JCTC Journal of Chemical Theory and Computation

PELE: Protein Energy Landscape Exploration. A Novel Monte Carlo Based Technique

Kenneth W. Borrelli, Andreas Vitalis, Raul Alcantara, and Victor Guallar*

Department of Biochemistry, Washington University School of Medicine, St Louis, Missouri 63108

Received July 26, 2005

Abstract: Combining protein structure prediction algorithms and Metropolis Monte Carlo techniques, we provide a novel method to explore all-atom energy landscapes. The core of the technique is based on a steered localized perturbation followed by side-chain sampling as well as minimization cycles. The algorithm and its application to ligand diffusion are presented here. Ligand exit pathways are successfully modeled for different systems containing ligands of various sizes: carbon monoxide in myoglobin, camphor in cytochrome P450cam, and palmitic acid in the intestinal fatty-acid-binding protein. These initial applications reveal the potential of this new technique in mapping millisecond-time-scale processes. The computational cost associated with the exploration is significantly less than that of conventional MD simulations.

1. Introduction

Obtaining a molecular-level understanding of protein biochemical and biophysical processes is a complex task requiring the characterization of long-time conformational rearrangements. Molecular dynamics methods, in which the evolution of the system is projected as a series of snapshots resulting from the integration of classical equations of motion, have been largely used with great success to model conformational rearrangement in biomacromolecules.[1−6] These simulations, however, are sensitive to the size of the system and to the total time of propagation, requiring significant computational resources to reach times on the order of hundreds of nanoseconds. Different approximations have recently been introduced to modify conventional molecular dynamic methods in order to capture large-scale conformational rearrangements: simplified electrostatics models,[7] multiple replica dynamics,[8−10] stochastic path approaches,[11] Monte Carlo/molecular dynamics combinations,[12] targeted constraint dynamics,[13] modified potential landscape,[14] and so forth. These methods offer a valuable approach to model long-time dynamics, although often limited to the constraint of providing initial and final states and to the employment (for some methods) of large biasing potentials. The modeling of long-time dynamics still remains a challenge.[15,16] Of particular interest is the elucidation of protein pathways, describing a biophysical or biochemical event, for which only the initial state is known. For this purpose, various approaches often referred to as energy landscape exploration techniques have been recently developed.[16] While using different sampling methods, the overall idea is to project a set of local minima describing the protein conformational changes. Among others, sampling methods include eigenvector following[17] and backbone dihedral pivoting.[18,19] Various simpler mechanical approaches, including normal-mode analysis,[20] essential dynamics analysis,[21] and the Gaussian network model,[22] are widely used to characterize long-range conformational changes and collective protein dynamics. These mechanical methods offer valuable information on conformational changes, requiring only the initial state, but at the cost of a loss in detailed atomic resolution. The combination of these approaches with atomic detailed landscape exploration methods, like the one proposed here, could certainly constitute a new generation of landscape exploration and dynamical tools.

Parallel to the development of the above techniques, there has been a significant effort toward the development of theoretical methods for protein structure prediction based on the use of rotamer libraries.[23−26] Recent work by Hellinga et al. has proven the possibilities of theoretical guided protein engineering when using protein structure prediction algo-

* Corresponding author. E-mail: guallarv@biochem.wustl.edu.

10.1021/ct0501811 CCC: $30.25
Published on Web 10/12/2005

rithms.[27,28] Jacobson et al. have recently developed PLOP, a program for protein modeling using all-atom energy functions and specialized sampling algorithms for side-chain and loop prediction.[29,30] The program uses an implicit surface-generalized Born (SGB) continuum solvent model[31] and the recently updated version of the OPLS-AA force field parametrization (OPLS2001). The latter was demonstrated[32] to provide significant improvements in side-chain prediction as the fit to quantum chemical data in experimentally relevant regions of phase space was improved. The sampling algorithms include the use of highly detailed rotamer state libraries for side-chain conformational searching, hierarchical screening methods based on steric overlap, and approximate electrostatics to rapidly eliminate obviously incorrect conformations. With these technological advances, rigorous sampling of side chains can be done at a rate of one side chain per second on a regular desktop (throughout the paper, CPU estimates refer to a 2.8 GHz Pentium IV PC).

Using these specialized protein structure sampling algorithms, we have developed a novel approach to explore the protein landscape dynamics. The program has been named PELE: Protein Energy Landscape Exploration. The method uses a localized steered perturbation coupled to side-chain prediction and minimization algorithms. The procedure includes the possibility to perform single-point mixed quantum mechanics/molecular mechanics (QM/MM) calculations[33] to update the charges of complex ligands or to obtain quick estimates of a biochemical reaction.

We have initially applied this method to the study of ligand diffusion on several enzymes. In all the cases studied, myoglobin, cytochrome P450, and the intestinal fatty-acid-binding protein, we have obtained entry/exit pathways for one representative ligand for each system: CO, camphor, and palmitate, respectively. All diffusive pathways are consistent with experimental data. The computational cost varies with the system, from 1 day for myoglobin to 4 days on a single CPU for P450 and the fatty-acid-binding systems. In this first paper, we present the methodology and results on camphor migration on P450cam. To demonstrate the potential of the method, a summary of the results on myoglobin and the intestinal fatty-acid-binding protein are also given; more extensive results on these two systems will be presented elsewhere.

Because of the abundant experimental information, camphor migration on P450cam is a well-defined test case. A recent study trapped two different synthetic molecular wires, indicating the importance of the motion of the F and G helix in the substrate entry pathway[34] (Figure 1). Furthermore, simulations using random steered molecular dynamics have also been performed on this system.[35] This method, however, required a large force pushing away the ligand from the heme center, resulting in multiple paths for ligand exit. Our results are consistent with the synthetic molecular wires' crystallographic results and reveal a detailed atomic description of the protein response along the entire migration pathway.

2. Methods

We have developed PELE, a new approach to map the energy landscape for long-time conformational rearrangement in proteins. This procedure incorporates Monte Carlo moves, rotamer library side-chain optimizations, truncated Newton minimizations, and Metropolis acceptance tests. The method is used to generate and propagate changes in a system by generating a series of structurally similar local minima that are then combined into a trajectory.

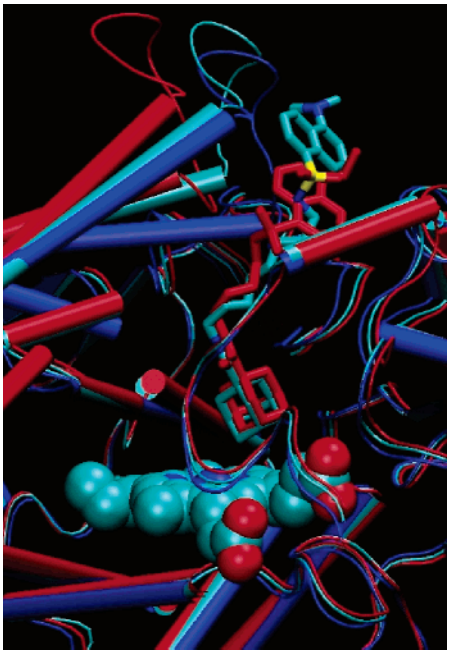

Figure 1. Local superposition of the crystal structures for P450cam with camphor (PDB: 1PCH; dark blue), the short D-4d wire ligand (PDB: 1RF9; red), and the longer 8D−d wire ligand. (PDB: 1RE9; light blue). A clear opening of the F/G helixes and the loop connecting them is observed for the wire ligands.

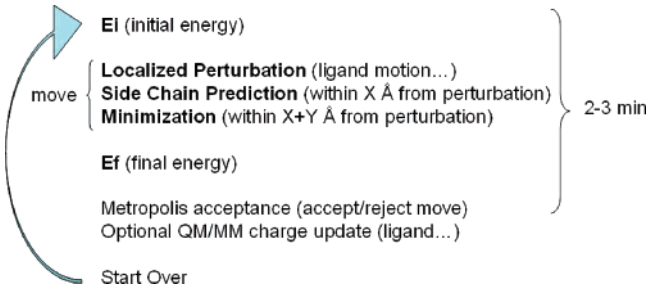

Figure 2. Sampling procedure used in PELE.

Sampling Procedure. *Local Perturbation*. The procedure employed (Figure 2) begins with the generation of a local perturbation of the ligand. Many ligands can be treated as rigid bodies so only three rotational and three translational degrees of freedom are required. Flexible ligands, such as palmitate, cannot be adequately described as a single rigid unit. Thus, the perturbation of these flexible ligands includes additional degrees of freedom from the dihedral angles of rotatable bonds, while the bond distances and bending angles are kept fixed. At this stage, the user has to provide a list of rotatable bonds for the ligand as well as OPLS-AA style parameters (at least atom types). Next, a series of filters are applied to determine if there is any steric contact between the ligand and the backbone of the protein and, in the case of flexible ligands, between the ligand and itself. If any such

contacts are found, the perturbation is rejected. A pair of nonbonded atoms were determined to create a steric clash if the distance between them was less than 70% of the sum of the van der Waals radii of the two atoms. A set of 400 perturbations are generated, and the one with the best scoring is selected. The scoring function used is an OPLS-AA force field in which only the ligand and the backbone of the protein were considered. The solvation effects were neglected as were the nonbonded interactions with residues more than 15 Å from the ligand in its position at the start of the step.

Side-Chain Sampling. The algorithm proceeds by optimally arranging all side chains local to the atoms perturbed in the first step with a rotamer library side-chain optimization at a rotamer resolution of 10°. The side-chain algorithm[32] uses steric filtering and a clustering method to reduce the number of rotamers to be minimized (side-chain minimization only). The typical time required for the side-chain sampling of 20−30 residues, corresponding to a sampling distance of 8 Å from a medium-size ligand, is in the 20−40 s range. Future versions will speed up the side-chain sampling procedure by adapting the side-chain rotamer resolution along the landscape exploration; side chains experiencing less conformational changes will decrease their rotamer resolution.

Minimization. The last step in every move involves the minimization of a region including, at least, all residues local to the atoms involved in steps 1 and 2. The truncated Newton minimization algorithm uses a multiscale protocol, 1−2 orders of magnitude more efficient than conventional approaches.[29] The minimization is intended to generate a backbone response to the initial local perturbation and to possible side-chain rearrangement in the first steps. The overall approach is designed on the basis of the assumption that side chains will act as protein sensors, responding to ligand motion, protein−protein interactions, biochemical activity, and so forth. The backbone will then follow the side-chain response.

The *X* and *Y* distances for the local side chain and minimizations (see Figure 2) are measured in Å and are chosen by the user. They determine the local areas for side chains and minimization. All residues with an atom within the *X* (or *Y*) distances from any ligand atom are included in the list of local residues.

These three steps compose a *move* which is accepted (defining a new minimum) or rejected on the basis of a Metropolis criterion for a given temperature

$$\Delta V < 0$$
$$\exp(-\Delta V/K_B T) < R \qquad (1)$$

that is, by a decrease of the potential surface, $\Delta V < 0$, or by satisfying the second criterion where K_B is the Boltzmann constant, T the temperature chosen for the simulation, and R is a random number with a [0, 1] range. The conformational changes are propagated to the nonlocal environment by means of the diffusion of the local perturbed region and by introducing a long-range step similar to the local step but involving a longer sampling radius. In the results presented here, the local step was confined to residues within 8 Å of the perturbation. The long-range step was conducted every 10 moves and extended out to 25 Å from the ligand. The overall time for each move depends on the chosen side chain and minimization region, and on ligand size. For the default values, 8 Å for side chains and 10 Å for minimization, a complete iteration required 2−3 min of CPU time for the tested ligands.

To perform efficient sampling of rare events, the direction of the perturbation is kept constant for a set number of steps making up a steering cycle. The translation is described by where $\vec{r}$ is the translational vector, λ is a variable describing

$$\vec{r} = (\lambda + R\lambda)\vec{r} \qquad (2)$$

the translational range, and R is, again, a random number. Thus, translations are always forced within the $[\lambda, 2\lambda]$ interval, with λ typical values between 0.05 and 0.2 Å (user-defined variable). $\vec{r}$ is previously aligned with the normalized steering vector using a Gaussian function with a tightness criterion, which ensures that the perturbation will be within 10−30° of the steered direction. A steered cycle of 20 indicates that after a period of 20 moves the steered direction might be randomly updated. The update takes place if the direction was an unsuccessful direction search (with a user-defined number of accepted steps or overall displacement). The Metropolis acceptance/rejection criterion for each move ensures that maintaining a perturbation direction does not result in a large energy increase. The result of this procedure is a series of local minima which are structurally highly correlated. However, large structural changes can be attained from one end of the trajectory to the other. In this way, the atomic-scale details of possible mechanisms can be explored.

The algorithm can be used to attain different goals by adjusting the steered and acceptance parameters. An equilibration procedure with a very short steered cycle and low Metropolis temperature can be used to produce local energy minimization and exploration. Such an equilibration procedure is necessary prior to any landscape exploration. A setup with a longer steered cycle and higher temperature can be used to explore possible pathways. Initially, the exploration runs are carried out at high Metropolis temperatures (1000 °K, for example), defining one or various ligand reaction coordinates. As emphasized by Mousseau et al., the Metropolis temperature does not correspond to a real thermal bath, the effective temperature being significantly lower.[36] The ligand reaction coordinates are reduced to an increase or decrease of a distance between the ligand center of mass and a target(s) (a set of *x*, *y*, and *z* coordinates). Various parallel trajectories at a lower Metropolis temperature are then started at the initial point. The information from the different nodes is shared using the MPI communications protocol. Whenever any trajectory is significantly farther along the reaction coordinate than any of the other trajectories, the trailing trajectory is abandoned and restarted from the position of the leading trajectory. We should emphasize that the higher-temperature exploration runs only determine the ligand target, a criterion by which the different nodes spawn their geometries. They do not introduce any further bias in the protein dynamics. Thus, trajectories produced at any given temperature are entirely produced from the starting structure at the given temperature, with no memory of any

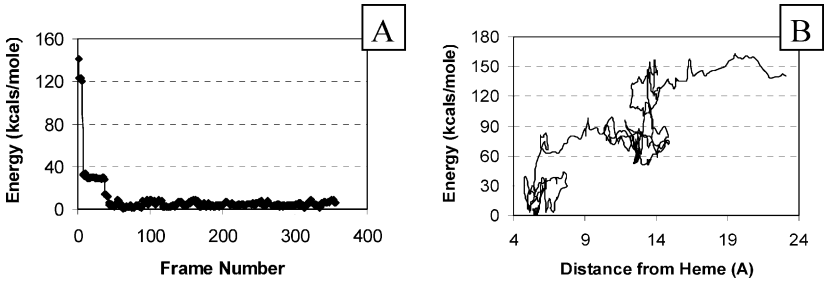

Figure 3. (A) Representative energy profile of an equilibration run. A stable minimum is found after about 50 steps and is maintained for another 300 steps without significant change. (B) Initial exploration run for P450cam run at 2000 K.

previous production at higher temperatures. Obviously, lower temperatures require longer production runs (or more processors). Using this parallel procedure, we have been able to run ligand escape simulations at room temperature for each of the systems studied.

Following recent studies[37,38] where we combined protein structure prediction algorithms with QM/MM techniques, the landscape exploration algorithm has been coupled with Qsite, a QM/MM program.[39] These techniques contain the elements necessary to properly describe the potential energy surfaces relevant to biochemical processes.[33] We have used single-point QM/MM calculations to obtain the initial atomic charges for the ligand and to account for the charge polarization of the ligand along its migration in the protein frame (using electrostatic potential charges). Single-point calculations are performed only after a large ligand rearrangement (every ~50 steps or longer), requiring on the order of 3−15 min of CPU time (3−4 min for camphor, ~15 for the fatty acid). To soften discontinuities in the potential surface, any new set of charges are scaled along six consecutive accepted sampling steps. We should emphasize that the overall charge is conserved. As seen in Figure 2, the charge update is always performed after the Metropolis test. Furthermore, an additional minimization follows any charge update; the initial and final energies for any Metropolis evaluation correspond to minimized structures with the same charge set. The QM region for all systems was reduced to the ligand. For the P450 and myoglobin systems, an additional initial QM calculation, including the heme group and the proximal residue in the QM region, provided the set of charges (fixed along the simulation) for the heme group. QM calculations are carried out using the B3LYP functional in combination with the 6-31G* basis set.

All classical calculations are done with an OPLS-AA description of the ligand and protein atoms and with an implicit SGB solvent model. As stated above, the atomic charges for the ligand (and the heme group in P450 and myoglobin) are obtained from QM/MM single-point calculations. The continuum solvation model has been extensively parametrized by Levy and co-workers to provide accurate predictions for a wide range of small molecule solvation free energies.[40] This parametrization includes the use of a new nonpolar term, which more accurately reproduces nonpolar solvation free energies for molecules of different shapes (e.g., cyclic and branched alkanes) as well as providing first-shell corrections to the generalized Born electrostatics. Finally, Friesner et al. have reparametrized key terms in order to improve conformational side-chain prediction as compared to experimental data.[31] A particular focus has been the ability to predict the formation of surface salt bridges, which are sensitive to the continuum solvation parameters for ionizable side chains. The unmodified SGB model overpredicts the formation of surface salt bridges by a substantial margin; reparametrization of the effective screened pair interaction term for these structures leads to much more accurate results. These improved Born models have been shown to give accurate forces and relative energies when compared with Poisson−Boltzmann and explicit solvent models.[10,41,42] In particular, reparametrized Born models perform well with sampling methods such as the one described in this work; the use of SGB with molecular dynamics requires a careful selection of a friction degree of freedom.[43−45]

PELE has been merged with PLOP and will be available free of charge for academic purposes from the UCSF site (http://francisco.compbio.ucsf.edu/~jacobson/plop_manual/plop_license.htm). To run the QM/MM options, it will require the installation of Qsite.[39]

3. Results and Discussion

Results focus mainly on camphor migration in cytochrome P450cam. The existence of crystallographic direct evidence of the ligand entry/escape channel makes P450cam an ideal test system. A summary of our results on CO and palmitate fatty acid diffusion on myoglobin and intestinal fatty-acid-binding protein, respectively, are also shown. These results indicate the possibilities of our methodology in treating small (CO) as well as complex (palmitate) ligands. More extensive results on these two systems will be presented elsewhere.

3.1. Ligand Migration on P450cam. The first step before launching an exploration run is to obtain an equilibrated local minimum. Such a structure is obtained by adjusting the local perturbation so that it always performs a random update of the direction search. Additionally, the steered cycle is reduced to <5. Thus, there is no successful search direction, reducing the landscape exploration to a local refinement of a given initial conformation. Panel A in Figure 3 introduces the energy profile for the camphor equilibration run. Similar energy profiles are obtained for each of the other systems. As seen in Figure 3, a stable plateau is obtained after 50 iterations. These equilibration runs required about 10 CPU-hours on a single processor.

After the equilibration run, we performed a quick landscape exploration by using high Metropolis temperatures.

At this stage, we usually keep the steered cycle at ~20 and adapt the localized perturbation amplitude and the temperature until an acceptance rate of >50% is reached. The initial exploration is intended to find different possible pathways. From the analysis of the different pathways, we select the optimal reaction coordinate to run a multiprocessor production run at lower Metropolis temperatures. This initial exploration is performed on a single processor and has an associated computational cost of 10−20 CPU-hours. For the P450cam system, we have performed 10 different exploration paths with Metropolis temperatures ranging from 500 to 2000 K. All paths, even at 2000 °K, resulted in camphor leaving the active site in a path closely following the wire ligands shown in Figure 1. No alternate pathway was observed, and no other protein cavities, distant from the exit pathway, were visited. In the random steered molecular dynamics studies,[35] multiple pathways were obtained. One of them closely corresponds to the pathway observed by us (in agreement with the crystallographic wire). However, the large force introduced, necessary to observe the camphor migration, apparently affects the potential landscape and results in alternate pathways. In other cases, myoglobin, for example, this procedure does predict ligand diffusion to a diverse set of experimentally observed inner cavities (details follow).

Panel B in Figure 3 displays an overlay of a few exit snapshots along one of the 10 different trajectories and the energy associated with the ligand motion. As a result of the large temperature acceptance associated with the exploration run, the energy profile is significantly large. Once the pathway(s) of interest have been identified, reaction coordinates can be developed that track the progress along these pathways. For the case of camphor, only a single pathway was observed, and hence, a simple reaction coordinate was implemented, which tracks the distance between the camphor center of mass and the iron of the heme. Using the parallel version of the code, trajectories were conducted using 6−10 processors at Metropolis temperatures of 750, 500, 400, and 250 K. The difference between the energy profiles of the last two temperatures were reduced only to 1−2 kcal/mol. As described in the Methods section, each of these trajectories is independent. By running multiple temperature trajectories, one can follow the convergence of the potential energy surface, as seen in Figure 4. Additionally, a future objective is to build the density of states by means of an energy histogram[46−48] along different bids of the reaction coordinate by means of multiple temperature trajectories. Each of these refinement runs required 2−4 days of CPU time (for each processor).

An update of the atomic ligand charges by means of single-point QM/MM calculations did not introduce any significant change in the ligand escape pathway and its potential surface. The atomic charge standard deviations were in the 0.02−0.06 range, with no abrupt change in atomic charges. Interestingly, the oxygen atom on camphor is the heavy atom with less charge fluctuation (with a 0.026 standard deviation). One would expect these fluctuations to be even less important when employing a polarizable force field for the protein atoms. Future studies need to address the importance of ligand polarization on a larger diversity of ligand and ligation

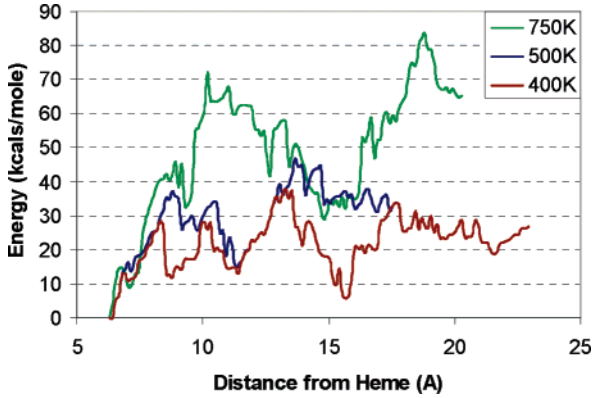

Figure 4. Energy profile associated with camphor exiting P450cam at three different temperatures of coordinated runs. Each successively lower temperature reduced the barriers associated with camphor leaving. At 400 K, energy barriers of ~35 kcal/mol were found (250 K gives essentially the same energy profile as 400 K).

pathways. A QM/MM description of the ligand atomic charges, for example, has recently been shown to considerably improve docking simulations.[49]

As seen in Figure 5, the exit pathway passes through the F−G loop, displacing both the F and G helix by a small amount, and is gated by the movements of several phenylalanine residues (F87, F98, and F193). Panel E in Figure 5 indicates a superposition of several camphor snapshots with the crystallographic wire along the exit pathway. The same panel also indicates the overall motion of the F and G helixes. Panels A−D highlight the molecular details of the phenylalanine motion along the camphor migration. The highest energy minimum, located at a camphor−heme distance of about 13 Å, is shown in panel B. At the present time, the landscape exploration algorithm only samples local minima. A work in progress is applying double-ended methods[50−52] for the characterization of the energy barriers. The ultimate goal is to obtain a network and a transition-state matrix characterizing the dynamics of the system in a fashion similar to recent studies on protein folding.[10,53−55] The highest energy minimum, shown in panel B of Figure 5, corresponds to the point in the trajectory where the camphor must pass between the two phenylalanine residues (F87 and F98). This large amplitude motion of the phenylalanine residues is also observed in the molecular dynamic studies by Wade et al.[35] Furthermore, the important role of these residues agrees with recent observations of the Batista's group at Yale University (personal communication). Their studies have observed a large coupling of these phenylalanine vibrations with the abstraction of a hydrogen atom from the camphor substrate by the putative active species, the oxy-ferryl heme species also known as compound I. Thus, the ligand diffusion (and possibly the enzymatic activity) seems to be mainly gated by the movements of these phenylalanine residues.

3.2. Ligand Migration on Other Systems: Myoglobin and Intestinal Fatty-Acid-Binding Protein. *Myoglobin.* The study of CO diffusion on globular proteins was the simplest and fastest application, requiring only temperatures of 300−400 K on a single processor, for producing successful migration pathways. In 20 exploration trajectories in the

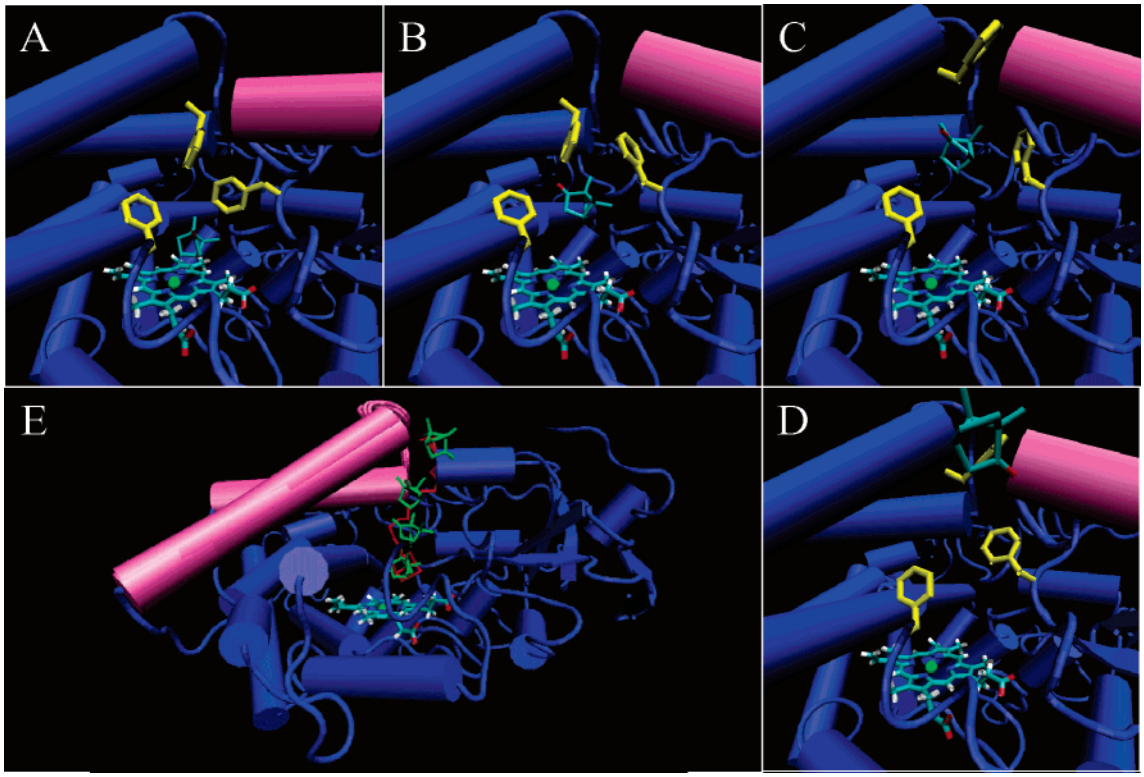

Figure 5. Exit pathway of camphor from P450cam as identified by PELE at room temperature (300 K). Panels A−D show four points along the exit pathway. The three phenylalanine residues experiencing large conformational changes, F87, F98, and F193 are shown in yellow. Panel B represents the structure at the largest energy barrier. Panel E displays an overall view of the pathway and is shown in green. The path identified by molecular wire studies is shown in red. Helices F and G as well as the F−G loop are shown in pink.

myoglobin system, 90% of the escape pathways involved pathways in the heme distal site. In most of the trajectories, the ligand visited different inner cavities (known as xenon cavities) before exiting the protein. The trace of the CO ligand is shown with a sphere representation for the C and O atoms in Figure 6. Such cavities have been clearly characterized previously.[56−60] The ligand exit, however, involved mainly the cavity connecting the distal site with the CD loop. Half of those pathways involve the opening of the distal histidine, a mechanism largely discussed in the literature.[61] The other half of the migration pathways involved the path shown in Figure 6 with an orange arrow. This migration pathway was observed in early calculations, using a time-dependent Hartree approximation, by Elber and Karplus.[59] This cavity contains three of the six phenylalanine residues present in myoglobin (shown with a white stick representation in Figure 6). This result is consistent with mutational studies by Scott et al. where mutation of these phenylalanine residues (and residues next to them) largely affected the entry and escape kinetics; mutation of all residues in the vicinity of the other cavities (Xe1, Xe4, etc.) did not show any significant effect on the ligand kinetics (see, for example, Figure 5 in Scott et al.[62]). Our results indicate that the CO stacking on top of the phenylalanine side chain participates actively in the migration pathway. Protein fluctuations in the CD loop region are essential for the escape of the ligand. The importance of the CD region on the globin heme-binding family has been corroborated by Fernandez-Alberti et al.[63] Using Gaussian network model analysis, the authors showed the existence of conserved low modes involving the CD helixes and the loop connecting them.

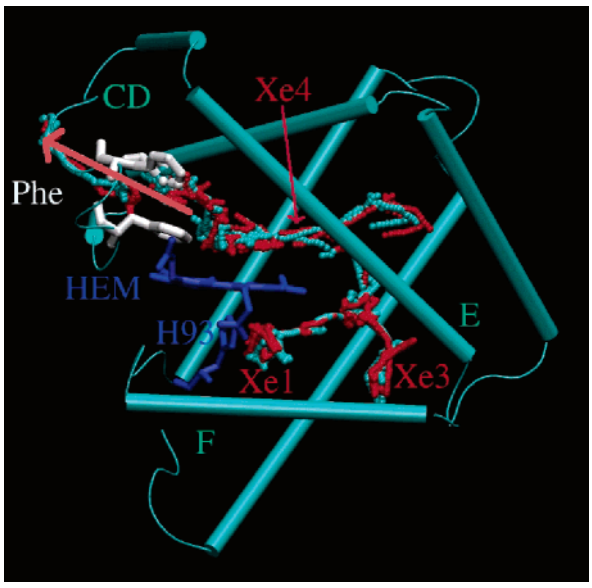

Figure 6. Carbon monoxide migration in myoglobin. The protein is shown in a cartoon representation with helices E and F and the loop CD being labeled. CO migration is shown in a sphere representation (red for oxygen and cyan for carbon), with some xenon cavities labeled in red. Phe 33, 43, and 46 are shown in white. The heme group and the proximal histidine, H93, are shown in dark blue. The exit pathway is underlined with a pink arrow.

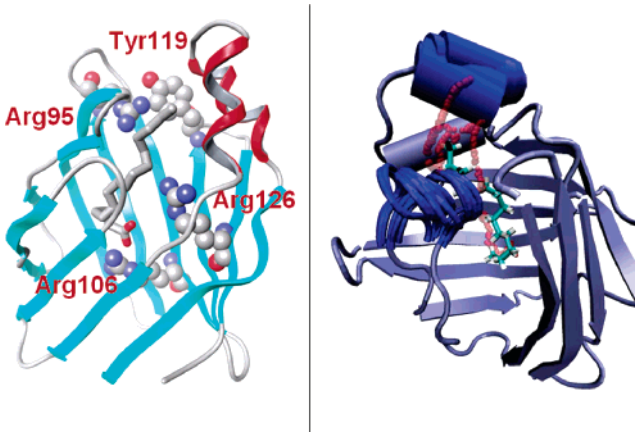

Figure 7. (Left) Important residues along the palmitate escape pathway. (Right) Main protein conformational changes obtained in our simulation along the exit pathway involving mainly helix αII and turns βC−βD and βE−βF.

Intestinal Fatty-Acid-Binding Protein. Palmitate diffusion in the intestinal fatty-acid-binding protein has been associated with the overall motion of the helix αII and turns between βC−βD and βE−βF.[64,65] Covalent pH-dependent cross-linking of these two secondary structure elements trapped palmitate. Figure 7 (right panel) indicates the main protein conformational changes associated with this process and a series of snapshots where several conformations of the fatty acid ligand are captured along the deligation pathway. The (long-range) protein conformational changes obtained in our simulation involve mainly helix αII and turns βC−βD and βE−βF, in great agreement with the experimental observations.[64,65] Our results not only confirm those experimental results but represent the first theoretical simulation where the escape of the palmitate ligand is observed. In particular, several polar residues in the interior of the protein appear to be crucial. These residues, shown in the left panel in Figure 7, induce a 180° turn on the fatty acid from its bound crystal structure, inducing an orientation where the carboxylic head of the ligand faces the solution. In particular, Arg126 appears to be a crucial residue in the proper alignment of the ligand toward the active site and in its initial rotation in the escape pathway. Mutational studies on Arg126 confirm the importance of this residue in the binding event.[66] The complex deligation pathway observed in our simulations is consistent with the experimental observation suggesting that lipid association is more complex than that described within a single equilibrium event.[65] Obtaining an exit pathway for this system involved 4 days of a single CPU time. This result evidences the possibilities of the method in sampling the energy landscape of a large ligand with several rotatable bonds.

4. Conclusions

We have introduced a novel landscape exploration technique based on Metropolis sampling and protein structure prediction algorithms. Ligand escape pathways in myoglobin, P450, and fatty-acid-binding protein have been presented and are consistent with experimental and theoretical data. Those processes, expanding from the microsecond to the millisecond time scale, require large protein conformational changes, which were captured by our methodology. Our approach, therefore, signifies an innovative methodology, capable of describing slow deligation events at a detailed atomic level. The landscape exploration output could be used as an input for several free-energy methodologies (umbrella sampling, etc). Present studies are focusing on obtaining such a description using only small modifications of the algorithms described herein. An accurate free-energy study, however, will necessarily involve a more rigorous solvation model. A mixed explicit/SGB solvation scheme has shown promising results.[67] Additionally, the algorithm proposed here cannot account for a detailed solvent−protein (or solvent−ligand) interaction, which might be essential in some particular protein and ligand dynamics. Further research in mixed solvation models is necessary to address this point. Current studies are also investigating the possibilities of landscape exploration into a wider scope of applications, including the protein response to a biochemical process as well as protein folding.

Acknowledgment. Financial support has been provided by startup package from Washington University School of Medicine. We are very grateful to Prof. Nathan Baker and Prof. Rohit Pappu for helpful discussions.

References

(1) Warshel, A.; Levitt, M. A. *J. Mol. Biol.* **1976**, *103*, 227−249.

(2) Karplus, M.; Petsko, G. A. *Nature* **1990**, *347*, 631−639.

(3) Tidor, B. *Curr. Biol.* **1997**, *7*, R525−R527.

(4) Kleinekathoefer, U.; Isralewitz, B.; Dittrich, M.; Schulten, K. *Biophys. J.* **2004**, *86*, 246A−246A.

(5) Zhou, R. H.; Huang, X. H.; Margulis, C. J.; Berne, B. J. *Science* **2004**, *305*, 1605−1609.

(6) Zhou, H. X.; Wlodek, S. T.; McCammon, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 9280−9283.

(7) Duan, Y.; Kollman, P. A. *Science* **1998**, *282*, 740−744.

(8) Daura, X.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. *J. Mol. Biol.* **1998**, *280*, 925−932.

(9) Nymeyer, H.; Gnanakaran, S.; Garcia, A. E. In *Numerical Computer Methods, Pt D*; Academic Press Inc.: San Diego, CA, 2004; Vol. 383, pp 119 ff.

(10) Andrec, M.; Felts, A. K.; Gallicchio, E.; Levy, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6801−6806.

(11) Elber, R.; Ghosh, A.; Cardenas, A. *Acc. Chem. Res.* **2002**, *35*, 396−403.

(12) Zhou, R. H.; Berne, B. J.; Germain, R. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 14931−14936.

(13) Schlick, T. *Biophys. J.* **2003**, *85*, 1−4.

(14) Hamelberg, D.; Mongan, J.; McCammon, J. A. *J. Chem. Phys.* **2004**, *120*, 11919−11929.

(15) Elber, R. *Curr. Opin. Struct. Biol.* **2005**, *15*, 151−156.

(16) Wales, D. J. *Energy Landscapes*; Cambridge University Press: Cambridge, MA, 2003.

(17) Wei, G. H.; Mousseau, N.; Derreumaux, P. *Proteins: Struct., Funct., Bioinf.* **2004**, *56*, 464−474.

(18) Madras, N.; Sokal, A. D. *J. Stat. Phys.* **1988**, *50*, 109−186.

(19) da Silva, R. A.; Degreve, L.; Caliri, A. *Biophys. J.* **2004**, *87*, 1567−1577.

(20) Brooks, B.; Karplus, M. *PNAS* **1983**, *80*, 6571−6575.

(21) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. *Proteins: Struct., Funct., Genet.* **1993**, *17*, 412−425.

(22) Bahar, I.; Atilgan, A. R.; Erman, B. *Folding Des.* **1997**, *2*, 173−181.

(23) Lovell, S. C.; Word, J. M.; Richardson, J. S.; Richardson, D. C. *Proteins: Struct., Funct., Genet.* **2000**, *40*, 389−408.

(24) Al-Lazikani, B.; Jung, J.; Xiang, Z. X.; Honig, B. *Curr. Opin. Chem. Biol.* **2001**, *5*, 51−56.

(25) Xiang, Z. X.; Honig, B. *J. Mol. Biol.* **2001**, *311*, 421−430.

(26) Dunbrack, R. L. *Curr. Opin. Struct. Biol.* **2002**, *12*, 431−440.

(27) Dwyer, M. A.; Looger, L. L.; Hellinga, H. W. *Science* **2004**, *304*, 1967−1971.

(28) Looger, L. L.; Dwyer, M. A.; Smith, J. J.; Hellinga, H. W. *Nature* **2003**, *423*, 185−190.

(29) Jacobson, M. P.; Friesner, R. A.; Xiang, Z. X.; Honig, B. *J. Mol. Biol.* **2002**, *320*, 597−608.

(30) Jacobson, M. P.; Pincus, D. L.; Rapp, C. S.; Honig, B.; Friesner, R. A. *Proteins* **2004**, *55*, 351−367.

(31) Ghosh, A.; Rapp, C. S.; Friesner, R. A. *J. Phys. Chem. B* **1998**, *102*, 10983−10990.

(32) Jacobson, M. P.; Kaminski, G. A.; Friesner, R. A.; Rapp, C. S. *J. Phys. Chem. B* **2002**, *106*, 11673−11680.

(33) Friesner, R. A.; Guallar, V. *Ann. Rev. Phys. Chem.* **2005**, *56*.

(34) Hays, A.-M. A.; Dunn, A. R.; Chiu, R.; Gray, H. B.; Stout, C. D.; Goodin, D. B. *J. Mol. Biol.* **2004**, *344*, 455−469.

(35) Ludemann, S. K.; Lounnas, V.; Wade, R. C. *J. Mol. Biol.* **2000**, *303*, 797−811.

(36) Wei, G. H.; Mousseau, N.; Derreumaux, P. *J. Chem. Phys.* **2002**, *117*, 11379−11387.

(37) Guallar, V.; Borrelli, K. W. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 3954−3959.

(38) Guallar, V.; Jacobson, M.; McDermott, A.; Friesner, R. A. *J. Mol. Biol.* **2004**, *337*, 227−239.

(39) *Qsite*; Schrödinger, Inc.: Portland, Oregon, 2001.

(40) Gallicchio, E.; Zhang, L. Y.; Levy, R. M. *J. Comput. Chem.* **2002**, *23*, 517−529.

(41) Alexey Onufriev, D. A. C.; Bashford, D. *J. Comput. Chem.* **2002**, *23*, 1297−1304.

(42) Wagoner, J.; Baker, N. A. *J. Comput. Chem.* **2004**, *25*, 1623−1629.

(43) Fan, H.; Mark, A. E.; Zhu, J.; Honig, B. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6760−6764.

(44) Zhu, J.; Alexov, E.; Honig, B. *J. Phys. Chem. B* **2005**, *109*, 3008−3022.

(45) Shen, T. Y.; Wong, C. F.; McCammon, J. A. *Biopolymers* **2003**, *70*, 252−259.

(46) McDonald, I. R.; Singer, K. *Discuss. Faraday Soc.* **1967**, *43*, 40−49.

(47) Mastny, E. A.; de Pablo, J. J. *J. Chem. Phys.* **2005**, *122*, 14109−6.

(48) Wang, F. G.; Landau, D. P. *Phys. Rev. Lett.* **2001**, *86*, 2050−2053.

(49) Cho, A. E.; Guallar, V.; Berne, B. J.; Friesner, R. A. *J. Comput. Chem.* **2005**, *26*, 915−931.

(50) Trygubenko, S. A.; Wales, D. J. *J. Chem. Phys.* **2004**, *120*, 2082−2094.

(51) Henkelman, G.; Jonsson, H. *J. Chem. Phys.* **2000**, *113*, 9978−9985.

(52) Czerminski, R.; Elber, R. *J. Chem. Phys.* **1990**, *92*, 5580−5601.

(53) Gallicchio, E.; Andrec, M.; Felts, A. K.; Levy, R. M. *J. Phys. Chem. B* **2005**, *109*, 6722−6731.

(54) Ozkan, S. B.; Dill, K. A.; Bahar, I. *Protein Sci.* **2002**, *11*, 1958−1970.

(55) Swope, W. C.; Pitera, J. W.; Suits, F.; Pitman, M.; Eleftheriou, M.; Fitch, B. G.; Germain, R. S.; Rayshubski, A.; Ward, T. J. C.; Zhestkov, Y.; Zhou, R. *J. Phys. Chem. B* **2004**, *108*, 6582−6594.

(56) Tilton, R. F. J.; Kuntz, I. D. J.; Petsko, G. A. *Biochemistry* **1974**, *19*, 9, 2849−2857.

(57) Teeter, M. M. *Protein Sci.* **2004**, *13*, 313−318.

(58) Ostermann, A.; Waschipky, R.; Parak, F. G.; Nienhaus, G. U. *Nature* **2000**, *404*, 205−208.

(59) Elber, R.; Karplus, M. *J. Am. Chem. Soc.* **1990**, *112*, 9161−9175.

(60) Bossa, C.; Anselmi, M.; Roccatano, D.; Amadei, A.; Vallone, B.; Brunori, M.; Di Nola, A. *Biophys. J.* **2004**, *86*, 414a−414a.

(61) Kottalam, J.; Case, D. A. *J. Am. Chem. Soc.* **1988**, *110*, 7690−7697.

(62) Scott, E. E.; Gibson, Q. H.; Olson, J. S. *J. Biol. Chem.* **2001**, *276*, 5177−5188.

(63) Maguid, S.; Fernandez-Alberti, S.; Ferrelli, L.; Echave, J. *Biophys. J.* **2005**, *89*, 3−13.

(64) Hodsdon, M. E.; Cistola, D. P. *Biochemistry* **1997**, *36*, 2278−2290.

(65) Ory, J.; Kane, C. D.; Simpson, M. A.; Banaszak, L. J.; Bernlohr, D. A. *J. Biol. Chem.* **1997**, *272*, 9793−9801.

(66) Richieri, G. V.; Ogata, R. T.; Kleinfeld, A. M. *Mol. Cell Biochem.* **1999**, *192*, 77−85.

(67) Yu, Z. Y.; Jacobson, M. P.; Josovitz, J.; Rapp, C. S.; Friesner, R. A. *J. Phys. Chem. B* 2004, ***108***, 6643−6654.

CT0501811