J|A|C|S ARTICLES

Published on Web 01/22/2010

Heme Proteins: The Role of Solvent in the Dynamics of Gates and Portals

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

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

Received November 19, 2009; E-mail: matteo.ceccarelli@dsf.unica.it

Abstract: Water plays a pivotal role in the correct functioning of proteins. Hydration is fundamental to their stability and flexibility, to folding process and specific functions, and to protein−protein interactions. In this work, the effects of solvation on proteins dynamics have been investigated by employing molecular dynamics simulations and using myoglobin as a model system. The investigation has been focused on solvent waters residing around/inside the protein, with average times of up to tens of nanoseconds, revealing that these *slow* waters may have significant effects on biological functioning of the protein. Our study pointed out that water is able to interact with proteins in diverse ways, leading to different kinds of perturbations in their intrinsic dynamic behavior. In particular, for myoglobin it was found that a water molecule can (i) "block" entry/escape of ligands to/from a particular docking site, (ii) act as a "wedge" modulating the dynamics of internal cavities, or (iii) join a "flow" of waters taking a ligand into (or "washing" a ligand away from) the protein interior. The information gathered in this work allowed us to provide a fingerprint of protein solvation state, the *hydration sites map*, which may represent a novel tool for comparing different forms/species of globular proteins.

Introduction

Water plays a pivotal role in many aspects of living organisms,[1] such as for correct behavior of proteins,[2,3] protein−protein interactions,[4,5] and membranes and DNA structure.[1,2,6] In particular, proteins are complex machines whose functions are strictly connected to their biological environment in which water molecules, as a solvent, have a major role.[7] Specifically, the hydration of protein surface is fundamental for the stability and flexibility of the protein itself, the folding process, molecular recognition, and enzymatic reactions.[1,2,6−12]

In this work, the chosen model is a small heme protein, myoglobin (Mb), the first protein solved at the atomic level and often employed to investigate the structure/function paradigm.[13−21] Concerning myoglobin structure and dynamics, it is now ascertained that migration of gaseous ligands (including O_2, NO, CO) from the solvent inside the protein (and vice versa) through an intricate network of cavities is crucial[18,22−29] for it to exert the biological roles of oxygen storage/transport and NO

[†] Department of Chemical Sciences, University of Cagliari.
[‡] Sardinian Laboratory for Computational Materials Science SLACS (IOM-CNR).
[§] Department of Physics, University of Cagliari.

(1) Makarov, V.; Pettitt, B. M.; Feig, M. *Acc. Chem. Res.* **2002**, *35*, 376–384.
(2) Ball, P. *Chem. Rev.* **2008**, *108*, 74–108.
(3) De Simone, A.; Dodson, G. G.; Verma, C. S.; Zagari, A.; Fraternali, F. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 7535–7540.
(4) Papoian, G. A.; Ulander, J.; Eastwood, M. P.; Luthey-Schulten, Z.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 3352–3357.
(5) Papoian, G. A.; Ulander, J.; Wolynes, P. G. *J. Am. Chem. Soc.* **2003**, *125*, 9170–9178.
(6) Pal, S. K.; Zewail, A. H. *Chem. Rev.* **2004**, *104*, 2099–2123.
(7) Zhang, L.; Wang, L.; Kao, Y.-T.; Qiu, W.; Yang, Y.; Okobiah, O.; Zhong, D. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 18461–18466.
(8) Wüthrich, K. *Angew. Chem., Int. Ed.* **2003**, *42*, 3340–3363.
(9) Kuntz, I. J.; Kauzmann, W. *Adv. Protein Chem.* **1974**, *28*, 239–345.
(10) Rupley, J. A.; Careri, G. *Adv. Protein Chem.* **1991**, *41*, 37–172.
(11) Teeter, M. M. *Annu. Rev. Biophys. Chem.* **1991**, *20*, 577–600.
(12) Tournier, A. L.; Jiancong, X.; Smith, J. C. *Biophys. J.* **2003**, *85*, 1871–1875.
(13) Springer, B.; Sligar, S.; Olson, J.; Phillips, G. *Chem. Rev.* **1994**, *94*, 699–714.
(14) Frauenfelder, H.; McMahon, B. H.; Fenimore, P. W. *Proc. Natl. Acad. Sci. U.S.A* **2003**, *100*, 8615–8617.
(15) Brunori, M.; Bourgeois, D.; Vallone, B. *J. Struct. Biol.* **2004**, *147*, 223–234.
(16) 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.
(17) Bellelli, A.; Blackmore, R.; Gibson, Q. H. *J. Biol. Chem.* **1990**, *265*, 13595–13600.
(18) Carlson, M.; Regan, R.; Gibson, Q. H. *Biochemistry* **1996**, *35*, 1125–1136.
(19) Gibson, Q. H.; Regan, R.; Elber, R.; Olson, J. S.; Carver, T. *J. Biol. Chem.* **1992**, *267*, 22022–22034.
(20) Elber, R.; Karplus, M. *J. Am. Chem. Soc.* **1990**, *112*, 9161–9175.
(21) Tilton, R.; Kuntz, I.; Petsko, G. *Biochemistry* **1984**, *23*, 2849–2857.
(22) Cohen, J.; Arkhipov, A.; Braun, R.; Schulten, K. *Biophys. J.* **2006**, *91*, 1844–1857.
(23) 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.
(24) Ostermann, A.; Waschipky, R.; Parak, F. G.; Nienhaus, G. U. *Nature* **2000**, *404*, 205–208.
(25) Dantsker, D. *J. Biol. Chem.* **2005**, *280*, 38740–38755.
(26) 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.
(27) Bossa, C.; Anselmi, M.; Roccatano, D.; Amadei, A.; Vallone, B.; Brunori, M.; Di Nola, A. *Biophys. J.* **2004**, *86*, 3855–3862.

 10.1021/ja909822d

scavenging,[15,22,30−32] as well as to prevent toxic action of other molecules such as CO. To study and understand the role of solvation in this relatively simple system,[14] it is certainly useful not only to shed light on processes that involve myoglobin but also, and more generally, to unveil the subtle interplay between a protein and the surrounding waters. As shown in a recent review, water confined in nonpolar cavities, as occurs in myoglobin, exhibits unusual properties.[33]

A first attempt to clarify the role of solvation on myoglobin functions was proposed by Pettitt and co-workers,[34] who in 2001 reported a ∼1-ns-long molecular dynamics simulation of sperm whale myoglobin. Previous investigations, despite the short simulation time, pointed out the importance of simulation analysis in rationalizing experimental data and provided a new picture of protein−solvent structure and dynamics.[35,36] The analysis of waters with residence times of less than 500 ps, down to tens of picoseconds, provided a detailed map of almost 300 hydration sites. More recently, Zhong et al.[7] experimentally confirmed the presence of many hydration sites at the myoglobin surface with solvent waters having residence times of up to 200 ps. In addition, NMR studies identified water residence times within the sub-nanosecond regime.[7,37,38] To date, however, a comprehensive systematic study of myoglobin hydration sites based on long-residence-time waters is still lacking. In this regard, water molecules buried in proteins should be considered an integral part of the folded structure. In fact, such structural water molecules can form strong H-bonds with polar groups of the surrounding amino acids and therefore are believed to tighten/loosen portions of the protein matrix.[39] This, in turn, should have important effects on protein dynamics and, hence, on protein functioning. Moreover, internal water molecules may be highly conserved in homologous proteins, as much as amino acidic residues.[11,39] Thus, investigations on protein hydration states and water−protein interaction should be regarded as complementary to single-point mutation studies.

Computer simulations have improved significantly in the past decade[40−47] and can now furnish atomic-level information about water−protein interactions on a relatively long time scale (see, for example, refs 3, 48−52), while the same may be not trivial to obtain with experimental techniques.[6,7,11,53] In this work, to complement our recent study on the dynamics of internal cavities of myoglobin,[54] the effect of solvation has been explored. In particular, a 60-ns-long molecular dynamics simulation of horse heart met-myoglobin was further analyzed, and the dynamics of waters residing around/inside the protein with average residence times of up to tens of nanoseconds was monitored. Together with the knowledge obtained previously,[54] the analysis of solvent revealed that myoglobin has in fact only a few stable hydration sites in which a water molecule can stay for times longer than 2 ns. Strikingly, all of these sites are close to protein/solvent *portals* observed in previous studies focused on the entry/escape and migration of various ligands in myoglobin.[46,55]

The information gathered in this work allowed us to build a fingerprint of protein solvation, the *hydration sites map*. This may be a reliable graphical tool to compare hydration states, for instance, of homologues, of different forms/species of the same protein, or of different proteins with a similar fold. Regarding the model employed here, this systematic investigation on the most stable water binding sites showed that solvent molecules strongly interact with myoglobin only in a few cavities. The analysis of water−protein interactions in these specific sites revealed that the solvent way of action is diverse, depending on the environment where the interaction takes place. In particular, water molecules have been found to (i) "block" ligands' entry/escape to/from a particular docking site, (ii) act as a "wedge" modulating the dynamics of internal cavities, or (iii) join a "flow" of waters taking a ligand into (or "washing" a ligand away from) the protein interior.

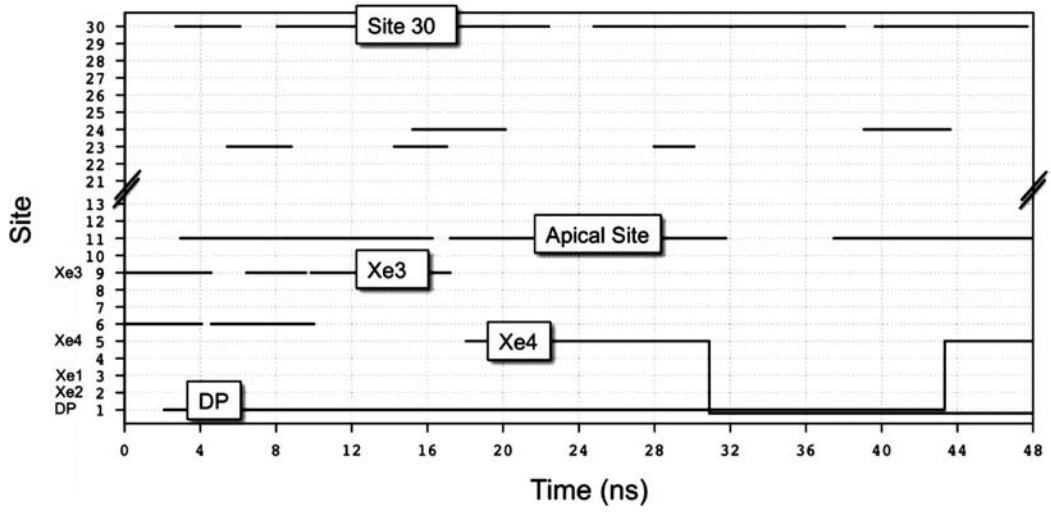
Finally and more generally, it can be speculated that other ways of action might also be possible. Thus, rethinking the buried waters as real protein residues, we believe that it may be very interesting to apply this approach to further study the role of solvation in the stability, flexibility, and dynamics of other globular proteins.

Computational Details

Molecular Dynamics. The analysis presented in this paper is based upon a molecular dynamics (MD) simulation of horse heart met-myoglobin that has been recently reported.[54] As in our previous studies,[46,56] all-atom MD simulations were performed with the ORAC program,[57] employing the Amber95 force field,[58,59] and TIP3P[60] for the protein (and heme group) and water, respectively. Myoglobin (horse heart met-myoglobin, PDB code 1YMB at 1.90 Å)[61] was solvated in an initially orthorhombic water box of 70 Å side length with 6679 water molecules (about 23 000 atoms). We

(28) 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*, 2612–2616.
(29) Dantsker, D.; Samuni, U.; Friedman, J. M.; Agmon, N. *BBA−Proteins Proteom.* **2005**, *1749*, 234–251.
(30) Cossins, A.; Berenbrink, M. *Nature* **2008**, *454*, 416–417.
(31) Brunori, M. *Trends Biochem. Sci.* **2001**, *26*, 209–210.
(32) Flögel, U.; Merx, M.; Gödecke, A.; Decking, U.; Schrader, J. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 735–740.
(33) Rasaiah, J. C.; Garde, S.; Hummer, G. *Annu. Rev. Phys. Chem.* **2008**, *59*, 713–740.
(34) Makarov, V.; Andrews, B.; Smith, P.; Pettitt, B. M. *Biophys. J.* **2000**, *79*, 2966–2974.
(35) Lounnas, V.; Pettitt, B. M.; Phillips, G. N., Jr. *Biophys. J.* **1994**, *66*, 601–614.
(36) Lounnas, V.; Pettitt, B. M. *Proteins* **1994**, *18*, 133–147.
(37) Otting, G.; Liepinsh, E.; Wuthrich, K. *Science* **1991**, *254*, 974–980.
(38) Wuthrich, K.; Billeter, M.; Guntert, P.; Luginbuhl, P.; Riek, R.; Wider, G. *Faraday Discuss.* **1996**, *103*, 245–253.
(39) Fischer, S.; Verma, S. C. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9613–9615.
(40) Giudice, E.; Lavery, R. *Acc. Chem. Res.* **2002**, *35*, 350–357.
(41) Orozco, M.; Perez, A.; Noy, A.; Luque, F. J. *Chem. Soc. Rev.* **2003**, *32*, 350–364.
(42) Gumbart, J.; Wang, Y.; Aksimentiev, A.; Tajkhorshid, E.; Schulten, K. *Curr. Opin. Struct. Biol.* **2005**, *15*, 423–431.
(43) Adcock, S.; McCammon, J. *Chem. Rev.* **2006**, *106*, 1589–1615.
(44) Marques, H.; Brown, K. *Coord. Chem. Rev.* **2002**, *225*, 123–158.
(45) Vargiu, A.; Ruggerone, P.; Magistrato, A.; Carloni, P. *Nucleic Acids Res.* **2008**, *36*, 5910–5921.
(46) Ceccarelli, M.; Anedda, R.; Casu, M.; Ruggerone, P. *Proteins* **2008**, *71*, 1231–1236.
(47) Nishihara, Y.; Hayashi, S.; Kato, S. *Chem. Phys. Lett.* **2008**, *464*, 220–225.
(48) Cheng, Y.-K.; Rossky, P. J. *Nature* **1998**, *392*, 696–699.
(49) Smolin, N.; Winter, R. *J. Phys. Chem. B* **2004**, *108*, 15928–15937.
(50) De Simone, A.; Dodson, G. G.; Fraternali, F.; Zagari, A. *FEBS Lett.* **2006**, *580*, 2488–2494.
(51) Sterpone, F.; Ceccarelli, M.; Marchi, M. *J. Mol. Biol.* **2001**, *311*, 409–419.
(52) Marchi, M.; Sterpone, F.; Ceccarelli, M. *J. Am. Chem. Soc.* **2002**, *124*, 6787–6791.
(53) 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.
(54) Scorciapino, M. A.; Robertazzi, A.; Casu, M.; Ruggerone, P.; Ceccarelli, M. *J. Am. Chem. Soc.* **2009**, *131*, 11825–11832.
(55) 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.
(56) Ceccarelli, M.; Ruggerone, P.; Anedda, R.; Fais, A.; Era, B.; Sollaino, M. C.; Corda, M.; Casu, M. *Biophys. J.* **2006**, *91*, 3529–3541.

Chart 1[a]



[a] Lines represent the presence of a *slow* water (with an average residence time longer than 2 ns) in a certain site (*y*-axis) at a given time (*x*-axis). Note that the water directly bound to the heme iron, which characterizes the myoglobin met-form, is not considered. Sites (*y*-axis) are labeled on the basis of their occurrence as reported previously.[52] Site 30 is not a "proper" cavity (cavity calculation does not find it) but a surface site. The *y*-axis has been cut for graphical reasons, since no slow waters were found in a wide range, between 12 and 22. Sites occupied for more than 30% of the simulation time are noted with an extra label in the graph. We also stress that *fast* waters are not displayed in this graph.

would like to remark that no water molecules were initially placed inside the protein matrix (excluding that directly bound to the heme-iron which characterizes the myoglobin met-form). Following a procedure previously employed,[54] an initial slow heating from 10 to 250 K was carried out. Then, 60 ns state-of-the-art MD simulation was performed in the NPT ensemble ($N = 22524$ atoms, $P = 1.0$ bar, $T = 300$ K). The Soft-Particle-Mesh-Ewald algorithm (64 grid points and order 5 with direct cutoff at 10.0 Å) was used to treat long-range electrostatic contributions in combination with a multiple-time-step algorithm.[63] The initial 12 ns were rejected, and the last 48 ns were used for the analysis.

As described elsewhere,[54] the VOIDOO software[62] was employed to identify and characterize the internal cavities of the protein. Cavity calculations were performed on both the starting X-ray structure and the ones obtained from the MD simulation. In order to provide a statistical description of the cavities' dynamics, a cluster analysis was applied to the large amount of data obtained. Each cluster showed a peculiar dynamic behavior that was carefully analyzed to determine the residues that constitute the cavity and, among them, those that mainly influence volume fluctuations. This in turn allowed us to determine which atoms form a passage between two cavities, defined as a *gate*,[54] or between the protein and the solvent, defined as a *portal*, consistent with previous work by Onufriev et al.[55] Finally, the *area* of these passages was estimated and plotted as a function of time as described previously.[54] Although estimation of *gate* area does not have a quantitative character, we would like to remark that comparison among the *gate* areas gives a qualitative indication of the relative probability for a ligand to hop between adjacent cavities.[54]

Water Analysis. The analysis of solvent molecules was performed employing the dedicated code implemented in the ORAC program.[57] As previously shown,[51,52] all the solvent molecules were analyzed to determine their interactions with the protein in terms of the time spent within a specified distance (1.1 times the sum of the van der Waals radii) of any atom of the protein without discontinuity, allowing a maximum interruption of 2 ps. The number of contacts with the protein and the duration of each contact (the residence time) were evaluated for all the water molecules. This allowed us to discriminate between two types of events: short water−protein interactions (rapid exchange with the solvent) and long interactions (slow diffusion on the surface or water trapped in a site). This analysis resulted in a list of all the water molecules with a non-zero residence time and a list of all the protein residues visited by a certain water. Since we were interested in those solvent molecules that interact with the protein for a relatively long time, we focused on those waters with residence times longer than 2 ns. These are hereinafter defined as *slow* waters. Twenty-three water molecules were found that satisfied these criteria, and their trajectory was analyzed in terms of (i) which of the internal sites was occupied, (ii) for how long this occurred, and (iii) the time range of the simulation in which the occupation event took place. All this information was used to build a plot showing the hydration sites of the protein and the occurrence of their hydrated state along the MD simulation. This plot is defined as a *hydration sites map*.

Additionally, the dynamics of *fast* waters (with average residence time shorter than 2 ns) was analyzed. Taking advantage of the information produced in our recent work about the position and volume of all the cavities of myoglobin,[54] we computed the distance *d* between the center of gravity of each water molecule and that of each cavity in all the structures obtained during the MD simulation. A water molecule was thus considered to be inside a cavity if *d* was smaller than the radius of a sphere corresponding to the volume of the cavity in that simulation frame. The number of water molecules was then plotted as a function of time, and the average was employed to complement the information about the opening/closing mechanism of *gate*s and *portal*s, the volumes of cavities, and the dynamics of key residues.

Results and Discussion

Hydration Sites Map. Following the procedure described in the Computational Details, a map of myoglobin hydration sites (hereinafter referred to as a *hydration sites map*) was built for those waters with an average residence time longer than 2 ns

(57) Procacci, P.; Darden, T. A.; Paci, E.; Marchi, M. *J. Comput. Chem.* **1997**, *18*, 1848–1862.
(58) 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.
(59) Giammona, D. A. Ph.D. Thesis, University of California, Davis, 1984.
(60) Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. *J. Chem. Phys.* **1983**, *79*, 926.
(61) Evans, S. V.; Brayer, G. D. *J. Mol. Biol.* **1990**, *213*, 885–897.
(62) Kleywegt, G. J.; Jones, T. A. *Acta Crystallogr.* **1994**, *D50*, 178–185.
(63) Carver, T. E.; Rohlfs, R. J.; Olson, J. S.; Gibson, Q. H.; Blackmore, R. S.; Springer, B. A.; Sligar, S. G. *J. Biol. Chem.* **1990**, *265*, 20007–20020.

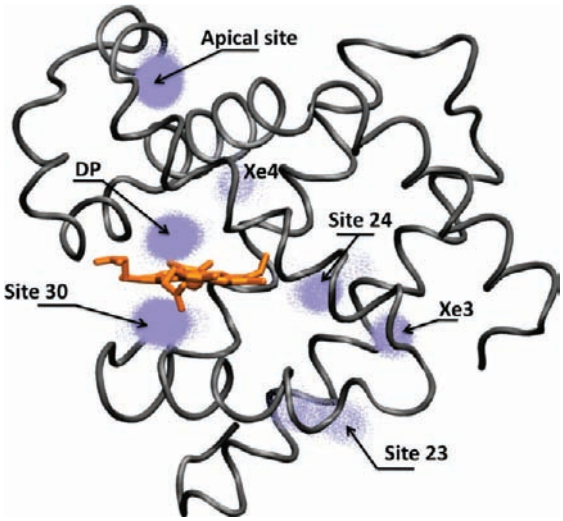


Figure 1. Myoglobin hydration sites. Blue points represent the position of oxygen atoms of those water molecules binding the protein with total occupation times longer than 2 ns (see Chart 1, Table S1, and Computational Details). The thickness of the cloud roughly correlates with the total occupation time. Due to poor occupation time, site 6 (between helices A and H, Table S1) is not shown.

(Chart 1). All the cavities found in the protein matrix are displayed on the *y*-axis (labeled on the basis of their occurrence as reported previously;[54] i.e., cavities 1 and 29 are the most and the least occurring cavities, respectively). Although site 30 is not a proper cavity (calculation does not identify this site as a cavity), water analysis revealed that this surface site is particularly favorable for the binding of a water molecule (see below), and it was thus included in the *hydration sites map*. We would like to remark that only the water molecule directly bound to the heme iron (myoglobin met-form) was placed inside the protein matrix in the initial structure used for simulation.

A first glance at Chart 1 reveals that only a few sites, i.e., distal pocket (DP), Xe3, Xe4, 6, 11, 23, 24, and 30, host *slow* water molecules, with five water molecules on average simultaneously inside the protein matrix (Figure 1). Among these sites, DP, Xe3, Xe4, site 30, and site 11 (named also as the *apical site*) are hydrated for a time longer than 30% of the entire simulation.

In order to fully characterize these hydration sites, the structural parameters of the most representative H-bonds formed by the water molecule with the residues of the cavity were collected and reported in Table S1.

DP hosts at least one solvent water for almost 100% of the simulation time (Figure 2A), stabilized by two H-bonds such as (wat)O−H···O(wat-Fe) and (wat)O···H−N$_\epsilon$(64H). Notably, in the time range 30−44 ns, DP hosts two water molecules, the second one entering from Xe4 (Figure 2B). Around 44 ns, one of these waters moves from DP to Xe4, where solvent molecules are present for about 30% of the simulation, forming (wat)-O−H···O(25G) and (wat)O−H···O(65G) H-bonds that stabilize the water molecule inside this internal cavity.

Xe3, a site relevant for ligand entry/escape,[46,55] contains a water molecule in three different simulation times, bound to the protein through H-bonds such as (wat)O−H···O(75I), (wat)O−H···O(13 V), and (wat)O···H−N(80G) (Table S1, Figure 1). Interestingly, after ∼16 ns (Chart 1), no *slow* waters were found. Such a surprising result inspired further investigation into the dynamics of Xe3, as discussed in detail in the following sections.

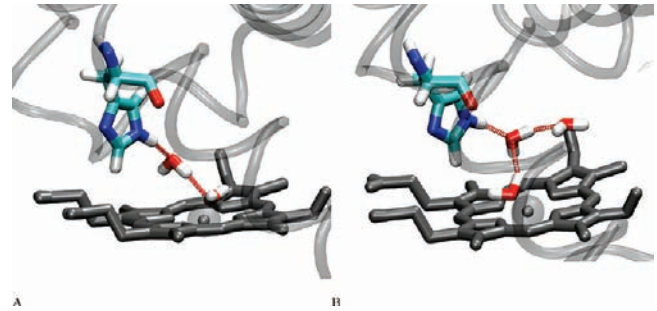


Figure 2. View of DP with (A) one solvent water (time ranges 2−30 ns and 44−48 ns) and (B) two solvent waters (time range 30−44 ns), other than that directly bound to the heme iron.

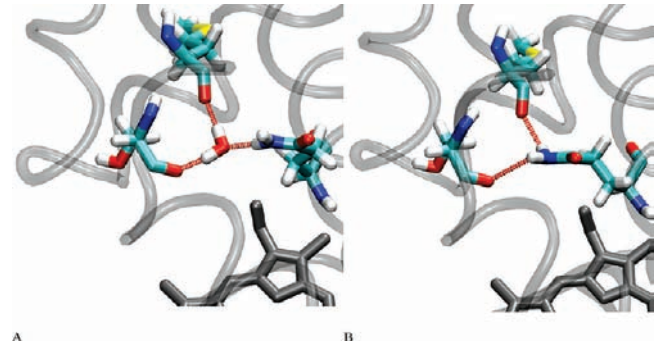


Figure 3. View of the apical site (A) with water involved in a hydrogen bond network with 26Q, 55M, and 58S and (B) without water, with 26Q forming two hydrogen bonds with 55M and 58S.

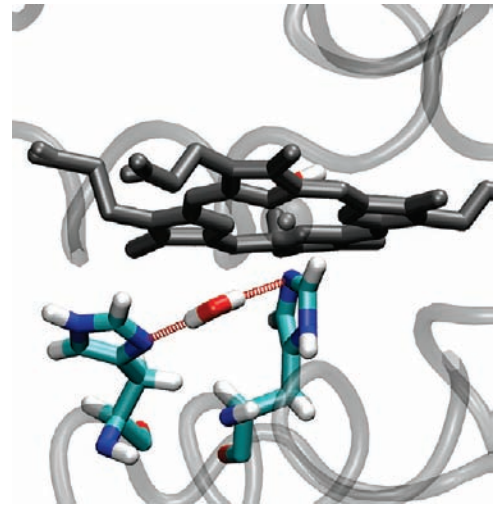
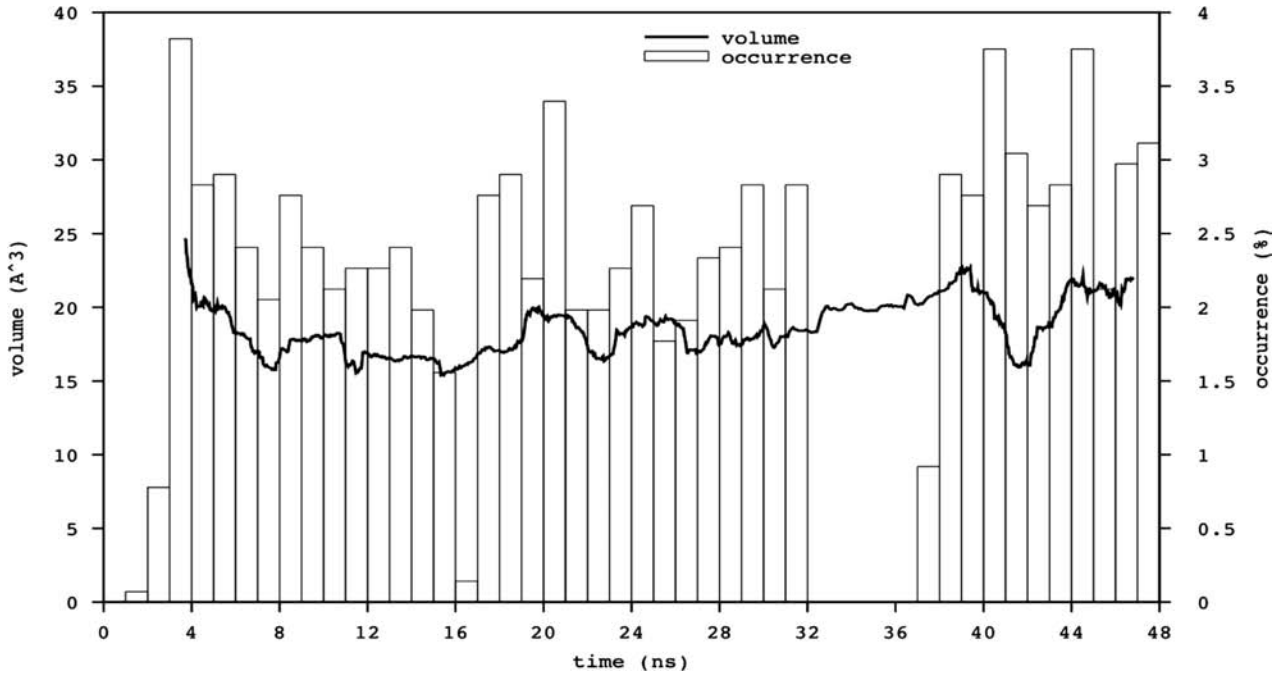


Figure 4. View of site 30. A water molecule is H-bonding 93H (proximal histidine) and 97H.

Similarly, the apical site (site 11) is relevant for ligands to enter/escape the protein (Figure 1).[55] In particular, cavity calculation carried out in our previous work[54] showed that this site may connect the external solvent to DP through cavities 17 and 8. In particular, the apical site hosts waters with long occupation times (about 12 ns each) for about 80% of the entire simulation, each water being stabilized by three stable H-bonds such as (wat)O−H···O(55M), (wat)O−H···O(58S), and (wat)O···H−N(26Q) (Table S1, Figure 3A).

Site 30 (Figures 1 and 4), located under the heme group, contains a water molecule for more than 80% of the simulation time. Each water remains in the site for about 4 ns, forming two hydrogen bonds, (wat)O−H···N$_\delta$(93H) and (wat)O−

Chart 2[a]



[a] Volume and occurrence of the apical site (cluster 11) are plotted as function of simulation time. Average volume remains almost constant during the whole simulation. Remarkably, the occurrence of the apical site falls to zero in all three time regions where no water molecules were found to be trapped in this site (Chart 1).

H···N(97H) (Table S1, Figure 4). We would like to remark that 93H is the proximal histidine, bound to the heme iron.

Waters were also found (for about 20% of the simulation time) in sites 23 and 24, which are small cavities with a poor occurrence (Figure 1). Their hydration may be of relevance since the former is located close to Xe3 between helices H and F, while the latter is close to Xe2 between helices G and H. Remarkably, sites 23 and 24 are close to *portals* 6 and 2 reported by Onufriev et al.,[55] where a total of nine events were observed: six escapes and three entries.

Finally, waters occupied for about 20% of the simulation time two slightly different positions in site 6, located between helices A and H, close to residues 119 and 127 (Table S1).

The systematic analysis of the hydration sites revealed different roles of solvent waters while interacting with the protein. In the following, detailed discussion focuses on the characterization of selected sites.

Distal Pocket: Waters as "Block". One water molecule enters DP around 2 ns, taking advantage of a slight displacement of 64H, which allows it to penetrate the pocket, where it is stable for almost the 100% of the simulation time, forming H-bonds with 64H and the water bound to iron (Table S1, Figure 2A). This finding is in agreement with previous studies that suggest a *portal* through residues 64, 67, and 154.[55] In addition, extensive experimental work[63,64] has shown that the polarity of 64H (more than its size) affects kinetic barriers to ligand entrance. According to our results, this may be caused by the ability of the distal histidine to form hydrogen bonds with a solvent water molecule, which can "block" ligand entrance. In fact, it has been suggested that binding of a ligand to the myoglobin iron center is only possible after expulsion of a water molecule stabilized by H-bonding to 64H.[65] Indeed, the comparison of several distal pocket mutants revealed an inverse correlation between water occupancy and ligand entry/escape rate.[66] The presence of a water molecule (other than that directly bound to the iron in the met-form) decreases the rate of ligand entry and lowers the overall affinity for ligands.[53]

In the specific case of the met-form, the presence of the water bound to the iron center may favor the entrance of other waters by providing the possibility of forming H-bonds that stabilize possible clusters. This hypothesis is also supported by a recent experimental study showing that the entropic barrier for a second water molecule to enter the distal pocket may be reduced by the presence of a first one.[67] Indeed, our simulation shows that DP is the only cavity where more than one *slow* water molecule can be present at the same time, i.e., two solvent waters and that bound to the heme iron (Figure 2B).

Apical Site (Site 11): Waters as "Wedge". Chart 2 displays the occurrence and the volume of the apical site as a function of time. The occurrence falls almost to zero in three simulation times, 0−2.5, 16−17, and 32−37 ns. Interestingly, these times perfectly overlap with those of Chart 1, showing the absence of a water molecule inside the cavity; i.e., when the apical site exists (its volume fluctuating around a volume of 20 Å^3), it contains a water molecule. In order to explore this point, a deep analysis of the site has been carried out. Among many structural parameters examined, the distance between three residues accounts for the dynamical behavior of the cavity, i.e., residues 26Q, 55M, and 58S. In particular, when the occurrence is close to zero, the distance between (26Q)N and (55M)O equals 3.2 Å on average. At this time, two hydrogen bonds are formed,

(64) Egeberg, K. D.; Springer, B. A.; Sligar, S. G.; Carver, T. E.; Rohlfs, R. J.; Olson, J. S. *J. Biol. Chem.* **1990**, *265*, 11788–11795.

(65) Brunori, M. *Biophys. Chem.* **2000**, *86*, 221–230.

(66) Similar results were observed for several gaseous ligands' binding.

(67) Goldbeck, R. A.; Bhaskaran, S.; Ortega, C.; Mendoza, J. L.; Olson, J. S.; Soman, J.; Kliger, D. S.; Esquerra, R. M. *Proc. Nat. Acad. Sci. U.S.A.* **2006**, *103*, 1254–1259.

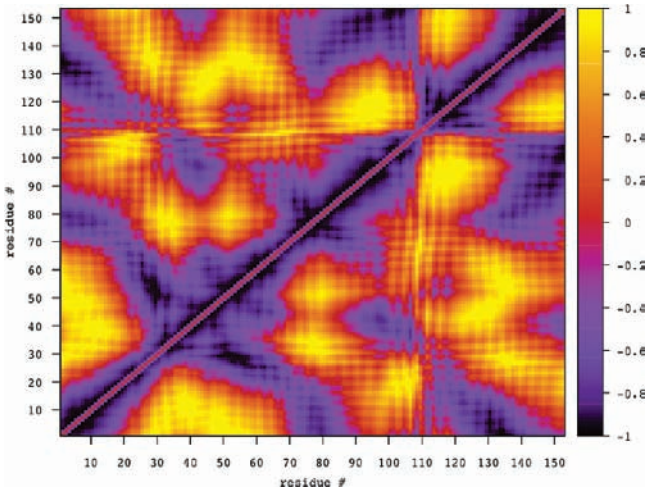


Figure 5. Correlation map of Cα fluctuations.

such as (26Q)N−H···O(55M) and (26Q)N−H···O(58S) (Figure 3B). When a water enters the site, acting as a sort of "wedge", the distance between these residues increases and a network of hydrogen bonds involving the water molecule is formed, such as (26Q)N−H···O(wat), (wat)O−H···O(55M), and (wat)O−H···O(58S), as displayed in Table S1 and Figure 3A. From our data it is not clear whether the molecule enters the protein and creates the cavity or, instead, the cavity opens and then a water molecule can slip inside. Yet, it is evident that the motions of residues 26Q, 55M, and 58S crucially depend on the presence of a water molecule, not only influencing the above-mentioned distances but also resulting in an anticorrelated displacement of their Cα's (Figure S1A). These motions are then transmitted to a large part of helices B, D, and E (Figure S1B). Indeed, the correlation map (Figure 5) shows a wide spot of anticorrelation between the residues 26 and 58, in a region comprising the residues from 18 to 30 and from 54 to 66. Notably, among these residues, 29L was recently shown to have a specific role in myoglobin.[54] Rotation of this amino acid around the dihedral N−Cα−Cβ−Cγ angle regulates the DP/Xe4 and 17/DP connections, i.e., 29L acts as a *switch* between these two *gate*s.[54] Further analysis of the trajectory revealed that an anticorrelation exists for this dihedral angle and the distance between Cα's of 29L and 55M, the cross-correlation coefficient being equal to −0.76 (Figure S2A). Thus, when the two helices approach because one water molecule leaves the apical site, 29L stays in the configuration corresponding to the close state of DP/Xe4 *gate*. When the water molecule enters the apical site and the helices go away from each other, 29L has the possibility of rotating, causing the DP/Xe4 to open. It is thus not surprising that a positive correlation exists between the DP/Xe4 *gate* area and 29−55 Cα distance (the cross-correlation coefficient is 0.60, corresponding to a medium-high correlation,[68] Figure S2B). It is worth remarking that the region of the apical site has already been indicated in previous studies as one possible *portal* connecting the solvent with the protein interior.[54,55] Our work also suggests that its hydration state may influence the dynamics of the *gates* between the main cavities of the distal region, thus affecting protein functioning. This hypothesis is supported by experimental studies which have shown that single-point mutation of 26Q (one of the residues belonging to the apical site) affects ligand migration kinetics in myoglobin.[22,69]

Xenon-3 Site: Waters as "Flow". As for the apical site, the analysis of Xe3 hydration and volume revealed interesting features that deserve a detailed description. As mentioned above, from ∼16 ns to the end of the simulation, no *slow* waters were found in Xe3. The analysis of cavity dynamics revealed that the volume doubles in the time range 17−24 ns, from an average value of ∼90 Å^3 to a peak of ∼190 Å^3 (Figure 6A). This occurs in only ∼7 ns, the contraction that follows being significantly slower (∼15 ns). Moreover, the cavity occurrence is inversely proportional to the volume, i.e., as volume increases, occurrence drops, suggesting that the cavity expands toward the external aqueous solution. Following our previous study,[54] the *portal* through which ligands may enter/escape Xe3 was determined. According to our results, residues Trp7, 79K, 80G, 82H, 137L, and 141D constitute it. Interestingly, experimental studies have shown that mutation of 137L affects the ability of the ligand to enter/escape the protein,[22,69] supporting our findings. Figure 6B,C shows the residues forming the *portal* before/after the opening of the cavity, which act as a sort of shutter to let molecules in/out of the protein. The *portal* area plotted as a function of time closely follows the trend of the volume of Xe3 (with a cross-correlation coefficient of +0.83, Figure S3A). These findings confirm that Xe3 opens toward the solvent, providing a clear rationale for the fact that this cavity is hardly found in X-ray structures and is often referred to as *vanishing*.[70,71] In addition, our results suggest that water molecules can enter the site when the cavity opens. This would also explain the reason why no stable water molecules can be found in Xe3 after 16 ns. To demonstrate this point, all the waters were analyzed irrespective of their residence times (see Computational Details). The number of waters inside Xe3 as a function of time is displayed in Figure S3B. As one may expect, the trend of this follows that of Chart 1 for times shorter than 16 ns. After this time (when the portal opens), a number of *fast* water molecules, with a maximum of seven (or four on average), enter the cavity. The average number of water molecules as a function of time closely follows the *portal* area trend, confirming the intuitive idea that a higher number of fast-exchanging water molecules can enter/escape the protein matrix when the Xe3 *portal* opens. The cross-correlation coefficient of the two curves is remarkably high, +0.73. Thus, data presented here reveal the mechanism through which Xe3 cavity exerts its function as *portal*, allowing a "flow" of water molecules to enter/escape the protein.[33,72]

Summary and Conclusions

In this work, molecular dynamics simulations of horse heart met-myoglobin were performed to assess the role of solvation in myoglobin functions. A *hydration sites map* of the protein

(68) Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed.; Lawrence Erlbaum Associates: Hillsdale, NJ, 1988.
(69) Huang, X.; Boxer, S. G. *Nat. Struct. Biol.* **1994**, *1*, 226–229.
(70) Teeter, M. M. *Protein Sci.* **2004**, *13*, 313–318.
(71) Tilton, R. F.; Singh, U. C.; Kuntz, I. D.; Kollman, P. A. *J. Mol. Biol.* **1988**, *199*, 195–211.
(72) We would like to remark that previous work by Kato and co-workers did not show any escape from the Xe3 cavity.[47] While the motions of explicit ligand were accelerated using metadynamics, molecular dynamics runs were performed for 10 ns (in 30 different simulations). One reason that may explain why those researchers did not observe escape via Xe3 may be that in their simulations the Xe3 cavity did not have enough time to open. Indeed, they noticed the presence of water molecules in Xe3 which may have prevented migration to Xe3. This is consistent with the closed state of Xe3 which we observed until 16 ns. Further factors may contribute to this discrepancy, which confirm one more time the complexity of the studied system.

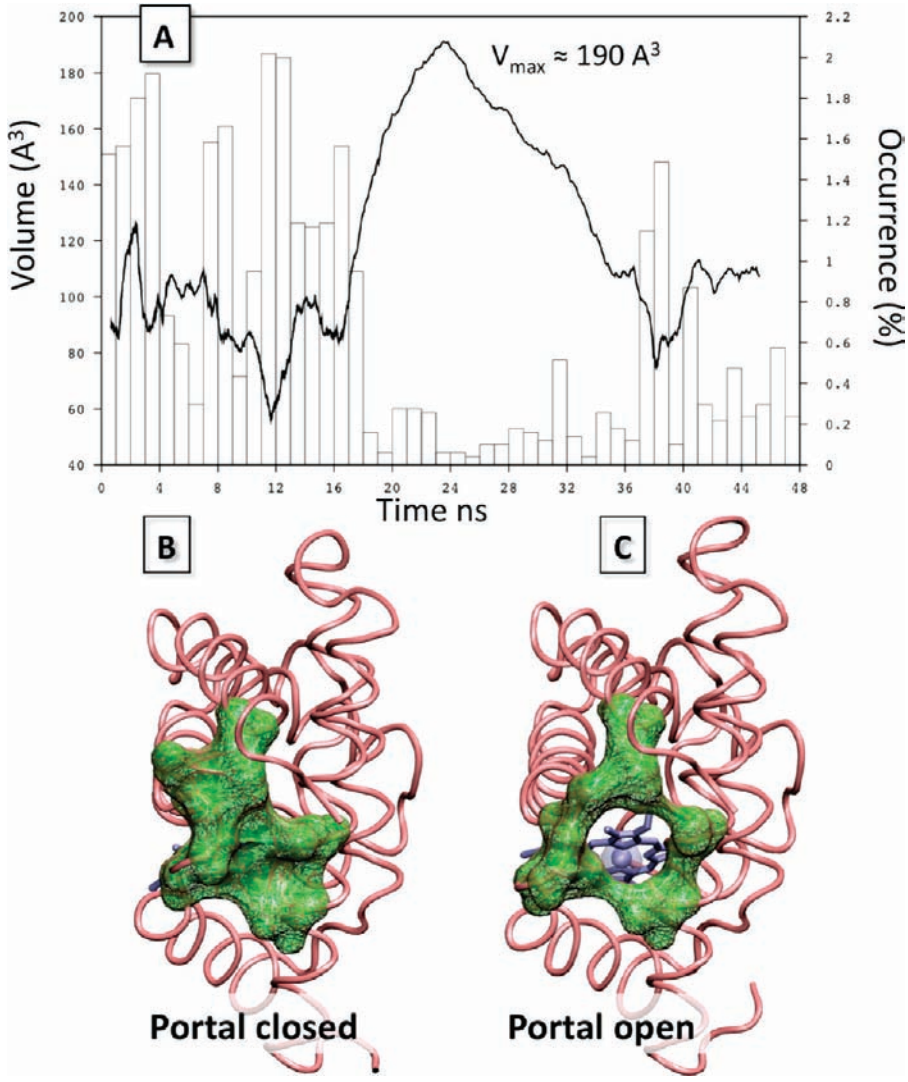


Figure 6. (A) Xe3 volume/occurrence as a function of time. At the bottom are shown two views of the Xe3 *portal* at two different simulation times: (B) 2 ns, *portal* closed; (C) 20 ns, *portal* open.

was built with the aim of obtaining a fingerprint of solvation that may help identify differences among the family members of myoglobin or, more generally, among other globular proteins. From this map, it appears evident that many sites identified through our procedure correspond (or are very close) to those previously indicated as possible solvent/protein *portals*:[55] the apical site, Xe3, 24, 23, and DP. In addition, mutations of certain residues that belong to the hydration sites are known to affect ligand enter/escape or migration inside the proteins,[22,69] suggesting that hydration may be crucial for the entry/escape of ligands in myoglobin and therefore fundamental for it to exert its biological functions.

From the present work, interesting evidence emerged: a relatively "simple" molecule like water can have many different subtle ways to interact with a macromolecule such as a protein, depending on the environment where it operates. Molecular dynamics simulations and a careful statistical analysis of solvent dynamics indeed showed that a water molecule can (i) "block" the entrance/escape to/from a site, (ii) act as a "wedge" supporting the dynamics of a cavity, or (iii) join a "flow" of other water molecules taking a ligand through (or "washing" a ligand away from) a protein *portal*.

In particular, the distal pocket was found to host at least one water molecule (in addition to the one bound to the heme iron) for almost the entire simulation, it being the only cavity that allowed a maximum of three waters inside (e.g., two solvent molecules together with the water bound to the heme iron, for about 10 ns), with 64H playing a pivotal role in stabilizing these molecules. Consistent with recent experimental studies,[53] our work confirms that water molecules can "block" ligand entrance into an internal cavity, thereby influencing the kinetic barrier of the entry.

In another protein region, namely the apical site, the motion of the three amino acids 26Q, 55M, and 58S constituting the *portal* was found to be highly correlated to the presence of a water molecule in the site. A water molecule trapped in the apical site was shown to work as a sort of "wedge" that forces *portal* residues to move apart, the latter being then followed by large parts of helices B, D, and E. Although it is not clear if the solvent molecule opens its way to the cavity or, rather, takes advantage of the cavity volume fluctuations, solvation is certainly crucial for the dynamics of the apical site. Indeed, this motion is transmitted through the protein backbone reaching 29L which, as shown in our previous work,[54] regulates the 17/

DP and DP/Xe4 *gates*. Therefore, hydration of the apical site may have an indirect role in ligand migration through the distal region and hence in protein biological functions. These findings are in very good agreement with the work of Fisher and Verma,[39] showing that water binding is accompanied by an increase in the vibrational entropy of a large part of the protein matrix; i.e., the overall flexibility of certain portions of the protein increases on binding.

Finally, the Xe3 cavity has been previously indicated as a possible *portal* for ligands to enter/escape the protein.[46,55] In this study, the mechanism of this *portal* is unveiled for the first time. It works as a *shutter*, with Trp7, Lys79, 80G, 82H, Leu137, and 141D residues involved. In addition, our results strongly suggest that the opening/closing mechanism is strictly related to cavity hydration: when the *portal* is closed, long-residence-time water molecules are found inside; in contrast, when the *portal* opens to the solvent, water molecules can easily enter/escape the cavity, with up to seven *fast* waters simultaneously passing through this *portal*. Such a "flow" of water molecules may have a double role: (i) to "wash" the protein, taking away a ligand from Xe3, and (ii) to transport a ligand inside Xe3, from which it may enter the protein.

In conclusion, three different ways of action of solvent water molecules have been revealed and their effects on myoglobin dynamics described in detail. In addition, we speculate that water may be involved in other kinds of interactions with different proteins or biological systems. In line with the growing concept of water as a fundamental component of cells,[2] our study confirms that the solvent can operate in different subtle ways, thereby having an active role in the biological functions of proteins.

Acknowledgment. Simulations were performed at the Caspur (Roma, Italy) and Cineca (Bologna, Italy). This work was supported by MIUR PRIN2006 (2006059902_003) project. The study makes use of results produced by the Cybersar Project managed by the Consorzio COSMOLAB. M. Ceccarelli 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".

Supporting Information Available: Table S1 and Figures S1−S3. This material is available free of charge via the Internet at http://pubs.acs.org.

JA909822D