PROTEINS: Structure, Function, and Genetics 10:70–80 (1991)

Computational Studies of Ligand Diffusion in Globins: I. Leghemoglobin

Ryszard Czerminski and Ron Elber
Department of Chemistry, M/C 111, The University of Illinois at Chicago, Box 4348, Chicago, Illinois 60680

ABSTRACT **The thermally assisted diffusion of a small ligand (carbon monoxide) through a protein matrix (lupine leghemoglobin) is investigated computationally. The diffusion paths are calculated by a variant of the time-dependent Hartree approximation which we call LES (locally enhanced sampling). The variant which was recently introduced by Elber and Karplus[1] is based on the classical TDSCF approximation of Gerber et al.[2] The simulation enables more significant search for diffusion pathways than was possible before. This is done by increasing the number of *ligand* trajectories using a single trajectory for the protein. We compare qualitatively diffusion rates in leghemoglobin and in myoglobin. The calculation shows that the diffusion in leghemoglobin is much faster than the diffusion in myoglobin, in agreement with experiment. The gate in leghemoglobin is opened by fluctuations at a close contact between the B/C and the G helices. The most relevant fluctuation is the *rigid* shift of the C helix with respect to the G helix. This path is not observed in a comparable calculation for myoglobin.[1] This finding is rationalized by the lack of the D helix in leghemoglobin and a significantly more flexible CE loop. Supporting experimental evidence for the importance of the CE loop in leghemoglobin can be found in the kinetics studies of Gibson et al.[28]**

Key words: locally enhanced sampling, molecular dynamics, ligand penetration

INTRODUCTION

The problem of how a small ligand can penetrate into a compact solid-like structure is of special interest in the molecular biology of globins.[3] The globins are oxygen-transport and oxygen-binding proteins with a binding site—the heme pocket which is not exposed to the solvent. In 1966 Perutz and Matthews[4] pointed out that the heme pocket is not accessible to the solvated ligand according to the X-ray structure of hemoglobin. Takano[5] reached a similar conclusion for myoglobin. No opening could be found in the X-ray structures which are, however, averaged over long times compared to the diffusion time scales. Transient pathways which can be induced (in principle) by protein fluctuations cannot be observed in the X-ray experiment.

The search for fluctuations relevant to escape routes attracted a great body of theoretical and experimental research. It is not possible to describe the current progress in the field in a short manuscript. We therefore refer the interested reader to a very recent review[6] and references therein. Most of the investigations focused on myoglobin for which a wealth of experimental data is available.

Ligand diffusion was studied theoretically for the first time by Case and Karplus.[7] Elaborations along the same reaction path were pursued by Case and McCammon[8] and Kottalam and Case.[9] There are two basic steps in such calculations. In the first step the computer is employed in a gross search for the exits. That is, the general direction of the plausible reaction coordinates of the ligand escape is investigated. In the next step the "low resolution" reaction coordinates found in the first step are refined and the rate constants are calculated.

To address the first problem in myoglobin Case and Karplus[7] calculated ligand trajectories in a *rigid* protein structure. The disadvantage of using a rigid protein structure is that transient paths which are generated by protein fluctuations are not observed. Nevertheless, based on the simulations, the next step of the calculation was pursued. A "minimum distance" path was proposed in which the side chains of Val E11 and His E7 change their orientation. The calculated rate constant is similar to the experimental measurements of Jongeward et al.[10] The reaction coordinate of Case and Karplus is an appealing, elegant, and probably the simplest solution to the problem. However, this work does not eliminate the possibility of alternate pathways.

Experiments of fluorescence quenching suggest that O_2 penetrates to the structure of almost all proteins. Hence, oxygen penetration is a general phenomenon which depends only slightly on the details

Received February 16, 1990; revision accepted July 3, 1990.
Address reprint requests to Dr. Ron Elber, Department of Chemistry, The University of Illinois at Chicago, M/C 111, P.O. Box 4348, Chicago IL 60680.

Abbreviations used: TDH, time-dependent Hartree; TDSCF, time-dependent self-consistent field; LES, locally enhanced sampling.

of the structure. There is no reason to assume a specific minimum distance path for myoglobin, since a large number of proteins independent on their specific structure make ligand penetration possible. Furthermore, recent simulations support the idea of multiple paths. The simulations of Tilton et al.[12,13] and of Elber and Karplus[1] are in qualitative agreement with the diffusion experiment, showing that a large portion of the protein is accessible to the small ligand. Tilton et al. performed two simulations. In the first,[12] the protein was fixed similarly to the study of Case and Karplus.[7] The ligand trajectories were however longer and the whole protein molecule was considered. Therefore a more complete ensemble (within the rigid protein approximation) was obtained. The spherical probe in the simulation of Tilton et al. was hopping between a network of cavities in the protein interior. It was not confined to the heme pocket and as a result a number of alternative escape routes was detected. Tilton et al. pursued a second simulation which included the motions of both the protein and the ligand. Hence, in the second simulation the transient paths generated by the protein fluctuations could be examined. The number of ligand trajectories using a moving protein was five. The five ligands present a sample which is too small to study the large number of possible "holes" in the protein structure.

In this paper we apply an automatic search technique which was developed recently by Elber and Karplus.[1] The technique was designed to search extensively for "holes" in a *fluctuating* protein structure. In this method a substantial number of ligand trajectories (~100) are obtained at a computational cost comparable to that of a single trajectory. Elber and Karplus[1] employed this method in myoglobin and obtained (with more statistics) results which are in agreement with the simulations of Tilton et al.[12,13] The two main results are (1) that a large internal volume of the protein molecule is accessible to the diffusing ligand; and (2) that a number of alternative pathways exists. In addition Elber and Karplus were able to analyze in detail diffusion mechanisms using a larger sample than was possible before. The protocol is expected to be useful also for other searches provided that enhanced sampling is required for a small part of the system (e.g., the ligand). We name this method locally enhanced sampling or LES. LES is closely related to the time-dependent Hartree approximation[14,15] and to the classical version of the TDSCF approximation which was introduced by Gerber et al.[2] In fact, it is a special implementation of the "trajectory bundle" idea in which we use bundles of trajectories for a small part of the system only.

Here we focus on a different globin (leghemoglobin) which has higher oxygen affinity and a higher association rate than the protein myoglobin. Stetzkowsky et al.[16] studied carbon monoxide and oxygen rebinding in soybean leghemoglobin using a 30 nsec flash. At room temperature it was not possible to measure the rate of ligand diffusion from the protein matrix to the solvent. That is, steady state is reached between a solvated ligand and a ligand in the protein matrix for times longer than the pulse width of 30 ns. At 260 K the diffusion through leghemoglobin (the "matrix process" of ref. 16) is 20 times faster than in myoglobin. The diffusion time scale at room temperature for myoglobin is of order of tens to hundred of nanoseconds.[10] We expect the diffusion in leghemoglobin to be significantly faster than that in myoglobin at room temperature.

The purpose of this paper is to use the LES protocol and to check if the qualitative differences between myoglobin and leghemoglobin which were observed experimentally are reproducible in the simulation. Indeed, we shall show that the present calculation yields a diffusion rate which is considerably larger than in myoglobin. Furthermore, based on the simulation, an atomic detail picture of the experimentally observed differences is proposed.

This paper is organized as follows: In the second section we review the physical ideas and the principles of the LES computational protocol. In the third section the simulation procedure and several tests of the simulation quality are described. The results for the diffusion paths are summarized in the fourth section, followed by concluding remarks.

METHOD

The straightforward search for possible diffusion pathways of a small ligand in a *fluctuating* protein is an expensive computation. In order to probe a significant number of exit pathways it is necessary to repeat a large number of trajectories of escaping ligands. The different trajectories correspond to the set of different initial conditions. Assume (a modest assumption) that one needs ≈100 trajectories to obtain a reasonable sample of diffusion paths. This corresponds to an increase of computational effort (compared to a single trajectory) by a factor of 100. Two orders of magnitude is a nonnegligible factor even considering the rapid growth in the present computer technology.

LES is a trick to get around this problem. We run a large number of ligand trajectories using a *single* trajectory for the protein. The ligand is small and the large number of protein degrees of freedom require most of the computer time. Therefore, a simulation with 100 ligands (but one protein!) is associated with a computational effort comparable to that of a single trajectory. The price paid is introduction of an approximation.

Consider an ensemble of N different protein molecules each with a ligand buried in the protein matrix. We seek the diffusion path which each of the N ligands will follow. The protein is a large molecule (~1,500 atoms) and therefore the small ligand (2

atoms) will not affect significantly the global thermal motions of the protein. Thus, if one examined *extended* protein properties which depend on the protein fluctuations (e.g., the all atoms rms), it is unlikely that the small ligand would have a significant effect on their numerical values. One kind of thermal fluctuations corresponds to the opening of escape pathways. In LES we assume that the protein fluctuations and more specifically the escape pathways are determined primarily by internal protein interactions. The protein fluctuations are influenced only to a minor extent by the forces which the ligand imposes on the protein. We call this idea the "passive ligand" assumption. Reaction path calculations (not employing LES) of the diffusion of carbon monoxide in another globin (erythrocruorin hemoglobin, Yip and Elber[17]) suggest that the ligand indeed plays the passive role that we assumed above. The assumption may cause difficulties if the ligand strongly perturbs its *local* neighborhood in the protein (the nearby residues). The escape path which we found in the present investigation is however indeed extended. We stress that as a tool to search for holes in the protein structure (rather than as a tool for the direct simulation of the dynamics) the LES is a useful approach even when the protein structure is perturbed locally. That is because the obtained trajectories can be refined to reaction coordinates[17] which will be then used in rate calculations.

When the effect of ligand motions on the protein fluctuations is small the protein fluctuations are close to those found in equilibrium. Hence a *single* trajectory can be employed to represent the range of the thermal motions of the protein. So, rather than having N trajectories for the N different protein molecules and their corresponding ligands we have only *one* trajectory to sample the equilibrium protein fluctuations. At the same time we have N different ligand trajectories representing the nonequilibrium part of the process. For the convenience of the reader the equations of motion of the LES protocol[1] are summarized in the appendix.

The computational benefits of using the above protocol are the following: A single protein trajectory is employed to run N (N of order of 100) ligand trajectories. Since the calculations associated with the protein are the numerically intensive part, the speed up compared to exact 100 protein trajectories is substantial (by a factor of several tens).

THE COMPUTATIONAL PROTOCOL

Equations (1) and (2) were implemented in a special version of the program CHARMM.[1,18] The ligands were defined as special residues which do not "see" each other and the proper force normalization (Eq. 2) was introduced to the fast energy routines. The initial coordinates were taken from the X-ray structure of leghemoglobin (Arutyunyan et al.[19]). (All the ligands occupy the same point in space at the beginning of the simulation.) The velocities are selected from the Boltzmann distribution (the velocity of each of the ligand replicas is selected independently). The simulations were performed with parameters appropriate for a heme in the deoxy state.[21]

The initial conditions were used in a standard molecular dynamics run. The trajectory was calculated in vacuo using the constant dielectric option of the CHARMM potential energy. The Verlet integrator[22] was used to propagate the structures in time, with a step size of 1 fsec. Bonds with polar hydrogens were fixed using the "SHAKE" algorithm[23] and the CH_n groups were treated in the extended atom model. The coordinate sets were saved at 0.1 psec intervals, and the assigned temperature was maintained by scaling the velocities after every 100 integration steps.

It is important to emphasize that the LES protocol is designed to overcome *entropic* barriers. An example for a process with an entropic barrier is of a ligand escape from a box with a small hole. The large number of ligand trajectories in LES will enhance the probability of sampling a ligand at the exit (hole). Unfortunately, an energy barrier is a different story. The Boltzmann weight: $\exp(-E/K_BT)$ enhances the sampling of configurations with low energies. At room temperature simulations, most of the ligand copies might have energies significantly lower than the barrier height.

It is not clear how to partition the required free energy of activation for ligand diffusion in proteins to enthalpic and to entropic components. The calculations of Case and McCammon[8] and Kottalam and Case[9] for a specific diffusion path in myoglobin gave only an entropic barrier. However, the experiments of Austin et al.[27a] and of Chatfield et al.[27b] suggests nonzero activation enthalpy.

We started the simulations assuming the existence of both types of barriers. We therefore assigned high energy to the ligand copies, similarly to the studies on myoglobin.[1,7,12,13] At certain time intervals random velocities were selected for the ligand copies according to predetermined temperature (see Table I for the periods and temperatures). After the selection usual molecular dynamics followed. During the simulation we found, however, that the high energy assignments are not necessary. In contrast to myoglobin, the diffusion in leghemoglobin was sufficiently fast and we could sample it at room temperature employing several picosecond runs. We therefore modified the protocol during the simulation. A significant part of our calculations is performed directly at room temperature.

Our atomic detail calculation is the first *direct* simulation of a full size ligand escaping from a fluctuating protein matrix. We were able to probe the escape of carbon monoxide molecule from leghemo-

TABLE I. A Summary of the Properties of Different Trajectories*

Trajectory	Number	Temperature	Time	Escapes
A	120	2700	4	0
		900	6	2
		900	6	5
		900	3	2
		300	26	12
B	120	300	12	55
C	120	100	3	0
		300	9	16
		300	9	6
D	60	100	2	0
		300	9	23
E	1	300	14	0
F†	120	100	2	0
		300	9	2
G‡	120	10000	55	1

*Trajectory A is analyzed in detail in the text. The column "number" is of the number of ligands, "time" gives the time interval between successive heat shocks (in picoseconds), and "escapes" corresponds to the number of ligands which left the protein matrix during this period. All the ligands left from the same exit. See text for details on the escape pathway. Since in each of the separate trajectories the protein fluctuates somewhat differently, the gate is opened at different times and to a different extent. This results in a different "rate." The sampling of protein fluctuations is not complete in a single run. We therefore did not attempt to calculate the rate from these calculations.

†F is a simulation with an X-ray structure[19] different from the one employed in the rest of the studies.

‡G is a rigid protein simulation.

globin using the LES method. This suggests that the room temperature free energy barrier is dominated by entropy. This result is similar to the activated dynamics calculations of Case and McCammon[8] and Kottalam and Case.[9] They calculated the rate of ligand diffusion in a different globin (myoglobin).

The properties of the new simulation technique (LES)[1] have not, as yet, been fully explored. We therefore tried a number of different simulation protocols in order to test the sensitivity of the calculation to the setup chosen, and to test the convergence of our results. We varied the following parameters in our calculations:

1. The number of ligand copies. The simulations pursued included 120 and 60 copies of the ligand.

2. Ligand velocities. As explained above we tried to "help" the ligand to escape. This help was later found unnecessary. (Nevertheless it is interesting to note that the ligand copies with the excess energy yielded the same reaction coordinate as the copies at room temperature.) We reinitialized the ligand velocities at different time intervals. They were sampled from the Boltzmann distribution and several different temperatures were considered: 300 K, 900 K, 2,700 K. See Table I for more details.

3. A rigid protein was examined in addition to the more realistic flexible protein. (In the rigid protein simulation the ligand temperature was maintained at 10,000 K to make escapes of ligand copies possible).

4. We considered two different X-ray structures (met-leghemoglobin and deoxy leghemoglobin) as a starting configuration for the protein coordinates.

One expects that if the diffusion paths sampled in the LES protocol were highly sensitive to the choice of initial conditions then the different parameters in (1)–(4) would have a significant effect on the results. It is therefore of considerable interest to note that the diffusion patterns were very similar in all the simulations and that only one dominant escape route was obtained, regardless of the protocol employed.

Our main interest focused of course on the ligand trajectories and the residues which were in a close contact with the ligand(s) during the simulation. It is important to remember, however, that LES is an approximate simulation technique. To check if the approximations are sound we first analyze the properties of the *protein* trajectory. We describe data for one of the runs only. This is since the analysis is standard and the results from different trajectories are similar. The length of the analyzed trajectory was 45 psec and it included 120 copies of the ligand (carbon monoxide).

During the simulation, seven χ_1 transitions occurred. The number of flips per residue and per unit time is comparable to what was found in other simulations.[25] In Figure 1, we show the time history of χ_1 of tryptophan 126. Tryptophan 126 is the last residue of the GH loop and it underwent a χ_1 transition in the A trajectory. Similarly to simulations of tryptophan ring flips in myoglobin,[24] the transition is rare (sampled only once in 45 psec) and fast. The transition happened over a 3 psec period between the 17th and 20th psec of the trajectory. Hence the properties of this ring flip are similar to other well-behaved trajectories. A standard test of a simulation quality is the rms (root mean square difference) between structures. In Figure 2 we show the rms difference between the structures along the trajectory and the X-ray coordinates. The results seem to converge after 15 psec on a value of ~2.5 Å for all atoms and ~2.1 Å for backbone only. Though relatively high, the deviations are still within the range of acceptability and are common during simulation of proteins. We further note that there is experimental evidence that leghemoglobin is more flexible than myoglobin.[26] Therefore, larger deviations from the X-ray data are to be expected. Another test of the simulation quality is shown in Figure 3, in which the time averaged rms is plotted. Periodic boundary conditions in time are employed.[25] The time averaged rms seems to approach a typical asymptotic

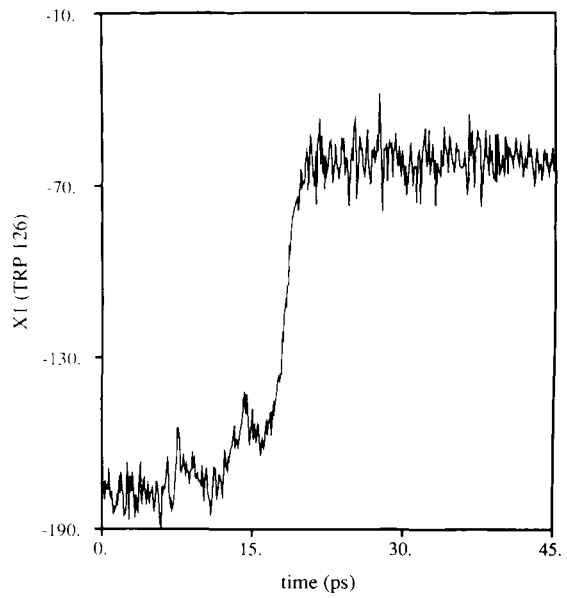

Fig. 1. The time dependence of χ_1 angle of tryptophan 126 in trajectory. A. Time is in picoseconds and angle in degrees.

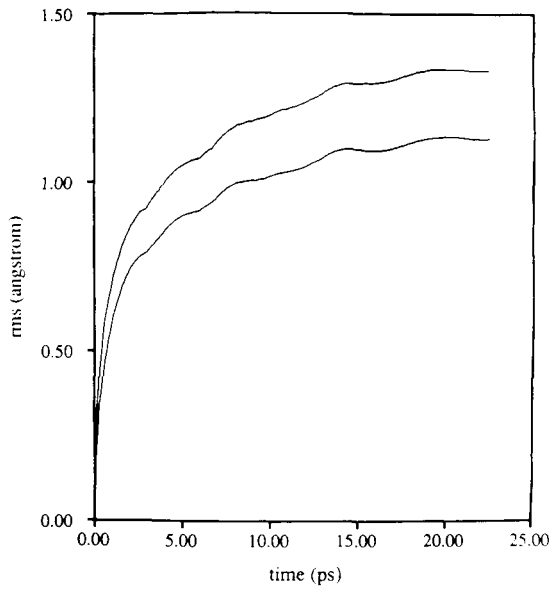

Fig. 3. The average over time of the rms for trajectory. A. Periodic boundary conditions are employed. Thus the last structure of the trajectory is followed by the first one. See text for more details.

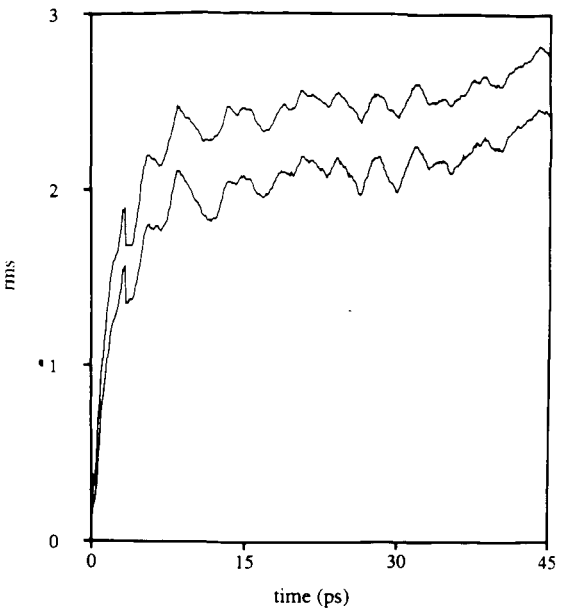

Fig. 2. Root mean square differences (rms) between the X-ray coordinates and the coordinates from the trajectory. The lower curve is the rms of the backbone atoms (C_α, C, N) and the upper curve is the rms of all the atoms.

value [1.3 Å for all atoms rms and 1.1 Å for backbone (C_α,N,C) rms].

The rms calculations and the side chain flip suggest that the simulation is typical. We are therefore ready to proceed and to present the results for the diffusion paths in the next section.

RESULTS AND DISCUSSIONS

Here we focus on the motion of the ligand through the protein matrix and on the coupling of the protein fluctuations to the ligand trajectories.

It has long been recognized that the reaction of a small ligand with heme proteins can be separated into several sequential steps. The low temperature recombination studies of Austin et al.[27a] suggest the following reaction scheme: (1) breaking the bond between the ligand and the heme iron; (2) ligand exit from the heme pocket (still in the protein interior); and (3) ligand diffusion from the protein to the solvent. Past studies for myoglobin demonstrated that the ligand is trapped for a considerable amount of time in several alternative sites in the protein interior.[1,12,13] The present study, which is focused on leghemoglobin, shows a considerably simpler picture; there we find only one cavity in which the ligand is trapped for a substantial length of time during the simulation—the heme pocket, and from which the escape (in contrast to a recent simulation in myoglobin[1]) is almost direct. In Figure 4 we show a snapshot in time of the heme environment. The structure is after 20 psec simulation of trajectory A (see Table I). The dark spots in the middle of the figure are the ligand copies which are still trapped in the heme pocket at this time (105 copies). The residues at the boundaries of the heme pocket are:

Phe-29 B9, Phe-44 CE1, His-63 E7, Ala-64 E8, Val-67 E11, Phe-68 E12, Val-110 G8, Ile-114 G12, Tyr-138 H12.

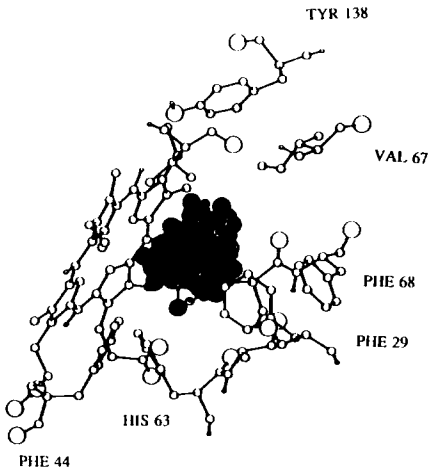

Fig. 4. A model of sticks and balls for the heme environment in leghemoglobin. The structure is after 20 psec simulation (trajectory A). The black circles are the ligand copies which are still in the pocket at this time (105 replica).

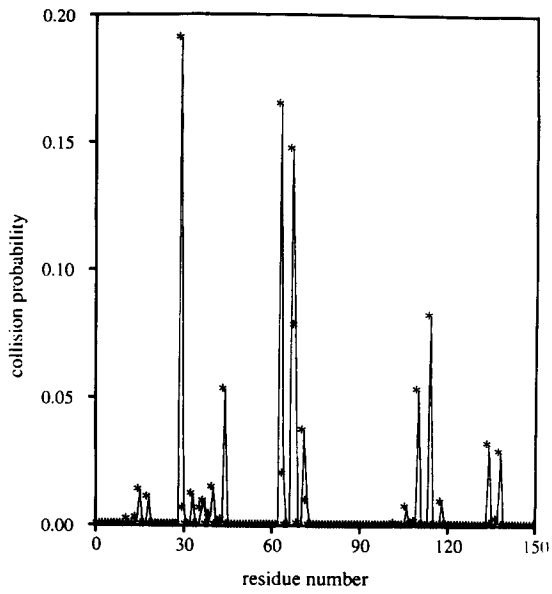

Fig. 5. A histogram plot of the collision probability between the ligand and the protein residues (not including the heme). A collision is assumed when the distance between any of the ligand atoms or any of the residue atoms is smaller than 4 Å. The histogram is an average over all times and over all the 120 ligand copies.

We defined the residues of the pocket by the close contacts with the ligands during the trajectory. We examined a large number of protein structures which fluctuate during the trajectories. A cutoff distance of 5 Å from the center of the ligand "cloud" trapped in a cavity was employed to define a cavity. A more generous definition with a cutoff distance of 7 Å adds the following residues to the list of heme pocket residues: Phe-30 B10, Val-33 B13, Val-71 E15, His-106 G4, Ala-113 G11.

In Figure 5 we show a histogram plot of the collision frequencies between the ligand and any of the protein residues. The frequencies are normalized to one to resemble probabilities. (Note that we did not include collisions with the heme which is an "obvious" collider.) An encounter is counted when the distance between at least one of the ligand atoms and at least one of the residue atoms is equal to or less than 4 Å. Only collision frequencies larger than 0.001 are shown. The histogram is from trajectory A using a double average: an average over time and an average over the different ligand copies. The other trajectories (B–G) yield similar results. The "spiky" character of the frequent collider histogram suggests that only a small fraction of the residues are of major importance in the diffusion process and we therefore list these residues in Table II. Of the total number of collisions with the ligand 94% are with 15 residues, less than 10% of the 153 residues of leghemoglobin. The diffusion process in the protein is therefore highly constrained. The ligand spends the largest portion of its time in the heme pocket. The three residues, each with more than 10% of the total number of collisions are Phe-29 B9, His-63 E7, and Val-67 E11. All the three are residues of the heme pocket. The "traditional" Case and Karplus reaction coordinate includes the motion of the distal histidine His E7 and that of the Val E11.[7] Our simulations are consistent with the idea that these residues are very important in stopping the ligand escape. If these residues would change their orientation significantly during our simulation it is likely that the ligand would escape that way. However, fluctuations of these residues which are sufficient to open escape pathways were not sampled in our simulations, rather, the ligand followed preferentially another exit route to be described later.

In several runs another cavity in the protein structure was sampled. Only a small number of trajectories entered it. We therefore called it "the minor cavity." Ligands that jumped to this cavity stayed there for the rest of the simulation. The hopping to this cavity is obviously a part of the complete diffusion process and as such it deserves consideration (even though it is not leading to ligand escape). In Figure 6 we show a model of balls and sticks of the minor cavity. For the replica we used dark circles to distinguish them from the protein atoms (open circles). In this snapshot in time (20 psec of trajectory A), there were 4 ligand copies inside the cavity. To find the residues around the cavity we employed the same definition for close contacts as the one used for the heme pocket. The closest "shell" of residues (cutoff of 5 Å from the center of the ligands in the pocket) is shown in Figure 6. The residues are also listed below:

TABLE II. Residues Which Collide With the Ligand During the LES Trajectories Together With Their Collision Probability*

Residue	Collision probability
Phe-29 B9	0.190
His-63 E7	0.165
Val-67 E11	0.147
Ile-114 G12	0.082
Phe-68 E12	0.078
Phe-44 C7	0.053
Val-110 G8	0.053
Val-71 E15	0.037
Trp-134 H8	0.032
Tyr-138 H12	0.029
Ala-64 E8	0.020
Ala-40 C3	0.015
Trp-15 A12	0.014
Val-33 B13	0.012
Phe-18 A15	0.011
Ala-37 B17	0.009
Tyr-72 E16	0.009
Ile-118 G12	0.009
His-106 G4	0.007
Phe-30 B10	0.006
Ile-36 B16	0.006
Ala-39 C2	0.004
Ser-14 A11	0.003
Val-11 A7	0.002
Leu-43 C6	0.002
Val-109 G7	0.002
Ala-137 H11	0.002

*For each structure a collision is counted if the distance between at least one atom of the ligand and at least one atom of the colliding residue is less or equal to 4 Å. The results are averaged over all ligand copies and all times.

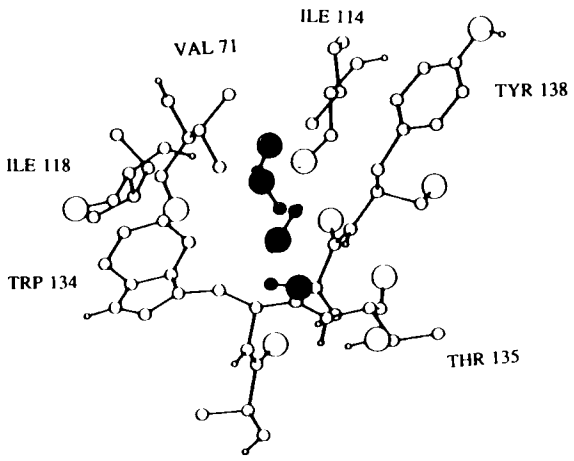

Fig. 6. A model of sticks and balls for the minor cavity. The dark circles are the ligands. See text for more details.

Tyr-134 H8, Thr-135 H9, Ala-137 H11, Tyr-138 H12, Ile-114 G12, Ile-118 G16, Val-71 E15, Ala-133 H7

The generous definition (cutoff 7 Å) includes in addition:

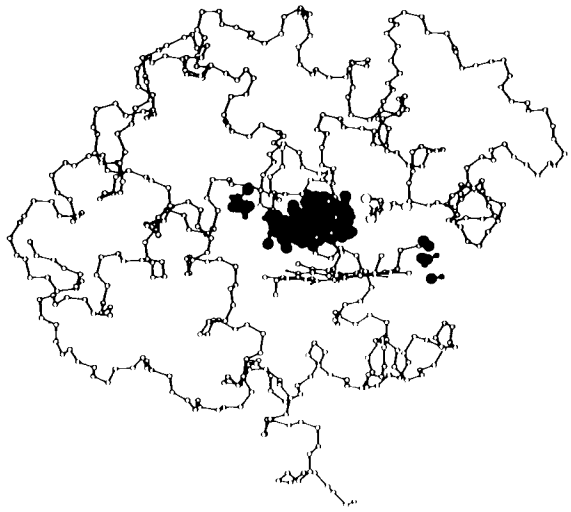

Fig. 7. A snapshot in time of the last structure of trajectory C. The plot includes the backbone of the protein (C_α, C, N), the heme, and the escaping copies (in dark circles).

Ile-136 H10, Asp-139 H13, Glu-140 H14, Leu-141 H15, Ala-142 H16, Lys-111 G9, Lys-119 G17, Phe-68 E12

The cavity is primarily between helices G and H. To hop from the heme cavity to the minor cavity the ligand is required to pass between the following residues: Phe-68 E12, Val-71 E15, and Ile-114 G12 which form the gate between the two cavities. This gate was found by graphical examination of the trajectories.

We next turned our attention to the pathway of the escaping ligands. In Figure 7 we show the last structure of trajectory C (after 21 psec simulation). The backbone of the protein (C_α, C, and N) is plotted as open circles and the ligand copies as dark circles. The black cloud above the heme corresponds to the copies that are still trapped in the heme pocket. The tiny cloud on the left side of the big one includes the ligands of the minor cavity and finally the dark circles level with the heme on the right are ligand copies on their way out. The protein orientation is such that the heme is (approximately) parallel to the x-axis and the G and H helices are close to the reader. The exit is at the beginning of the G helix and near the bend between the B and the C helices. The picture of the heme pocket stuffed with ligand copies, the minor cavity with a minute number of copies and a small number of copies streaming out, repeated itself in other situations with different parameters. In Figure 8 the last structure of the A trajectory is plotted. All the ligands are shown. However, in contrast to the previous picture only the

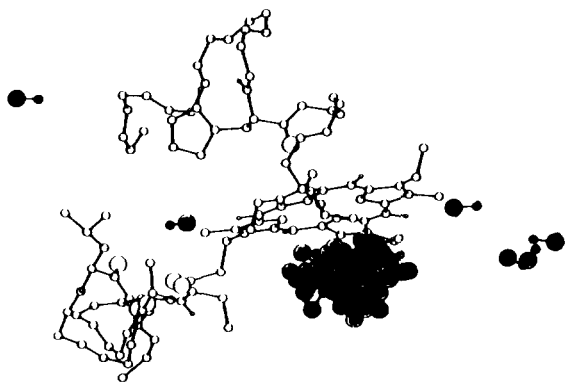

Fig. 8. The last structure of trajectory A showing all the ligands and the protein residues along the escape pathway. This includes the heme, the backbone of residues 36 to 45 (part of the B and C helices), the backbone of residues 100 to 110 (part of F and G helices), and the side chains His-106 G4 and Ala-37 B17.

protein parts relevant to the escape gate are plotted. The relevant protein parts include the heme, the backbone of residues 36 to 45 (parts of the B and C helices), the backbone of residues 100 to 110 (part of F and G helices), and the side chains of the frequent colliders at the gate. The residues at close contact (less than 5 Å) are Ala-37 B17, Ala-40 C3, Leu-43 C6, and His-106 G4. Larger cutoff for collision definition (7 Å) yields additional contacts of ligands with Val-102 FG2, Val-109 G7, Ile-36 B16, and Phe-44 C7.

We comment that the internal volume of the protein molecule explored by different ligands is similar for all the ligands. Specifically, ligands which escape and ligands which do not escape explore the same cavities in the protein structure.

Until now we focused on the motions of the ligand only. It is of considerable interest to understand the coupling between the protein fluctuations and the motions of the different ligand copies. This question is far from trivial and occupied us for days of continuous staring at graphic visualization of molecular dynamics trajectories. There are a number of contributions to the protein fluctuations in general and to the transient opening of the path in particular. The side chain fluctuations as well as the shift of secondary structure units play an important role in reducing the energy barriers for ligand escape. We found, however, that the largest structural fluctuations in the neighborhood of the gate correspond to secondary structure motions (displacements of helices). If one accepts the reasonable suggestion (which we cannot prove) that the largest geometrical displacements in the neighborhood of the gate are the ones which are the most important in reducing the diffusion energy barrier, then the energy barrier is controlled by the shift of the C helix. To study the formation dynamics of the transient hole we show in Figure 9a–c three structures from trajectory A at times of 3, 6, and 9 psec. Also shown in Figure 9d is the overlap of Figure 9a and 9c. The overlap is such that the rms of the G helix and of the heme is minimized. It is therefore useful to demonstrate the shift of the C helix. The figures include only the parts of the protein which participate in the opening of the gate (same as Fig. 8). In (a) we show a closed structure. In the closed structure the shortest distance between any of the atoms of His-106 G4 and any of the atoms of Ala-37 B17 is 6.19 Å which is the same as in the X-ray data. In (b) the ligands are "waiting" at the door. The structure is partially open. When the heme and the portion of the F/G helices in the figure are overlapped we find that the B/C bend is displaced away from the F/G by ~1 Å. In (c) we realize that the displacement was increased to ~2 Å. The distance between His-106 G4 and Ala-37 B17 is 8.6 Å. The gate is completely open and the ligand copies flow out.

We further note that even in the rigid protein simulations (see Table I) the ligand escapes from the same exit. The simulation conditions were of course more extreme: i.e., the ligand temperature was 10,000 K and only a single ligand escaped after 55 psec of the molecular dynamics simulation. However, the fact that the escape route was the same suggests that the role of protein fluctuations in leghemoglobin is to enlarge existing holes and not to generate new ones.

We should like to devote the remainder of the section to possible problems and to the things which may go wrong in the simulation. The first obvious difficulty is the lack of water in the simulation. This may affect the diffusion to an unknown extent. It may change the average structure as well as the magnitude of the protein fluctuations. Simulation of the dynamics of solvated leghemoglobin is currently under progress (W. Nowak and R. Elber, work in progress). Another possible problem is related to the LES protocol. It is an approximate procedure which may enhance the importance of open gates. For example, at a gate which is already open there is no barrier for the next ligand to pass through. The ligands may escape in a correlated way in which the escaping ligand keeps the gate opened for the next one. It is important to emphasize that the "low-frequency" motion of the C helix is indeed slow and only a few vibrational periods were sampled during the trajectories described above. The LES protocol is quite useful in this case because it increases the probability that a ligand will be at the right place (the opened gate) at the right time. However, due to

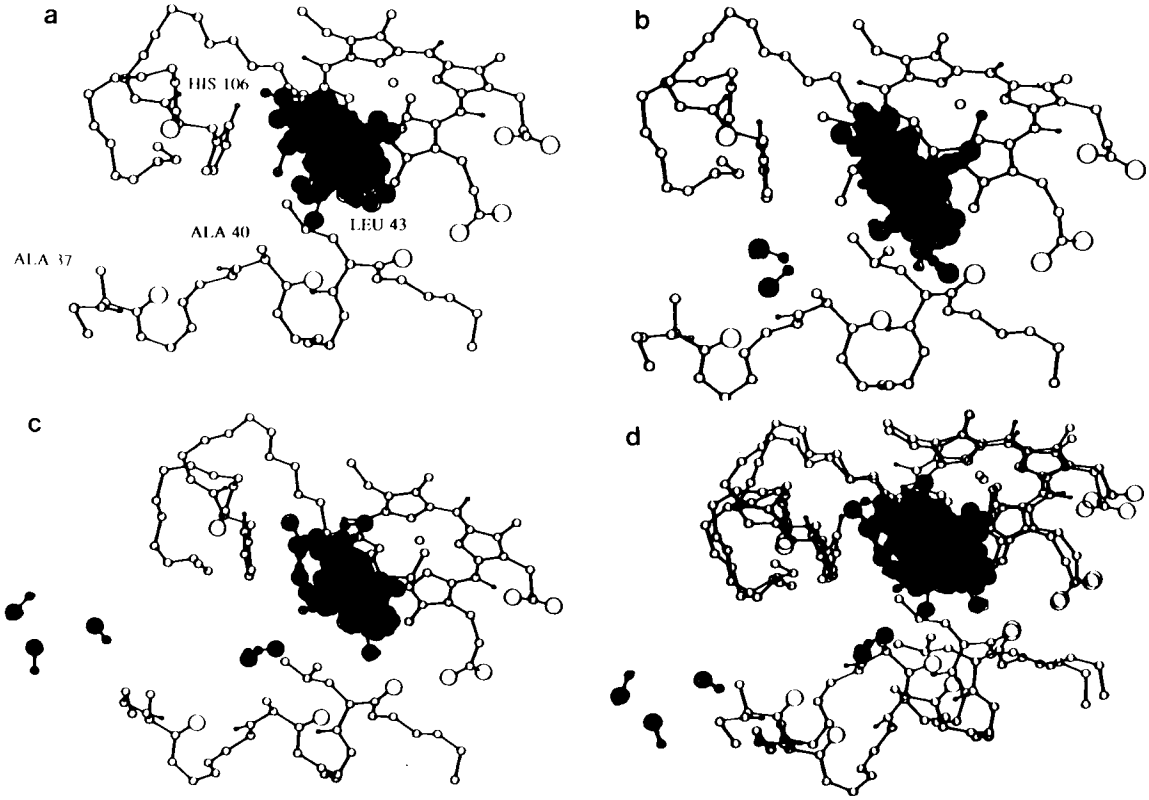

Fig. 9. The time evolution of the ligand position and the protein conformational changes relevant to the ligand escape. The same residues as in Figure 8 are shown. The structure as a function of time **(a)** 3 psec, **(b)** 6 psec, **(c)** 9 psec; **(d)** the structures at 3 and 9 psec in which the heme and the G helix are overlapped. See text for more details.

the incomplete sampling of protein fluctuations it is not possible at present to extract directly the diffusion rate from the simulation results. Only gross estimate for the diffusion time scale may be obtained. Care must be taken in the interpretation of the results. For example the 55 ligands which escape from the protein matrix during the B run did it over a relatively short time period (histogram of escape times is shown in Fig. 10). Once the gate was opened the ligands escaped in a *collective and correlated way*.

Finally, if the formation of the hole depends on a rare fluctuation in the protein structure, there is a good chance that we shall miss it. This is since we enhance the sampling of the ligand only. More complete sampling of the diffusion pathways may require enhanced sampling of other protein residues too.

Thus the possible difficulties are (1) missing of some of the paths and (2) inaccurate estimates of the relative importance of paths. We emphasize, however, that the path which we obtained by direct simulation is energetically reasonable and a straightforward solution to the problem. We believe

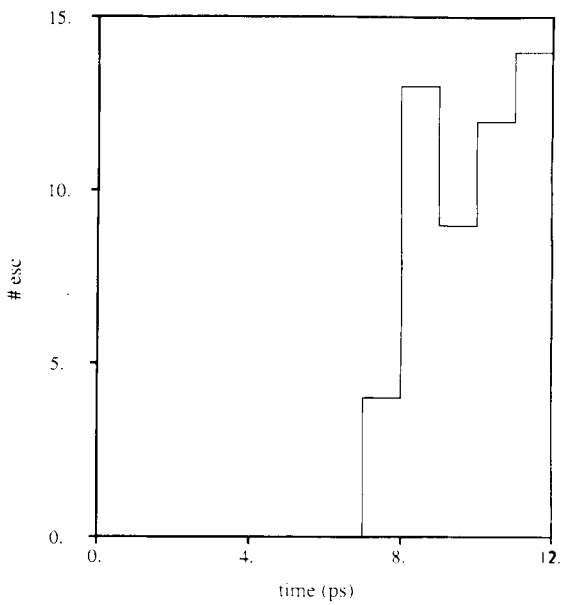

Fig. 10. A histogram plot of the number of ligands which escape at a specific time from the protein matrix for the B run. The box size is 1 psec.

therefore that a complete picture of the process should include the path at the C helix which we found in addition to other paths which may possibly exist. We believe that our path is reasonable also because the path calculated was insensitive to the changes in the computational protocols that were used.

FINAL REMARKS

We employed a new simulation methodology[1] for an automatic search of escape pathways in a thermally *fluctuating* protein (leghemoglobin). The proposed method for search of paths is more efficient by a factor ~10–100 as compared to straightforward simulations. For comparison, Tilton et al.[13] employed five ligands in a straightforward simulation of ligand diffusion in myoglobin. We considered 661 ligand trajectories in leghemoglobin which is a protein of a similar size.

An intriguing result is that in contrast to the complex network of cavities found in myoglobin[1,12,13] we observed only one dominant path. The path involves the shift of the C helix and enables ligand escape directly from the heme pocket. The distal histidine "door" was not opened during the simulation. The simulation indicates also that diffusion rate from leghemoglobin is significantly faster than from myoglobin in accord with experiment. Simulation of the diffusion in a solvated protein and the calculation of the refined reaction coordinate are currently in progress. Gibson et al. studied the rebinding kinetics of CO, O_2 and NO in a series of leghemoglobins.[28] One conclusion was that the CE loop and the residues E11 and G3 play an important role in determining the kinetics. These observations are in in agreement with the diffusion mechanism which we proposed.

Finally we should like to stress the difference between our model and some models based on the distal histidine.[3] Our model is of global secondary structure motion. The complete C helix is shifted. The distal histidine path is based on local side chain transitions. We note that the low-frequency modes of proteins are indeed delocalized and it is therefore likely that energy inexpensive motions will include these extended coordinates. Side chain transitions are known to be rare and of short duration. It is not obvious that the ligand will wait long enough to take advantage of the rare flip. In our and others' view[1,12,13] even for myoglobin it is still not clear if the reaction coordinate(s) is(are) really dominated by side chain—local fluctuations. Future experiments testing alternatives routes (fortified by theoretical studies) may provide answers to the last question.

APPENDIX: EQUATIONS OF MOTION FOR LES

The coordinates of the ith ligand are denoted by $\mathbf{R}_i^l$ and the coordinates for the protein are $\mathbf{R}^p$. The potential energy is $V(\mathbf{R}_i^l,\mathbf{R}^p)$. The differential equations of motion for any of the ligands are

$$\mathbf{M}_l\frac{d^2\mathbf{R}_i^l}{dt^2}=-\nabla_{\mathbf{R}_i^l}V(\mathbf{R}_i^l,\mathbf{R}^p) \tag{1}$$

$\mathbf{M}$ is the mass matrix. For the protein we have

$$\mathbf{M}_p\frac{d^2\mathbf{R}^p}{dt^2}=-\frac{1}{N}\sum_{i=1}^{N}\nabla_{\mathbf{R}^p}V(\mathbf{R}_i^l,\mathbf{R}^p) \tag{2}$$

ACKNOWLEDGMENTS

This research was supported by NIH Grant GM41905 and by NIH Equipment Grant RR04884 to RE. RE is a Camille and Henry Dreyfus New Faculty.

REFERENCES

1. Elber, R., Karplus, M. Enhanced sampling in molecular dynamics: Use of the time-dependent Hartree approximation for a simulation of carbon monoxide diffusion through myoglobin. J. Am. Chem. Soc., in press.
2. Gerber, R.B., Buch, V., Ratner, M.A. Time-dependent self consistent field approximation for intramolecular energy transfer. I Formulation and application to dissociation of van der Waals molecules. J. Chem. Phys. 77:3022–3030, 1982.
3. Perutz, M.F. Mechanisms of cooperativity and allosteric regulations in proteins. Quart. Rev. Biophys. 22:2, 1989.
4. Perutz, M.F., Mathews, F.S. An X-ray study of azide methemoglobin. J. Mol. Biol. 21:199–202, 1966.
5. Takano, T. Structure of myoglobin refined at 2.0Å resolution. J. Mol. Biol. 110:537–568, 1977; 110:569–584, 1977.
6. For a recent review see Brunori, M., Coletta, M., Ascenzi, P., Bolognesi, M. Kinetic control of ligand binding processes in hemoproteins. J. Mol. Struct. 42:175–193, 1989, and references therein.
7. Case, D.A., Karplus, M. Dynamics of ligand binding to heme proteins. J. Mol. Biol. 132:343–368, 1979.
8. Case, D.A., and McCammon, J.A. Dynamical simulations of oxygen binding to myoglobin. Ann. N.Y. Acad. Sci. 482: 222–230, 1986.
9. Kottalam, J., Case, D.A. Dynamics of ligand escape from the heme pocket of myoglobin. J. Am. Chem. Soc. 110: 7690–7697, 1988.
10. Jongeward, K.A., Magde, D., Taube, D.J., Marsters, J.C., Traylor, T.G., Sharma, V.J. Picosecond and nanosecond germinate recombination of myoglobin with CO, O_2, NO, and isocyanides. J. Am. Chem. Soc. 110:380–387, 1988.
11a. Lakowicz, J.R., Weber, D. Quenching of protein fluorescence by oxygen. Detection of structural fluctuations in proteins on the nanosecond time scale. Biochemistry 12: 4171–4179, 1973.
Calhoun, D.B., Englander, S.W., Wright, W.W., Vanderkooi, J.M. Quenching of room temperature protein phosphorescence by added small molecules. Biochemistry 27: 8466–8474, 1988.
11b. Barboy, N., Feitelson, J. Diffusion of small molecules through the structure of myoglobin. Environmental effects. Biochemistry 28:5450–5456, 1989.
12. 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. Computational studies of the interaction of myoglobin and xenon. J. Mol. Biol. 192:443–456, 1986.
13. Tilton, R.F., Jr., Singh, U.C., Kuntz, I.D., Jr., Kollman, P.A. Protein ligand dynamics: A 96 picosecond simulation of a myoglobin-xenon complex. J. Mol. Biol. 199:195–211, 1988.
14. For a textbook see De-Shalit, A., Feshbach, H. "Theoretical Nuclear Physics, Vol I: Nuclear Structure." New York: Wiley, 1974, chapter VII.
15. For a recent paper see Balian, R., Veneroni, M. Lyapunov stability and Poisson structure of the thermal TDHF and RPA equations. Ann. Phys. 195:325–355, 1989.

16. Stetzkowski, F., Banerjee, R., Mardens, M.C., Beece, D.K., Bowne, S.F., Doster, W., Eisenstein, L., Frauenfelder, H., Reinisch, L., Shyamsunder, E., Jung, C. Dynamics of dioxygen and carbon monoxide binding to soybean leghemoglobin. J. Biol. Chem. 260:8803–8809, 1985.
17. (a) Elber, R., Karplus, M. A method for determining reaction path in large molecules: Application to myoglobin. Chem. Phys. Lett. 139:375, 1987; (b) Czerminski, R., Elber, R. Reaction path study of conformational transitions and helix formation in a tetrapeptide. Proc. Natl. Acad. Sci. U.S.A. 86:6963–6967, 1989; (c) Ulitsky, A., Elber, R. A new technique to calculate steepest descent paths in flexible polyatomic systems. J. Chem. Phys. 92:1510–1511, 1990; (d) Czerminski, R., Elber, R. Reaction path study of conformational transitions in flexible systems: Applications to peptides. J. Chem. Phys. 92:5580, 1990; (e) Czerminski, R., Elber, R. Self avoiding walk between two fixed points as a tool to calculate reaction paths in large molecular systems. The proceeding of Sanibel Symposia 1990. Int. J. Quant. Chem., in press; (f) Yip, V., Elber, R. Calculation of a reaction coordinate for ligand diffusion in erythrocruorin hemoglobin, to be published.
18. Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., Karplus, M. CHARMM: A program for macromolecular energy minimization and dynamics calculations. J. Comput. Chem. 4:187–217, 1983.
19. Arutyunyan, E.G., Kuranova, I.P., Vainstein, B.K., Steigmann, W. X-ray structural investigation of leghemoglobin. Structure of acetate-ferileghemoglobin at a resolution of 2.0 angstrom. Sov. Phys. Crystallogr. 25:43–58, 1980.
20. Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T. "Numerical Recipes." Cambridge: Cambridge Univ. Press, 1986, chapter 10.
21. Arata, A., Kuczera, K., Karplus, M., to be published.
22. Verlet, L. Computer experiments on classical fluids I. Thermodynamical properties of Lennard-Jones molecules. Phys. Rev. 159:98–104, 1967.
23. Ryckaert, J.P., Ciccotti, G., Berendsen, H.J.C. Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. J. Comput. Phys. 23:327–341, 1977.
24. An example for simulation of tryptophan ring flips: Henry, E.R., Hochstrasser, R.M. Molecular dynamics simulations of fluorescepce polarization of tryptophans in myoglobin. Proc. Natl. Acad. Sci. U.S.A. 84:6142–6146, 1987.
25. Rojewska, D., Elber, R. Molecular dynamics study of secondary structure motions in proteins: Application to myohemerythrin. Proteins 7:265, 1990.
26. Kong, S.B., Cutnell, J.D., La Mar, G.N. Proton nuclear magnetic resonance study of the dynamic stability of the heme pocket of soybean leghemoglobin a. J. Biol. Chem. 258:3843–3849, 1983.
27. (a) Austin, R.H., Beeson, K.W., Eisenstein, L., Frauenfelder, H., Gunsalus, I.C. Dynamics of ligand binding to myoglobin. Biochemistry 14:5355–5373, 1975; (b) Chatfield, M.D., Walda, K.N., Magde, D. Activation parameters for ligand escape from myoglobin proteins at room temperature. J. Am. Chem. Soc. 112:4680, 1990.
28. Gibson, Q.H., Wittenberg, J.B., Wittenberg, B.A., Bogusz, D., Appleby, C.A. The kinetics of ligand binding to plant hemoglobins. J. Biol. Chem. 264:100–107, 1989.