 7, Leu 76, His 82, Leu 132, and Ala 134) as an intermediate on the way to the EF/G cavity in several cases. Previously, that pathway was also detected in the LES study of CO diffusion.[4] Two variants of the last step for ligand escape from the EF/G cavity were observed. Four copies escaped between the EF loop (Leu 76) and the H helix (Leu 135), similar to trajectory 1 in ref 4 (Figure 9). Four copies exit between the GH loop (Pro 120, Gly 121, Asn 122, and Phe 123) and the beginning of the A helix (Trp 7, Gln

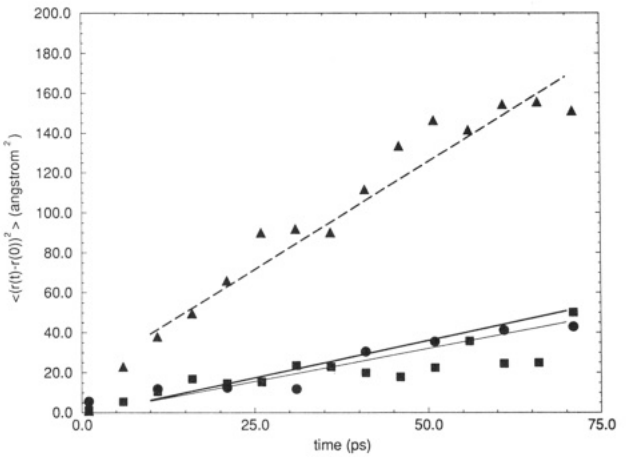

Figure 8. Mean square displacement for the Ar atom moving inside the heme pocket calculated using the single-copy (circles), cLES (squares), and LES (triangles) protocols. The lines represent the least-squares fit to the computed data.

TABLE 1: Components of the Diffusion Tensor (in Å²/ps) for Ar Motion in the Heme Pocket Computed Using cLES and LES

	LES			cLES		
	x	*y*	*z*	*x*	*y*	*z*
x	0.23	0.08	0.00	0.08	0.03	0.00
y	0.08	0.39	0.00	0.03	0.13	0.00
z	0.00	0.00	0.04	0.00	0.00	0.03

8, Leu 9, and Val 10). The last route however was more complicated. In several cases, the ligand populated another small cavity located between the A helix and the GH loop near Val 15 and Asn 122, Phe 123, and Pro 120. Of the other two exits detected in the LES simulations, one is near the distal histidine. That pathway was observed also in the cLES simulations and will be discussed later. In the course of the LES simulations, one copy escaped within ~15 ps, passing in the direction of the BC loop (Ser 35). The average time required for the other copies to escape was ~130 ps. Therefore, the last trajectory might be a rare event connected (possibly) with inappropriate initial equilibration.

In the cLES simulations, the final step of escape leading to the exterior of the protein was sometimes different from that in LES. In seven escapes (out of nine), the copy exists from the EF/G near Leu 76 and Leu 135. The alternative final step between the GH loop and the A helix (which was detected in the LES simulation) was not observed. The motion of the ligand from the heme to the EF/G was similar in both simulations. Since the temperature of the ligand in the LES simulation is considerably higher, LES is less selective in picking plausible diffusive pathways as compared to cLES. In two other events of escape (detected using the cLES protocol), the copy exits from the heme cavity, passing near His 64. Thermal fluctuations of the distal histidine were considered to have a significant effect on ligand escape from the heme cavity.[1,7,8] A side-chain transition of the His 64 opens a direct pathway for the ligand leading from the heme cavity to the exterior of the protein.[4,7] However, molecular dynamic studies of myoglobin suggested that this transition is a rare event and the contribution of other pathways is not less important.[2-4] In our study with explicit water shell, χ_1 transitions of His 64 were not observed. The His 64 fluctuations are primarily of the χ_2 dihedral angle (Figure 10). The small size of the Ar atom may facilitate the escape through this narrow gate, following small His 64 side-chain fluctuations. A similar pathway (near His 64) was also detected in previous simulations of ligand diffusion.[1,2,4]

If we assume a hopping mechanism between the protein internal cavities, a useful function will be the distribution of times required for the ligand to make a "jump" between two sites. A "jump"

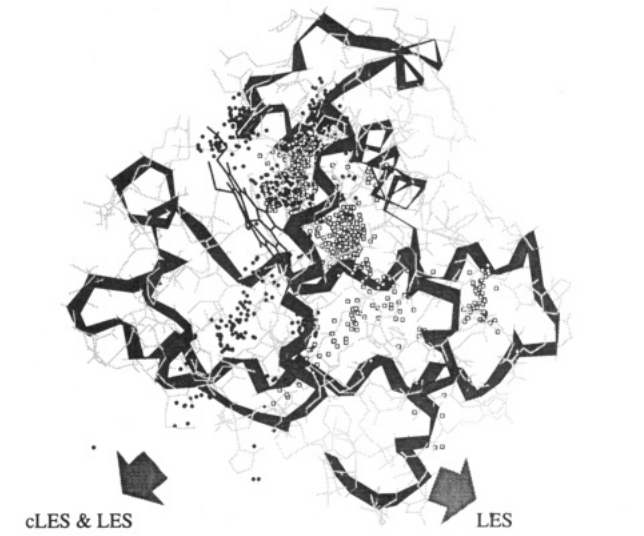

Figure 9. Different exits from the protein are illustrated by the ligand trajectories. cLES (LES) copies are represented by filled circles (open squares).

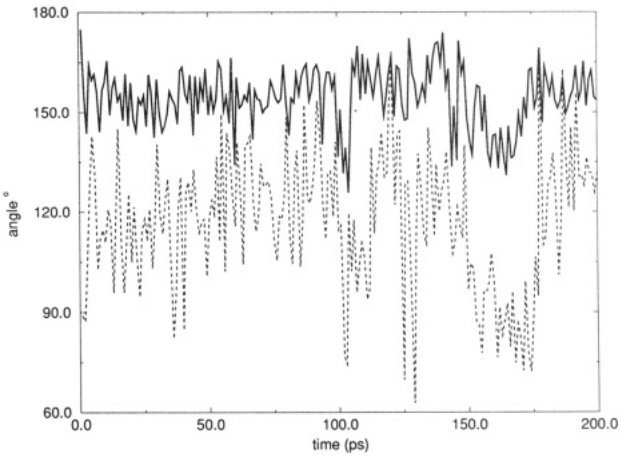

Figure 10. Fluctuations of the χ_1 (solid line) and χ_2 (dashed line) angles for His 64 as a function of time. The Ar atom exits passing near His 64 at ~130 ps.

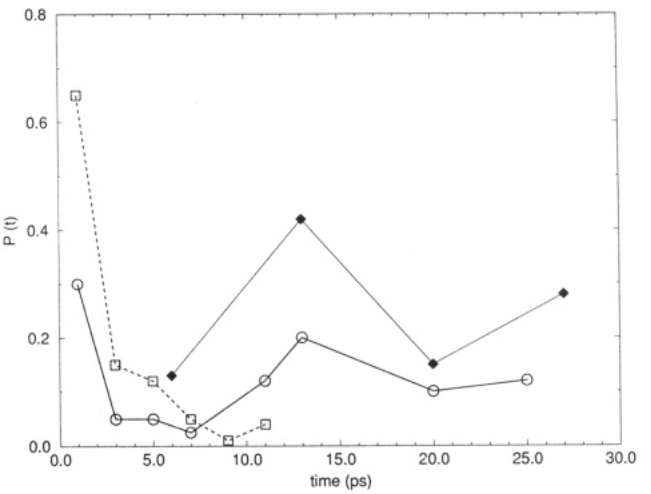

Figure 11. Distribution of the ligand "jumps" computed using the single-copy (filled squares), cLES (circles), and LES (open squares) protocols.

is defined as a 3-Å shift of the center of mass of three consecutive positions of the Ar atom (copy). Following visual inspection of the trajectory, this definition seems useful. The distribution is presented in Figure 11. For the LES copies, the probability to make a second "jump" immediately after the first is over 60% and it rapidly decays as a function of time. In the cLES

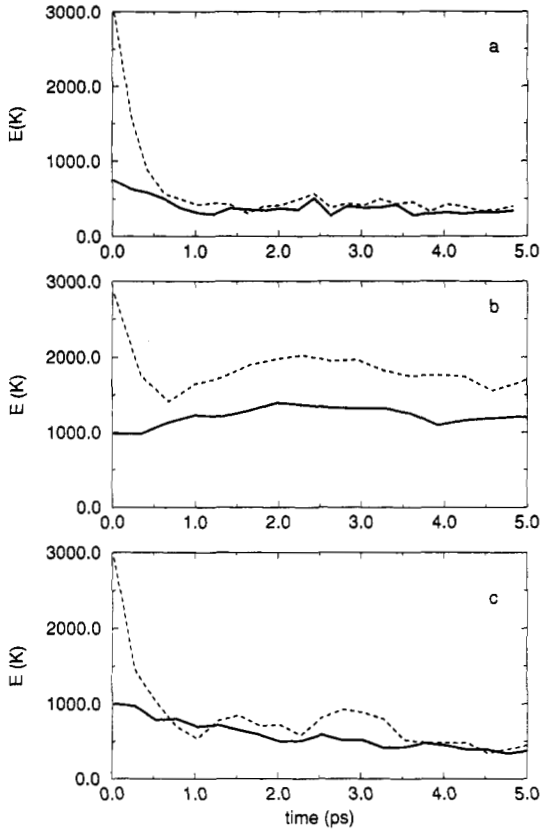

Figure 12. Relaxation of the translational (dashed line) and vibrational (solid line) energy of photodissociated nitric oxide in myoglobin: (a) single copy, (b) LES, and (c) cLES.

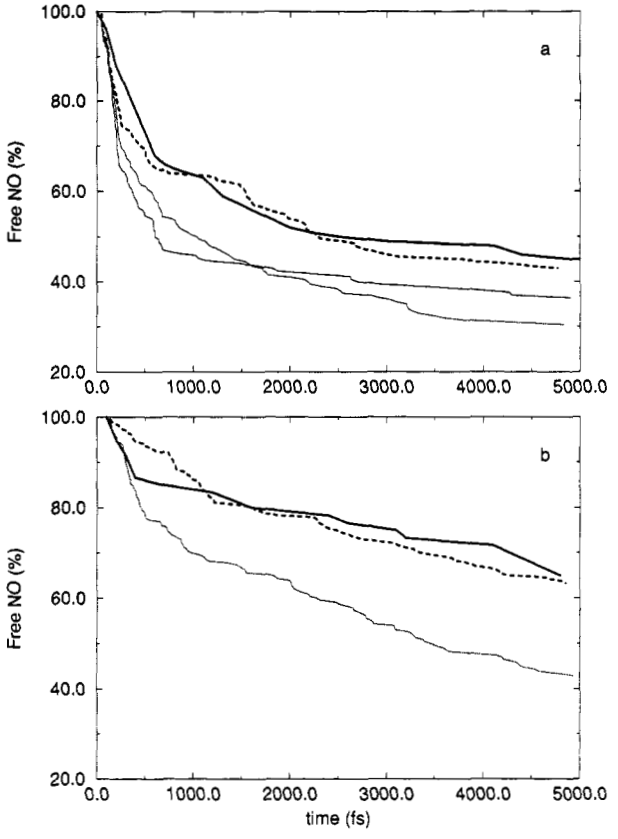

Figure 13. Geminate recombination curves for NO: (a) wild type (Leu 29) and (b) Val 29 mutant; solid line, single-copy calculation, dashed line, cLES, thin solid line, LES, and long dashed line, LES with velocity scaling (for the wild form of myoglobin).

simulations, two possibilities exist, to make a second "jump" immediately (~30%) or with at least a 15-ps delay (~20%). The cLES data are more similar to the results of the single-copy simulation, suggesting that cLES describes better (than LES) the mechanism of ligand diffusion through myoglobin. The average time between the "jumps" is ~10.5 ps for cLES and ~3 ps for LES. It is ~16.5 ps in single-copy simulations. However, the small number of "jumps" observed in the exact trajectories did not provide sufficient data to perform more quantitative analyses.

4.2. NO Recombination Study. We investigated the recombination rate of NO in the wild type (Leu 29) and in a mutant form (Val 29) of myoglobin. The length of the trajectories was limited to 10 ps, which is considerably shorter than the typical time required for relaxation of the protein matrix.[4,9,34] However, this time is sufficient for the relaxation of the ligand degrees of freedom in cLES (Figure 12c). The cLES protocol reproduced the fast thermalization of the ligand coordinates observed in single-copy calculations (Figure 12a). No significant relaxation of ligand degrees of freedom was observed when the LES protocol was employed (Figure 12b). The fast relaxation of the excess of translational energy in a period of less than 0.5 ps in the single-copy and cLES simulations reflects the effective energy transfer through the collision mechanism. The rapid thermalization of the ligand which requires only a few collisions was proposed previously in an ultrafast spectroscopy study of the CO motion after dissociation.[36] It was also observed in molecular dynamic simulations of the process by Straub and Karplus.[9]

The recombination curves calculated from 20 cLES and 20 LES trajectories are presented in Figure 13a,b. Additionally, the results from a single-copy simulation[11] are plotted. For the wild type, the data obtained by LES with velocity scaling are presented in Figure 13a. cLES reproduces the correct order of recombination rates for both myoglobin forms, and the results are in good agreement with theoretical estimates computed by

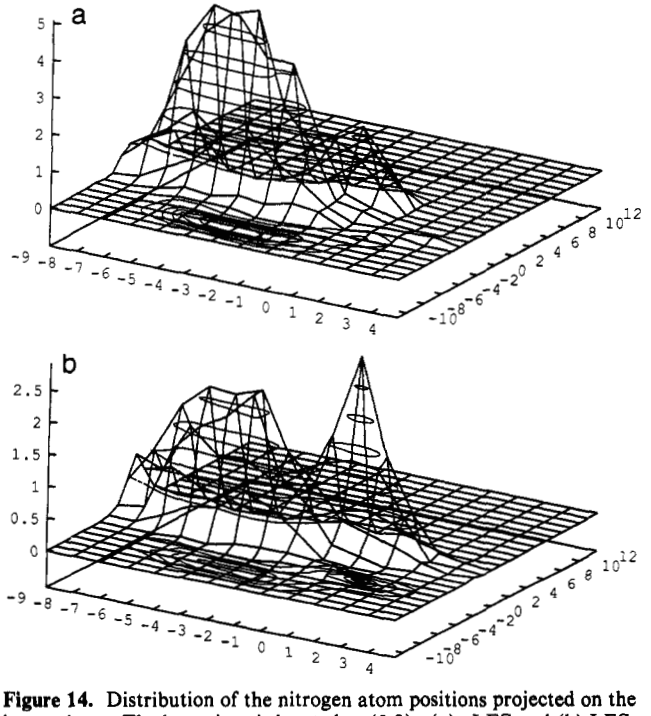

Figure 14. Distribution of the nitrogen atom positions projected on the heme plane. The heme iron is located at (0,0): (a) cLES and (b) LES.

single-copy trajectories.[11] The regular LES and the LES with velocity scaling give higher rates of rebinding.

We comment that application of the enhanced sampling approximation made it possible to obtain the data for this study using only 20 trajectories with 10 copies of the ligand compared to using 100 exact trajectories as computed in ref 11. That corresponds to ~4–5 times savings in computer time.

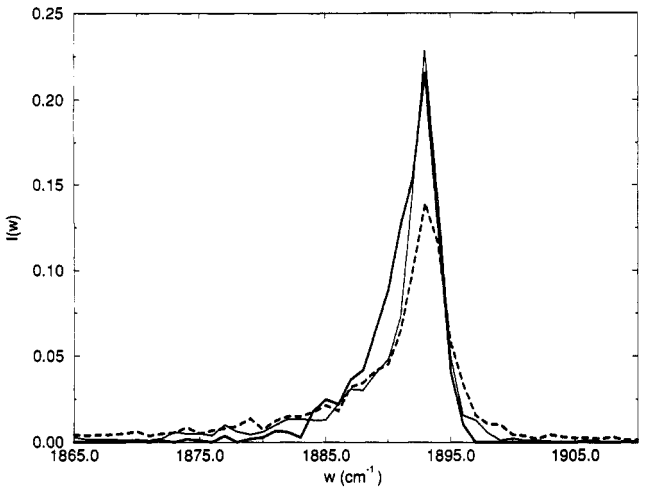

Figure 15. Computed distribution of the vibrational frequencies for the unbound NO moving in the heme pocket. The data are calculated using the single-copy (thick line, cLES (thin line), and LES (dashed line) protocols.

To characterize the ligand motion in the heme cavity, we computed the distribution of the positions of the nitrogen atom mapped on the heme plane. In Figure 14, that distribution is presented for the cLES and LES protocols. The ligand coordinates were taken from the first 5 ps of dynamics for ligand copies that did not rebind. The heme iron is located at the center of the coordinates. The distribution has two major peaks. These peaks correspond to two localized sites inside the heme cavity. The first one is near the center of the heme, His 64 and Phe 46. The second location is in the heme-cavity region near Leu 29, Val 68 on the boarder of the heme, and the AB/G cavities. In the simulation of the Ar diffusion, the Ar atom populated the AB/G cavity which is located ~3 Å away (toward the AB loop) from the NO site. The larger size of the diatomic molecule compared to that of Ar and the short time of the simulation make it impossible to detect NO transitions from the heme pocket to the AB/G cavity (the Ar atom was detected in the AB/G cavity only after 80 ps of simulation using the cLES and LES protocols). The site at the center of the heme cavity was not populated significantly since it is in a favorable position for ligand recombination.

The bond axis of the NO copies has no preferred orientation, suggesting the lack of specific binding. The distribution of vibrational frequencies for the NO was determined along calculated trajectories as follows: each 200 fs of dynamics, the vibrational energy for each ligand copy was calculated. The vibrational potential energy includes the interactions of the ligand with the protein and was computed at five points, varying only the bond length. The five energy values (for each copy) were fitted against a Morse function using the Powell minimization procedure.[22] The spectroscopy parameters ω_e and $\omega_e x_e$ were finally computed along the LES and cLES trajectories. The trajectories computed by Li et al.[11] were used to calculate the Morse parameters for single-copy dynamics. The resulting distribution of ω ($\omega = \omega_e - (1/2)\omega_e x_e$) is shown in Figure 15. The peak of the distribution is positioned at ~1893 cm^{-1} which is approximately 10 cm^{-1} shifted from the value corresponding to NO in the gas phase (ω_o = 1903 cm^{-1} [35]). The frequency distributions obtained by all protocols are similar—they decay uniformly. Therefore, they do not reflect the occurrence of several maxima distinguished in vibrational spectra of a diatomic molecule positioned in the heme pocket (B-states).[34-36] However, even in this simple model there is a considerable contribution of red- and blue-shifted frequencies. The frequency distributions were similar when computed for a ligand moving in the heme cavity and in the B/G cavity (observed in a single-copy simulation[11]). An improved form of the NO–protein potential is required to reproduce and assign the set of different vibrational frequencies observed experimentally.[34,36]

5. Conclusions

A protocol based on the locally enhanced sampling approximation in conjunction with a binary collision model (cLES) was applied to study diffusive motion of a small ligand in myoglobin. cLES as a protocol to study dynamics makes it possible to benefit from the enhanced sampling for the ligand and the acceptable accuracy for the dynamics. The violation of energy partition in LES made the ligand copies less sensitive to the protein environment, and higher energy paths were sampled. These difficulties are corrected (to a significant extent) in the cLES protocol. In a previous investigation of ligand diffusion in myoglobin, only LES was employed.[4] Compared to the older investigation, the present study is different in a number of aspects.

In addition to LES, a better approximation (cLES) is employed, and a detailed comparison to exact trajectories is made.

(b) The potential used is more recent.[22]

(c) A solvation shell around myoglobin is used in the present simulation. The solvent was not included in those of ref 4.

(d) The temperature of the ligand is considerably lower, and the simulation protocol is more straightforward (i.e., no very frequent velocity scaling).

Two different processes were examined, Ar diffusion and NO recombination in myoglobin. The main results of the simulations are that both ligands (Ar and NO) diffuse via a hopping mechanism, i.e., by jumping between localized sites. The ligand motion was considerably less localized when the LES protocol was used. The diffusion constant corresponding to the Ar motion near the heme was computed. The cLES estimate of this constant reproduces the data obtained from single-copy simulations ($D_{cLES} \approx 0.07 \pm 0.02$ Å^2/ps, $D_{single} \approx 0.07 \pm 0.04$ Å^2/ps). LES provided a considerably higher value, reflecting higher mobility of the ligand copies ($D_{LES} \approx 0.22 \pm 0.03$ Å^2/ps).

The data on rebinding curves obtained using the cLES protocol are in good agreement with exact simulations reported previously.[11] However, the amount of computational time required to obtain the cLES results was considerably smaller (by four to five times) compared to the resources employed in the exact simulations.[11]

Analysis of the set of Ar trajectories that reached the exterior of the solvated protein was performed. The major route included a transition from the heme cavity to the semicavity located between the EF loop and the beginning of the A and H helices. However, the final step observed in the LES calculations involved two type of "doors" in the protein structure: one between the EF loop and the H helix and another between the GH loop and the A helix. Only the first type was observed in the cLES simulations which may suggest it as the lower energy path. The solvation shell included in the simulations restricted considerably the mobility of the protein molecule. No escape events were detected in the exact simulations which lasted a total of ~2.8 ns. Ten events of escape were detected using LES and nine events using cLES during only 0.8 ns of simulations for each protocol (10 Ar copies were employed in each of the four trajectories for the cLES and LES study).

The Ar diffusion is a thermal process. Another process addressed in this paper is the short-time motion of a ligand (nitric oxide) that is not in a thermal equilibrium. In both limits, the cLES protocol makes it possible to obtain results which were in good agreement with the exact simulations at a considerable savings of computer time (when our CPU resources were sufficient to use the exact simulations to obtain the reference data).

Acknowledgment. This research was supported by NIH Grant GM41905 to R.E. R.E. is a University of Illinois West Scholar and Alon fellow in the Hebrew University.

Appendix

The LES equations of motion for the N_L ligand copies and N_P protein atoms are defined as follows:

$$\frac{M_{L1}}{N_L}\frac{d^2x_{Lj}}{dt^2} = -\frac{1}{N_L}\frac{\partial U(x_{p1},\ldots,x_{pN_P},x_{Lj})}{\partial x_{Lj}} \quad \text{(A1)}$$

$$M_j\frac{d^2x_{pj}}{dt^2} = -\frac{1}{N_L}\sum_{i=1}^{N_L}\frac{\partial U(x_{p1},\ldots,x_{pN_P},x_{Li})}{\partial x_{pj}} \quad \text{(A2)}$$

where the subscripts p and L are used for protein and ligand degrees of freedom. $\{x\}$ is a set of Cartesian coordinates of protein atoms and ligand copies used in LES. x_{Lj}, M_{L1} are the coordinates (or velocities) of the ligand copy and the mass of the ligand. M_j is the mass of the protein particle j. The LES Lagrangian has the following form:[21]

$$L^{\circ}_{LES} = \sum_{j=1}^{N_P}\frac{M_j}{2}\left(\frac{dx_{pj}}{dt}\right)^2 + \frac{1}{N_L}\sum_{i=1}^{N_L}\frac{M_{L1}}{2}\left(\frac{dx_{Li}}{dt}\right)^2 - \frac{1}{N_L}\sum_{i=1}^{N_L}U(x_{p1},\ldots,x_{pN_p},x_{Li}) \quad \text{(A3)}$$

The coordinates (and velocities) denoted as $\{q\}$ for a collision representation for a monoatomic ligand (N_L copies) and the nth protein atom are

$$q_{Lj} = x_{Lj} - x_{pn} \quad \text{(A4)}$$

$$q_{cm} = \frac{1}{M_n + M_{L1}}\left(M_n x_{pn} + \frac{1}{N_L}\sum_{j=1}^{N_L}M_{L1}x_{Lj}\right) \quad \text{(A5)}$$

$$q_{pj} = x_{pj} - x_{pn} \quad \text{(A6)}$$

where the subscripts p and L are used for protein and ligand degrees of freedom. $\{x\}$ is a set of Cartesian coordinates of protein atoms and ligand copies used in LES. x_{Lj}, M_{L1} are the coordinates (or velocities) of the ligand copy and the mass of the ligand. x_{pn}, M_n are the coordinates (or velocities) and the mass of the protein atom participating in the collision. x_{pj} denotes a coordinate (or velocity) of any other protein particle.

For a diatomic ligand, the collision coordinates are

$$q_{L1j} = x_{L1j} - x_{L2j} \quad \text{(A7)}$$

$$q_{L2j} = \frac{1}{(M_{L1}+M_{L2})}(M_{L1}x_{L1j} + M_{L2}x_{L2j}) - x_{pn} \quad \text{(A8)}$$

$$q_{cm} = \frac{1}{M_n + M_{L1} + M_{L2}}(M_n x_{pn} + \frac{1}{N_L}\sum_{j=1}^{N_L}(M_{L1}x_{L1j} + M_{L2}x_{L2j})) \quad \text{(A9)}$$

$$q_{pj} = x_{pj} - x_{pn} \quad \text{(A10)}$$

where M_{L1}, M_{L2} are the atomic masses for the diatomic ligand. Coordinates with L1 and L2 subscripts are used for the first and second atom in the diatomic molecule.

The Lagrangian (L_{coll}^n) describing the collision of the n-th protein atom with a monoatomic ligand is

$$L_{coll}^{n} = \frac{1}{N_L}\sum_{\substack{j=1\\ j\neq n}}^{N_P}\sum_{i=1}^{N_L}\frac{M_j}{2}\left(\frac{dq_{pj}}{dt} + \frac{dq_{cm}}{dt} + \kappa_1\frac{dq_{Li}}{dt}\right)^2 + \frac{M_{cm}}{2}\left(\frac{dq_{cm}}{dt}\right)^2 + \frac{1}{N_L}\frac{M_r}{2}\sum_{j=1}^{N_L}\left(\frac{dq_{Lj}}{dt}\right)^2 - \frac{1}{N_L}\sum_{j=1}^{N_L}U(q_{p1},\ldots q_{pN_P},q_{Lj},q_{cm}) \quad \text{(A11)}$$

where M_{cm} is a sum of the M_{L1} and M_n, $\kappa_1 = -\frac{M_{L1}}{M_{cm}}$, $M_r = \frac{M_{L1}M_n}{M_{L1}+M_n}$. The first term in kinetic energy includes the contribution from all protein atoms except the colliding one. The second and third terms represent the energy of the $n$ith atom and a set of ligand copies.

The L_{coll}^n describing the collision of the protein atom n with a diatomic ligand has the following form:

$$L_{coll}^{n} = \frac{1}{N_P}\sum_{\substack{j=1\\ j\neq n}}^{N_P}\sum_{i=1}^{N_L}\frac{M_j}{2}\left(\frac{dq_{pj}}{dt} + \frac{dq_{cm}}{dt} + \kappa_1\frac{dq_{L2i}}{dt}\right)^2 + \frac{M_{cm}}{2}\left(\frac{dq_{cm}}{dt}\right)^2 + \frac{1}{N_L}\frac{M_{r1}}{2}\sum_{j=1}^{N_L}\left(\frac{dq_{L1j}}{dt}\right)^2 + \frac{1}{N_L}\frac{M_{r2}}{2}\sum_{j=1}^{N_L}\left(\frac{dq_{L2j}}{dt}\right)^2 - \frac{1}{N_L}\sum_{j=1}^{N_L}U(q_{p1},\ldots q_{pn-1},q_{pN_P},q_{L1j},q_{L2j},q_{cm}) \quad \text{(A12)}$$

where M_{cm} is the sum of the M_{L1}, M_{L2}, and M_n, $\kappa_1 = -\frac{M_{L1}+M_{L2}}{M_{cm}}$, $M_{r1} = \frac{(M_{L1}+M_{L2})M_n}{M_{cm}}$, $M_{r2} = \frac{M_{L1}M_{L2}}{M_{L1}+M_{L2}}$. The first term in kinetic energy includes the contribution from all protein atoms exept the colliding one. The second, third, and fourth terms are the energy of the nth protein atom and the set of ligand copies. In the transformation from LES to the collision representation at time t_c, the values of the velocities of the enhanced part and the center of mass of the colliding particles $\frac{dq_{cm}}{dt}$ satisfy the following equations:

$$E_{coll}^{protein}(t_c) = E_{LES}^{protein}(t_c)$$
$$E_{coll}^{ligand}(t_c) = E_{LES}^{ligand}(t_c) \quad \text{(A13)}$$

where $E_{LES}^{protein}(t_c)$, $E_{LES(t_c)}$ are the kinetic energies at time t_c for the protein and the ligand in the LES approximation. $E_{LES}^{protein}(t_c)$, $E_{coll}^{ligand}(t_c)$ are the kinetic energies for the protein and the ligand using the collision approximation with a Lagrangian (eqs A11 and A12).

References and Notes

(1) Case, D. A.; Karplus, M. *J. Mol. Biol.* **1979**, *132*, 343.
(2) Tilton, R. F.; Kuntz, I. D., Jr.; Petsko, G. A. *Biochemistry* **1984**, *23*, 2849.
(3) Tilton, R. F.; Singh, U. C.; Weiner, S. J.; Connoly, M. L.; Kuntz, I. D., Jr.; Kollman, P. A.; Max, N.; Case, D. A. *J. Mol. Biol.* **1986**, *192*, 443.
(4) Elber, R.; Karplus, M. *J. Am. Chem. Soc.* **1990**, *112*(*25*), 9161.
(5) Czerminski, R.; Elber, R. *Proteins* **1991**, *10*, 70.
(6) Nowak, W.; Czerminski, R.; Elber, R. *J. Am. Chem. Soc.* **1991**, *113*, 5628.
(7) Kottalam, J.; Case, D. A. *J. Am. Chem. Soc.* **1988**, *110*, 7690.
(8) Case, D. A.; McCammon, J. A. *Ann. N.Y. Acad. Sci.* **1986**, 483.
(9) Straub, J. E.; Karplus, M. *Chem. Phys.* **1991**, *158*, 221.
(10) Gibson, Q. H.; Regan, R.; Elber, R.; Olson, J. S.; Carver, T. E. *J. Biol. Chem.* **1992**, *267*, 22022.

(11) Li, H.; Elber, R.; Straub, J. *J. Biol. Chem.* **1993**, *268*, 7908.
(12) Henry, R.; Levitt, M.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 2034.
(13) Shaad, O.; Zhou, H. X.; Szabo, A.; Eaton, W. A.; Henry, E. R. *Proc. Natl. Acad. Sci. U.S.A.*, submitted for publication.
(14) Sassaroly, M.; Rousseau, D. L. *J. Biol. Chem.* **1986**, *261*, 16292.
(15) Gerber, R. B.; Buch, V.; Ratner, M. A. *J. Chem. Phys.* **1982**, *77*, 3022.
(16) Verkhivker, G.; Elber, R.; Nowak, W. *J. Chem. Phys.* **1992**, *97*, 7838.
(17) Roitberg, A.; Elber, R. *J. Chem. Phys.* **1992**, *95*, 9277.
(18) Roitberg, A.; Elber, R. In *Protein Structure Determination*; Merz, K., Le Grand, S., Eds.; in press.
(19) Simmerling, C.; Elber, R. Submitted for publication.
(20) Straub, J.; Karplus, M. *J. Chem. Phys.* **1991**, *94*, 6737.
(21) Ulitsky, A.; Elber, R. *J. Chem. Phys.* **1993**, *98*, 3380.
(22) Elber, R.; Roitberg, A.; Simmerling, C.; Verkhivker, G.; Goldstein, R.; Ulitsky, A.; Li, H. "*Moil: A molecular dynamics program with emphasison reaction path calculations and conformational searches in proteins*". NATO conference proceedings on *Statistical Mechanics and Protein Structure*, Ed. Doniach, S., Plenum press, 1993, (in press).
(23) Phillips, G. N.; Arduini, R. M.; Springer, B. A.; Sligar, S. G. *Proteins* **1990**, *7*, 358.
(24) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
(25) Andersen, H. C. *J. Comput. Phys.* **1982**, *52*, 24.
(26) Johnston, H. J. *Gas Phase Reaction Theory*; Ronald: New York, 1966; p 74.
(27) Verlet, L. *Phys. Rev.* **1967**, *159*, 98.
(28) Landau, L. *Sov. Phys.* **1932**, *1*, 89.
(29) Zener, C. *Proc. R. Soc. Edinburg* **1932**, *A137*, 696.
(30) Steinbach, P. J.; Loncharich, R. J.; Brooks, B. R. *Chem. Phys.* **1991**, *158*, 383.
(31) Verkhivker, G.; Elber R.; Gibson, Q. H. *J. Am. Chem. Soc.* **1992**, *114*, 7866.
(32) Stetzkowski, F.; Banerjee, R.; Mardens, M. C.; Breece, D. K.; Bowne, S. F.; Dorster, W.; Frauenfelder, H.; Reinisch, L.; Shyamsunder, E.; Jung, C. J. *J. Biol. Chem.* **1985**, *260*, 8803.
(33) Gibson, Q. H.; Wittenberg, J. B.; Wittenberg, B. A.; Bogusz, D.; Appleby, C. A. *J. Biol. Chem.* **1989**, *264*, 100.
(34) Young, R. D.; Frauenfelder, H.; Johnson, J. B.; Lamb, D. C.; Nienhaus, G. U.; Philipp, R.; Scholl, R. *Chem. Phys.* **1991**, *158*, 315.
(35) Herzberg, G. *Molecular Spectra and Molecular Structure*, 2nd ed.; D. Van Nostrand Company: New York, 1967; p 558.
(36) Anfinrud, P. A.; Han, C.; Hochstrasser, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 8387.