Theor Chem Acc (2011) 130:1105–1114
DOI 10.1007/s00214-011-1041-6

REGULAR ARTICLE

MD simulations of plant hemoglobins: the hexa- to penta-coordinate structural transition

Mariano Andrea Scorciapino · Claire Wallon · Matteo Ceccarelli

Received: 7 June 2011 / Accepted: 12 September 2011 / Published online: 4 October 2011

Abstract Hemoglobins are ubiquitous proteins found in bacteria, plants, and animals with diverse functions other than the classical transport/storage of oxygen. Different functions are expected to correspond to substantially different structures, such as the hexa- and penta-coordination of the iron atom. It is now widely believed that pentacoordinate hemoglobins evolved from the hexacoordinate ones, both in plants and in animals. Since plant hemoglobins evolved more recently than in animals, they represent a simpler and thus useful system to investigate protein sequence/structure features that specifically supported, guided by molecular evolution, the capacity for oxygen transport. In the present work, we selected a fully hexacoordinate globin, AHb2 from Arabidopsis thaliana and the pentacoordinate oxygen-transporting LegHb from yellow lupin, that share a high degree of sequence identity. Our aim is to identify the structural determinants for oxygen transport by analyzing the structural/dynamical differences of a hexacoordinate and a pentacoordinate globin using all-atom molecular dynamics simulations. Using comparative MD simulations, we were able to go beyond the simple sequence alignment, pointing out important differences between these two hemoglobins especially at the level of the CD region, whose dynamics was found, in turn, to be strongly correlated with that of the distal region.

Keywords Molecular dynamics · Oxygen transport · Plant hemoglobins

Dedicated to Professor Akira Imamura on the occasion of his 77th birthday and published as part of the Imamura Festschrift Issue.

Electronic supplementary material The online version of this article (doi:10.1007/s00214-011-1041-6) contains supplementary material, which is available to authorized users.

M. A. Scorciapino
Dipartimento di Scienze Chimiche, University of Cagliari, Monserrato, Italy

M. A. Scorciapino · M. Ceccarelli
Istituto Officina del Materiali, SLACS, CNR, Monserrato, Italy

C. Wallon · M. Ceccarelli (✉)
Dipartimento di Fisica, University of Cagliari, Monserrato, Italy
e-mail: matteo.ceccarelli@dsf.unica.it

1 Introduction

The globins family represents a nice example of proteins that, through evolution, diversified their functions. Hemoglobin (Hb), probably the most important protein of this family, is an oxygen binding protein that is present in most living organisms. Hbs were originally described in animals, and in particular, in vertebrates they have been shown to facilitate oxygen transport in blood. However, animals, plants, fungi, and bacteria express a number of different Hbs [1], many of which have functions other than the oxygen transport [2, 3]. These different functions correspond to substantially different structures, some of these globins being monomeric, some others being characterized by a multimeric quaternary structure. Recent genomic studies [4] proposed Hbs to have originated in a bacterial ancestor, which probably needed to sequester reduced iron on one side and to control excessive oxygen concentrations, lethal for the anaerobic life, on the other. Another possible function may have been the detoxification of nitric oxide through its oxidation in an oxygen-rich successive environment. From bacteria, possibly through lateral gene transfer or some other mechanism, this Hb ancestor appeared in animals (~500 Myears) and more recently in plants (~200 Myears) [1, 5, 6]. It has been recognized that natural selection and evolution through

mutations, on the one hand, improved/extended the protein's function, and, on the other, led to overlapping functions, sometimes in such a manner so as to hide structure/function correlations [7]. Since plant Hbs evolved more recently than in animals, they represent a thought to be simpler and thus useful system to investigate protein sequence/structure features that specifically supported, guided by molecular evolution, the capacity for oxygen transport.

The classical pentacoordinate oxygen transport Hbs are now widely considered to have evolved from hexacoordinate Hbs both in plants and in animals. In the hexacoordinate non-oxygen transport Hbs, the distal histidine reversibly coordinates the ligand binding site, i.e., the heme iron. The net result is a competition between the histidine and the ligand to bind the sixth iron coordination site, affecting both ligand kinetic rates and affinity that, in turn, might be no longer optimal for oxygen transport. Thus, the structural passage from non-oxygen to oxygen transport function seems to be linked to globins capability of avoiding the endogenous hexacoordination, i.e., the coordination between the heme iron and the distal histidine. Nevertheless, this passage had to take place without the distal histidine deletion since it is well documented to have a primary role in the optimal stabilization of the bound ligand, such as the oxygen, in the pentacoordinate Hbs [8]. Despite the high resolution achieved by X-ray crystallography, no sequence/structure features specifically supporting the heme hexa- to penta-coordination transition could be yet recognized. Only one comparison has appeared in the literature, between the endogenous hexacoordinated rice Hb1 and the exogenous hexacoordinated barley Hb (with CN^- as bound ligand) [6]. What is questionable in that paper, as was also recognized by the authors themselves, is whether the identified differences, found mainly in the CD and EF regions, are actually correlated with the endogenous-exogenous forms or they are merely due to inherent differences in the two polypeptide chains, the rice and barley Hbs, respectively.

Thus, in the present work we selected a fully hexacoordinate globin, the non-symbiotic class 2 Hb from *Arabidopsis thaliana* (AHb2) and the pentacoordinate oxygen-transporting LegHb from yellow lupin, as they share a high degree of sequence identity and similarity (58 and 87%, respectively). While LegHb is known to be monomeric, it has not yet been clarified whether AHb2 has a quaternary structure (dimeric, such as AHb1 from rice) or a monomeric one [9]. Our goal in the present paper has been to identify the structural determinants that distinguish AHb2 from LegHb, i.e., a hexacoordinate from a pentacoordinate globin using molecular dynamics simulations. We compare the monomeric units of the two proteins.

Thanks to the introduction and development of molecular modeling in the last 30 years, the term "structure" has now evolved to have a wider (and more general) meaning. Immediately after the first protein simulation [10], it was clear that proteins are not at all static structures, and the extent of spontaneous fluctuations can reach the value of interatomic distances: dynamics of proteins is as important as their structures, being altogether the essential elements that tune/determine their functions [11–13].

Since AHb2 has not yet been crystallized, molecular modeling was the method of choice in order to compare the structure/dynamics of the two proteins under investigation. Using MD simulations, we could go beyond the simple sequence alignment, identifying key residues, especially looking at the CD and EF regions, expected to tune the mentioned structural change [6].

Today, the understanding of protein structure/function relationship represents an extraordinary challenge to protein engineers, who aim at reproducing or inducing new functions in synthetic polypeptide chains [7]. The structure and the chemistry underlying hexacoordination are the subject, indeed, of increasing attention by the scientific community due to potential roles in sensing and detoxifying NO. Altogether, our results on the structural determinants for oxygen transport provide more structural and dynamical information, which can be used in designing new synthetic proteins with applications, for instance, as substitutents in blood substitutes.

2 Methods

We started MD simulations on LegHb from its high-resolution X-ray structure (PDB code 2GDM at 1.7 Å) [14]. AHb2 initial structure was obtained by homology modeling starting from its amino acids sequence (GenBank id: U94999) [5]. The list of the five highest homologous sequences generated by the program PDBBlast2 contained only subunits of hemoglobin tetramers, while in a recent paper [13] AHb2 was derived from the X-ray structure of AHb1 of rice (1D8U solved at 2.4 Å) [9]. After constructing different AHb2 models on the basis of the structures from the five sequences generated by PDBBlast2 and that from AHb1 of rice using modeler [15], we analyzed and compared all of them with the Qmean server [16]. The highest score corresponded to the structure derived from AHb1 of rice (80% similarity and average deviation of 1.7 Å), and thus it has been used as the initial structure for our AHb2 simulations.

Both proteins were separately solved in an orthorhombic water box with an initial side length of 69 Å and containing ~9,900 water molecules (~31,000 atoms). After 1,000 steps of steepest descent minimization, we performed 50 ps of simulations at 20 K in the NVT ensemble, followed by a slow heating (600 ps) in the NPT ensemble, from 10 to

300 K at 1.0 bar. During heating, we maintained C-alpha atoms close to their initial position by applying harmonic restraints (10 kcal/mol/Å^2), reaching a relaxed box size with an average side length of 66.8 Å. Then, we decreased the restraints to zero in two steps (2 ns), and 108 ns long MD simulations have been finally performed for each of the two proteins. A posteriori, the AHb2 simulation was run for an additional 220 ns in order to increase the statistics (see below). Moreover, the AHb2 sequence was modified at the level of the CD region (from residue 41 to 60) in order to obtain an AHb2 with the same CD region as LegHb. In particular, G45D, R51K, D52G, S53T, D54S and H58Q mutations were introduced, leading to the AHb2-1a mutant. The latter was relaxed in a water box, and 240 ns long MD was run as described for the two wild-type proteins.

The Soft-Particle-Mesh-Ewald (SPME) algorithm (64 grid points and order 4 with direct cutoff at 10.0 Å and switch at 6.0 as suggested in [17] to better conserve the total energy) was used to treat long-range electrostatic contributions in combination with a multiple-time-step algorithm. The initial 9 ns were rejected, while the last 96 ns were used for the analysis. MD simulations were performed with the NAMD program (version 2.7), employing the Amber99SB-ildn [18] force-field and TIP3P [19] for the protein and water, respectively. For the heme group in the deoxy state, we used the amber parametrization corrected by the value "hard" as reported in [20].

The VOIDOO software [21] was employed to localize the internal voids and calculate their volume. As described in detail elsewhere [22], a cluster analysis was applied to the cavity calculation results in order to provide a statistical description of their dynamics and occurrence.

Secondary structure analysis was carried out with Simulaid using the DSSP program [23]. Finally, we used the PISQRD web server to analyze the protein internal fluctuations [24]. The used algorithm takes the 10 eigenvectors with the lowest frequencies and decomposes them into the motions of quasi-rigid subunits.

3 Results

We simulated the deoxy form of both proteins. It is important to stress here that in our model of AHb2 the distal histidine H66 was deliberately left unbound (not bonded) with the heme iron with the epsilon-N protonated. Thus, AHb2 was simulated in the pentacoordinate form even if this protein is known to be mostly present in the hexacoordinate state [25]. This particular choice was aimed at comparing AHb2 structural/dynamical features to those of a mostly pentacoordinated and symbiotic LegHb, excluding the influence of such a bond and focusing on the intrinsic thermal fluctuations, known to play a fundamental role in ligand migration and thus respiratory globins functioning [22, 26–28].

Figure 1 shows the sequence alignment and the secondary structure comparison between AHb2 and LegHb as obtained with Simulaid [23]. Capital letters from A to H are used to indicate the presence of that specific residue in one of the eight expected helices, respectively, when the secondary structure is maintained for at least 75% of total simulation time. Lowercase letters have been alternatively used to indicate a threshold of 50%. Table 1 summarizes these results including identity, high- and low-similarity percentage for each of the protein portions. Nine well-defined regions could be identified with a relatively high homology rate. Sequence alignment required only five insertions (AHb2 has five residues more than LegHb), i.e., three in the N-terminus tail, one within the relatively long

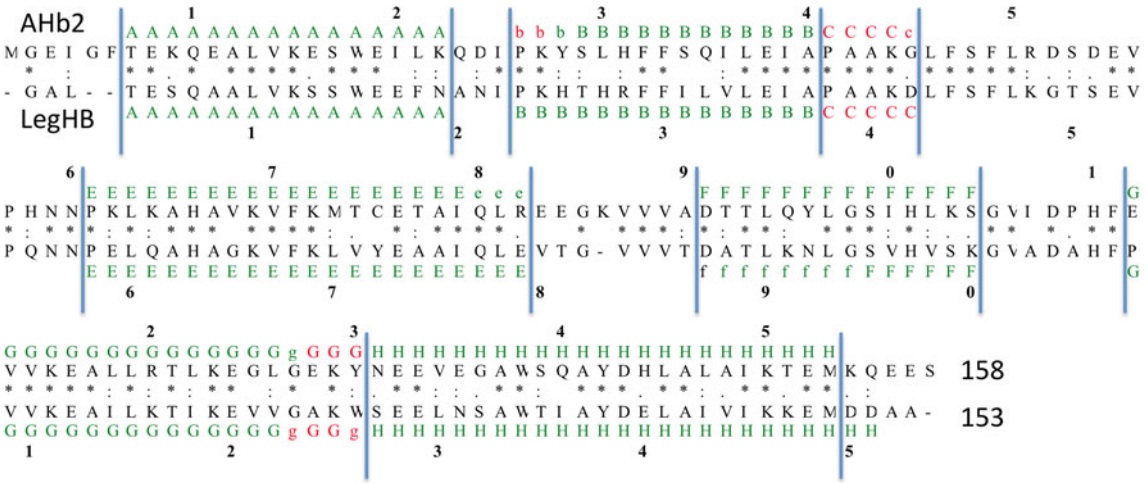

Fig. 1 Sequence alignment and secondary structure comparison as obtained with Simulaid. Helices are indicated either in *green* (4-helix) or in red (3-helix), with *capital/lowercase letters* corresponding to residues present is in the helix conformation for more than 75 and 50% of total simulation time, respectively. *Asterisk* indicates identity; double and single dots indicate high and low similarity, respectively

Table 1 Protein regions defined by residues number together with identity and similarity percentages as obtained by alignment and secondary structure analysis (see Fig. 1)

	AHb2	LegHb	Number of residues	Identity[a,b]	Similarity-high[a,c]	Similarity-low[a,d]
A	7–22	4–19	16	63% (10)	75% (12)	88% (14)
B	26–40	23–37	15	53% (8)	80% (12)	80% (12)
C	41–45	38–42	5	80% (4)	80% (4)	100% (5)
D region	46–60	43–57	15	67% (10)	87% (13)	100% (15)
CD	41–60	38–57	20	70% (14)	85% (17)	100% (20)
E	61–82	58–79	22	64% (14)	82% (18)	91% (20)
EF loop	83–90	80–86	7	57% (4)	71% (5)	71% (5)
F	91–104	87–100	14	50% (7)	79% (11)	93% (13)
G	112–130	108–126	19	58% (11)	84% (16)	84% (16)
H	131–153	127–149	23	57% (13)	74% (17)	96% (22)
All	158	153	153	58% (88)	76% (117)	87% (133)

[a] number of residues corresponding to the reported percentage is indicated in brackets

[b] amount of conserved residues

[c] amount of conserved residues plus high-similarity mutations

[d] amount of conserved residues plus all similarity mutations (both high and low)

EF loop and one at the C-terminus. The most evident difference between the two proteins is the relatively low stability of the first half of the F helix in the LegHb with respect to AHb2. The EF loop appeared to be more flexible in the former, the first half of the F helix being almost unfolded for about half of the total simulation time. It is also interesting to note that, comparing the two proteins, the CD region is the most conserved one, which is, indeed, the only one reaching 100% similarity (Table 1).

Backbone RMSD was calculated without taking into account, especially in the case of AHb2, the relatively long and flexible N- and C- termini. Then, in order to perform a more meaningful comparison between the results obtained for the two proteins, otherwise referred either to a homology model (AHb2) or to a X-ray structure (LegHb), we decided to evaluate the RMSD with respect to a more representative reference structure extracted from the MD trajectory. A cluster analysis was applied to all conformers in order to identify the one with the highest number of neighbors. As the effective distance, we calculated the weighted sum of (1) the RMSD from the starting structure, (2) the distance between the distal histidine (C-alpha) and the iron and (3) the distance between the CD region (through C-alpha position of residue 54 and 51 for AHb2 and LegHb, respectively) and the iron. We anticipate here that these properties have been selected during the analysis to represent structural determinants describing the relaxation of the two proteins.

Figure 2 shows the comparison between backbone RMSD of the two proteins computed for (a) the entire backbone except the termini, (b) CD region only and (c) the whole protein except the CD region. LegHb RMSD was found to be almost constant along the simulation in each case, with relatively low fluctuations about ca. 0.1 nm. This clearly indicates that the starting X-ray structure was actually a good model for LegHb in solution. On the other hand, AHb2 RMSD fluctuates to a larger extent. However, RMSD analysis has shown that these fluctuations are mainly due to some conformational rearrangements of the CD region, which, starting from our model, needed about 50 ns to relax in solution, as shown by the almost constant RMSD reached during the second half of the simulation (Fig. 2b). This relaxation time was comparable to that (of similar conformational rearrangements of the CD region) in LegHb. Indeed, during the first 50 ns, the D region of AHb2 is characterized by a 3-helix conformation starting from residue 55 (Fig. S1), resembling, for instance, the D-helix of myoglobin or hemoglobin β subunits, which is lost upon CD region relaxation.

The residue RMSFs support these results too (Fig. 3). As expected, AHb2 shows higher fluctuations at the termini than LegHb. The EF loop and the first part of F helix are the regions with the highest fluctuations in LegHb, in agreement with secondary structure analysis. Even more important, as shown by backbone RMSD analysis, AHb2 shows higher fluctuations in the CD region than LegHb. It is also interesting to note that these differences can be also seen when RMSF is calculated for the second half of the simulations (Fig. S2), i.e., even if CD region of AHb2 has relaxed (Fig. 2), this protein region has higher flexibility than that observed for LegHb.

Relaxation of the CD region in AHb2 results in the entire distal portion of the protein being more flexible and mobile. In a concerted fashion, the CD region unfolded and extended toward the solvent, and moved away from the

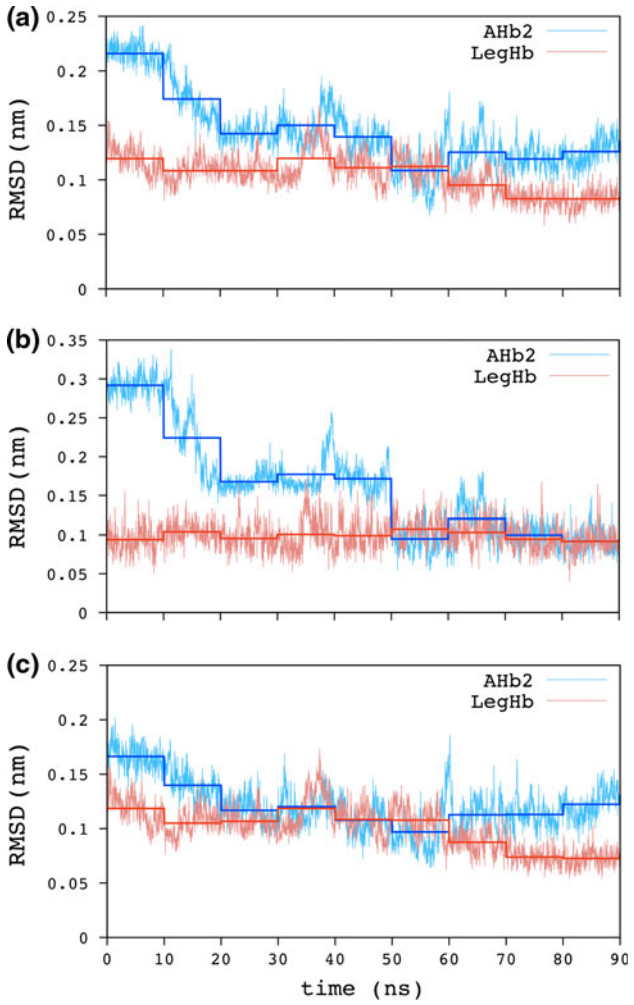

Fig. 2 Backbone RMSD calculated with respect to the conformer with the highest number of neighbors for **a** the entire backbone except the termini, **b** CD region only, and **c** the whole backbone except the CD region and the termini. Blocks averages computed for 10 ns long blocks are also shown

distal heme region, as shown in Fig. 4a by the changes in the distance between residue 54 (located in the middle of the CD region) and the heme iron, and by the changes in the CD backbone RMSD during the same MD simulation. The last part of the C helix and the first part of the E helix tended to approach each other, reaching an almost constant distance (between Cαs of residue 47 and 62) during the CD region relaxation (Fig. 4b). Figure 4c shows the correlation between these two distances. Finally, these motions influence also the distance between the distal histidine (residue 66) and the heme iron as shown in Fig. 4d, resulting in the distal region for AHb2 being more compact than found for the symbiotic LegHb. This is particularly evident when comparing the distribution of this distance for the two proteins under investigation (Fig. 5).

Application of the PiSQRD server [24], using the same more representative structure as in the RMSD calculations,

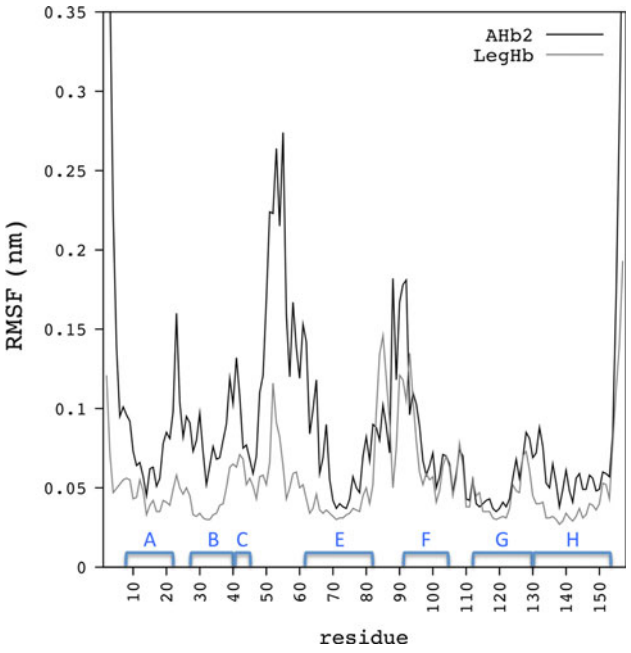

Fig. 3 Residue RMSFs calculated from Cα coordinates. Residue numbers correspond to AHb2 amino acids sequence, while LegHb has been aligned with respect to the latter on the basis of Fig. 1

revealed that both proteins can be divided into three main domains, defined as protein regions whose collective motions reproduce 80% of the motion of the 10 lowest eigenvectors. Figure 6 shows the 3D structure of both AHb2 and LegHb where these domains are differently colored. Interestingly, both proteins are characterized by a well-defined CD region, which, as expected, is correlated with the distal one, the beginning of the E-helix and to the B-helix (red in Fig. 6). However, while LegHb shows a proximal domain (yellow) distinguishable from the other domain comprising the A-helix, the second half of the E- and G- helices, and the first half of the H one (blue), AHb2 present a unique main central domain (blue) plus the two relatively long termini (yellow and orange). Intriguing differences have been also found through the cavities analysis. As described in the computational details, VOIDOO [20] has been employed to look for internal voids in all conformers recorded every 30 ps along the two simulations and to estimate their volumes and positions. Finally, a cluster analysis with respect to their position inside the protein was applied to all the cavities, allowing for their classification in terms of occurrence and average volume. In Fig. 6, the cavities with a minimum occurrence of 40% are shown as spheres, whose volumes are equal to the average values found with the cluster analysis. Occurrence and average volumes of these main cavities are reported in Fig. 7. Both proteins do not have relatively high occurring cavities on the proximal side of the heme group while all of the main voids are localized on the distal one, the CD region

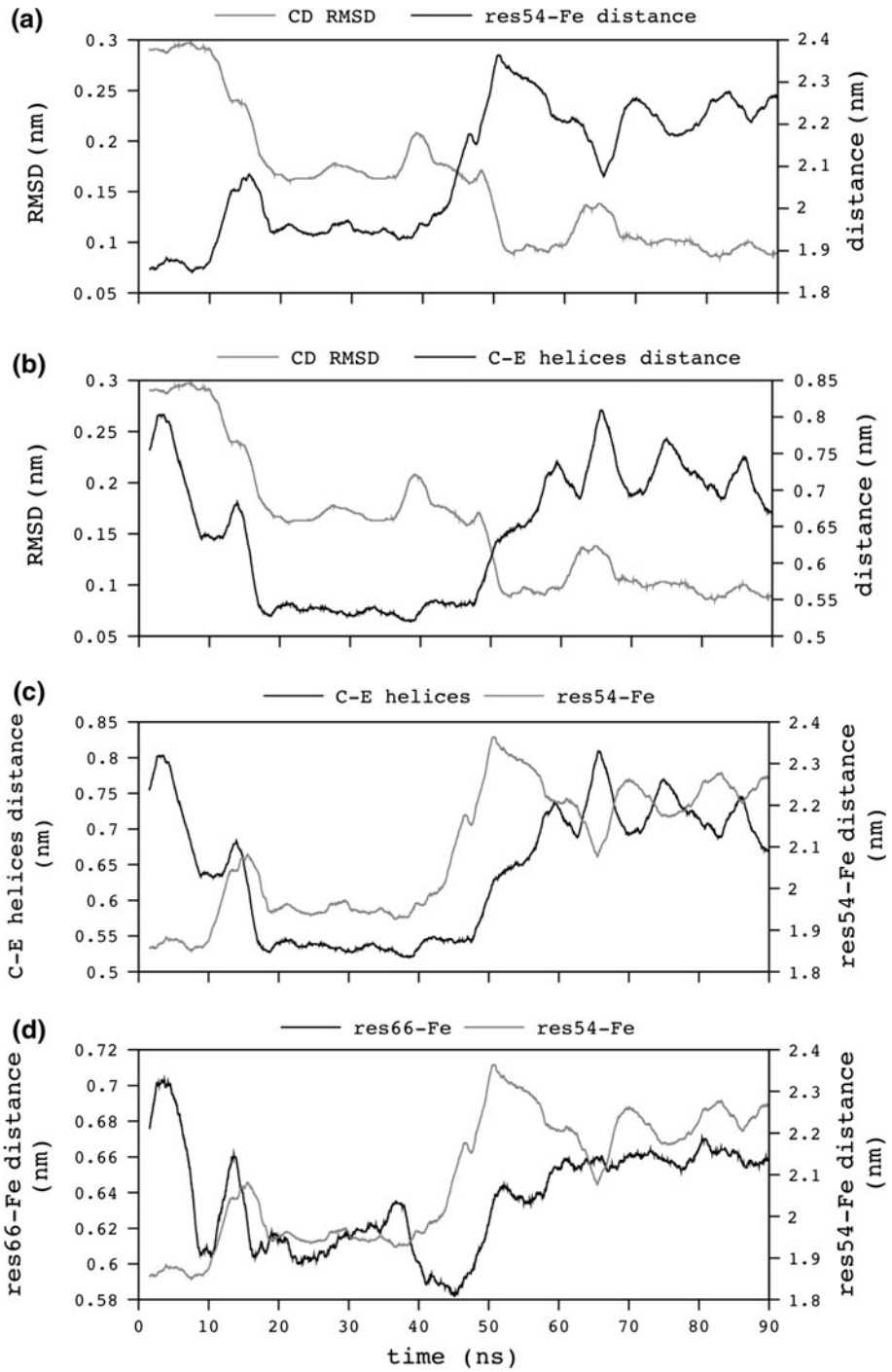

Fig. 4 Comparison between particular atoms distances in AHb2 chosen to show the concerted motions in the CD and the distal regions. **a** CD backbone RMSD and the distance between the Cα of residue 54 from the heme iron; **b** CD backbone RMSD and the distance between the Cα of residues 47 (located in the end of C-helix) and 62 (located at the beginning of E-helix); **c** the distance between the Cα of residue 54 from the heme iron and the distance between the Cα of residues 47 and 62; **d** the distance between the Cα of residue 54 and of residue 66 (distal histidine) from the heme iron

and close to G and H helices. The most interesting difference is the absence of a distal pocket in the AHb2, in agreement, as mentioned, with a more compact distal region compared to LegHb, which, on the contrary, has the distal pocket as the most occurring and large cavity. Definitely, the latter appears to be more similar to myoglobin, having also two of the main cavities corresponding to the so-called Xe4 cavity (between B and G helices) and to the apical site (between B-helix and the CD loop) [22, 29]. A more detailed inspection of the distal pocket residues revealed that while myoglobin has the 29L(B10) separating the active site from the so-called Xe4 cavity, both AHb2 and LegHb have the larger phenylalanine, 33F(B8) and 30F(B8), respectively. Almost 60 years of research on myoglobins, both experimental and computational, led to identify a significant number of key residues shown to be relevant to the protein functions [22, 28, 30–35]. Among them, the B10 residue has been shown to play a fundamental role in ligand migration kinetics. In particular, (1) it acts as a physical barrier as the ligand enters/exits the protein from

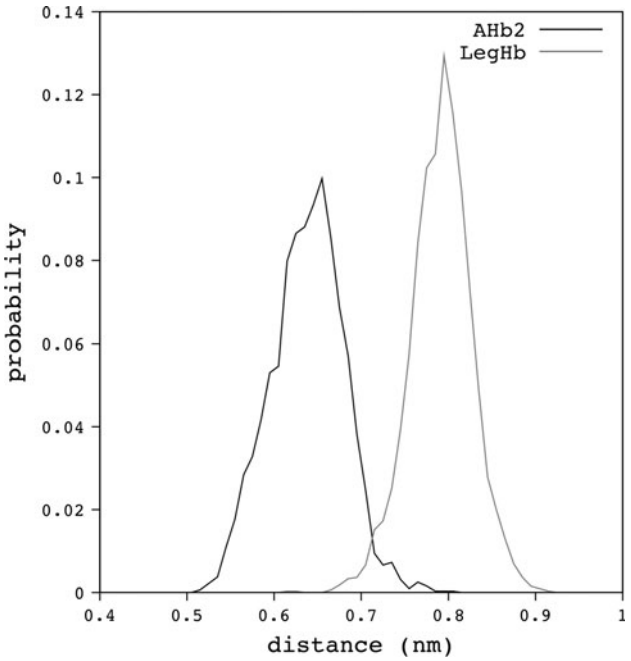

Fig. 5 Comparison between the distribution of distance between distal histidine Cα from the iron, calculated for AHb2 and LegHb

the distal pocket [33]; (2) its mutation can lead to significant changes in the overall binding kinetic rate [35] inhibiting the ligand entrance inside the protein [32]; and (3) mutations of the B10 residue show stronger effects than mutations of the distal histidine itself [35]. Definitely, it was found that geminate ligand recombination and thermal dissociation rates have an opposite correlation with the size of this particular residue, confirming that this residue plays a crucial role in both ligand escape to the solvent and migration inside the protein [35]. Even if, as mentioned, both AHb2 and LegHb have a Phe instead of a Leu at the level of the gate separating the distal pocket from the Xe4 cavity, the present work revealed intriguing differences between these two plant hemoglobins as far as the side chain fluctuations of this fundamental residue. The 30F of LegHb has the side chain oriented toward the heme iron, i.e., acting as a barrier between the active site and the inner Xe4 cavity. On the other hand, due to the more compact distal region found in our model of pentacoordinate deoxy-AHb2, the 33F side chain moved far from the distal pocket upon the above described structure relaxation, pointing toward the G-helix and allowing the E-helix and the C- one to come closer. The presence of a phenylalanine at that position and the fact that it is oriented internally seems to favor the ligand internal trapping in AHb1, as shown by FTIR spectroscopy upon mutation Phe->Leu [36]. On the contrary, AHb2 does not show any ligand internal trapping, in agreement with our simulations that predict the outward orientation of 33F.

Finally, as described in the Methods, deoxy-AHb2 molecular dynamics was run for an additional 220 ns in order to investigate backbone fluctuations excluding the first part of the MD where, as discussed hereinbefore, protein relaxation was predominant. Moreover, the deoxy-AHb2-1a mutant was prepared on the basis of the AHb2 3D structure and amino acids sequence, except for the CD region (from residue 41 to 60). The latter was mutated, indeed, in order to have exactly the same sequence as in the Leg-Hb. Simulation of this mutant was aimed at checking whether the CD region dynamics depends mainly on the primary sequence of the region itself or upon the particular protein model employed. Comparison of these two simulations with that of LegHb supported the discussion presented above, the most interesting and significant results coming from the RMSF shown in Fig. 8. The AHb2 wild-type shows higher fluctuations in the CD region than LegHb as already observed, while the EF loop and the first part of F helix are the regions with the highest fluctuations in the latter. Interestingly, while the AHb2-1a mutant RMSF profile closely matches the wild-type one from the N-terminus to approximately the first half of the B-helix, as well as from the second half of the E-helix till the C-terminus, fluctuations of the mutated region follows the same trend observed for the LegHb. Furthermore, a closer inspection of the data reveals that this similarity of CD

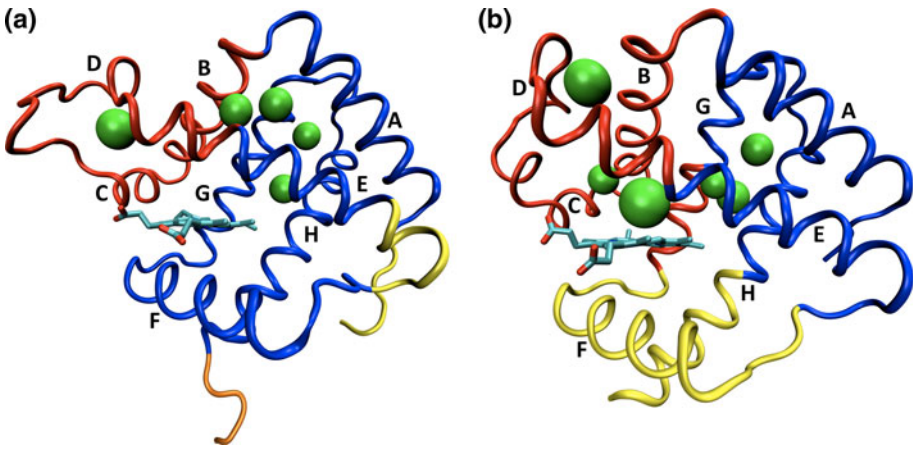

Fig. 6 3D structure of **a** AHb2 and **b** LegHb with differently colored domains as found with the PiQSRD. The *green* spheres represent the most occurring cavities (>40%) found with a cluster analysis performed on VOIDOO results (see Fig. 7). Regions labels are reported, according to Fig. 1

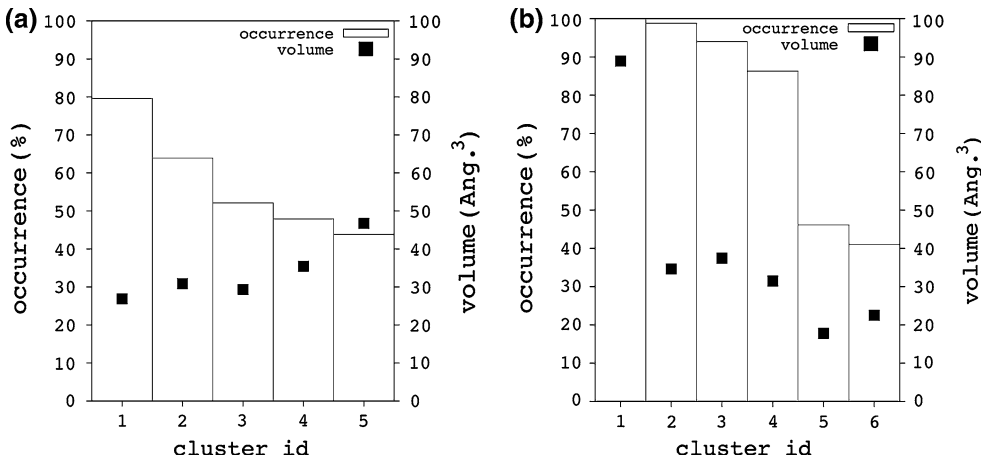

Fig. 7 Main cavities occurrence and average volume for **a** AHb2 and **b** LegHb resulting from VOIDOO and the successive cluster analysis

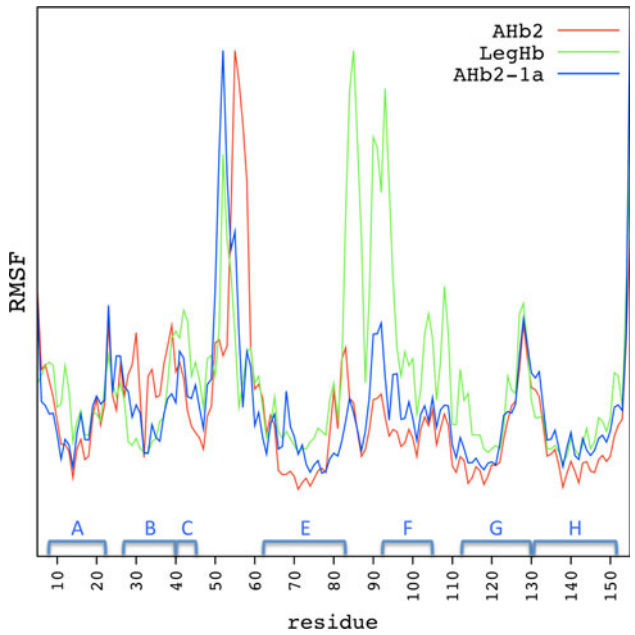

Fig. 8 Residues RMSF calculated from Ca coordinates. Residue numbers correspond to AHb2 amino acids sequence, while LegHb has been aligned with respect to the latter on the basis of Fig. 1. AHb2 RMSF have been computed from a second MD in order to eliminate relaxation contributions. LegHb data are exactly the same as in Fig. 3. AHb2-1a is the AHb2 mutant built in order to have exactly the same CD region as in the LegHb. The three data sets have been normalized

region dynamics between AHb2-1a mutant and LegHb is actually not restricted to the six mutations, but is observed from the second half of the B-helix to the beginning of the E-one, which is exactly one of the domains identified by the PiSQRD application.

4 Summary and conclusions

Interesting structural and dynamical differences emerged from the present study. While the symbiotic LegHb, which is known to mainly act as an oxygen transporter, shows structural and dynamical features comparable to myoglobin, AHb2 presents remarkable differences. It is known to be mainly present in the hexacoordinate form; however, we wanted to investigate it in the pentacoordinate deoxy-form in order to directly compare the results to those obtained on LegHb. Even in the absence of the covalent bond between the distal histidine and the heme iron, AHb2 structure relaxed bringing the E-helix, thus the distal histidine, closer to the heme itself than observed for LegHb. It seems that AHb2 has actually an intrinsic tendency to fold with a relatively compact distal region as shown by the absence of the distal pocket. The present investigation suggests that this state cannot be reached if the CD region does not unfold and extend away from the heme, pointing out a strong correlation between the CD and the distal region dynamics, also confirmed by PiSQRD analysis.

Recent FTIR vibrational spectroscopic analysis supports this correlation [36]. As reported, the higher flexibility of the CD region in the AHb2, with respect to the mostly pentacoordinate AHb1, suggests a piston-like movement for the helix E, allowing the distal histidine binding to the iron. Furthermore, the compact structure of AHb2 does not favor the access of ligands to secondary sites after photolysis [36], in agreement with our results on the internal cavities, i.e., the existence of internal voids with a very low occurrence compared to LegHb, while the difference on the average volumes is negligible. The difficulty for the highly compact AHb2 fold to create permanent internal cavities could also influence ligands escape, which is known to be regulated by protein fluctuations, as shown by ligand rebinding kinetics upon proteins encapsulation in silica gels [13].

CD region has been also shown to have a deep influence on the heme retention and folding stability [37] in globins with possible implications of its hydration state [29]. Other careful investigations are needed to shed light on this issue, but, from the present work, the structure and dynamics of the CD region appeared to be dramatically different when a

hexacoordinate plant Hb such as AHb2 has been compared to the pentacoordinate LegHb. This is certainly one of the protein portions to look at when facing with the evolutionary transition from hexa- to pentacoordinate Hb in plants.

Another very interesting difference between these two plant hemoglobins has been found at the level of the gate separating the distal pocket from the inner Xe4 cavity. In particular, the side chain of 30F resulted to be oriented toward the heme iron, closely resembling the B10 residue of myoglobin, acting as a barrier between the two sites and strongly affecting ligands binding and migration kinetics. On the other hand, in the AHb2, the 33F side chain moved far from the distal pocket upon the above described structure relaxation, contributing to the more compact fold of the distal region. This is in agreement with the absence of internal ligand docking sites for AHb2, as shown by FTIR spectroscopy [36].

Altogether, the results of the present work point out a specific protein region (and particular key residues) as one of the main structural/dynamical differences between the hexa- and the pentacoordinate hemoglobin investigated. AHb2, which belongs to the former, presents a really dynamic CD region correlated with a more tight distal portion, suggesting that ligands might hardly come in/out of the protein interior from the widely cited histidine gate [38]. Probably, ligands might enter this kind of globins using some other routes as also suggested by the different orientation found for the 33F side chain. AHb2's tighter global folding possibly indicates a structure more prone to ligands sequestering than delivering, even if specifically devoted studies should be needed to prove this point. On the other hand, the pentacoordinate globins, like LegHb and myoglobin, are known to be oxygen-transporting proteins [39]. From the present comparative study, the less flexible CD region (but we cannot exclude other concomitant factors that need further studies to be unveiled) seems to be one of the main factors hindering the distal region to collapse, keeping the distal histidine sufficiently far from the heme iron to host a ligand, and leaving the distal region itself more free to fluctuate, which is fundamental to exert the desired role of an oxygen storage as well as delivery system [11].

Acknowledgments We acknowledge the CYBERSAR and CINECA award, the latter under the ISCRA initiative for the availability of high-performance computing resources and support. CW thanks the ERASMUS program for supporting her visit to Department of Physics, University of Cagliari.

References

1. Hunt PW, Watts RA, Trevaskis B, Llewelyn DJ, Burnell J et al (2001) Expression and evolution of functionally distinct haemoglobin genes in plants. Plant Mol Biol 47:677–692
2. de Sanctis D, Pesce A, Nardini M, Bolognesi M, Bocedi A et al (2004) Structure-function relationships in the growing hexacoordinate hemoglobin sub-family. TBMB 56:643–651. doi:10.1080/15216540500059640
3. Burmester T, Weich B, Reinhardt S, Hankeln T (2000) A vertebrate globin expressed in the brain. Nature 407:520–523. doi:10.1038/35035093
4. Vinogradov SN, Hoogewijs D, Bailly X, Mizuguchi K, Dewilde S et al (2007) A model of globin evolution. Gene 398:132–142. doi:10.1016/j.gene.2007.02.041
5. Trevaskis B, Watts RA, Andersson CR, Llewellyn DJ, Hargrove MS et al (1997) Two hemoglobin genes in Arabidopsis thaliana: the evolutionary origins of leghemoglobins. Proc Natl Acad Sci USA 94:12230–12234
6. Hoy JA, Robinson H, Trent JT, Kakar S, Smagghe BJ et al (2007) Plant hemoglobins: a molecular fossil record for the evolution of oxygen transport. J Mol Biol 371:168–179. doi:10.1016/j.jmb.2007.05.029
7. Koder RL, Anderson JLR, Solomon LA, Reddy KS, Moser CC et al (2009) Design and engineering of an O(2) transport protein. Nature 458:305–309. doi:10.1038/nature07841
8. Martí MA, Capece L, Bikiel DE, Falcone B, Estrin DA (2007) Oxygen affinity controlled by dynamical distal conformations: the soybean leghemoglobin and the Paramecium caudatum hemoglobin cases. Proteins 68:480–487. doi:10.1002/prot.21454
9. Hargrove MS, Brucker EA, Stec B, Sarath G, Arredondo-Peter R et al (2000) Crystal structure of a nonsymbiotic plant hemoglobin. Structure 8:1005–1014
10. McCammon JA, Gelin BR, Karplus M (1977) Dynamics of folded proteins. Nature 267:585–590
11. Frauenfelder H, Chen G, Berendzen J, Fenimore PW, Jansson H et al (2009) A unified model of protein dynamics. Proc Natl Acad Sci 106:5129–5134. doi:10.1073/pnas.0900336106
12. Nishihara Y, Kato S, Hayashi S (2010) Protein collective motions coupled to ligand migration in myoglobin. Biophys J 98:1649–1657. doi:10.1016/j.bpj.2009.12.4318
13. Bruno S, Faggiano S, Spyrakis F, Mozzarelli A, Cacciatori E et al (2007) Different roles of protein dynamics and ligand migration in non-symbiotic hemoglobins AHb1 and AHb2 from Arabidopsis thaliana. Gene 398:224–233. doi:10.1016/j.gene.2007.02.042
14. Harutyunyan E, Safonova T, Kuranova I, Popov A, Teplyakov A et al (1995) The structure of deoxy-leghemoglobin and oxy-leghemoglobin from lupin. J Mol Biol 251:104–115
15. Fiser A, Sali A (2003) Modeller: generation and refinement of homology-based protein structure models. Meth Enzymol 374:461–491. doi:10.1016/S0076-6879(03)74020-8
16. Benkert P, Künzli M, Schwede T (2009) QMEAN server for protein model quality estimation. Nucleic Acids Res 37:W510–W514. doi:10.1093/nar/gkp322
17. Morrone JA, Zhou R, Berne BJ (2010) Molecular dynamics with multiple time scales: how to avoid pitfalls. J Chem Theory Comput 6:1798–1804. doi:10.1021/ct100054k
18. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL et al (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins 78:1950–1958. doi:10.1002/prot.22711
19. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79:926. doi:10.1063/1.445869
20. Henry ER, Levitt M, Eaton WA (1985) Molecular dynamics simulation of photodissociation of carbon monoxide from hemoglobin. Proc Natl Acad Sci USA 82:2034–2038
21. Kleywegt GJ, Jones TA (1994) Detection, delineation, measurement and display of cavities in macromolecular structures. Acta Crystal D Biol Crystal 50:178–185. doi:10.1107/S0907444493011333

22. Scorciapino MA, Robertazzi A, Casu M, Ruggerone P, Ceccarelli M (2009) Breathing motions of a respiratory protein revealed by molecular dynamics simulations. J Am Chem Soc 131:11825–11832. doi:10.1021/ja9028473
23. Mezei M (2010) Simulaid: a simulation facilitator and analysis program. J Comput Chem 31:2658–2668. doi:10.1002/jcc.21551
24. Aleksiev T, Potestio R, Pontiggia F, Cozzini S, Micheletti C (2009) PiSQRD: a web server for decomposing proteins into quasi-rigid dynamical domains. Bioinformatics 25:2743–2744. doi:10.1093/bioinformatics/btp512
25. Bruno S, Faggiano S, Spyrakis F, Mozzarelli A, Abbruzzetti S et al (2007) The reactivity with CO of AHb1 and AHb2 from Arabidopsis thaliana is controlled by the distal HisE7 and internal hydrophobic cavities. J Am Chem Soc 129:2880–2889. doi:10.1021/ja066638d
26. Dantsker D, Samuni U, Friedman J, Agmon N (2005) A hierarchy of functionally important relaxations within myoglobin based on solvent effects, mutations and kinetic model. Bba-Proteins Proteom 1749:234–251. doi:10.1016/j.bbapap.2005.04.002
27. Schotte F, Lim M, Jackson TA, Smirnov AV, Soman J et al (2003) Watching a protein as it functions with 150-ps time-resolved X-ray crystallography. Science 300:1944–1947. doi:10.1126/science.1078797
28. Brunori M, Bourgeois D, Vallone B (2004) The structural dynamics of myoglobin. J Struct Biol 147:223–234. doi:10.1016/j.jsb.2004.04.008
29. Scorciapino MA, Robertazzi A, Casu M, Ruggerone P, Ceccarelli M (2010) Heme proteins: the role of solvent in the dynamics of gates and portals. J Am Chem Soc 132:5156–5163. doi:10.1021/ja909822d
30. Springer BA, Egeberg KD, Sligar SG, Rohlfs RJ, Mathews AJ et al (1989) Discrimination between oxygen and carbon monoxide and inhibition of autooxidation by myoglobin. Site-directed mutagenesis of the distal histidine. J Biol Chem 264:3057–3060
31. Huang X, Boxer SG (1994) Discovery of new ligand binding pathways in myoglobin by random mutagenesis. Nat Struct Biol 1:226–229
32. Olson J (2007) Ligand pathways in myoglobin: a review of trp cavity mutations. TBMB
33. Uchida T, Ishimori K, Morishima I (1997) The effects of heme pocket hydrophobicity on the ligand binding dynamics in myoglobin as studied with leucine 29 mutants. J Biol Chem 272:30108–30114
34. Cohen J, Arkhipov A, Braun R, Schulten K (2006) Imaging the migration pathways for O2, CO, NO, and Xe inside myoglobin. Biophys J 91:1844–1857. doi:10.1529/biophysj.106.085746
35. Gibson QH, Regan R, Elber R, Olson JS, Carver TE (1992) Distal pocket residues affect picosecond ligand recombination in myoglobin. An experimental and molecular dynamics study of position 29 mutants. J Biol Chem 267:22022–22034
36. Nienhaus K, Dominici P, Astegno A, Abbruzzetti S, Viappiani C et al (2010) Ligand migration and binding in nonsymbiotic hemoglobins of Arabidopsis thaliana. Biochemistry 49:7448–7458. doi:10.1021/bi100768g
37. Whitaker TL, Berry MB, Ho EL, Hargrove MS, Phillips GN et al (1995) The D-helix in myoglobin and in the beta subunit of hemoglobin is required for the retention of heme. Biochemistry 34:8221–8226
38. Elber R (2010) Ligand diffusion in globins: simulations versus experiment. Curr Opin Struct Biol 20:162–167. doi:10.1016/j.sbi.2010.01.002
39. Garrocho-Villegas V, Gopalasubramaniam SK, Arredondo-Peter R (2007) Plant hemoglobins: what we know six decades after their discovery. Gene 398:78–85. doi:10.1016/j.gene.2007.01.035