J|A|C|S ARTICLES

Published on Web 08/04/2009

Breathing Motions of a Respiratory Protein Revealed by Molecular Dynamics Simulations

Mariano Andrea Scorciapino,[†,‡] Arturo Robertazzi,[‡] Mariano Casu,[†] Paolo Ruggerone,[‡,§] and Matteo Ceccarelli*[,‡,§]

Department of Chemical Sciences, University of Cagliari, Cittadella Universitaria, I-09042 Monserrato (Ca), Italy, Sardinian Laboratory for Computational Materials Science (SLACS), INFM-CNR, Cittadella Universitaria, I-09042 Monserrato (Ca), Italy, and Department of Physics, University of Cagliari, Cittadella Universitaria, I-09042 Monserrato (Ca), Italy

Received April 20, 2009; E-mail: matteo.ceccarelli@dsf.unica.it

Abstract: Internal cavities, which are central to the biological functions of myoglobin, are exploited by gaseous ligands (e.g., O_2, NO, CO, etc.) to migrate inside the protein matrix. At present, it is not clear whether the ligand makes its own way inside the protein or instead the internal cavities are an intrinsic feature of myoglobin. To address this issue, standard molecular dynamics simulations were performed on horse-heart met-myoglobin with no ligand migrating inside the protein matrix. To reveal *intrinsic* internal pathways, the use of a statistical approach was applied to the cavity calculation, with special emphasis on the major pathway from the distal pocket to Xe1. Our study points out the remarkable dynamical behavior of Xe4, whose "breathing motions" may facilitate migration of ligands through the distal region. Additionally, our results highlight a two-way path for a ligand to diffuse through the proximal region, possibly allowing an alternative route in case Xe1 is occupied. Finally, our approach has led us to the identification of key residues, such as leucines, that may work as *switches* between cavities.

Introduction

Myoglobin (Mb) is a relatively small globular metalloprotein (153 residues) mostly found in heart and skeletal muscles of humans and other animal species.[1−4] Historically, Mb was the first protein to be solved at atomic level.[4] Because of its size, Mb is often employed as a model system for larger and more complex globular proteins.[1−4] However, for a long time, myoglobin has been considered as simply an oxygen storage and transport system[1−4] in which the entry/exit path is regulated by the so-called "histidine gate".[5] While this process involves direct migration through a single route to the heme binding site, a more intricate picture has emerged from later studies[6−10] showing that the gas molecule can migrate inside the protein matrix through a complex network of cavities before or after forming a bond with iron. This has led to the hypothesis that the internal cavities may be crucial for the multiple biological roles of myoglobin. For instance, it has been shown that this protein can bind/store other small gas ligands (e.g., CO, NO, Xe, etc.) and that Mb may have biological functions other than that of oxygen transport.[11−13] It is astonishing that one of the most studied proteins in history can still provide surprises for the scientific community.[1−4,11,12]

Although migration of ligands from the solvent to the interior of the protein (and vice versa) is crucial for it to exert the biological role,[7,14−22] none of the more than 250 crystallographic structures of Mb shows an obvious path connecting the external

† Department of Chemical Sciences, University of Cagliari.
‡ Sardinian Laboratory for Computational Materials Science.
§ Department of Physics, University of Cagliari.

(1) Springer, B.; Sligar, S.; Olson, J.; Phillips, G. *Chem. Rev.* **1994**, *94*, 699–714.
(2) Frauenfelder, H.; McMahon, B. H.; Fenimore, P. W. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 8615–8617.
(3) Brunori, M.; Bourgeois, D.; Vallone, B. *J. Struct. Biol.* **2004**, *147*, 223–234.
(4) Kendrew, J. C.; Dickerson, R. E.; Strandberg, B. E.; Hart, R. G.; Davies, D. R.; Philips, D. C.; Shore, V. C. *Nature* **1960**, *185*, 422–427.
(5) Perutz, M. F.; Mathews, F. S. *J. Mol. Biol.* **1966**, *21*, 199–202.
(6) Bellelli, A.; Blackmore, R.; Gibson, Q. H. *J. Biol. Chem.* **1990**, *265*, 13595–13600.
(7) Carlson, M.; Regan, R.; Gibson, Q. H. *Biochemistry* **1996**, *35*, 1125–1136.
(8) Gibson, Q. H.; Regan, R.; Elber, R.; Olson, J. S.; Carver, T. *J. Biol. Chem.* **1992**, *267*, 22022–22034.
(9) Elber, R.; Karplus, M. *J. Am. Chem. Soc.* **1990**, *112*, 9161–9175.
(10) Tilton, R.; Kuntz, I.; Petsko, G. *Biochemistry* **1984**, *23*, 2849–2857.
(11) Cossins, A.; Berenbrink, M. *Nature* **2008**, *454*, 416–417.
(12) Brunori, M. *Trends Biochem. Sci.* **2001**, *26*, 209–210.
(13) Flögel, U.; Merx, M.; Gödecke, A.; Decking, U.; Schrader, J. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 735–740.
(14) Cohen, J.; Arkhipov, A.; Braun, R.; Schulten, K. *Biophys. J.* **2006**, *91*, 1844–1857.
(15) 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.
(16) Ostermann, A.; Waschipky, R.; Parak, F. G.; Nienhaus, G. U. *Nature* **2000**, *404*, 205–208.
(17) Dantsker, D. *J. Biol. Chem.* **2005**, *280*, 38740–38755.
(18) Schotte, F.; Lim, M. H.; Jackson, T. A.; Smirnov, A. V.; Soman, J.; Olson, J. S.; Phillips, G. N.; Wulff, M.; Anfinrud, P. A. *Science* **2003**, *300*, 1944–1947.
(19) Bossa, C.; Anselmi, M.; Roccatano, D.; Amadei, A.; Vallone, B.; Brunori, M.; Di Nola, A. *Biophys. J.* **2004**, *86*, 3855–3862.
(20) Bossa, C. *Biophys. J.* **2005**, *89*, 465–474.
(21) 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. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 2612–6.
(22) Dantsker, D.; Samuni, U.; Friedman, J. M.; Agmon, N. *Biochim. Biophys. Acta* **2005**, *1749*, 234–251.

solvent to the protein interior.[10,23] It has therefore been suggested that thermal fluctuations can create a thick network of internal voids connected by a dynamical pathway through which gaseous ligands may migrate before reaching the solvent.[3,7,10,14,18,19,22]

In this regard, computer simulations represent the method of choice for studying this idea since they provide atomic-level information about protein thermal motions that may be difficult to obtain with experiments. Recent improvements,[24−31] have allowed molecular dynamics (MD) simulations to be extensively employed to study the migration of various ligands inside myoglobin.[14,19,20,23,31−34]

In particular, Bossa et al.[19] employed standard molecular dynamics simulations to obtain details about the diffusion of CO inside myoglobin, which suggested that protein fluctuations may be influenced by the presence of the ligand, which opens/closes the passage between adjacent cavities to promote its diffusion. In an effort to reach a more statistical description of this process, Onufriev and co-workers[23] performed a large number of simulations, and their results showed multiple paths for CO to enter/exit the protein. In order to achieve a quantitative free-energy description of these processes, other approaches have been applied.[14,31,32,34] For instance, Kato and co-workers[34] used metadynamics[29] to reconstruct a free-energy surface for CO migration, providing an overall picture that is in fair agreement with previous work of some of the authors.[31]

In order to obtain a general description of ligand pathways, Schulten and co-workers[14,32] proposed an implicit-ligand sampling method to study Mb and other members of the globin superfamily. Notably, they confirmed the importance of thermal motions for ligand migration through internal pathways, which were found to be mainly dependent on the amino acid sequence rather than on the secondary and tertiary structures. Moreover, they suggested that the chemical nature of the ligand might affect the free energies of cavity occupation.

In such an intricate scenario, the interesting picture emerges that myoglobin can work as a cyclic microscopic engine,[35] i.e., that the "breathing motions" of the internal cavities[21] are capable of pushing the ligand away from the heme binding site. However, two main open points need to be addressed: (i) the microscopic mechanism by which the protein regulates the internal migration of ligands and (ii) the influence of the ligand on the protein dynamics. These two issues raise the following unsolved question: does the ligand open its own way inside the protein, or is the dynamical path an *intrinsic* feature of myoglobin, or could a combination of these two factors (i.e., both induced and intrinsic behavior) be involved?[3,21]

In this work, we performed standard MD simulations of horse-heart met-myoglobin (h-MMb) with no ligand migrating inside the protein in order to identify a possible intrinsic contribution of the protein to ligand migration. This approach indeed ensures that the observed motions are due neither to the presence of a specific ligand (as necessary for some experiments; e.g., even xenon, considered a neutral probe, can affect the myoglobin structure)[33] nor to a bias term of the theoretical method but solely to spontaneous protein fluctuations. Since it is clear that internal cavities are crucial for the biological function of myoglobin, we focused our investigation on their structural and dynamical properties. The use of a statistical approach guided us to achieve an atomic-level comprehension of the intrinsic mechanism underlying the dynamics of the network of cavities, with special emphasis on the role of key residues in the ligand migration process.

Computational Details

Molecular Dynamics. As in our previous studies,[31,36] all-atom MD simulations were performed with the ORAC program,[37] employing the Amber95 force field[38,39] and TIP3P[40] for the protein (and heme group) and water, respectively. Horse-heart met-myoglobin (PDB code 1YMB at 1.90 Å)[41] was solvated in an initial orthorhombic water box with a side length of 70 Å and containing 6679 water molecules (∼23 000 atoms). In accordance with a previously employed procedure,[31] an initial slow heating from 10 to 250 K was carried out, after which a 60 ns state-of-the-art MD simulation was performed on the *NPT* ensemble at 300 K and 1.0 bar. SPME (64 grid points and order 5 with the direct cutoff at 10.0 Å) was used to treat long-range electrostatic contributions in combination with a multiple-time-steps algorithm.[42] The initial 12 ns was rejected, and the last 48 ns was used for the analysis.

Cavity Calculation and Cluster Analysis. The VOIDOO software[43,44] was employed to calculate the internal cavities, defined as those occupied by a spherical probe of radius 1.2 Å rolling on a predefined grid with dimensions of 0.2 Å. After the position and size of the internal cavities were roughly identified, a refinement was carried out by employing an iterative procedure: the grid size was progressively reduced until the volume V_i was found to be equivalent to volume V_{i-1} within a difference smaller than 1%. Such calculations were performed on both the X-ray structure and those obtained from the MD simulation. In the latter, structures were saved every 5 ps for a total of 9600 different conformers (corresponding to 48 ns). For each structure, the center of gravity of each cavity was determined with respect to an internal coordinate system centered on the heme group, along with its volume and the atoms of the residues constituting it. We note that such a procedure, carried out with as-fine-as-possible parameters, was the bottleneck of our study, as it was much more computationally expensive than the simulation itself. In order to provide a statistical description of the cavity dynamics, a cluster analysis was then applied to the large amount of data obtained from cavity calculation. This approach

(23) 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.
(24) Giudice, E.; Lavery, R. *Acc. Chem. Res.* **2002**, *35*, 350–357.
(25) Orozco, M.; Perez, A.; Noy, A.; Luque, F. J. *Chem. Soc. Rev.* **2003**, *32*, 350–364.
(26) Gumbart, J.; Wang, Y.; Aksimentiev, A.; Tajkhorshid, E.; Schulten, K. *Curr. Opin. Struct. Biol.* **2005**, *15*, 423–431.
(27) Adcock, S.; McCammon, J. *Chem. Rev.* **2006**, *106*, 1589.
(28) Marques, H.; Brown, K. *Coord. Chem. Rev.* **2002**, *225*, 123–158.
(29) Laio, A.; Gervasio, F. L. *Rep. Prog. Phys.* **2008**, *71*, 22.
(30) Vargiu, A.; Ruggerone, P.; Magistrato, A.; Carloni, P. *Nucleic Acids Res.* **2008**, *36*, 5910–5921.
(31) Ceccarelli, M.; Anedda, R.; Casu, M.; Ruggerone, P. *Proteins* **2008**, *71*, 1231–1236.
(32) Cohen, J.; Schulten, K. *Biophys. J.* **2007**, *93*, 3591–3600.
(33) Anedda, R.; Era, B.; Casu, M.; Fais, A.; Ceccarelli, M.; Corda, M.; Ruggerone, P. *J. Phys. Chem. B* **2008**, *112*, 15856–15866.
(34) Nishihara, Y.; Hayashi, S.; Kato, S. *Chem. Phys. Lett.* **2008**, *464*, 220–225.
(35) Agmon, N. *Biophys. J.* **2004**, *87*, 1537–1543.
(36) Ceccarelli, M.; Ruggerone, P.; Anedda, R.; Fais, A.; Era, B.; Sollaino, M. C.; Corda, M.; Casu, M. *Biophys. J.* **2006**, *91*, 3529–3541.
(37) Procacci, P.; Darden, T. A.; Paci, E.; Marchi, M. *J. Comput. Chem.* **1997**, *18*, 1848–1862.
(38) Cornell, W.; Cieplak, P.; Bayly, C.; Gould, I.; Merz, K.; Ferguson, D.; Spellmeyer, D.; Fox, T.; Caldwell, J.; Kollman, P. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
(39) Giammona, D. A. Ph. D. Thesis, University of California, Davis, CA, 1984.
(40) Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. *J. Chem. Phys.* **1983**, *79*, 926–935.
(41) Evans, S. V.; Brayer, G. D. *J. Mol. Biol.* **1990**, *213*, 885–897.
(42) Marchi, M.; Procacci, P. *J. Chem. Phys.* **1998**, *109*, 5194–5202.
(43) Kleywegt, G. J.; Jones, T. A. *Acta Crystallogr.* **1994**, *D50*, 178–185.
(44) Uppsala Software Factory. VOIDOO. http://xray.bmc.uu.se/usf/voidoo.html (accessed July 29, 2009).

was applied to the positions of cavities found along the simulation (9600 conformers), with an rmsd of 2.0 Å and minimum occurrence of 10%. As discussed in the Results, each cluster had a peculiar dynamical behavior that was carefully analyzed to determine (i) which residues formed the cavity and (ii) which residues were responsible for the contraction/expansion of its volume. Knowledge of the residues forming adjacent cavities led us to determine which atoms formed a passage, defined here as a *gate*. Subsequently, the *gate area* was estimated as follows: (i) the center of gravity of the atoms defining the gate was calculated; (ii) the mean gate-atom to center-of-gravity distance (r) was calculated; (iii) the gate area was approximated as the area of a circle with radius r; and (iv) the gate area was calculated for the entire simulation and plotted as a function of time. We are aware that the area calculated in this manner is an approximation and that its absolute value has no quantitative meaning. Nevertheless, a comparison of the gate areas provides a qualitative indication of the probability for a ligand to hop between adjacent cavities.[45] This approach allowed us to gather information about the mechanism of the network of dynamical cavities as well as to reveal possible intrinsic ligand pathways.

Results

As described in Computational Details, a standard MD simulation was performed for a total of 60 ns (of which the initial 12 ns was rejected and the remaining 48 ns was analyzed) using the crystallographic structure of h-MMb (PDB code 1YMB)[41] as the starting configuration. In line with our previous study,[31] the overall structure was preserved during the simulation: (i) the XRMS (the displacement of the whole protein from the X-ray structure) was 1.23 Å on average [Figure S1 in the Supporting Information (SI)], and (ii) the calculated thermal fluctuations were in good agreement with the experimental B-factors (Figure S2 in the SI) and close to the fluctuations calculated for sperm-whale myoglobin.[31]

Clusters of cavities determined by the use of the cluster analysis (see Computational Details) are defined hereinafter as *cavities* or *sites*. Figure 1 displays all 23 cavities found inside the protein matrix. For each of these, both the occurrence and the mean volume were calculated; these values are displayed in Figure 2 (cavities were labeled on the basis of their occurrence).

In the first instance, we determined the position of the xenon binding sites previously revealed by X-ray crystallography.[10] For these, we followed the nomenclature of Tilton et al.[10] (i.e., as found in sperm-whale met-myoglobin). The Xe cavities were numbered from 1 to 4 depending on the occupancies (e.g., cavity Xe1 is the most stable binding site for Xe).[10]

The occurrences of the Xe cavities were close to 100% (i.e., the Xe cavities were found in almost all of the computed configurations) in all cases except that of Xe3, whose occurrence was less than 40% (Figure 2). This finding is not surprising, since it has been shown that Xe3 is an elusive cavity that in some cases was not even found in the X-ray structures.[46,47]

In line with previous studies,[14,19,31,32] several *new* sites were also found that are not observed in the crystal structures but have occurrences and/or volumes comparable (or even larger)

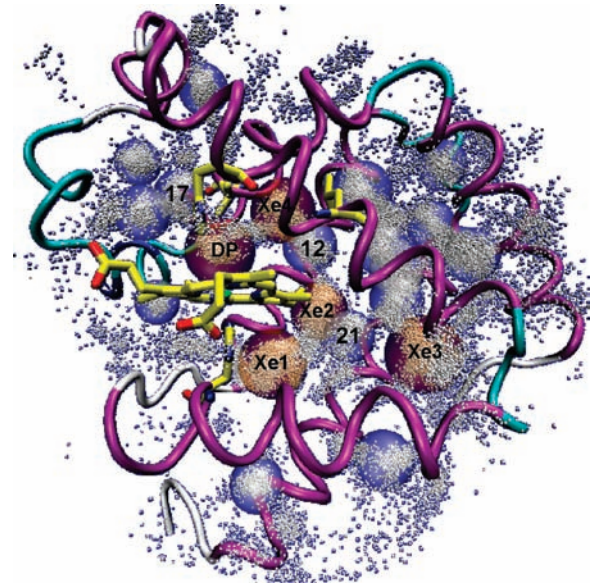

Figure 1. Myoglobin structure diplaying all of the cavities [large spheres; the distal pocket (DP), Xe1−Xe4, 12, 17, and 21 cavities are labeled] as well as all of the voids (small gray spheres) found using the VOIDOO software.[43,44]

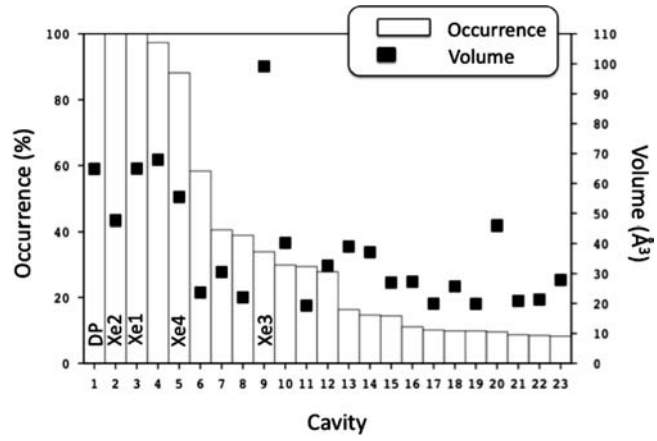

Figure 2. Occurrences and mean volumes of all 23 cavities found in the myoglobin structure. Cavities are labeled on the basis of their occurrence, e.g., cavity 1 occurred most and cavity 23 least. Cavities 10−23, whose occurrences were less than ∼30%, were defined as *transient* cavities, consistent with the previously proposed concepts of transient pathways and voids.[20,23,32]

than those of the Xe cavities (Figures 1 and 2). Some of these are close to the *phantom* voids reported previously,[19] while others appear in only a very few structures (e.g., cavities 10−23, which had occurrences less than ∼30%). We define these as *transient* cavities, consistent with the previously proposed concepts of transient voids and transient pathways.[20,23,32]

Because of the large amount of data, this work focused only on those cavities close to the heme region, i.e., the distal pocket (DP), Xe4, Xe2, Xe1, and cavities 12, 17 and 21, which belong to the major ligand pathway in myoglobin.[3,8,21,34,35,46,48−50] The

(45) Quantitative comparison of the gate areas and free-energy profile of a ligand moving inside the protein is not trivial. However, the trend of the average gate areas is in fair agreement with the activation free energies of CO migration inside horse-heart myoglobin obtained by Kato and co-workers[34] (e.g., the smaller the gate area, the larger the activation free energy, and vice versa).

(46) Teeter, M. M. *Protein Sci.* **2004**, *13*, 313–318.
(47) Tilton, R. F.; Singh, U. C.; Kuntz, I. D.; Kollman, P. A. *J. Mol. Biol.* **1988**, *199*, 195–211.
(48) Uchida, T.; Ishimori, K.; Morishima, I. *J. Biol. Chem.* **1997**, *272*, 30108–30114.
(49) Ishikawa, H.; Uchida, T.; Takahashi, S.; Ishimori, K.; Morishima, I. *Biophys. J.* **2001**, *80*, 1507–1517.
(50) Olson, J.; Soman, J.; Phillips, G. *IUBMB Life* **2007**, *59*, 552–562.

Table 1. Mean Volumes and Constituting Residues of the Studied Cavities

cavity	mean volume (Å^3)	residues[a]
DP	65	**29L^{I}**−32L−**33F^{I}**−**43F^{I}**−**64H^{I}**−**68V^{I}**−**107I^{III}**−heme
Xe1	65	89L−90A−93H−104L−**138F^{III}**−142I−**146Y^{III}**−heme
Xe2	48	72L−104L−**107I^{III}**−**108S^{III}**−111I−135L−**138F^{III}**−139R−heme
Xe4	55	**24H^{III}**−25G−28V−**29L^{I}**−65G−**68V^{I}**−**69L^{III}**−**107I^{III}**−111I−heme
12	33	**68V^{I}**−**69L^{III}**−72L−**107I^{III}**−111I−heme
17	20	**29L^{I}**−**33F^{I}**−**43F^{I}**−**46F^{I}**−60D−**61L^{III}**−**64H^{I}**−65G
21	21	**71A^{IV}**−72L−75I−86L−89L−**138F^{III}**

[a] Residues whose mutation is known to affect the kinetics of ligand diffusion inside the protein are rendered in bold. Superscripts refer to the classification scheme of Schulten and coworkers:[14] **group I**, distal pocket; **group III**, residues that line a constriction between adjacent cavities; **group IV**, residues found on the periphery of the protein that point toward the external solvent. Residues in **group II** belong to regions that are not discussed in this work.

mean volume (Table 1, Figure 2), the volume distribution,[51,52] and the temporal evolution of the volume (Figure 3) for each of these sites are described in detail. The dynamics of the cavities will be discussed further in terms of expansion/contraction and gate-area fluctuations[45] (area vs time plots are collected in Figure S3 in the SI), with special focus on those residues having a significant role in these processes. Table 2 lists all of the studied gates and the corresponding constituting atoms and mean areas.

Distal Region. This includes DP, Xe4, Xe2, and cavities 12 and 17 (Figure 1, Table 1). The dynamics of the entire region is dominated by the peculiar behavior of Xe4 (Figures 4 and 5), which undergoes an outstanding expansion/contraction in a stepwise manner; this stepwise process involves DP and cavity 12 as well. Detailed analysis of our data suggests that the Xe4 expansion/contraction is mainly regulated by the opening/closing of the DP/Xe4 and Xe4/12 gates, which in turn is determined by two amino acids, 29L and 69L, whose swinging motions around the Cα−Cβ bonds showed the most relevant trends among all the monitored residues. In the following paragraphs, the four-step Xe4 expansion/contraction mechanism is fully described.

Step 1: Opening of the DP/Xe4 Gate. The Xe4 volume increases from ∼40 Å^3 to an intermediate value of ∼60 Å^3 (Figure 4), mainly as a result of the opening of the DP/Xe4 gate, whose area increases by ∼30%. This in turn is caused by rotation of 29L, i.e., the N−Cα−Cβ−Cγ dihedral angle goes from about −70 to −150° (Figure S4 in the SI).

Thus, 29L is the main residue responsible for the opening of Xe4 toward the DP and also plays a role in the dynamics of gate DP/17. Upon rotation, the 29L side chain comes closer to 64H, thereby causing the DP/17 gate to close (Figure 5). The area fluctuations of the DP/17 and DP/Xe4 gates had a cross-correlation coefficient of −0.65 (corresponding to a medium-high anticorrelation),[53] which suggests that 29L can act as a switch for opening/closing of these two gates alternatively.

Step 2: Opening of the Xe4/12 Gate. As in the case of step 1, rotation of a leucine affects the opening/closing mechanism of a gate between two cavities. Here, the 69L N−Cα−Cβ−Cγ dihedral angle goes from about −70 to −140° (Figure S4 in the SI), which is the main cause of the opening of the Xe4/12 gate, whose area increases by ∼25%. Once the dihedral angles of both 29L and 69L reach their minimum values, Xe4 undergoes a further expansion, reaching a maximum volume of ∼80 Å^3. To this end, Xe4 is connected with the DP on one side and is open toward cavity 12 on the other, forming an outstanding channel that links these three cavities (Figure 5). This behavior is mirrored by the dynamical evolution of the Xe4 center of gravity, which shifts toward the DP and then toward Xe2 upon rotation of 29L and 69L, respectively (Figure S5 in the SI).

Interestingly, further analysis of the temporal evolution of the dihedral angles of 29L and 69L showed that the motion of these two amino acids is not completely independent, as the cross-correlation of the dihedral angles was 0.5, a value corresponding to a medium correlation.[53]

Step 3: Closure of the DP/Xe4 Gate. The mechanism of Xe4 cavity contraction mirrors that of the expansion. The Xe4 volume decreases from ∼80 to ∼60 Å^3, mainly because of 29L, which returns to its original value, thereby closing the Xe4/DP gate and in turn opening the DP/17 gate.

Step 4: Closure of the Xe4/12 Gate. Leucine 69L returns to its initial value, causing a further decrease in the Xe4 volume to the original value of ∼40 Å^3 via closure of the Xe4/12 gate (Figure 5).

Importantly, the opened/closed states of the Xe4 cavity are mirrored by the two distinct volume distributions shown in Figure 3. We also must point out that the discussion of the Xe4 expansion/contraction mechanism focused on the first 24 ns of the simulation. In the remainder, the Xe4 volume fluctuated around the value of ∼45 Å^3 until ∼40 ns, when another expansion began (Figure 3).

In regard to the 12/Xe2 gate, it is interesting that its area is mainly correlated with the torsion of another leucine, i.e., 72L [with a minor contribution by 111I (see below)]. In particular, when the 12/Xe2 gate area increases (by ∼30 and ∼60% in two different simulation time ranges), cavity 12 doubles in size (up to 50 Å^3) and the Xe2 occurrence drastically decreases (Figure S6 in the SI).

These data suggest a dynamical connection between Xe4 and Xe2 mediated by cavity 12, which is one of those transient cavities mentioned above. In other words, the volume of cavity 12 is swapped between Xe4 and Xe2, which means that this cavity acts as a sort of *volume carrier* that connects Xe4 and Xe2 (Figure 6). This was observed only twice during the simulation time (Figure S6 in the SI), suggesting that this connection may be a bottleneck of the migration from the distal region to the proximal region.

Proximal Region. This region is dominated by the dynamical behavior of 138F and three spatially close leucines, namely,

(51) It is worth mentioning that the volume distribution of all the studied cavities was found to be asymmetric, with a longer tail on the high-volume side. It was found that a log-normal distribution gave the best to fit these data (better than Gaussian-like functions). Indeed, previous studies have strongly suggested that such behavior is often adopted in nature (see, for instance, ref 52).

(52) Limpert, E.; Stahel, W. A.; Abbt, M. *Bioscience* **2001**, *51*, 341–352.
(53) Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed.; Lawrence Erlbaum Associates: Hillsdale, NJ, 1988.

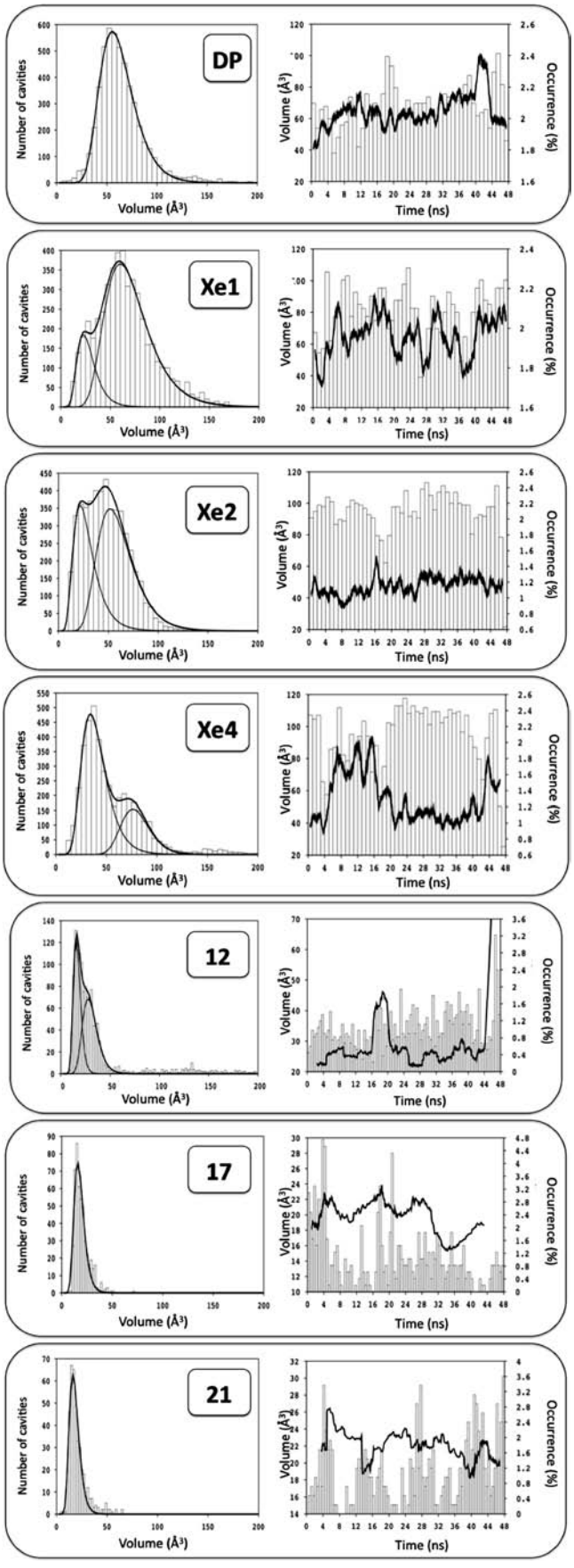

Figure 3. Volume distributions[51,52] and plots of volume (and occurrence) vs time for all of the studied cavities. Volumes are depicted as dashed lines and occurrences as vertical boxes.

Table 2. Mean Gate Areas[45] and Constituting Residues (Atoms)

gate	mean gate area (Å²)	residues[a] (atoms)
DP/17	30	**29L^{I}** (Cγ−Cδ1−Cδ2) **33F^{I}** (Cε1−Cε2−Cζ) **64H^{I}** (Cδ2−Nε2)
DP/Xe4	33	**29L^{I}** (Cγ−Cδ1−Cδ2) **68 V^{I}** (Cβ−Cγ1−Cγ2) **107I^{III}** (Cδ)
Xe4/12	42	**68 V^{I}** (Cβ−Cγ1) **69L^{III}** (N−Cβ−Cγ−Cδ1−Cδ2) **107I^{III}** (Cγ−Cδ)
12/Xe2	29	72L (Cδ2) **107I^{III}** (Cγ2) 111I (Cδ1) heme (CBB)
Xe2/Xe1	32	104L (Cγ−Cδ1−Cδ2) **138F^{III}** (Cζ−Cε1−Cε2) heme (CAB−CBB)
Xe2/21	22	72L (Cγ−Cδ1−Cδ2) **138F^{III}** (Cζ−Cε1−Cε2)
Xe1/21	33	9L (Cγ−Cδ1−Cδ2) **138F^{III}** (Cζ−Cε1−Cε2) heme (CMB)

[a] Residues whose mutation is known to affect the kinetics of ligand diffusion inside the protein are rendered in bold. Superscripts refer to the classification scheme of Schulten and coworkers:[14] **group I**, distal pocket; **group III**, residues that line a constriction between adjacent cavities; **group IV**, residues found on the periphery of the protein that point toward the external solvent. Residues in **group II** belong to regions that are not discussed in this work.

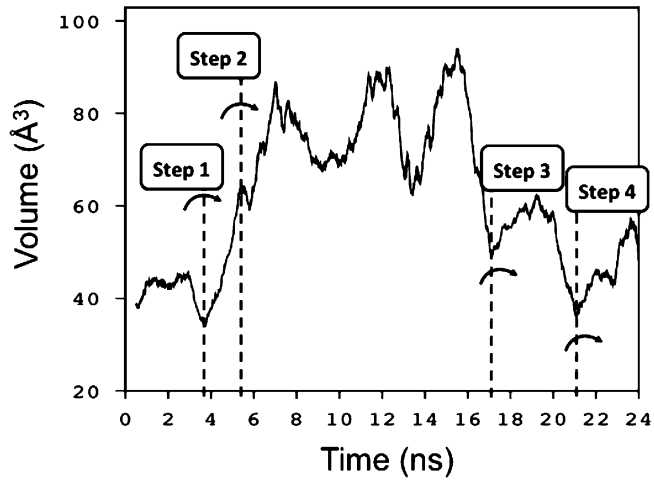

Figure 4. Xe4 volume as a function of time (the first 24 ns is shown). Steps 1−4 are highlighted (see the text for further details).

72L (helix E), 89L (helix F), and 104L (helix G), which point toward the aromatic ring of 138F (Figure 7).[54]

In particular, both the Xe2/12 and Xe2/21 gates are mainly regulated by 72L, whose dynamical behavior has been briefly discussed above. 72L rotation (Cγ−Cβ−Cα−C goes from ∼50 to ∼85°) causes the Xe2/12 gate area to increase by ∼20%. In rare events, this motion is coupled with that of 111I (Cγ1−Cβ−Cα−N goes from ∼85 to ∼20°), leading to a further increase up to 60%. As mentioned, 72L also influences the opening/closing mechanism for the Xe2/21 gate. In particular, the gate

(54) Though it is beyond the focus of this work, we will briefly mention that 135L regulates the gate between Xe2 and cavity 4, close to the phantom cavities previously observed. This is consistent with previous theoretical studies of ligand migration from the proximal region to the phantom region via a bottleneck close to Xe2 (see, for instance, refs 23 and 31).

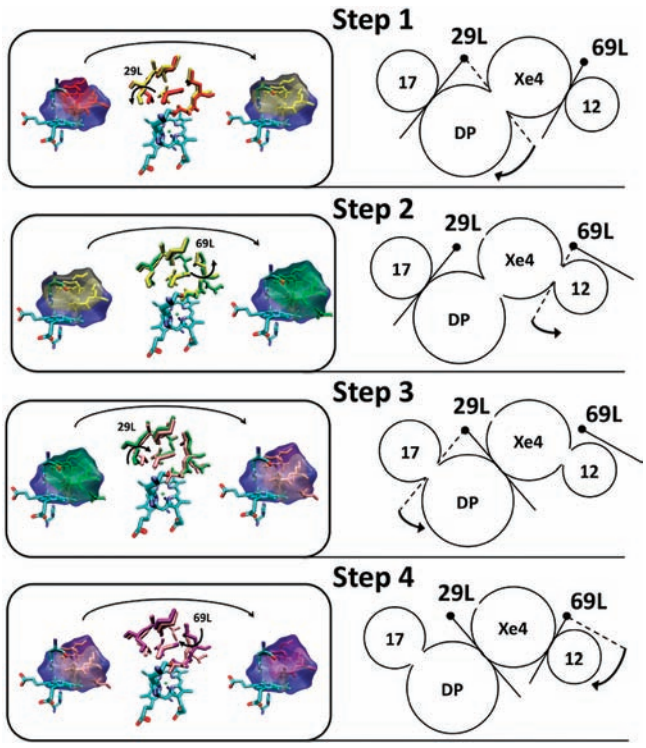

Figure 5. Xe4 contraction/expansion mechanism, illustrated by structures and pictorial schemes of steps 1−4. Step 1: 29L rotates, the Xe4 volume expands toward the DP, and the 17/DP gate closes. Step 2: 69L rotates, the Xe4 volume expands toward cavity 12, and a channel connecting these three cavities is formed. Step 3: 29L rotates back, the Xe4 volume decreases, and the 17/DP gate opens. Step 4: 69L rotates back, the Xe4 volume returns to its minimum value, and the gate with cavity 12 is closed.

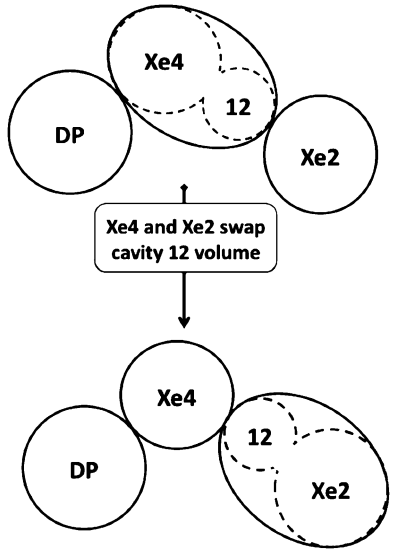

Figure 6. Pictorial view of dynamical volume exchange between Xe4, cavity 12, and Xe2: (i) Xe4 swallows cavity 12 volume; (ii) Xe4 releases cavity 12 volume to Xe2.

area correlates with the distance between the 72L methyl carbons and the 138F aromatic ring: when this distance is at its minimum (∼4 Å), the gate area is ∼18 Å^2, whereas when this distance is at its maximum (∼6 Å), the gate area is ∼26 Å^2. It is worth mentioning that the distances at the minima are consistent with *weak* C−H$\cdots\pi$ interactions.[55−58] Studies based on ab initio calculations are required to further investigate this point.

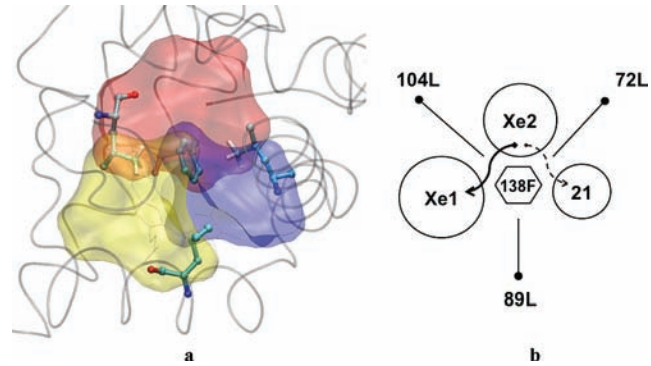

Figure 7. (a) View of the 3D spatial arrangement of 72L, 89L, and 104L around 138F in the proximal region of myoglobin. Color code: yellow, Xe1; red, Xe2; blue, cavity 21. (b) Schematic diagram of the two-way path from Xe2 to Xe1 and from Xe2 to 21.

A similar scenario was observed for the Xe1/21 and Xe2/Xe1 gate areas, which mainly depend on the distance from 138F to the 89L and 104L methyl carbons, respectively. The cross-correlation coefficients of these distances were calculated for all the possible couples (Table S1 in the SI), and the only significant anticorrelation (−0.77) found was between 138F−72L and 138F−104L, i.e., when the Xe2/Xe1 gate area is at its maximum, that of Xe2/21 is at its minimum value, and vice versa. In addition, the Xe2/Xe1 gate area is typically larger than that of Xe2/21, and cavity 21 is a void with a very low occurrence. All of these findings suggest, in line with previous studies,[3,21,32,50] that the direct passage from Xe2 to Xe1 is the most likely path. Since Xe1 is a preferential docking site for binding of gaseous ligands[3,10,33] and possibly is involved in myoglobin biological functions such as NO scavenging, Xe1 may be occupied.[12,13] Our results indicate that even if this happens, a passage from Xe2 to cavity 21 is still available for ligands to cross the proximal region, moving, for instance, from the distal region to Xe3 (Figure 7).[59]

In regard to the dynamical behavior of the Xe1 volume, analysis of its temporal evolution and gate areas confirmed that the volume depends mainly on the opening/closing of the gates connecting Xe1 to Xe2 and Xe1 to cavity 21 (Figure S7a,b in the SI). Interestingly, we found that proper torsions of 142I reduce the Xe1 cavity volume by rotating its side chain inside the cavity, i.e., Cδ_1 moves toward the center of gravity of the cavity, thereby reducing its volume by ∼20% (Figure S7c). The complex Xe1 dynamics, which depends on the 142I motions as well as on Xe1/21 and Xe2/Xe1 fluctuations, is mirrored by the two distinct volume distributions shown in Figure 3.

Discussion

We begin by remarking that the positions of the sites with the highest occurrence were found to be in agreement with previous studies reporting xenon[10] and phantom cavities.[19] In further support of the validity of this study, a significant number of key residues identified using our approach have been shown

(55) Wang, W.; Pitoňák, M.; Hobza, P. *ChemPhysChem* **2007**, *8*, 2107–2111.
(56) Shibasaki, K.; Fujii, A.; Mikami, N.; Tsuzuki, S. *J. Phys. Chem. A* **2006**, *110*, 4397–4404.
(57) Brandl, M.; Weiss, M. S.; Jabs, A.; Suhnel, J.; Hilgenfeld, R. *J. Mol. Biol.* **2001**, *307*, 357–377.
(58) Ran, J.; Wong, M. *J. Phys. Chem. A* **2006**, *110*, 9702–9709.
(59) Anselmi, M.; Dinola, A.; Amadei, A. *Biophys. J.* **2008**, *94*, 4277–4281.

to be relevant to the protein functioning (Tables 1 and 2).[3,8,14,22,48−50,60−63] For the sake of clarity, we have followed the scheme of Schulten and co-workers,[14] who used Huang and Boxer's experimental mutations[60] to classify these residues into four groups, depending on their effect and/or their position inside the protein (Tables 1 and 2).

Implicit Path from the DP to Xe1. First, we will discuss the delicate issue of the so-called "histidine gate".[5] Early experimental studies based on X-ray data suggested a peculiar role for 64H, whose twisting motion was indicated as the main one responsible for the opening/closing mechanism of the "histidine gate". This would allow a ligand to reach the distal pocket and eventually bind the iron center of the heme group. However, this process is still matter of study, as no conclusive results have been reported.[9,20,23,35,64] For instance, while Bossa[20] reported the spontaneous swinging of the distal histidine in Mb on a 10−100 ns simulation time scale, Schulten's shorter MD simulations of sperm whale and horse Mb did not show this phenomenon.[32] Perhaps this is not surprising, as it has been suggested that the "histidine gate" requires a complex motion of the protein (e.g., involving helix E and the CD turn) that may be completed on a time scale of hundreds nanoseconds.[3] Nevertheless, thanks to the reconstruction of the implicit ligand potential mean force maps, Schulten and co-workers were able to unveil a possible ligand escape path from the DP.[14,32] Our simulations fit in this scenario well, as they do not show the "histidine gate"[65] but rather indicate that a ligand may enter/exit the DP via a gating process regulated by rotation of 29L, in which 64H plays only a minor role. In particular, our results suggest that 29L is the residue mainly responsible for the DP/17 connection, in line with several experimental studies showing the following: (i) 29L acts as a physical barrier[48] as the ligand enters/exits the protein from the distal pocket; (ii) mutation of 29L can lead to significant changes in the overall binding kinetic rate,[8] inhibiting the ligand entrance inside the protein;[50] and (iii) mutations of 29L show stronger effects than mutations of 64H.[8]

Notably, our data add another intriguing feature to this important residue. As shown in Figure 5, 29L has the peculiar role of *switch*: it is capable of opening the DP/17 or DP/Xe4 gates exclusively (i.e., when the former is open, the latter is closed, and vice versa). This is consistent with previous experimental studies suggesting that geminate ligand recombination and thermal dissociation rates have an opposite correlation with the size of the residue at the 29-position, confirming that this residue plays a crucial role in both escape to the solvent and migration inside the protein.[8,35] More generally, experimental mutations of residues that according to our calculations belong to the DP/17 and DP/Xe4 gates are expected to affect the recombination kinetics as well as the migration rate through the protein matrix (Tables 1 and 2).[14,60]

Once the ligand enters Xe4, the expansion/contraction that this site undergoes may facilitate the diffusive motion of the ligand inside the protein matrix. In line with the recently proposed concepts of myoglobin as a cyclic engine[35] and the breathing motions of internal cavities,[21] we suggest that the Xe4 expansion/contraction may repeat during the protein dynamics, as this process is mainly regulated by the two residues 29L and 69L. While longer simulations are needed to confirm beyond doubt the *exact* sequence of the Xe4 opening/closing mechanism, the most important finding is that 29L and 69L rotations *can* regulate the DP/Xe4 and Xe/12 gates.

The Xe4 expansion/contraction revealed by our simulations is also consistent with the findings of Kato and co-workers,[34] who observed two different minima for Xe4, defined as Xe4(1), which is closer to the DP, and Xe4(2), which is closer to Xe2. In particular, our results suggest that a *mediated connection* may occur between Xe4 and Xe2, which are separated by one of the mentioned transient cavities: transient cavity 12 is first swallowed by Xe4 and then released to Xe2, thereby acting as a volume carrier linking these two Xe cavities (Figure 6). This scenario perfectly matches previous theoretical findings suggesting that Xe4−Xe2 passage is (i) the limiting step of the migration process of CO inside horse-heart Mb[34] and (ii) a higher energy path for O_2 inside horse-heart Mb.[32]

After entering the DP and crossing Xe4 and Xe2, the ligand may move to the proximal region, where the most striking feature observed in this work is the spatial arrangement of three leucines around 138F (Figure 7). According to our results, this 3D network may also have a specific role. In particular, a ligand coming from Xe2 may take one of two different paths: it may either go directly to Xe1 (the most likely path) or use cavity 21 as an intermediate docking site to continue, for instance, to Xe3.[59] This two-way path is mainly regulated by 138F together with both 72L and 104L, which account for the opening/closing of the Xe2/21 and Xe2/Xe1 gates, respectively. Notably, the role of 138F and 104L in the opening/closing mechanism of the Xe2/Xe1 gate is supported by two experimental observations: (i) mutation of 138F affects the internal migration rates[14,60] of gas ligands inside Mb (Tables 1 and 2), and (ii) L104T mutation decreases the Xe1 volume and possibly closes the way to Xe2.[66] In addition, these results can now explain why mutation of 71A, which belongs to cavity 21, affects geminate rebinding rates (Tables 1 and 2).[14,60]

Such an intriguing mechanism may be a crucial feature that allows MB to exert the role of NO scavenger, for which a sequential binding of two ligands, namely, O_2 and NO, is required.[12,13,67,68] Indeed, it has been suggested that in many heme proteins, such as myoglobins,[12,13] truncated hemoglobins,[68] and nonsymbiotic hemoglobin AHb1,[69] reaction of NO may be achieved via ligand docking to Xe1. Even if this site is occupied, our calculations show that there is an alternative path available for the ligand to migrate through the proximal region, for instance, to move from the distal region to Xe3.[59]

Role of Leucines. In agreement with Schulten's findings that leucines have the largest propensity to form O_2 pathways, we found *at least* one leucine in every gate between the studied

(60) Huang, X.; Boxer, S. G. *Nat. Struct. Biol.* **1994**, *1*, 226–229.
(61) Olson, J. S.; Mathews, A.; Rohlfs, R.; Springer, B.; Egeberg, K.; Sligar, S.; Tame, J.; Renaud, J.; Nagai, K. *Nature* **1988**, *336*, 265–266.
(62) Springer, B.; Egeberg, K.; Sligar, S.; Rohlfs, R.; Mathews, A.; Olson, J. S. *J. Biol. Chem.* **1989**, *264*, 3057–3060.
(63) Dantsker, D.; Roche, C.; Samuni, U.; Blouin, G.; Olson, J. S.; Friedman, J. M. *J. Biol. Chem.* **2005**, *280*, 38740–38755.
(64) Scott, E. *J. Biol. Chem.* **2001**, *276*, 5177–5188.
(65) We cannot exclude the possibility that this result was obtained because our model is based on a met-myoglobin. The water molecule bound to the iron center of the heme group may influence the histidine gate.
(66) Brunori, M. *Biophys. Chem.* **2000**, *86*, 221–230.
(67) Brunori, M.; Giuffrè, A.; Nienhaus, K.; Nienhaus, G. U.; Scandurra, F. M.; Vallone, B. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 8483–8488.
(68) Bidon-Chanal, A.; Martí, M. A.; Crespo, A.; Milani, M.; Orozco, M.; Bolognesi, M.; Luque, F. J.; Estrin, D. A. *Proteins: Struct., Funct., Bioinf.* **2006**, *64*, 457–464.
(69) Abbruzzetti, S.; Grandi, E.; Bruno, S.; Faggiano, S.; Spyrakis, F.; Mozzarelli, A.; Cacciatori, E.; Dominici, P.; Viappiani, C. *J. Phys. Chem. B* **2007**, *111*, 12582–12590.

cavities. For instance, 29L acts as a switch that exclusively opens/closes the DP/17 and DP/Xe4 gates; 69L is mainly responsible for functioning of the Xe4/12 gate; 72L was found in the 12/Xe2 gate; and the Xe2/Xe1, Xe2/21, Xe1/21 gates are regulated by a mechanism involving 104L, 72L, and 89L (with a minor role) together with 138F.

Clearly, more simulations and experimental studies are needed to prove beyond doubt that leucines can play the specific role of switches between cavities. Nevertheless, it has been experimentally shown that mutation of 68V and 107I (according to our calculations, these residues belong to the DP/Xe4, Xe4/12, and 12/Xe2 gates) to leucine significantly lowered the kinetic barrier for the ligand migration, mainly because of the flexibility of the isobutyl side chain rather than simply the size.[49,70] Finally, we calculated the conservation of the key leucines (i.e., 29L, 69L, 72L, 89L, 104L) across the myoglobin family using BLAST[71] (over 250 proteins). Interestingly, this equals ~80% on average, a value comparable to that for distal and proximal histidines and significantly higher than the average conservation of all 153 residues (~67%).

Conclusions

In this work, molecular dynamics simulations of horse heart met-myoglobin (with no migrating ligands inside the protein matrix) were performed to unveil an *intrinsic* dynamic path for ligands to migrate through the protein. Cavity calculations coupled with cluster analysis allowed us to gather atomic-level information about the dynamical network of cavities, thereby determining the key residues that play a role in ligand migration.

Our results show an impressive behavior of the Xe4 cavity, whose expansion/contraction may facilitate internal migration of ligands through the distal region. Consistent with the recently proposed concepts of cyclic engines and breathing motion of internal cavities, we propose that the Xe4 mechanism repeats during the protein thermal motions and is mainly controlled by the two residues 29L and 69L. In particular, 29L regulates the opening/closing of the 17/DP and DP/Xe4 gates and 69L the opening/closing of the Xe4/12 gate. Other leucines, namely 72L, 89L (with a minor role), and 104L, together with 138F, regulate the dynamical connections involving Xe1, Xe2, and cavity 21. Our results show a two-way path for a ligand to diffuse through the proximal region, possibly allowing an alternative route in case Xe1 is occupied. This feature may be crucial for the sequential binding of two gas molecules, as required for NO scavenging.

All of these findings suggest the role of *switch* for leucines in the connections between the cavities. Although further evidence is required to confirm the active role of this residue type in myoglobin (and perhaps in other proteins with similar functions), the importance of leucines in Mb function has already been shown by previous studies.[32,48−50] Indeed, a search across the Mb family carried out in this work confirmed that such crucial leucines (i.e., 29L, 69L, 72L, 89L, and 104L) are well-conserved over 250 analyzed sequences. Further simulations and experimental studies are required to conclusively prove these points, yet the intriguing hypothesis of leucines as switches between cavities presented in this work is strongly based on previous studies.

The novelty of our work is that we have been able to highlight an intrinsic dynamical path for possible ligand migration through the protein from the distal to the proximal region and to provide a statistical description of the network of cavities, with a particular focus on a relatively small number of residues that play a role in this process. Knowledge of this intrinsic path, i.e., the protein's fingerprint, is the first step toward the understanding the influence of (i) the ligand (and its chemical nature), (ii) the primary structure, and (iii) the myoglobin form (deoxy-, met-, oxy-) on the dynamics of the internal pathways. Now that the key residues have been identified, a further step is to investigate a way of controlling their motions, thereby tuning the mechanism of Xe4 and the connections among Xe1, Xe2, and the 21 cavity.

Finally, in light of the broad agreement achieved with previous studies, the general conclusion of this work is that cavity calculations coupled with cluster analysis may provide great support to the investigation of internal voids inside proteins and their fluctuations. This is of primary interest for the study of all those proteins for which experimental data are not yet available and whose function is based on an internal dynamical pathway.

Acknowledgment. Simulations were performed at the Caspur (Roma, Italy) and Cineca (Bologna, Italy). This work was supported by the Italian Ministry of University and Research (MIUR) PRIN2006 Project (2006059902_003). This study makes use of results produced by the Cybersar Project managed by the Consorzio COSMOLAB, a project cofunded by the MIUR within the Programma Operativo Nazionale 2000-2006 "Ricerca Scientifica, Sviluppo Tecnologico, Alta Formazione per le Regioni Italiane dell'Obiettivo 1 (Campania, Calabria, Puglia, Basilicata, Sicilia, Sardegna) Asse II, Misura II.2 Società dell'Informazione, Azione a Sistemi di calcolo e simulazione ad alte prestazioni" and Cosmolab (Cagliari, Italy) computer centers. M.C. thanks CNR for funding A.R. through the "Seed" Projects−INFM call for Young Researchers 2008, entitled "Binding/release of oxygen in hemoglobin through molecular simulations". We thank A. Giorgetti (University of Verona) for discussions.

Supporting Information Available: Figures S1−S7 and Table S1. This material is available free of charge via the Internet at http://pubs.acs.org.

JA9028473

(70) Quillin, M. L.; Li, T.; Olson, J. S.; Phillips, G. N.; Dou, Y.; Ikeda-Saito, M.; Regan, R.; Carlson, M.; Gibson, Q. H.; Li, H. *J. Mol. Biol.* **1995**, *245*, 416–436.

(71) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. *J. Mol. Biol.* **1990**, *215*, 403–410.