J|A|C|S
ARTICLES

Published on Web 12/29/2009

Mapping the Network of Pathways of CO Diffusion in Myoglobin

Luca Maragliano,*,† Grazia Cottone,‡ Giovanni Ciccotti,§ and Eric Vanden-Eijnden||

Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, Illinois 60637, Department of Physical and Astronomical Sciences and CNISM, University of Palermo, Palermo, Italy, Physics Department and CNISM Unit of Rome 1, University of Rome "La Sapienza", Rome, Italy, and Courant Institute of Mathematical Sciences, New York University, New York, New York 10012

Received July 14, 2009; E-mail: maraglia@uchicago.edu

Ⓦ **This paper contains enhanced objects available on the Internet at http://pubs.acs.org/jacs.**

Abstract: The pathways of diffusion of a CO molecule inside a myoglobin protein and toward the solvent are investigated. Specifically, the three-dimensional potential of mean force (PMF or free energy) of the CO molecule position inside the protein is calculated by using the single-sweep method in concert with fully resolved atomistic simulations in explicit solvent. The results are interpreted under the assumption that the diffusion of the ligand can be modeled as a navigation on the PMF in which the ligand hops between the PMF local minima following the minimum free energy paths (MFEPs) with rates set by the free energy barriers that need to be crossed. Here, all the local minima of the PMF, the MFEPs, and the barriers along them are calculated. The positions of the local minima are in good agreement with all the known binding cavities inside the protein, which indicates that these cavities may indeed serve as dynamical traps inside the protein and thereby influence the binding process. In addition, the MFEPs connecting the local PMF minima show a complicated network of possible pathways of exit of the dissociated CO starting from the primary docking site, in which the histidine gate is the closest exit from the binding site for the ligand but it is not the only possible one.

Introduction

Myoglobin (Mb) has been recognized as an ideal test case for understanding how dynamical aspects of protein function are related to its structure.[1,2] Mb is a small globular protein with well-known atomistic structure—it is the first protein whose structure was resolved by X-ray diffraction[3]—which is involved in oxygen transport and storage in various animal species, including humans. Mb binds small ligands such as O_2, CO, and NO at the iron atom located at the center of its active site, the heme. What makes the dynamical aspects of this binding process nontrivial is that the heme is buried in the protein interior, and by observing the molecular surface, no clear pathway to it is visible. This started a quest for the ligand route to (and from) the binding site. To date, this route remains unknown in Mb as well as other ligand-binding proteins.[4,5] For many years, the standard interpretation of the experimental results on Mb has been that there is only one pathway for ligand escape toward the solvent, via the so-called histidine gate located close to the heme. However, pioneering molecular dynamics (MD) simulation[6] in which the motion of the CO molecule was artificially accelerated showed that the ligand can diffuse in a nontrivial way inside the protein, visiting several hydrophobic cavities before finding its way to the solvent. These cavities are named Xe1−Xe4 (see Figure 1), since they were found to be xenon binding sites in a Xe-bound Mb study.[7] Years later, groundbreaking experiments determined the structure of Mb at different times after photolysis of CO, finding indeed the ligand molecule in some of these cavities.[8−10] Such results led to the idea that the existence of multiple cavities or packing defects inside the protein must play a functional role in the kinetics of ligand binding.[11,12] For instance, it has been speculated that the cavities could affect the rate of ligand delivery by trapping it

† University of Chicago.
‡ University of Palermo.
§ University of Rome "La Sapienza".
|| New York University.

(1) Frauenfelder, H.; McMahon, B. H.; Fenimore, P. W. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 8615–8617.
(2) Brunori, M.; Bourgeois, D.; Vallone, B. *J. Struct. Biol.* **2004**, *147*, 223–234.
(3) Kendrew, J.; Dickerson, R.; Strandberg, B.; Hart, R.; Davies, D.; Phillips, D.; Shore, V. *Nature* **1960**, *185*, 422–427.
(4) Cohen, J.; Olsen, K. W.; Schulten, K. *Methods Enzymol.* **2008**, *437*, 437–455.
(5) Baron, R.; Riley, C.; Chenprakhon, P.; Thotsaporn, K.; Winter, R. T.; Alfieri, A.; Forneris, F.; van Berkel, W. J. H.; Chaiyen, P.; Fraaije, M. W.; Mattevi, A.; McCammon, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 10603–10608.
(6) Elber, R.; Karplus, M. *J. Am. Chem. Soc.* **1990**, *112*, 9161–9175.
(7) Tilton, R. F.; Kuntz, I. D.; Petsko, G. A. *Biochemistry* **1984**, *23*, 2849–2857.
(8) Srajer, V.; Teng, T. Y.; Ursby, T.; Pradervand, C.; Ren, Z.; Adachi, S. I.; Schildkamp, W.; Bourgeois, D.; Wulff, M.; Moffat, K. *Science* **1996**, *274*, 1726–1729.
(9) Schotte, F.; Lim, M.; Jackson, T. A.; Smirnov, A. V.; Soman, J.; Olson, J. S.; Phillips, G. N.; Wulff, M.; Anfinrud, P. A. *Science* **2003**, *300*, 1944–1947.
(10) Ostermann, A.; Waschipky, R.; Parak, F. G.; Nienhaus, G. U. *Nature* **2000**, *404*, 205–208.

10.1021/ja905671x

transiently[11,13] and influence its binding affinity by acting as a storage facility for ligand molecules, thereby increasing their concentration.[14] Understanding the exact role of these packing defects is further complicated by the fact that they are not static structural features. Indeed, it has been observed in MD simulations[15,16] and experiments[17] that their sizes can vary over time. To settle these questions, it is important to locate the cavities inside Mb and map the paths connecting them with accuracy. Indeed, xenon-based experiments have raised concern on the extent to which the molecular structure under study is altered, as discussed in ref 18. These experiments are also likely to miss cavities that could accommodate smaller ligands than xenon itself, and they give little information about the possible paths connecting the cavities.

Computational techniques can help here. MD simulations, in particular, offer a way to investigate the pathways followed by the ligand which is more detailed than the one accessible via experiments. There is a difficulty, however, namely, that the time scales of ligand diffusion inside the protein go from tens to thousands of nanoseconds. Reaching such time scales is still a challenge for standard MD simulations. Only recently, few studies were able to produce trajectories following the diffusion of ligands on long time scales using extensive computations[19,20] and/or simplifying modeling assumptions.[5,15] Unfortunately, in the case of Mb, even the most extensive brute-force MD simulations to date done in ref 20, where about 7 μs of cumulative simulation time was generated, likely do not have enough statistics for accurate quantitative estimation of relevant observables.

In view of these difficulties, a valuable alternative to standard MD simulations is potential of mean force (PMF), or free energy, calculations.[21,22] Given a set of collective variables, which are some specific functions of the coordinates of the system under study, the PMF is by definition proportional to the logarithm of the equilibrium probability density function of these variables. Therefore, the PMF gives insight about the important states in the space of the collective variables by identifying those that are more likely to be populated and their relative lifetimes. If, in addition, the collective variables describe properly the reactive event (in the sense made precise in refs 23−25), then the PMF also gives information about the kinetics of the process. Indeed, the dynamics of the reaction can then be reduced to a continuous-time Markov chain, i.e., a random walk on a graph whose states are the local PMF minima, whose edges are the minimum free energy paths (MFEPs) connecting them, and in which the PMF barriers between states along the edges determine the transition rates between these states.

In this study, we investigate the process of CO diffusion in the Mb interior and toward the solvent at room temperature, using a fully atomistic model. We calculate the PMF landscape using the coordinates of the ligand molecule as collective variables, identify all possible metastable states as minima on this PMF, and compute the MFEPs connecting them. The MFEPs and the free energy barriers among them give information about the kinetics of the CO diffusion process from the binding site toward the solvent, under the assumption that the process can be described as the dynamical evolution of the ligand molecule on its PMF. As we will see a posteriori, this assumption is reasonable.

To compute the three-dimensional PMF map of the CO location inside Mb, we apply the single-sweep method introduced in ref 29 (see also refs 22, 30, and 31). In this method, the PMF landscape of a set of collective variables is first quickly explored by using the temperature-accelerated molecular dynamics (TAMD) technique.[32] Then it is efficiently reconstructed globally from a set of local calculations of its gradient at points chosen along the accelerated trajectory. Once the PMF is calculated, we compute MFEPs on this landscape with the string method[25,33,34] and thereby identify the CO migration paths. Our results show a network of possible pathways for the dissociated CO from its starting location in the cavity right above the heme (the distal pocket, DP) toward the Mb interior and the solvent. These paths connect the xenon cavities and a few more smaller cavities that were found also in recent studies. The shortest of these paths can be associated with the histidine gate and is also the only one without intermediate minima. We also find other possible gates for the CO to exit/enter Mb which are located far from the heme.

The idea of mapping a PMF to explore CO diffusion in Mb has been exploited before. Although standard PMF calculation methods such as umbrella sampling[35] are still too costly for the time and space scale of the process,[36] recently an interesting ad hoc method was developed in ref 37 in which the presence of the ligand is treated as a perturbation to the dynamics of the

(11) Brunori, M.; Gibson, Q. H. *EMBO Rep.* **2001**, *2*, 674–679.
(12) Teeter, M. *Protein Sci.* **2004**, *13*, 313–318.
(13) Frauenfelder, H.; McMahon, B. H.; Austin, R. H.; Chu, K.; Groves, J. T. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 2370–2374.
(14) Chu, K.; Vojtchovsky, J.; McMahon, B. H.; Sweet, R. M.; Berendzen, J.; Schlichting, I. *Nature* **2000**, *403*, 921–923.
(15) Bossa, C.; Anselmi, M.; Roccatano, D.; Amadei, A.; Vallone, B.; Brunori, M.; Di Nola, A. *Biophys. J.* **2004**, *86*, 3855–3862.
(16) Scorciapino, M. A.; Robertazzi, A.; Casu, M.; Ruggerone, P.; Ceccarelli, M. *J. Am. Chem. Soc.* **2009**, *131*, 11825–11832.
(17) Tomita, A.; Sato, T.; Ichiyanagi, K.; Nozawa, S.; Ichikawa, H.; Chollet, M.; Kawai, F.; Park, S. Y.; Tsuduki, T.; Yamato, T.; Koshihara, S. y.; Adachi, S. i. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 2612–2616.
(18) Anedda, R.; Era, B.; Casu, M.; Fais, A.; Ceccarelli, M.; Corda, M.; Ruggerone, P. *J. Phys. Chem. B* **2008**, *112*, 15856–15866.
(19) Elber, R.; Gibson, Q. H. *J. Phys. Chem. B* **2008**, *112*, 6147–6154.
(20) Ruscio, J. Z.; Kumar, D.; Shukla, M.; Prisant, M. G.; Murali, T. M.; Onufriev, A. V. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 9204–9209.
(21) Roux, B. *Comput. Phys. Commun.* **1995**, *91*, 275–282.
(22) Vanden-Eijnden, E. *J. Comput. Chem.* **2009**, *30*, 1737–1747.
(23) Geissler, P. L.; Dellago, C.; Chandler, D. *J. Phys. Chem. B* **1999**, *103*, 3706–3710.
(24) Vanden-Eijnden, E. In *Computer Simulations in Condensed Matter Systems: From Materials to Chemical Biology*; Ferrario, M., Ciccotti, G., Binder, K., Eds.; Springer: New York, 2007; Vol. 1, pp 453−493.
(25) Maragliano, L.; Fischer, A.; Vanden-Eijnden, E.; Ciccotti, G. *J. Chem. Phys.* **2006**, *125*, 024106−024115.
(26) Vitkup, D.; Petsko, G. A.; Karplus, M. *Nat. Struct. Biol.* **1997**, *4*, 202–208.
(27) Meuwly, M.; Becker, O. M.; Stote, R.; Karplus, M. *Biophys. Chem.* **2002**, *98*, 183–207.
(28) Hummer, G.; Schotte, F.; Anfinrud, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 15330–15334.
(29) Maragliano, L.; Vanden-Eijnden, E. *J. Chem. Phys.* **2008**, *128*, 184110–184120.
(30) Monteferrante, M.; Bonella, S.; Meloni, S.; Vanden-Eijnden, E.; Ciccotti, G. *Sci. Model. Simul.* **2008**, *15*, 187–206.
(31) Monteferrante, M.; Bonella, S.; Meloni, S.; Ciccotti, G. *Mol. Simul.* **2009**, *35*, 1116–1129.
(32) Maragliano, L.; Vanden-Eijnden, E. *Chem. Phys. Lett.* **2006**, *426*, 168–175.
(33) E, W.; Ren, W.; Vanden-Eijnden, E. *Phys. Rev. B* **2002**, *66*, 052301−052305.
(34) E, W.; Ren, W.; Vanden-Eijnden, E. *J. Chem. Phys.* **2007**, *126*, 164103–164111.
(35) Torrie, G. M.; Valleau, J. P. *Chem. Phys. Lett.* **1974**, *28*, 578–581.
(36) Banushkina, P.; Meuwly, M. *J. Phys. Chem. B* **2005**, *109*, 16911–16917.
(37) Cohen, J.; Arkhipova, A.; Braun, R.; Schulten, K. *Biophys. J.* **2006**, *91*, 1844–1857.

protein. To make the calculation feasible, however, electrostatic interactions between the ligand and the protein had to be neglected. In another study,[38] the PMF was obtained from a long metadynamics[39] trajectory. With respect to these techniques, the single-sweep method has the advantage that it keeps the PMF exploration and its reconstruction separate and bases the reconstruction not on direct sampling of the PMF but rather on interpolation from the mean forces computed locally. This is especially suitable in problems such as CO escape from Mb, where the volume of the product state is so large that it is too demanding to wait for the ligand to come back to the reactant, as is in principle needed in sampling strategies. In addition, the single-sweep method is based on a set of completely independent simulations that can be trivially distributed over multiple processing nodes. These features make the single-sweep method efficient enough to be used in conjunction with fully atomistic models in explicit solvent as we do here.

Theory: Potential of Mean Force Reconstruction with the Single-Sweep Method

The single-sweep method involves three stages. First, a TAMD[32] simulation is performed to explore the unknown PMF landscape in some collective variables. Second, the mean force is computed locally at points chosen along this trajectory. Third, the PMF is reconstructed globally from these mean forces, using a variational procedure and a representation in terms of radial basis functions (RBFs). Most of the computational effort of the method is required at the stage of mean force calculations. However, since these calculations are completely independent of each other, they can be easily distributed on multiple processing nodes.

The three stages of the single-sweep method are briefly summarized below. For more detailed discussions, we refer the interested reader to refs 22, 29, and 32. We consider a molecular system with n degrees of freedom whose position in configuration space $\Omega \subseteq \mathbb{R}^n$ will be denoted by x. We also introduce a set of N collective variables $\theta(x) = (\theta_1(x), ..., \theta_N(x))$, functions of the coordinates x (typical choices are distances, angles, or coordination numbers). If $V(x)$ denotes the potential energy of the system and $1/\beta = k_B T$, where k_B is the Boltzmann constant and T the system's temperature, the PMF $A(z)$ in the variables $\theta(x)$ is defined as

$$A(z) = -\beta^{-1} \log \int_\Omega e^{-\beta V(x)} \delta(\theta(x) - z)\, dx \qquad (1)$$

where $z \in \mathbb{R}^N$. Equivalently, the probability density function (PDF) of the variables $\theta(x)$ is, up to a proportionality constant, $e^{-\beta A(z)}$. A quantity that will have an important role in what follows is the negative gradient of the PMF, the mean force $f(z) = -\nabla_z A(z)$. At variance with the PMF, this is a quantity which can be computed locally at point z via calculation of an expectation.[40,41]

TAMD for Exploring the Unknown PMF Surface. TAMD was introduced in ref 32 as a method to efficiently explore the PMF landscape of a set of collective variables. The starting point is to consider an extended system where the z variables are taken as dynamical ones, coupled to the original ones via the extended potential $U_\kappa(x, z) = V(x) + {}^1/_2\kappa|\theta(x) - z|^2$, where $\kappa > 0$ is an adjustable parameter. The extended system is evolved according to

$$\begin{cases} \mathbf{m}\ddot{x} = -\nabla_x V(x) - \kappa \sum_{\alpha=1}^{N} (\theta_\alpha(x) - z_\alpha)\nabla_x \theta_\alpha(x) \\ \qquad\qquad + \text{thermostat terms at } \beta^{-1} \\ \gamma \dot{z} = \kappa(\theta(x) - z) + \sqrt{2\gamma\bar{\beta}^{-1}}\eta(t) \end{cases} \qquad (2)$$

where $\mathbf{m}$ is the mass matrix of the original system and $\eta(t)$ is a Gaussian process with mean 0 and covariance $\langle \eta_\alpha(t)\, \eta_{\alpha'}(t') \rangle = \delta_{\alpha\alpha'}\, \delta(t - t')$. The evolution equation for $z(t)$ in eq 2 involves a friction coefficient $\gamma > 0$ and an artificial temperature $1/\bar{\beta}$ ($\neq 1/\beta$). It was shown in ref 32 that, by adjusting the parameter κ so that $z(t) \approx \theta(x(t))$, and the friction coefficient γ so that z moves slower than x, one can generate a trajectory $z(t)$ which effectively moves at the *artificial* temperature $1/\bar{\beta}$ on the PMF computed at the *physical* temperature $1/\beta$. Then, by taking $1/\bar{\beta} > 1/\beta$, the $z(t)$ trajectory will visit rapidly the relevant regions of the z-space where the PMF is within $1/\bar{\beta}$ of its minimum, even when these regions are separated by barriers which the physical system would take a long time to cross at the physical temperature $1/\beta$. In other words, eq 2 can be used to rapidly sweep through z-space.

Mean Force Calculations. Along the accelerated $z(t)$ trajectory, we select a set of points (also called *centers*) at which to compute the mean force. A possible, but not unique, way of doing this is to start with $z_1 = z(0)$ and then deposit a new center z_k along $z(t)$ each time $z(t)$ reaches a point which is more than a prescribed distance d away from all the previous centers. To compute the mean forces f_k, at each of the centers z_k we launch a simulation of eq 2 with $z(t) = z_k$ fixed and compute

$$f_k = \frac{1}{T}\int_0^T \kappa(\theta(x(t)) - z_k)\, dt \qquad (3)$$

The calculation of each one of these time averages is independent of all the others. For this reason, they can be trivially distributed on computing clusters. Equation 3 has the advantage of being simple, but it introduces an error due to the finiteness of κ. This error can be eliminated by using constrained simulations and using the blue-moon estimator for the mean force.[40,41]

Free Energy Representation via RBFs. Once the centers $z_1, ..., z_K$ have been deposited and the estimates $f_1, ..., f_K$ of the mean force at these centers have been obtained, we use these data to reconstruct the PMF $A(z)$ globally by using an RBF representation for $A(z)$ with centers at $z_1, ..., z_K$:[42,43]

$$A(z) = \sum_{k=1}^{K} a_k \varphi_\sigma(|z - z_k|) + C \qquad (4)$$

where C is a constant used to adjust the overall height of $A(z)$ but is otherwise irrelevant, $|...|$ denotes the Euclidean norm in $\mathbb{R}^N$, and $\varphi_\sigma(u)$ is an RBF with shape parameter σ; a convenient choice is to use the Gaussian packet $\varphi_\sigma(u) = \exp(-u^2/2\sigma^2)$, though other RBFs can be used as well.[29] In eq 4, the heights a_k and the RBF width $\sigma > 0$ are adjustable parameters which

(38) Ceccarelli, M.; Anedda, R.; Casu, M.; Ruggerone, P. *Proteins: Struct., Funct., Genet.* **2008**, *71*, 1231–1236.
(39) Bussi, G.; Laio, A.; Parrinello, M. *Phys. Rev. Lett.* **2006**, *96*, 090601.
(40) Carter, E. A.; Ciccotti, G.; Hynes, J. T.; Kapral, R. *Chem. Phys. Lett.* **1989**, *156*, 472–477.
(41) Ciccotti, G.; Kapral, R.; Vanden-Eijnden, E. *ChemPhysChem* **2005**, *6*, 1809–1814.
(42) Schaback, R. *Adv. Comput. Math.* **1995**, *3*, 251–264.
(43) Buhmann, M. D. *Radial Basis Functions: Theory and Implementations*; Cambridge University Press: Cambridge, U.K., 2003.

we determine by minimizing over a_k and σ the following objective function, which measures the discrepancy between the negative gradient of eq 4 at the centers z_k and $-\nabla_z A(z_k) = -\sum_{k'=1}^{K} a_{k'} \nabla_z \varphi_\sigma(|z_k - z_{k'}|)$ and the mean force f_k estimated at these centers:

$$E(a,\sigma) = \sum_{k=1}^{K} \left| \sum_{k'=1}^{K} a_{k'} \nabla_z \varphi_\sigma(|z_k - z_{k'}|) + f_k \right|^2 \qquad (5)$$

The minimization of $E(a,\sigma)$ in eq 5 can be performed as follows. For fixed σ, the a_k* minimizing $E(a,\sigma)$ solve a linear algebraic system whose coefficients have explicit expressions in terms of $\nabla_z \varphi_\sigma(|z_k - z_{k'}|)$ and f_k. The parameter σ can then be determined by minimizing $E(a^*(\sigma),\sigma)$. Overall, the procedure is simple and inexpensive since the determination of a_k* at fixed σ is straightforward and cheap and can be easily repeated to perform the one-dimensional minimization over σ. The relative residual of eq 5 at the minimum also gives an estimate of the accuracy of the result.

The RBF representation in eq 4 appears as a natural and convenient choice for the PMF for several reasons. First, in these representations the centers z_k do not have to lie on a regular grid, which permits use of mean force data sparsely collected. Second, RBFs can be used in any dimension. Third, these representations have very good convergence properties; i.e., a small number of centers gives an accurate representation of $A(z)$.

Computational Methods

Simulation Setup. The initial coordinates for the protein were obtained from the crystallographic data of sperm whale myoglobin (PDB code 2MB5).[44] A configuration of the protein equilibrated in water was taken from previous work.[45] All hydrogen atoms were explicitly included, and histidines 24, 64, and 93 were single-protonated at N_δ. All water molecules were removed, and the protein was rotated so as to align the eigenvectors of the mass-weighted covariance matrix with the Cartesian frame. This orientation was kept fixed during all simulations via the implementation of holonomic constraints (see the next section for the motivation and the Supporting Information for details on the constraint). A rectangular simulation cell was built around the molecule and filled with water by using a pre-equilibrated water configuration as the building block. Water molecules whose atoms were within 1.8 Å distance from any protein atom were removed, and nine of them were replaced with chlorine ions to ensure neutrality of the box. No water molecule was inserted inside the protein. The final size of the system is 20 357 atoms (1 Mb molecule, 1 CO molecule, 5941 water molecules, and 9 ions). All MD simulations were performed using the DL-PROTEIN package,[46] a modified version of DL-POLY.[47] The CHARMM22 force field has been used,[48] with the TIP3P model for the water molecules.[49] Periodic boundary conditions were used,[50] and van der Waals interactions were cut off beyond a distance of 9 Å. Electrostatic interactions were calculated with the Ewald sums using the SPME method;[51] the direct sum cutoff was 9 Å, the Ewald α parameter was set to 0.358, and 8 order cubic splines were used for interpolation on a grid of 64^3 points. All chemical bonds were kept fixed by using the SHAKE algorithm.[52] The equations of motion were integrated with the velocity Verlet scheme,[50] with a time step of 1 fs. The system was equilibrated at 300 K over 200 ps at constant pressure ($P = 1$ bar) and temperature ($T = 300$ K) by using the Nosé−Hoover technique,[53] with coupling constants 0.4 and 5 ps for the thermostat and the barostat, respectively.

In this work, we model CO dissociation as done in previous computational studies.[15,19,20,26−28] The associated and dissociated states are described by two different potential forms. The switch from the first to the second is realized instantaneously and permanently. The system is not allowed to reassociate. Thus, given a configuration extracted from the equilibration simulation of the associated system, we delete the C−Fe bond, and we change the potential energy for the heme internal interactions from carboxy to deoxy. The parameters required to describe the heme−CO interactions and configuration in the deoxy state were taken from the literature.[27]

The system in the deoxy state was equilibrated at 300 K over 250 ps at constant pressure ($P = 1$ bar) and temperature ($T = 300$ K) by using the Nosé−Hoover technique,[53] with coupling constants 0.4 and 5 ps for the thermostat and the barostat, respectively. During this trajectory, the CO molecule never left the DP. After this stage of equilibration, we started the TAMD simulations.

TAMD Simulations. We picked as collective variables the Cartesian coordinates of the center of mass of the CO molecule. The advantage of such variables is that they are very simple, and their PMF map can be readily superposed on the atomic structure of the molecule for visualization. On the other hand, they are not invariant with respect to rigid rotations and translations of the Mb molecule that might possibly occur. For this reason, all calculations presented here were performed at fixed orientation and center of mass of the protein. This condition was realized via the definition and implementation of a holonomic constraint (details are reported in the Supporting Information).

Once the CO molecule reaches the solvent, the probability that it re-enters Mb is low. Hence, to obtain an extensive exploration of the Mb interior, we performed different, independent TAMD simulations, which we stopped when the CO was in the solvent. About 10 independent TAMD trajectories were performed. We stress again that the TAMD simulations are used here for the purpose of quick exploration, and not sampling, of the PMF surface. All trajectories started with the same configuration, but with different momenta (and different noise in the collective variables' equations of motion). The parameters in all TAMD calculations were $\kappa = 200$ kcal/(mol Å^2) and $\gamma^{-1} = 0.004$ ps. The Nosé−Hoover equations for the NVT ensemble[54−56] were used for the coordinates and momenta of Mb and water molecules, with time step $\Delta t = 1$ fs, and coupling term with the thermostat $\tau = 0.4$ ps. Other simulation settings were as described in the section "Simulation Setup". To check that the accelerated motion of the CO molecule does not alter the structure of the protein, we monitored the root mean square displacement (rmsd) of Mb backbone atoms from the starting configurations, which was around 1 Å for all trajectories, never exceeding 1.5 Å.

In all TAMD simulations the CO molecule migrated from the distal pocket to the protein interior and then into the solvent. The length of the TAMD trajectories before exit was always on the order of 100 ps, while only two reached 200 ps. During its migration inside the protein, the ligand visited all of the xenon

(44) Cheng, X.; Schoenborn, B. P. *Acta Crystallogr., B* **1990**, *46*, 195.
(45) Cottone, G.; Cordone, L.; Ciccotti, G. *Biophys. J.* **2001**, *80*, 931–938.
(46) Melchionna, S.; Cozzini, S. *The DLPROTEIN 2.1 Manual*; University of Rome "La Sapienza": Rome, Italy, 2001.
(47) Todorov, I.; Smith, W. *Philos. Trans. R. Soc. London, A* **2004**, *362*, 1835–1852.
(48) MacKerell, A. D.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
(49) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
(50) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford Science: Oxford, U.K., 1987.
(51) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
(52) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. *J. Comput. Phys.* **1977**, *23*, 327–341.
(53) Melchionna, S.; Ciccotti, G.; Holian, B. L. *Mol. Phys.* **1993**, *78*, 533–544.
(54) Nosé, S. *Mol. Phys.* **1984**, *52*, 255–268.
(55) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695–1697.
(56) Tuckerman, M. E.; Liu, Y.; Ciccotti, G.; Martyna, G. J. *J. Chem. Phys.* **2001**, *115*, 1678–1702.

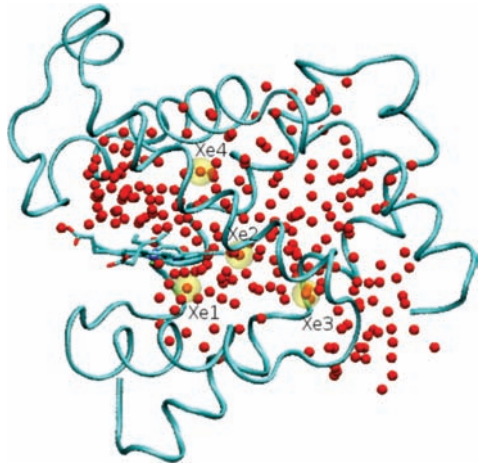

Figure 1. Set of 239 positions of the CO center of mass (red spheres) inside Mb, obtained from TAMD simulations, and used as locations to compute the mean forces. The protein's backbone is represented as ribbons, the heme residue is represented as sticks, and the locations of the xenon cavities are represented as yellow spheres.

cavities and a few more cavities also found in recent studies.[15,17,37] In two of the simulations, the ligand exited to the solvent directly from the distal pocket, passing close to histidine 64 (H64), i.e., through the so-called histidine gate.

Mean Force Calculations and PMF Reconstruction. By joining together all the TAMD trajectories, we obtained a list of values for the CO center of mass coordinates inside Mb. To reconstruct the PMF map, a set of values for the mean forces is needed, for example, computed at centers extracted from the list. As already done in previous studies,[29−31] we used a distance criterion to choose the centers; i.e., we took z_1 as the first element in the list and then ran along the list and extracted a new center each time an element was more than a prescribed distance d away from all the previous centers. By setting $d = 2.5$ Å, we obtained 239 centers. Figure 1 shows as red spheres the locations inside Mb of the 239 centers; the protein backbone is represented as ribbons, the heme residue is represented as sticks, and the locations of the xenon binding cavities are represented as yellow spheres. This and all other images in the paper were produced using the VMD program.[57]

Mean forces were computed at these centers by simulating eq 2 with $z(t) = z_k$ fixed and $\kappa = 200$ kcal/(mol Å^2) and using eq 3. All other simulation settings were as described in the section "TAMD Simulations". Each mean force calculation lasted 500 ps, at which length we observed convergence of the mean forces and of the PMF barriers after reconstruction. The cumulative simulation time is hence 120 ns, although we stress again that, all mean force calculations being independent of each other, they were distributed on different processing nodes. The PMF reconstruction was performed using Gaussian RBFs. The value of the optimal σ was 2.97 Å, at which the relative residual, defined as $[E(a,\sigma)/\Sigma_k|f_k|^2]^{1/2}$, is 0.64.

MFEPs as Ligand Migration Paths. Once the PMF surface is known, pathways for CO diffusion inside Mb are identified as MFEPs on the surface.[25] An MFEP is defined as the curve whose tangent is always parallel to $\mathbf{M}\nabla A(z)$, where $\mathbf{M}$ is a metric tensor[25] which, for the case of linear collective variables chosen here, is constant and diagonal. Hence, in the present case, the MFEPs coincide with the steepest descent paths from saddle points on the PMF surface. Since we have obtained an analytical approximation of this surface, such paths can be efficiently computed with the zero temperature string (ZTS) method.[33,34] Given an initial guess for a curve on the PMF surface, the ZTS method finds the closest MFEP by moving a discrete set of points on the curve by steepest descent on the PMF landscape, at the same time keeping the points at constant distance from each other. The procedure requires the first derivatives of the underlying PMF surface, which can be easily obtained from eq 4, with the optimal a_k set and σ determined by the minimization procedure. Note that no more MD calculations are required at this stage.

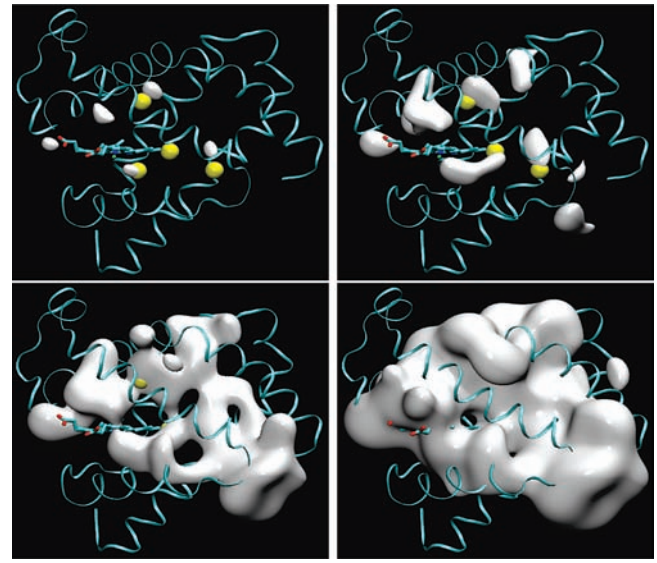

Figure 2. Four isosurfaces of the three-dimensional PMF map of the CO center of mass inside Mb, superposed on the protein structure. Energy levels are, from left to right and top to bottom, 1.5, 3.5, 5.5, and 8.5 kcal/mol with respect to the global minimum in the Xe4 cavity. The protein's backbone is represented as ribbons, the heme residue is represented as sticks, and the locations of the xenon binding cavities are represented as yellow spheres.

Results and Discussion

PMF Map for CO Diffusion in Mb. Figure 2 shows four isosurfaces of the three-dimensional PMF map as obtained from our calculations. For illustrative purposes, we superpose the isosurfaces on a representative structure of Mb, extracted from one of our mean force simulations, and hence aligned consistently with the PMF map. A movie with a 360° view of the map at different energy levels (from 0.5 to 10 kcal/mol) is available online as a Web-enhanced object. The global minimum of the map is in correspondence with the Xe4 cavity. Energy levels in Figure 2 are, from left to right and top to bottom, 1.5, 3.5, 5.5, and 8.5 kcal/mol with respect to the global minimum. The overall shape of the isosurfaces of our map is in good qualitative agreement with similar maps computed in refs 37 and 38 and with the dissociated CO trajectories computed in refs 6, 15, 19, and 20.

Other local minima are in the DP and in the Xe1, Xe2, and Xe3 cavities. The energy of the DP minimum is 0.65 kcal/mol higher than that of Xe4. Xe1 and Xe3 are almost at the same energy as the DP, and Xe2 is 1.75 kcal/mol higher than Xe4. To check the quality of our reconstruction in the DP, we ran an unbiased simulation of the dissociated Mb:CO system with the CO in the distal pocket, with all simulation settings as described in the section "Simulation Setup". After about 2 ns, the distributions of the CO center of mass coordinates are monomodal, with peaks at $x = 6.25$, $y = 2.08$, and $z = -0.49$, to be compared with the location of the local minimum in the DP of the PMF map, $x = 7.85$, $y = 1.86$, $z = 0.04$.

(57) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38.

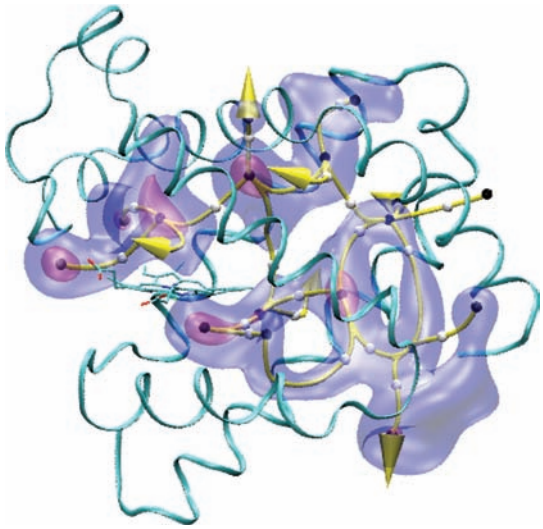

Figure 3. CO migration pathways inside Mb. The yellow curves are MFEPs computed with the string method on the CO PMF map and identify possible CO routes. Two isosurfaces of the PMF map are shown (red, 2.0 kcal/mol; blue, 5.0 kcal/mol; with respect to Xe4). The white and black spheres represent, respectively, the locations of the energy barriers and the local minima along the pathways. The yellow arrows represent the locations of the CO exits to the solvent as observed in the TAMD simulations. The protein's backbone is represented as ribbons and the heme as sticks.

The locations of the minima in our map well correlate with results from previous experimental and theoretical work. Indeed, the CO molecule was found in the Xe1 cavity by time-resolved crystallography, in native Mb, in the pioneering study of ref 8 and in the L29W Mb mutant in ref 9. Other crystallographic studies of the photolyzed L29W[10] mutant found the CO trapped in the Xe4 site. In a more recent study on Mb crystals at low temperature[17] the CO was found to populate, at different times after photolysis, all xenon binding cavities. Previous theoretical studies also found CO PMF minima (or density peaks) in xenon cavities and the DP.[6,15,19,20,28,37,38,58,59]

In addition to the xenon binding sites and DP, our map shows other features that are similar to those obtained in other computational studies.[15,20,37] In particular, we observe two local minima above Xe3 that are located in correspondence to cavities visited by the dissociated CO trajectory in refs 15 and 20 (they were called Ph1 and Ph2 in ref 15) and were also present in the map of ref 37. The ring-shaped structure above Xe3 is very similar to that obtained in ref 37. The shape of the isosurfaces in the DP region matches the plot of the positions occupied by dissociated CO in the extensive simulations of ref 20, where the ligand also visited the largest of the two minima we find in the proximity of the heme, close to the protein surface. Finally, the region above Xe4 toward the solvent was visited by the dissociated trajectories in refs 15, 19, and 20.

Pathways of CO Migration inside Mb. To accurately locate pathways and compute energy barriers for CO migration inside Mb, we use the string method to calculate MFEPs (i.e., the curves whose tangent is always parallel to $\nabla A(z)$, see the section "MFEPs as Ligand Migration Paths") on the reconstructed PMF surface. These MFEPs are identified as migration pathways for CO inside Mb. Figure 3 shows our results. MFEPs are shown as yellow curves. Two isosurfaces of the PMF map are represented (red, 2 kcal/mol; blue, 5 kcal/mol; with respect to Xe4). White and black spheres represent, respectively, the locations of energy barriers and local minima along the pathways. Starting from the DP, a network of possible pathways is accessible to the dissociated CO molecule. The locations of the pathways we found is in excellent agreement with previous studies. In the massive simulations of dissociated CO in ref 20, nine different gates were identified for the ligand to exit/enter Mb. Each one of these gates is connected with one of our pathways. Moreover, most of the trajectories followed by the CO molecules in ref 20 are along some of our pathways.

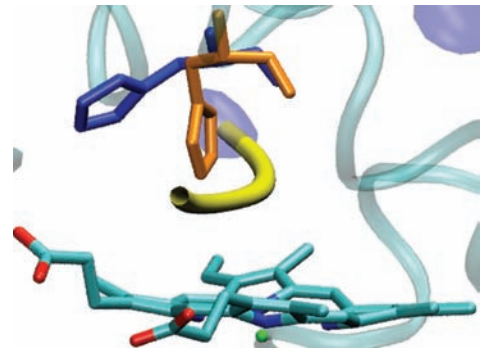

Figure 4. Illustration of the histidine gate as results from our calculations. The yellow curve is the MFEP. Orange and blue sticks represent, respectively, the side chain of H64 in the crystal of Mb with CO bound and in a deoxy-Mb configuration from our simulations.

On the experimental side, many of the residues that were found by random mutagenesis[60] to affect the ligand-binding kinetics lie along our paths. In particular, all of the energy barriers we find are close to at least one of the residues identified in ref 60. This result, as already observed in refs 37 and 20, is an important step toward a more atomistically based interpretation of the mutagenesis results. The largest cluster of such residues is around the DP, where indeed we observe three different possible escape routes toward the solvent, all of them in proximity of at least one of the kinetically relevant residues. Among these paths is the one toward the so-called histidine gate, which for many years has been considered the only one possible for CO to enter/exit Mb.[61] From our results, this is the shortest path connecting the DP and the solvent. It is also the only direct one, i.e., without intermediate minima along it. This might reflect the importance of this path in the escape process. Figure 4 shows a detailed view of the histidine gate path from our calculations. The yellow curve is the MFEP. In orange, we show the position of the H64 side chain in the crystal of Mb with CO bound, while in blue we show its position as we find it in the dissociated system configurations. It can be seen that, in the deoxy Mb configuration, the rotation of the side chain toward the solvent creates room for the pathway, opening the gate. This same mechanism has been proposed by several authors.[61,62]

It is important to compare the different PMF barriers that the CO molecule has to cross when moving along the paths. Figure 5 shows again the migration pathways of Figure 3, this time colored according to the value of the PMF along them. Energy barriers and local minima are represented as spheres,

(58) Kiyota, Y.; Hiraoka, R.; Yoshida, N.; Maruyama, Y.; Imai, T.; Hirata, F. *J. Am. Chem. Soc.* **2009**, *131*, 3852–3853.
(59) Nutt, D. R.; Meuwly, M. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 5998–6002.
(60) Huang, X.; Boxer, S. G. *Nat. Struct. Biol.* **1994**, *1*, 226–229.
(61) Scott, E. E.; Gibson, Q. H.; Olson, J. S. *J. Biol. Chem.* **2001**, *276*, 5177–5188.
(62) Perutz, M. F. *Trends Biochem. Sci.* **1989**, *14*, 42–44.

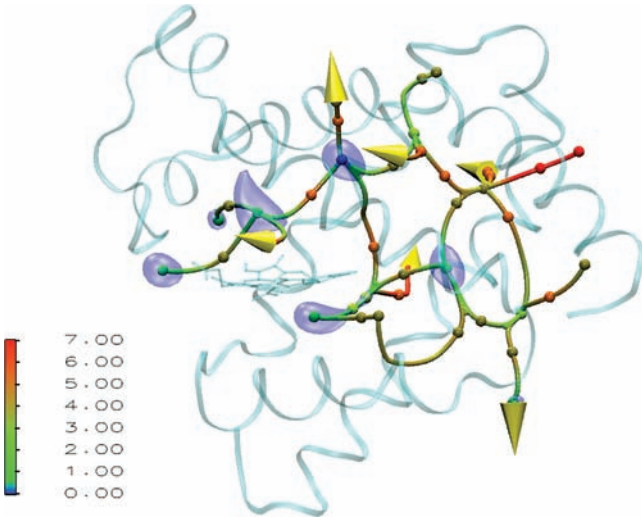

Figure 5. CO migration pathways inside Mb, colored according to the value of the PMF along them (kcal/mol). The Xe4 minimum is taken as the zero of the energy. Local minima and energy barriers along the paths are represented as spheres. Yellow arrows represent the locations of CO exits to the solvent as observed in the TAMD simulations. The protein's backbone is represented as ribbons and the heme as sticks. The isosurface at 2 kcal/mol is also shown in arbitrary color.

Table 1. Free Energy Barriers (kcal/mol) between Pairs of Minima

	(DP, HG)[a]	(DP, Xe4)	(Xe4, Xe2)	(Xe2, Xe1)	(Xe2, Xe3)
ΔA_1[b]	6.7	4.5	5.3	0.2	2.7
ΔA_2	0.1	5.1	3.5	1.3	3.9

[a] Histidine gate (HG) path. DP stands for distal pocket. [b] $\Delta A_{1,2}$ are values measured from the first and second minima of each pair, respectively.

also colored according to their value. A movie with a 360° view of Figure 5 is available online as a Web-enhanced object. The values of some of the barriers are also reported in Table 1. A full list of energy values for barriers and minima is reported in the Supporting Information.

It can be seen that, starting from the DP, the CO has several possible routes. The path to the histidine gate involves a barrier of 6.7 kcal/mol. The energy barrier along the path connecting the DP to Xe4 is 4.5 and 5.1 kcal/mol, starting from DP and Xe4, respectively, close to what was found by Cohen et al.,[37] who measured for the same barriers 3.5 and 4.5 kcal/mol, respectively. From our results, once the ligand migrates from the DP to Xe4, it can take several possible paths. One of these brings it to Xe1, where fewer paths are available and the ligand should reside for longer times. This is in qualitative agreement with experimental results where the lifetime of the signal associated with the ligand in Xe1 is found to be longer than that in Xe4.[9,63]

By looking at the full network of paths, barriers, and minima in Figure 5, we can speculate not only on the mechanism of CO migration inside Mb, but also on the possible gates toward the solvent. Indeed, there is still debate on whether in physiological conditions the ligand exits to the solvent from the DP only, or also from sites far from the heme. Kinetic models used to interpret experimental data are based on the definition of different "states" according to whether the ligand is in the solvent, in the DP, or inside one of the cavities.[63] Although the presence of hydrophobic cavities inside Mb has been known for a long time,[7] only recently experimental evidence was collected for exit pathways involving them.[17,63] Such possible routes were suggested already in the pioneering simulations of Elber and Karplus[6] and confirmed in more recent detailed simulation studies.[20,37] Our results also support this view. Indeed, we find that from each xenon cavity there is a possible route toward the solvent. To reach the farthest cavities from the DP, the ligand has to find a way through the network of paths, crossing multiple barriers. This will slow the diffusion process and can possibly explain the diffuse ligand density that makes it difficult to identify the CO in cavities far from the DP in photolysis experiments at room temperature.

A few of the centers used in the region of the histidine gate were obtained from the TAMD trajectory after it exits the Mb protein. As a result we can infer the free energy of CO in the solvent from the value of the free energy at these points. This free energy is about 9 kcal/mol, i.e., almost 8.5 above the value of the energy in the DP, which is not unreasonable. In particular, it is very close to the barrier from the DP to the solvent via the histidine gate found by Cohen et al.[37] (7.5 kcal/mol). Interestingly, from our results the free energy barrier from the DP toward Xe4 is lower than the one toward the solvent, also in agreement with ref 37. This implies that, starting from the DP, the diffusion of CO toward the interior of the protein is more likely than its direct exit to the solvent from the histidine gate. This finding confirms the importance of the cavities in the process of CO release and binding by Mb. A precise quantification of the rate of exit from Mb to the solvent along the network of pathways found here will be the subject of future work, using methods developed by some of us.[64,65]

Concluding Remarks

The process of CO diffusion in myoglobin at room temperature was investigated by reconstructing, via the single-sweep method and fully atomistic MD simulations, a three-dimensional PMF landscape for the ligand position inside the molecule. This landscape shows minima in correspondence with the DP and xenon cavities and in other pockets more recently observed in other studies.[15,37] We have also computed the MFEPs on this landscape, which give the possible migration paths of CO inside the protein and toward the solvent and permit estimation of the energy barriers involved in this diffusion process. Our results, in line with previous experimental and theoretical findings, show that a network of possible pathways is accessible to the dissociated ligand in the protein interior. Furthermore, we find that from each internal cavity there is a possible gate toward the solvent. Among the paths connecting the DP and the solvent, the one going through the histidine gate is the shortest and the only direct one, i.e., without intermediate minima along it. This might reflect its importance in the ligand escape/entering process.

The basic assumption of our work is that the CO diffusion process inside myoglobin can be described as the navigation of the ligand molecule over its free energy landscape, which accounts via thermal averages for the influence of the protein molecules as well as the solvent. In this picture, we rely on the calculation of a few statistical quantities needed to explain the mechanism of CO diffusion, rather than simple observation of

(63) Nishihara, Y.; Sakakura, M.; Kimura, Y.; Terazima, M. *J. Am. Chem. Soc.* **2004**, *126*, 11877–11888.
(64) Vanden-Eijnden, E.; Venturoli, M.; Ciccotti, G.; Elber, R. *J. Chem. Phys.* **2008**, *129*, 174102.
(65) Vanden-Eijnden, E.; Venturoli, M. *J. Chem. Phys.* **2009**, *19*, 194101.

a few escape trajectories. This type of approach is not only cheaper computationally, but also potentially more precise since it permits averaging of the features of the process that are deemed not interesting and focus on its essential aspects. We believe that it opens the way to investigate gaseous diffusion in more complex, larger enzymes, with a number of cavities and pathways larger than those of myoglobin,[4,5] or other processes of biological interest, such as enzyme−substrate interactions or conformational transitions in larger proteins, relying upon state-of-the-art, fully atomistic MD simulations, but with an accessible computational effort. It also opens the way to generalizations in which we would include in the PMF selected residues,[66] or bound water molecules,[67,68] or even some information about the solvent distribution.[69] In this case collective variables associated with these other degrees of freedom should be considered jointly with the CO coordinates. If the number of these additional collective variables is large, it will not be possible to reconstruct their free energy landscape globally, as we do here, but MFEPs on this landscape can still be calculated using pathway optimization algorithms in high dimensions such as the string method.[25,70]

Acknowledgment. We thank Ron Elber and Benoit Roux for useful discussions and encouragement. Calculations were mostly performed using the General Cluster from NYU High Performance Computing (HPC) Resources. We thank Joseph Hargitai and all HPC support for their help. Part of the calculations were performed on the grid facility of the COMETA consortium. The work of E.V.-E. was partially supported by NSF Grants DMS-0718172 and DMS-0708140 and by ONR Grant N00014-08-1-6046.

Supporting Information Available: Description of the constraint on protein rototranslational motion, table with free energy values for local minima and barriers, and complete ref 48. This material is available free of charge via the Internet at http://pubs.acs.org.

JA905671X

(66) Bossa, C.; Amadei, A.; Daidone, I.; Anselmi, M.; Vallone, B.; Brunori, M.; Di Nola, A. *Biophys. J.* **2005**, *89*, 465–474.
(67) Imai, T.; Hiraoka, R.; Kovalenko, A.; Hirata, F. *J. Am. Chem. Soc.* **2005**, *127*, 15334–15335.
(68) Goldbeck, R. A.; Pillsbury, M. L.; Jensen, R. A.; Mendoza, J. L.; Nguyen, R. L.; Olson, J. S.; Soman, J.; Kliger, D. S.; Esquerra, R. M. *J. Am. Chem. Soc.* **2009**, *131*, 12265–12272.
(69) Imai, T.; Hiraoka, R.; Seto, T.; Kovalenko, A.; Hirata, F. *J. Phys. Chem. B* **2007**, *111*, 11585–11591.
(70) Miller, T. F., III.; Vanden-Eijnden, E.; Chandler, D. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 14559–6002.