JCTC Journal of Chemical Theory and Computation

Article

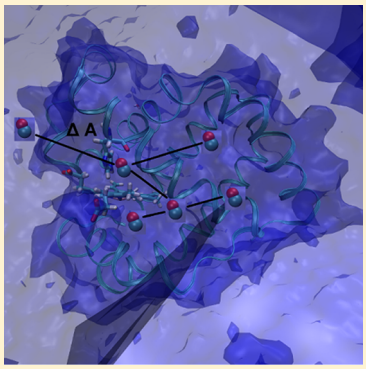
pubs.acs.org/JCTC

# A Computational Study of Water and CO Migration Sites and Channels Inside Myoglobin

Mauro Lapelosa and Cameron F. Abrams*

Department of Chemical and Biological Engineering, Drexel University, Philadelphia, Pennsylvania 19104, United States

**ABSTRACT:** Pathways are computed for transport of $H_2O$ and CO in myoglobin (Mb), using the single sweep and zero-temperature string methods in a fully atomistic, explicitly solvated model system. Our predictions of sites and barriers in the pathways for CO transport agree with previous studies. For $H_2O$, we predict a binding site in the distal pocket (DP), in agreement with crystallographic observations, and another one close to Leu 29, which explains the importance of this residue in controlling the pocket's hydrophobicity, as well as disordered minima in the largely apolar xenon cavities. In particular, $H_2O$ can occupy and transition among the xenon cavities, Xe4, Xe2, and Xe3. Our results support the hypothesis that the thermodynamically most favorable entry/exit portal for $H_2O$ is the so-called histidine gate (HG), the same as for CO. This result, along with the observation of water occupation of both DP and apolar Xe cavities, suggest that water and small gas molecules like CO compete for access to the protein interior, and therefore models of gas molecule transport within proteins should also explicitly consider water transport.




## INTRODUCTION

Diffusion of small molecules inside proteins is essential for many biological processes. Many proteins are targets of small compounds that through binding modulate their biological activity. Many such proteins function as carriers, binding to ligands with high affinity and selectivity, and consequently are able to release them in a controlled manner. A better understanding of the process of diffusion of a small molecule from the solvent to interior pockets of a protein would contribute to our understanding of the function of these proteins. Calculation of free energy changes associated with diffusion processes remains a significant challenge, and it can be directed for example to improve structure-based design of ligands with potential to diffuse and bind with enhanced selectivity and affinity to the protein target.

Myoglobin (Mb), an intracellular oxygen storage protein, has been the center of attention for many years in the study of diffusion pathways and mechanisms of ligand binding. Numerous experimental and computational techniques have been used to study the diffusion of small ligands inside Mb.[1−17] X-ray structures and spectroscopic studies conducted at a variety of temperatures and pressures have shown the existence of hydrophobic cavities with a volume ranging from 30 to 100 $Å^3$ inside Mb, where CO, NO, and $O_2$ can migrate.[14,18] Cavities in Mb (Figure 1) include the so-called distal pocket (DP) and four distinct apolar xenon sites, termed Xe1, Xe2, Xe3, and Xe4.[12,19−24] Crystal structures provide evidence of water in DP.[25] When no ligand is bound to the heme iron cation, $H_2O$ can enter DP and compete with CO in rebinding to the heme group, possibly causing CO to be trapped in Xe1 with $H_2O$ forming noncovalent interactions with the imidiazole ring of His 64.[12] The competition of CO with $H_2O$ in DP implies that $H_2O$ displacement is required for CO binding to the heme.[26]

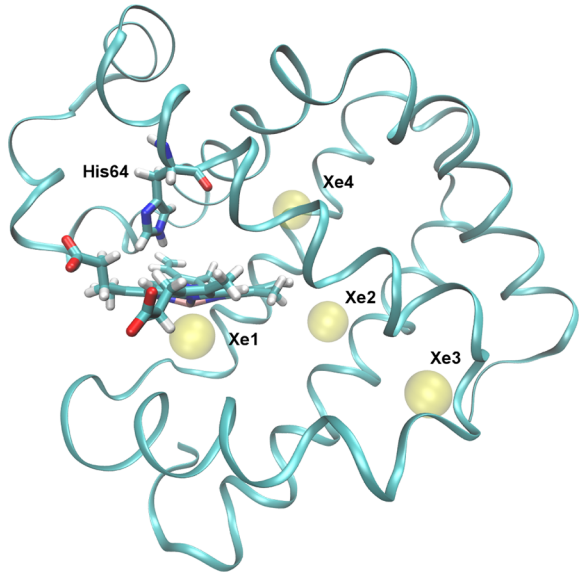



**Figure 1.** Representation of Mb with distal His 64 of E7 (stick), heme group (stick), and xenon adsorption sites (yellow spheres) located in the interior of the protein (ribbon).

Complementing experimental studies were several computational studies which focused on the diffusion of CO inside Mb.[5−8,10,13,16,27,28] Those studies addressed the diffusion of CO entry and exit from Mb focusing only on CO but do not consider the role of $H_2O$, which is known to affect the association rate constant of CO.[26] In particular, changes in

**Received:** October 16, 2012
**Published:** January 17, 2013





dx.doi.org/10.1021/ct300862j | J. Chem. Theory Comput. 2013, 9, 1265−1271

distal water occupancy can critically influence the kinetics of binding of CO, and mutations of E7 introducing apolar residues in DP can shift the water DP-to-solvent equilibrium.[25] Recent work has been done to develop a fingerprint of Mb solvation sites, both external and internal, although the free energy changes associated with water migration in Mb were not calculated.[29]

Standard MD techniques have the potential to shed light on ligand migration, but even long MD simulations face difficulties collecting good statistics due to undersampling transitions between minima in some reasonable free energy landscape.[30] This is because diffusion of ligand molecules in and out of protein cavities can be excessively slow due to energetic barriers associated with the motion of individual residues. Recently, Maragliano et al.[10] used the single sweep method to compute the free energy associated with CO in Mb and further used the string method on the free energy hypersurface to compute pathways of minimum free energy connecting low energy sites. In this work, we employ the single-sweep/string method protocol to study the diffusion of water in Mb.

We find here that the free energy for $H_2O$ has many local minima on the surface of Mb, in DP and in internal cavities. Our findings also support the notion that water and CO preferentially migrate from the DP to the solvent through the HG site. We note that by comparison to the CO free energy, water and CO migrate to different sites and have quite different networks of interconnecting pathways.

## THEORY AND METHODS

**Single-Sweep PMF Reconstruction and Pathway Characterization.** The single-sweep method is used to map free energy landscapes in collective variables (CVs). Briefly, single-sweep uses temperature-accelerated MD (TAMD)[31,32] to rapidly explore the chosen CV space, and samples in that space are harvested and used as locations where mean forces are accumulated for later reconstruction into an approximate free energy. In TAMD, the system is extended to include $z$ variables that act as anchor points for CVs, whose mapping functions are $\theta(x)$:

$$U_k(x, z) = V(x) + \frac{1}{2}k|\theta(x) - z|^2 \tag{1}$$

where $k > 0$ is an adjustable parameter.

The extended system evolves according to the following dynamics:

$$m_x\ddot{x} = -\nabla_x V(x) - k\sum_{\alpha=1}^{N}(\theta_\alpha(x) - z_\alpha)\nabla_x\theta_\alpha(x) + \text{thermostat terms at } \beta^{-1} \tag{2}$$

$$\overline{\gamma}\dot{z} = k(\theta(x) - z) + \sqrt{2\gamma\overline{\beta}^{-1}}\eta(t) \tag{3}$$

where $m_x$ is the mass and $\eta(t)$ is a Gaussian process with mean 0 and covariance $\langle\eta_\alpha(t)\eta_{\alpha'}(t')\rangle = \delta_{\alpha\alpha'}\delta(t - t')$. $\overline{\gamma}$ is a friction coefficient, and $1/\overline{\beta}$ is an artificial temperature. By taking $1/\overline{\beta} > 1/\beta$, the system is able to cross barriers along the trajectory of $z(t)$ and visit relevant regions of CV space which are not accessible at the physical temperature $1/\beta$ on typical MD time scales.

Centers for mean force calculations are harvested from a collection of TAMD simulations, or "sweeps," to cover all the relevant regions of the PMF. The only criterion for allowing a sampled configuration to be a center is that it has to be no closer than some chosen threshold distance from any other center in CV space. The mean force $f_k$ is computed at each center $z_k$, by running TAMD while fixing $z(t) = z_k$:

$$f_k = \frac{1}{T}\int_0^T k(\theta(x(t)) - z_k)\,dt \tag{4}$$

After the mean forces $f_1,...,f_K$ are calculated, the 3D-PMF is reconstructed using a radial basis function (RBF) representation of $A(z)$ for the centers $z_1,...,z_K$.

$$A(z) = \sum_{\alpha=1}^{N} a_k\varphi_\sigma(|z - z_k|) + C \tag{5}$$

The arbitrary origin of $A(z)$ is adjusted using the constant $C$. Euclidean norm is indicated as $|...|$ in $R^N$, and $\varphi_\sigma(u)$ is an RBF with length-scale parameter $\sigma$. We used $\varphi_\sigma(u) = \exp(-u^2/2\sigma^2)$, where $\sigma$ can be adjusted manually to provide a best-fit of the free energy gradients, and the optimal values are determined using a procedure as described previously.[10] Briefly, a gradient error estimate $E$ is used to find optimal parameters $a_k$

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

The minimization procedure involved solving a linear system in which coefficients can be obtained explicitly, and it has been described in details in previous work.[32] The second step of the approach is to compute minimal free energy pathways (MFEPs) connecting local minima using the zero-temperature string method (ZTS).[33−36]

**Protocol.** The initial structure of Mb with CO was retrieved from the PDB (PDB ID 2MB5). The starting structure for Mb with $H_2O$ was obtained from Mb with CO by mutating the CO into $H_2O$ and retaining the original oxygen atom coordinates. The crystallographic water molecules located outside the protein matrix, and ions, were removed. Ionization states were assigned considering neutral pH. Histidines were not protonated, and necessary hydrogen atoms were added. The centers of mass of residues 10, 100, and 150 were restrained near their original coordinates using an isotropic quadratic function with force constant $k_f$ = 1.0 kcal/mol/Å$^2$ to fix the orientation of the protein during the simulations. A rectilinear box with explicit water was built to solvate the protein. The system included 2533 protein atoms, 18 515 water molecules for solvation, one ion, and three atoms or two atoms of the small ligand. The MD simulations were carried out using NAMD v. 2.8[37] using the CHARMM27 force field[38] with TIP3P water.[39] Periodic boundary conditions were used, and long-range electrostatics were computed using particle-mesh Ewald summation with grid spacing of 1 Å. van der Waals interactions were cut off beyond a distance of 9 Å, and covalent bonds were kept fixed by using the SHAKE algorithm. A Langevin thermostat was used to control the temperature at 310 K with a coupling constant of 5 ps$^{-1}$, and the integration time step was 2 fs. The system was equilibrated for 2 ns at constant pressure ($P$ = 1.01 bar) and temperature ($T$ = 310 K) by using a Langevin Nosé−Hoover piston with a coupling constant of 1 ps for the thermostat and 0.5 ps for the barostat. The box dimensions were 65 × 60 × 60 Å$^3$ (21 049 atoms) for equilibration and production runs. For the simulations in which we recompute the CO pathways, we model the CO dissociated state, and the system is not allowed to bind the Fe and switch

to the associated state (C−Fe). The parameters of the potential to describe the dissociated state are retrieved from previous work.[40] During pre-single-sweep MD equilibration, the ligand was always located inside DP for both cases of CO and $H_2O$. TAMD simulations started after the equilibration.

We selected three CVs for both CO and $H_2O$: the Cartesian coordinates of the center of mass of the ligand molecule. We assumed that the diffusion of the small molecule can be analyzed using a small number of CVs, and the choice of those variables is based on the capability of mapping the resulting PMF on the initial structure and the relative ease in tracking the small molecule to detect when it has exited the protein to the solvent. Orientational and translational displacement of the protein was avoided using harmonic restraints described above. The exploration of the unknown free energy landscape is performed using 10 independent TAMD simulations, in each of which the CO or $H_2O$ after sampling interior cavities of Mb escaped to the solvent. We have used as parameters for all TAMD simulations $\kappa$ = 200 kcal/(mol Å$^2$), $\overline{\beta}^{-1}$ = 5 kcal/mol, $\overline{\gamma}$ = 50 ps$^{-1}$, and a time step of 2 fs. We tried also to run simulations with $\overline{\beta}^{-1}$ = 10 kcal/mol and $\overline{\beta}^{-1}$ = 20 kcal/mol, keeping the other parameters fixed, but CO and $H_2O$ then exited the protein too rapidly without sampling the cavities or locations between them.

All the TAMD trajectory frames were combined together to obtain a single sweep of the center of mass coordinates. This trajectory is used to select the centers in CV space, where the mean forces were calculated to reconstruct the PMF. We selected the centers starting from $z_1$ and depositing a center along the trajectory such that no two centers are closer than 2.5 Å from each other. At the end, we deposited 250 centers for CO and 130 for $H_2O$, respectively (Figure 2).

We compute the mean forces at the centers with a simulation in which we set $z(t) = z_k$ and $k$ = 200 kcal/(mol Å$^2$) described in eq 4. The mean forces were accumulated for 500 ps in order to ensure convergence of the mean force. These calculations were distributed in parallel since they are independent from each other. Gaussian RBFs were used for the reconstruction of the PMF with an optimal value of $\sigma$ = 3 Å, and we obtained a relative residual error of 0.6. VMD was used to produce the protein graphical representations.[41]

## RESULTS

**$H_2O$ and CO PMF Maps in Mb.** We determined 3-D PMFs in two separate sets of single sweep calculations for CO and $H_2O$. Both CO and $H_2O$ cross the protein interior on a 1 ns time scale to reach the solvent from the initial configuration in DP in all TAMD simulations launched. At the beginning of the simulations, $H_2O$ was confined by residues Leu 29, Phe 43, His 64, Val 68, and Ile 107. Energetic barriers arise from the side chains' orientation of key residues like Leu 29, Leu 32, Phe 33, Arg 45, His 64, Ile 107, and Ile 111 in DP that limit the motion of small molecules among cavities inside the protein in unbiased simulations. Overall, we observe that our TAMD simulations are successful in overcoming the energetic barriers and sampling relevant regions of the 3D-PMF going from DP to the apolar cavities and to reach the solvent.

Free energy minima on the 3D-PMF in the case of $H_2O$ represent locations of hydration sites in the protein. Free energy minima for water are found on the surface of the protein and also inside the core of the protein matrix in Xe2, Xe3, Xe4, and DP (Table 1). We found a deep minimum on the surface of the protein next to Glu 109, which is a conserved residue and

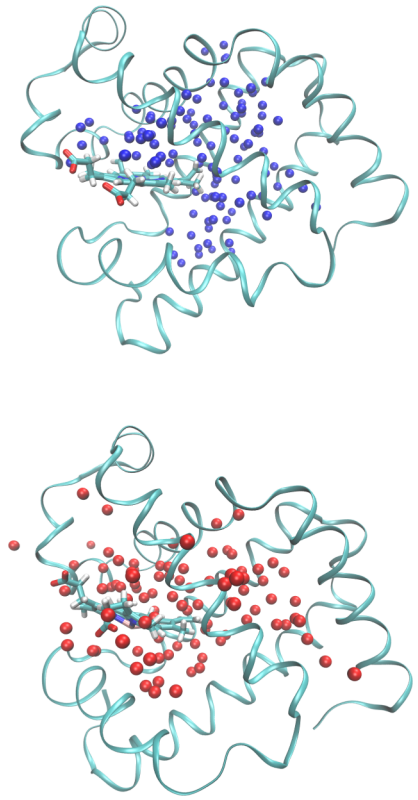


**Figure 2.** A total of 130 centers for $H_2O$ (top) and 250 for CO (bottom) as blue spheres ($H_2O$) or red spheres (CO) located inside Mb. The protein backbone is shown in cyan ribbon, the heme group as bonds.

**Table 1. Local Minima Sites for Water Identified Using Single Sweep**

| site | energy (kcal/mol) | residues |
|---|---|---|
| 1 | −8.0 | Glu109 |
| 2 | −7.9 | His64 |
| 3 | −7.0 | Glu105 |
| 4 | −6.5 | Lys79 |
| 5 | −5.9 | His24 |
| 6 | −5.7 | Arg45 |
| 7 | −5.7 | Phe33 |
| 8 | −5.6 | Gln26 |
| 9 | −5.4 | Arg139 |
| 10 | −5.2 | Glu18 |
| 11 | −4.8 | Arg139 |
| 12 | −4.4 | Glu85 |
| 13 | −4.3 | His24 |

located on the surface (Figure 3), and it represents the global minimum. Free energy minima are located in close proximity to His 64 in DP with a free energy of −7.9 kcal/mol and close to His 24 in Xe4 with a free energy of −5.9 kcal/mol, relative to the unbound ligand state. Water evidently forms polar interactions with those specific residues. Free energy local minima of water inside DP are shown in Figure 4. The 3D-PMF for $H_2O$ clearly indicates the presence of hydration sites inside the DP pocket. One local free energy minimum is found in proximity of N$\varepsilon$ of His 64 in E7, and it is compatible with the hydration site shown in the crystal structure of Mb with water located in DP.[25] A second minimum can be found in the interior of the protein on the other side of the side chain of Leu

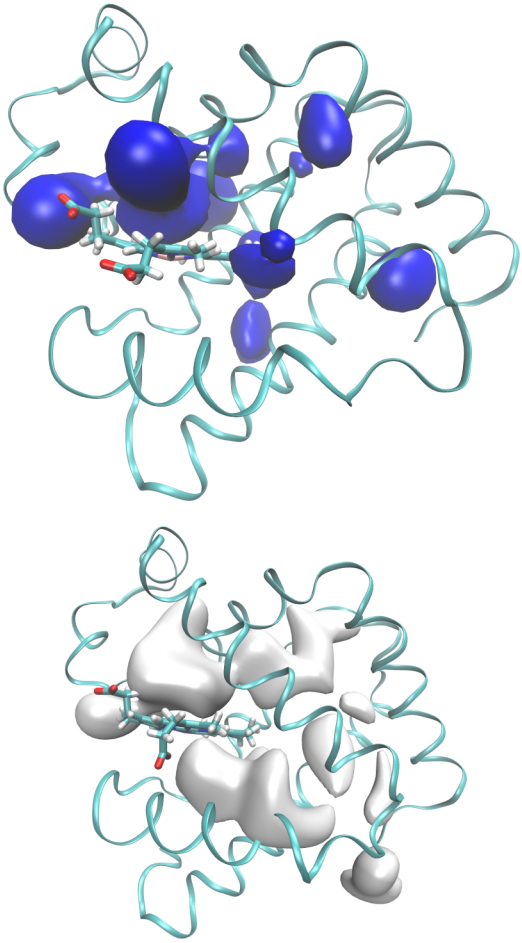


**Figure 3.** Isosurfaces of the 3D-PMF of Mb for $H_2O$ (top) and CO (bottom). Surfaces are represented at 7 kcal/mol with respect to their global energy minima. The global minimum for water is in proximity of Glu 109 on the surface of Mb, and the global minimum for CO is in Xe4.

29 from the His 64 minimum. The shape of the 3D-PMF (Figure 4) in the DP shows that water in DP can exchange positions within the pocket even if it occupies on average the site close to N$\varepsilon$ of His 64.

The 3D-PMF surface is also calculated for CO (Figure 3). We perform those calculations to assess the quality of our PMF reconstruction with respect to a previous study[10] and to compare the $H_2O$ PMF with that of CO produced using the same methods and simulation codes. One minimum is found in DP which is extended, with another minimum in the Xe4 pocket. Our findings reveal here that the local structure of the free energy in DP differs for CO and $H_2O$ (Figure 3), and $H_2O$ has a minimum toward the interior of Mb facing Xe4 that may signify its easy displacement in that direction. The minimum located in Xe2 is shifted slightly by 1 Å comparing the Xe2 site determined in the crystal structures. Finally, the Xe3 cavity also has a minimum. Thus, the minima for CO included all the known apolar cavities where CO was found: DP, Xe1, Xe2, Xe3, and Xe4. In particular, the shape of local free energy minima reasonably correlates with experimental defined internal Xe sites to which CO can migrate. The 3D-PMF shows the deepest free energy minima in Xe4. Overall, the minima are located in the apolar cavities and DP, and they reproduce the locations found by experiments on CO

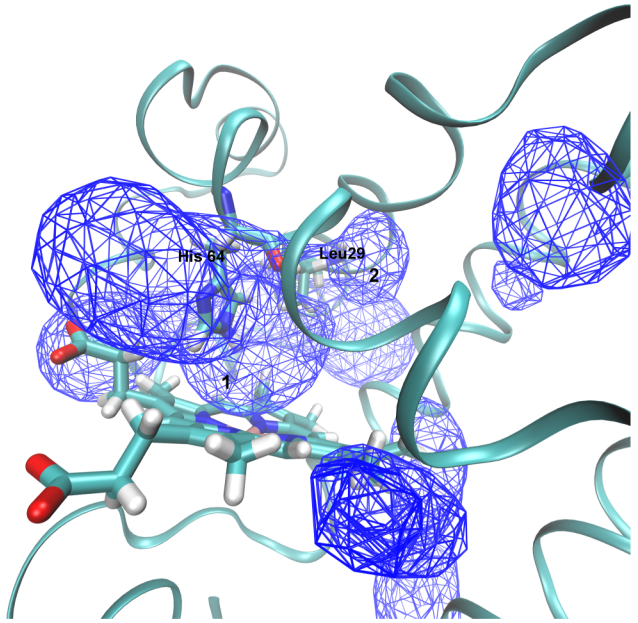



**Figure 4.** Closeup view of local free energy minima of the PMF for $H_2O$ close to DP of Mb. One local free energy minimum is found in proximity of N$\varepsilon$ of His 64 in E7. The isosurface of the PMF is shown as a mesh, while the heme group, His 64, and Leu 29 are represented as sticks and the protein in ribbon. The isosurface is 8 kcal/mol from the global minimum. The two minima are labeled with a number inside the mesh.

diffusion.[42] These results are in agreement with many computational studies directed to map the diffusion of CO inside Mb.[4−8,10,13,28]

**Diffusion Pathways.** After locating the free energy minima, we focus our attention on the diffusion pathways and entry/exit portals for $H_2O$ and CO. We first make the reasonable assumption that in diffusing through protein channels, ligands will roughly follow pathways of minimal free energy, so identifying diffusion routes is the same as finding MFEPs on the PMFs. The small molecule in the initial configuration is located in the DP, and the pathway to allow migration to the solvent covers a distance of 30−40 Å. From the local free energy minima, we search for the dominant pathways for $H_2O$ and CO on their respective 3D-PMFs using the ZTS method as described in the Theory and Methods section. MFEPs for $H_2O$ show that energetic barriers and dominant pathways as well free energy difference between a pair of Xe sites can be summed up for each barrier to determine the relative total energy barrier considering all the minima in the MFEPs (Table 2). Figure 5 shows free-energy profiles along MFEPs for water connecting solvent with Mb internal cavities. Figure 6 shows the dominant pathways as red curves connecting DP to HG, Xe4, Xe2, Xe1, and Xe3. The isosurface of the 3D-PMF is shown (gray, 7 kcal/mol with respect to Xe4) with gray spheres indicating the local minima. Figure 7 shows the dominant pathways for water in a

**Table 2. Total Free Energy Barriers Calculated for Local Minima Sites in Mb**

| ligand | DP-HG[a] | DP-Xe4[a] | Xe4−Xe2[a] | Xe2−Xe1[a] | Xe2−Xe3[a] |
|---|---|---|---|---|---|
| $\Delta A_{H_2O}$ | 7.2 | 4.9 | 7.5 | | 5.5 |
| $\Delta A_{CO}$ | 7.0 | 7.0 | 8.5 | 1.2 | 3.6 |

[a]In kcal/mol.


dx.doi.org/10.1021/ct300862j | J. Chem. Theory Comput. 2013, 9, 1265−1271

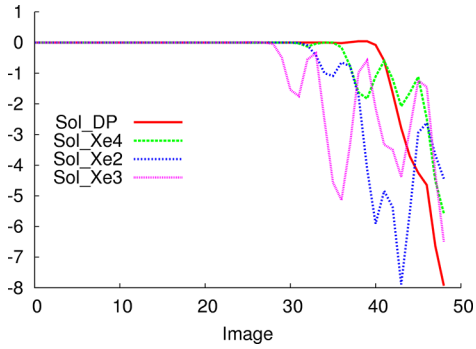



**Figure 5.** Free-energy profiles along MFEPs for water migration from the solvent to internal cavities. The difference is calculated with 50 images using the string method. Solvent−DP(red), solvent−Xe4 (green), solvent−Xe3 (purple), and solvent−Xe2 (blue) are represented.

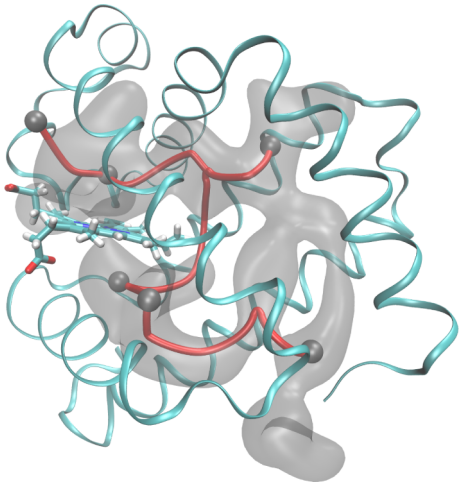


**Figure 6.** MFEPs for CO from DP−HG, DP−Xe4, Xe4−Xe2, and Xe2−Xe3. Local minima along the pathways (red) are presented as spheres (gray). The isosurface is drawn considering an energy level of 7 kcal/mol from the global minimum in Xe4.

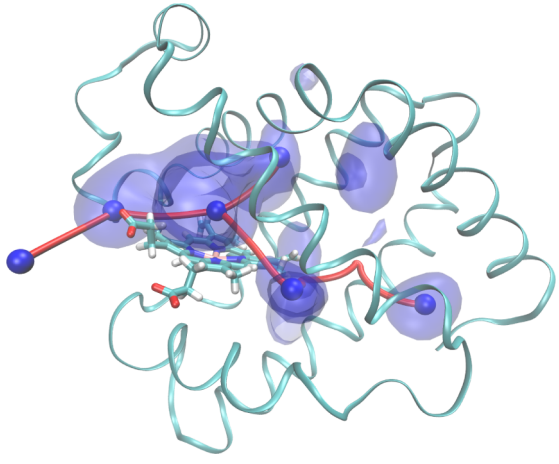


**Figure 7.** MFEPs for water from DP−HG, DP−Xe4, Xe4−Xe2, and Xe2−Xe3. Local minima are blue spheres, and pathways are red tubes. The isosurfaces are drawn with an energy level of 7 kcal/mol from the global minimum at Glu 109.

similar format as the CO pathways. The dominant pathways for CO computed in this study start in DP and go to HG, Xe4, Xe2, and Xe3 with local minima in blue spheres. For water, the red tubes connecting the internal cavities are organized differently with respect to CO, indicating two distinct networks for the two molecules.

For water, the difference in free energy from solvent to DP through HG is 8.0 kcal/mol (Figure 5), and there is a small barrier to entering the protein. Water to enter Xe4 has to overcome a barrier of 2.1 kcal/mol. In contrast, the entrance from Xe2 has three barriers: the first two are about 0.5 kcal/mol, and the third one is 5.0 kcal/mol. Migration of water inside the protein from Xe3 is unlikely due to many high barriers that need to be overcome to enter the protein matrix. We can establish that the most likely route from entry of water inside Mb is from HG to DP.

Surprisingly, however, we found that $H_2O$ traverses a high barrier prior even to entering the DP compared with the barriers of internal apolar sites. The barriers we computed suggest that $H_2O$ exits the binding pocket more easily to enter other cavities than it does to exit the protein. DP-Xe4 migration shows a barrier of 4.9 kcal/mol compared with 8.0 kcal/mol for $H_2O$ to leave through HG. Water can shuttle to other internal cavities from Xe4. The local pathway from Xe4 to Xe2 has a barrier of 7.5 kcal/mol. Xe2−Xe3 migration has a barrier of 5.5 kcal/mol.

We measure a barrier of 7.0 kcal/mol from HG to DP for CO. In contrast, the energy barrier for migration from DP to Xe4 has two minima with a difference of 5.0 kcal/mol for the first one and 2.0 kcal/mol for the second one. The Xe4−Xe2 migration crosses two barriers, 4.2 and 4.3 kcal/mol, which together constitute the largest energetic barrier in pathways of diffusion. From Xe2 to Xe1, the barrier is 1.2 kcal/mol. Free energy barriers from Xe2 to Xe3 are 1.3 and 2.3 kcal/mol, respectively. In a recent work, Bossa et al.[27] measured a PMF barrier of ∼2.6 kcal/mol from Xe4 to DP for CO, computed from a long umbrella sampling simulation with a flat umbrella potential of which about 3 ns is spent by CO at the barrier. Those results were reproduced by Cohen et al.[7] with a different technique. These methodologies are based on a different approach. In the first work, PMF is calculated using umbrella sampling,[27] and in the second one, implicit ligand sampling is used.[7] But, overall our results are similar to theirs.

Also, we can construct channels of migration inside the protein matrix for CO and $H_2O$ interconnecting DP, Xe4, Xe2, Xe1, and Xe3 (Figures 6 and 7). Differences between CO and $H_2O$ are shown; in general, we can say that CO follows different dominant routes of migration inside Mb than does water, and the Xe1 pocket is not visited by water.

## DISCUSSION

The computational method presented here provides free energies and small ligand diffusion pathways in Mb. Overall, the free energy minima for CO in DP, Xe4, Xe2, Xe1, and Xe3 are located in similar sites to other studies[7,13,27] and agree with crystallographic and spectroscopic evidence of wild type structures and mutants of Mb in the apolar cavities.[2,42−46] The PMF calculation for $H_2O$ adds a critical missing piece in the general picture of CO diffusion. $H_2O$ free energy local minima are located in DP, suggesting that water has an important role in DP both in controlling the HG entry of CO and in creating competition for CO binding to His 64 in the pocket. The presence of a minimum for water close to Leu 29 can explain how the displacement of water can take place in DP to allow the rebinding of the CO ligand to the heme. Indeed,


dx.doi.org/10.1021/ct300862j | J. Chem. Theory Comput. 2013, 9, 1265−1271

the replacement of Leu 29 with a less hydrophobic residue leads to the invasion of water in DP due to the decrease of hydrophobicity.[47] Leu 29 forms a separation barrier between the two energy minima in DP, and the presence of this barrier allows us to conclude that wetting/dewetting transitions with high water mobility likely exist in Mb. Although the hydration sites cannot be correlated with the residence time of water,[48] DP contains hydrophobic side chains capable of increasing the residence time for water.[48] The hydration water sites detected by X-ray diffraction experiments[25] systematically overlap onto calculated high-density regions within the approximations of Debye−Waller factors generally used in crystallographic refinement, which assume fully harmonic oscillations of small molecules in crystals. The presence of free energy minima for $H_2O$ in the DP of Mb matched qualitatively previous experimental work,[26] confirming hydration properties of an unliganded DP pocket, and the high occupancy of water that agrees with $H_2O$ hydrogen bonding to Ne of His 64.[25,49]Furthermore, our results argue that water can be present in the Xe2 of Mb even if it is not seen in the crystal structures, suggesting further that waters in these cavities are more disordered with shorter residence times compared with water in DP.[48] The presence of minima located in Xe4 and Xe3 can be understood by the polar side chains of His 24 and His 82 that may form hydrogen bonds with the water molecule. The 3D-PMF is different for CO and $H_2O$, in that the minima for CO are deep inside the apolar cavities due to the hydrophobic nature of residues in the protein matrix. Findings of local minima for $H_2O$ support the general model drawn by the Brunori and Gibson scheme in which water binding to His 64 is an intermediate state,[21] but also the DP can be characterized by dry and wet regions. Our prediction that water has an ordered behavior in DP and is instead disordered in the other cavities (Xe4, and Xe2, and Xe3) is important because it implies that water in cavities needs to be modeled in high resolution structures.[50,51]

Considering all the pathways and barriers we can speculate that the entry portal for $H_2O$ is the HG, but then the migration to other internal cavities like Xe4 and Xe2 is possible. MFEPs show a low free energy barrier for $H_2O$ from DP to Xe4 which suggests that when CO returns to DP to bind the heme, water can be displaced in the DP-Xe4 direction and then exit the protein. On the other hand, the migration of $H_2O$ from Xe4 to Xe2 seems to be unlikely due to a high free energy penalty, and instead the transition from DP to Xe2 and then from Xe1 and to the solvent is likely to be one dominant pathway. CO binding is thermodynamically favored in DP compared with $H_2O$, which can be also captured transiently. Indeed, an apolar gas like CO is favored entropically to move out of the water solvent, but the gain is compensated in the transfer to a closed internal cavity inside a protein. Once the CO moves to apolar cavities, $H_2O$ binding to His 64 is possible. If CO returns to DP to rebind or exit the protein matrix, the water needs to be displaced, and it is possible for this displacement to carry water into disordered internal sites.

It is not yet possible to experimentally map the free energy of ligands inside proteins at ambient conditions, so we cannot compare our CO and $H_2O$ free energies directly to experimental results. However, as pointed out by Elber,[9] we can make approximations of the escape time CO from DP through the HG route using simple transition-state theory. With a frequency of 2 $ps^{-1}$, our DP-to-HG barrier for CO implies an escape time of about 200 ns, and that for water is about 300 ns. Kinetics of geminate rebinding[15,52] suggest the CO escape time is a bit lower, about 100 ns. However, we stress that a more robust comparison of kinetics can be made by forgoing TST in favor of simulation methods such as milestoning,[53] an avenue we are currently pursuing.

Finally, we briefly discuss our choice of CVs. We followed the example of Maragliano et al.[10] in choosing as CVs the Cartesian coordinates of the ligand center of mass. Within the context of any CV-based enhanced sampling method, this of course means that there might be other slow variables that are missed, and in particular, one might be concerned here with transitions among side-chain rotameric states which may influence transport pathways.[5,54] It is possible to include such variables in a single-sweep analysis, but adding a fourth dimension to CV space would increase the number of mean-force computations by a large factor, making it somewhat challenging. Alternative approaches to detecting and handing hidden barriers once one has made an obvious choice for CVs should be pursued. At this stage, because single-sweep evidently does well at identifying key sites, these results provide a baseline against which more sophisticated methods that include additional slow variables could be compared.

## CONCLUSIONS

We used the single-sweep/string method to compute diffusion pathways for water and CO in Mb. Local free energy minima for water are located in DP and agree with experimental results, and local minima are observed in Xe4, Xe2, and Xe3, which are regions in which water was not resolved in the crystal structures. From the PMF maps of water and CO, we can assume the presence of competition for the apolar sites inside Mb. Given this, it is likely that transport of water and CO inside Mb are interconnected, and experimental diffusion rates are also reciprocally affected by the presence of those two molecules. This motivates future work in which the direct dependency of diffusion rates of CO on the presence of water inside the protein can be addressed.

## AUTHOR INFORMATION

### Corresponding Author

*E-mail: cfabrams@drexel.edu.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This work has been supported by the National Institute of Health under Grant No. R01GM100472. The calculations reported in this work have been performed at the TACC Supercomputing Center using XSEDE allocation (TG-MCB070073N).

## REFERENCES

(1) Anselmi, M.; Di Nola, A.; Amadei, A. *Biophys. J.* **2008**, *94*, 4277−4281.

(2) Aranda, R.; Levin, E. J.; Schotte, F.; Anfinrud, P. A.; Phillips, G. N. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2006**, *62*, 776−783.

(3) 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.

(4) Bossa, C.; Amadei, A.; Daidone, I.; Anselmi, M.; Vallone, B.; Brunori, M.; Di Nola, A. *Biophys. J.* **2005**, *89*, 465−474.

(5) Case, D. A.; Karplus, M. *J. Mol. Biol.* **1979**, *132*, 343−368.

(6) Ceccarelli, M.; Anedda, R.; Casu, M.; Ruggerone, P. *Proteins: Struct., Funct., Bioinf.* **2008**, *71*, 1231−1236.
(7) Cohen, J.; Arkhipov, A.; Braun, R.; Schulten, K. *Biophys. J.* **2006**, *91*, 1844−1857.
(8) Elber, R.; Gibson, Q. H. *J. Phys. Chem. B* **2008**, *112*, 6147−6154.
(9) Elber, R. *Curr. Opin. Struct. Biol.* **2010**, *20*, 162−167.
(10) Maragliano, L.; Cottone, G.; Ciccotti, G.; Vanden-Eijnden, E. *J. Am. Chem. Soc.* **2010**, *132*, 1010−1017.
(11) Levy, R.; Sheridan, R.; Keepers, J.; Dubey, G.; Swaminathan, S.; Karplus, M. *Biophys. J.* **1985**, *48*, 509−518.
(12) Olson, J. S.; Soman, J.; Phillips, G. N. *IUBMB Life* **2007**, *59*, 552−562.
(13) 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.
(14) Schotte, F.; Soman, J.; Olson, J. S.; Wulff, M.; Anfinrud, P. A. *J. Struct. Biol.* **2004**, *147*, 235−246.
(15) Scott, E. E.; Gibson, Q. H.; Olson, J. S. *J. Biol. Chem.* **2001**, *276*, 5177−5188.
(16) Straub, J. E.; Karplus, M. Molecular Dynamics of Carbon Monoxide After Photodissociation from Myoglobin. *AIP Conf. Proc.* **1991**, 147−152.
(17) Wang, P.-h.; Blumberger, J. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 6399−6404.
(18) Brunori, M. *Trends Biochem. Sci.* **2001**, *26*, 209−210.
(19) Brunori, M.; Cutruzzolà, F.; Savino, C.; Travaglini-Allocatelli, C.; Vallone, B.; Gibson, Q. H. *Biophys. J.* **1999**, *76*, 1259−1269.
(20) Brunori, M.; Vallone, B.; Cutruzzola, F.; Travaglini-Allocatelli, C.; Berendzen, J.; Chu, K.; Sweet, R. M.; Schlichting, I. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 2058−2063.
(21) Brunori, M.; Gibson, Q. H. *EMBO Rep.* **2001**, *2*, 674−679.
(22) Collman, J. P.; Decréau, R. A.; Dey, A.; Yang, Y. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 4101−4105.
(23) Ostermann, A.; Waschipky, R.; Parak, F. G.; Nienhaus, G. U. *Nature* **2000**, *404*, 205−208.
(24) Tetreau, C.; Blouquit, Y.; Novikov, E.; Quiniou, E.; Lavalette, D. *Biophys. J.* **2004**, *86*, 435−447.
(25) Quillin, M. L.; Arduini, R. M.; Olson, J. S.; Phillips, G. N. *J. Mol. Biol.* **1993**, *234*, 140−155.
(26) Goldbeck, R. A.; Bhaskaran, S.; Ortega, C.; Mendoza, J. L.; Olson, J. S.; Soman, J.; Kliger, D. S.; Esquerra, R. M. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 1254−1259.
(27) Bossa, C.; Anselmi, M.; Roccatano, D.; Amadei, A.; Vallone, B.; Brunori, M.; Di Nola, A. *Biophys. J.* **2004**, *86*, 3855−3862.
(28) Elber, R.; Karplus, M. *J. Am. Chem. Soc.* **1990**, *112*, 9161−9175.
(29) Scorciapino, M. A.; Robertazzi, A.; Casu, M.; Ruggerone, P.; Ceccarelli, M. *J. Am. Chem. Soc.* **2010**, *132*, 5156−5163.
(30) Lapelosa, M.; Gallicchio, E.; Levy, R. M. *J. Chem. Theory Comput.* **2012**, *8*, 47−60.
(31) Abrams, C. F.; Vanden-Eijnden, E. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 4961−4966.
(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.
(34) E, W.; Ren, W.; Vanden-Eijnden, E. *J. Chem. Phys.* **2007**, *126*, 164103.
(35) Maragliano, L.; Fischer, A.; Vanden-Eijnden, E.; Ciccotti, G. *J. Chem. Phys.* **2006**, *125*, 24106.
(36) Maragliano, L.; Vanden-Eijnden, E. *Chem. Phys. Lett.* **2007**, *446*, 182−190.
(37) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(38) Foloppe, N.; MacKerell, J. *J. Comput. Chem.* **2000**, *21*, 86−104.
(39) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
(40) Kuczera, K.; Kuriyan, J.; Karplus, M. *J. Mol. Biol.* **1990**, *213*, 351−373.
(41) Humphrey, W. *J. Mol. Graphics* **1996**, *14*, 33−38.
(42) Tilton, R. F.; Kuntz, I. D.; Petsko, G. A. *Biochemistry* **1984**, *23*, 2849−2857.
(43) Bourgeois, D.; Vallone, B.; Arcovito, A.; Sciara, G.; Schotte, F.; Anfinrud, P. A.; Brunori, M. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 4924−4929.
(44) Hummer, G.; Schotte, F.; Anfinrud, P. A. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 15330−15334.
(45) Srajer, V.; Teng, T.; Ursby, T.; Pradervand, C.; Ren, Z.; Adachi, S.; Schildkamp, W.; Bourgeois, D.; Wulff, M.; Moffat, K. *Science* **1996**, *274*, 1726−1729.
(46) Srajer, V.; Ren, Z.; Teng, T. Y.; Schmidt, M.; Ursby, T.; Bourgeois, D.; Pradervand, C.; Schildkamp, W.; Wulff, M.; Moffat, K. *Biochemistry* **2001**, *40*, 13802−13815.
(47) Uchida, T. *J. Biol. Chem.* **1997**, *272*, 30108−30114.
(48) Makarov, V. A.; Andrews, B. K.; Smith, P. E.; Pettitt, B. M. *Biophys. J.* **2000**, *79*, 2966−2974.
(49) Phillips, S. E. *J. Mol. Biol.* **1980**, *142*, 531−554.
(50) 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.
(51) Yu, B. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 103−108.
(52) Scott, E. E.; Gibson, Q. H. *Biochemistry* **1997**, *36*, 11909−11917.
(53) Faradjian, A. K.; Elber, R. *J. Chem. Phys.* **2004**, *120*, 10880−10889.
(54) Kottalam, J.; Case, D. A. *J. Am. Chem. Soc.* **1988**, *110*, 7690−7697.