J Mol Model (2007) 13:715–723
DOI 10.1007/s00894-007-0203-x

ORIGINAL PAPER

Locally enhanced sampling molecular dynamics study of the dioxygen transport in human cytoglobin

Slawomir Orlowski · Wieslaw Nowak

Received: 21 November 2006 / Accepted: 20 March 2007 / Published online: 15 May 2007

Abstract Cytoglobin (Cyg)—a new member of the vertebrate heme globin family—is expressed in many tissues of the human body but its physiological role is still unclear. It may deliver oxygen under hypoxia, serve as a scavenger of reactive species or be involved in collagen synthesis. This protein is usually six-coordinated and binds oxygen by a displacement of the distal HisE7 imidazole. In this paper, the results of 60 ns molecular dynamics (MD) simulations of dioxygen diffusion inside Cyg matrix are discussed. In addition to a classical MD trajectory, an approximate Locally Enhanced Sampling (LES) method has been employed. Classical diffusion paths were carefully analyzed, five cavities in dynamical structures were determined and at least four distinct ligand exit paths were identified. The most probable exit/entry path is connected with a large tunnel present in Cyg. Several residues that are perhaps critical for kinetics of small gaseous diffusion were discovered. A comparison of gaseous ligand transport in Cyg and in the most studied heme protein myoglobin is presented. Implications of efficient oxygen transport found in Cyg to its possible physiological role are discussed.

Keywords Cytoglobin · Locally enhanced sampling · Molecular dynamics · Ligand diffusion

Electronic supplementary material The online version of this article (doi:10.1007/s00894-007-0203-x) contains supplementary material, which is available to authorized users.

S. Orlowski · W. Nowak (✉)
Theoretical Molecular Biophysics Group, Institute of Physics,
Nicolaus Copernicus University,
ul. Grudziadzka 5,
87-100 Torun, Poland
e-mail: wiesiek@fizyka.umk.pl

Introduction

Hemoglobin (Hb) and myoglobin (Mb) are two of the most intensively studied proteins [1–9]. Both are members of the vertebrate heme globin family. Globins are heme-containing respiratory proteins that bind, reversibly, oxygen by means of Fe ion present in the heme group. Recent studies have extended this family by two new members: neuroglobin (Ngb) and cytoglobin (Cyg) [10–13].

Human Cyg shares about a 25% of amino acid identity with vertebrate Mb and Hb and only 16% with Ngb (Fig. 1). Bioinformatics analysis shows that it is distantly related to other heme globins. Cyg is expressed in many tissues, such as connective tissue fibroblasts and neurons. It is also present in chondroblasts and osteoblasts as well as in hepatic stellate cells. The physiological role of Cyg remains unknown [14]. Some hypotheses indicate its role as an NO or reactive oxygen species scavenger [15], oxygen storage and sensor [16, 17], or even suggest dioxygenase enzymatic activity [18]. Perhaps the best documented is, at this moment, Cyg involvement in collagen synthesis [18, 19].

Spectroscopic analysis has shown that Ngb and Cyg are the first examples of vertebrate globins which, in the absence of gaseous ligands (O_2 or CO), have a hexacoordinated iron centre [20]. In the absence of a ligand, the distal sixth coordination position of the heme Fe ion is occupied by the histidine HisE7 residue. On the proximal side of the heme group there is a standard bond between the iron ion and the conserved histidine HisF8. This may indicate a three-state mechanism of ligand replacement [12]. Cytoglobin displays a classical "globin fold" three-on-three alpha-helical sandwich. Compared to other vertebrate globins, sequence of the Cyg is unusually long, containing 190 residues (Fig. 1). In comparison to Hb, Mb and Ngb, Cyg shows the extended N- and C-term (about 40 residues)

Fig. 1 A comparison of sperm whale myoglobin, human cytoglobin and neuroglobin sequences. ClustalX [45] and Jalview [46] programs were used to prepare the alignment and the figure. *Dark rectangles* indicate conserved residues. Numbering of amino acids corresponds to Cyg

region, but the spatial structure of the core region (about 150 residues) remains almost unchanged. Analysis of the crystal structure shows a wide channel leading from the heme cavity to the solvent [21, 22].

While there are numerous computational studies of carbon monoxide and oxygen diffusion in Mb [23–27], no molecular dynamics (MD) studies of Cyg has been published so far. It is not clear whether O_2 may enter the Cyg protein matrix, and what are preferred binding sites in this protein. The static analysis of hypothetical cavities in Cyg has been performed based on the X-ray structure [21, 22], but dynamical data were not available until now. In the present paper, O_2 diffusion events were studied on an over 60 ns MD simulations timescale. In addition to the standard MD simulation method with one copy of the diffusing dioxygen molecule, an advanced technique of Locally Enhanced Sampling (LES) [28–31] with 5–15 copies of O_2 was also employed. These data may help to understand the physiological role of Cyg and bring new information on the architecture and function of this ancient protein.

Methods

A structure of human cytoglobin 1V5H [32] was obtained from Protein Data Bank (PDB) [33]. To perform molecular dynamics simulations, the NAMD 2.6b1 [34] program with CHARMM27 all-atom force field and locally enhanced sampling (LES) algorithm [28–31] as implemented in this code were used. Enhanced space in this approach contained coordinates of the ligand modeled, i.e. dioxygen. Parametrization of this ligand was taken from OM atoms of the CHARMM27 forcefield: l_{bond}=1.23 Å, ε_o=−0.12 kcal mol^{-1}, $R_{min/2}$=1.7A, |q| = 0.02 e. Four main model simulations, with 1, 5, 10 and 15 copies of dioxygen involved, were preformed (later named LES1, LES5, LES10, LES15, respectively). Each trajectory was 2.2 ns long. The Cyg molecule was immersed ino a TIP3 water box (61 ×45 × 60 Å). The distance from the protein surface to the box border was at least 8 Å. Periodic boundary conditions were used. The cutoffs for non-bonding (van der Waals and electrostatic) interactions were 12 Å. The switch distance was 8 Å, and 1.0 1–4 scaling factor was used.

The simulations were performed in three steps. In the first step, the protein was kept frozen and the equilibration of water molecules was performed for 100 ps of the Langevin dynamics (Langevin dumping was 5 ps^{-1}). Heating from 0 K to 300 K was carried out during the next 100 ps of a standard MD run. Finally, the main 2 ns of the Langevin dynamics simulations were preformed for each LES model. Structures used in these simulations had over 19,000 atoms. For the analysis, VMD [35] and DeepView [36] codes were used. A list of the all amino acids involved in short range interactions with the ligand was obtained by automatic checking for collisions between the ligand (or its LES copy) and all the atoms from the protein. It was assumed that a protein–ligand collision occurs when the distance between the ligand and an atom from Cyg is lower than (or equal to) 2.5 Å.

Cavities and gates in the Cyg model were quantitatively determined. Distances from O_2 ligands to all close-lying amino acids were carefully monitored. Then, the span of cavities (criteria of frequent collisions were used) were established. Common (for neighboring cavities) residues were selected as candidates for gates. Short O_2 passage time was also a factor in providing argument for the existence of the "gate".

The initial structure imported from the PDB (1V5H, ferric form of Cyg) had a six-coordinated heme iron centre and did not have a dioxygen ligand inside the heme pocket. The sixth coordination position was occupied in this structure by distal histidine HisE7. Several steps were performed to obtain Cyg with a five-coordinated iron and O_2 ligand in the heme distal cavity necessary for oxygen transport studies. The bond between the iron ion from the heme group and HisE7 residue was cut and the whole system (Cyg+water) was equilibrated during 300 ps of MD. The residue HisE7 moved away from the heme group and O_2 molecule was placed in the space released by this spontaneous HisE7 "swing". Inspection of O_2 locations in the other heme globins (Hb and Mb) helped to insert manually O_2 to such a prepared Cyg model. We note that careful localization of the ligand in the heme cavity is not necessary for MD simulations of ligand diffusion. Such a final structure of Cyg with 5-coordinated iron ion was used as a starting point in all simulations presented here.

All MD simulations were performed on a dual-core Athlon64 cluster running under Linux Fedora operation system.

In the LES method [28], the simulated structure is divided into two subsystems. In this case, a protein+water box compose one, big subsystem, and the dioxygen ligand build the second, small subsystem. Then, several non-interacting copies of the small subsystem (copies of the

ligand—so called LES copies) were generated. These atoms are subject to nonbonding van der Waals and electrostatic interactions from the big subsystem. Protein experiences an average potential from all copies of the ligand. LES atoms do not "see" each other and can occupy the same space. This approach substantially increased sampling of the selected regions of conformational space and reduces barrier heights [29]. It may be used for free energy calculations [30] and small ligand diffusion paths studies [28, 31].

A brief formal description of the LES method may be found in [37]. One LES trajectory with 10 copies gives, in principle, information equivalent to data obtained from one order of magnitude longer classical MD (i.e. LES1) simulations. In this way, the 2 ns trajectories with 1, 5, 10 and 15 copies of O_2 can bring the same information as 62 ns of standard MD simulation. While it is known that the original LES method [28] may lead to violation of correct equipartition of energy among protein and the enhanced part of the conformational space [38], it is not a serious problem in the present study, since temperature was carefully checked and kept constant by a Langevin dynamics protocol. The temperature was roughly maintained across all the atoms through the addition of friction and random forces. Plots of temperatures along all LES trajectories are presented in Figs. S6, S7, S8, S9 in the Electronic Supplementary Material.

Results and discussion

The inspection of root-mean-square (rms) distances from the minimized Cyg model shows that all 2 ns trajectories are reasonably stable (average rms<1.5 Å, LES15 is the least stable, see Fig. S1 presented in the Electronic Supplementary Material). The main question addressed in this paper is as follows: what are possible diffusion paths for small gaseous ligands, such as O_2, NO or CO, inside Cyg matrix? Is there only one dominant path or rather are several ways of ligand entry/exit possible?

The Cyg protein fold and cavities identified are shown in Figs. 2 and 3. All oxygen diffusion paths started from the heme pocket of Cyg (named here CavHem). The main residues roughly limiting its span are: Phe60, Lys127, Phe50, His81, heme, Lys78. The detailed amino acid composition of this pocket is presented in Fig. 4 and a spatial arrangement is shown in Fig. S10 (Electronic Supplementary Material).

Recent X-ray studies of human Cyg structure (a double mutant C38S, C83S, named CYGB*), filled with xenon atoms under pressure, revealed the presence of four Xe occupancy sites [22]. Three bound Xe atoms are observed in the region of the large internal pocket and one (Xe4) in the region of the narrow tunnel located close to the G and H helices.

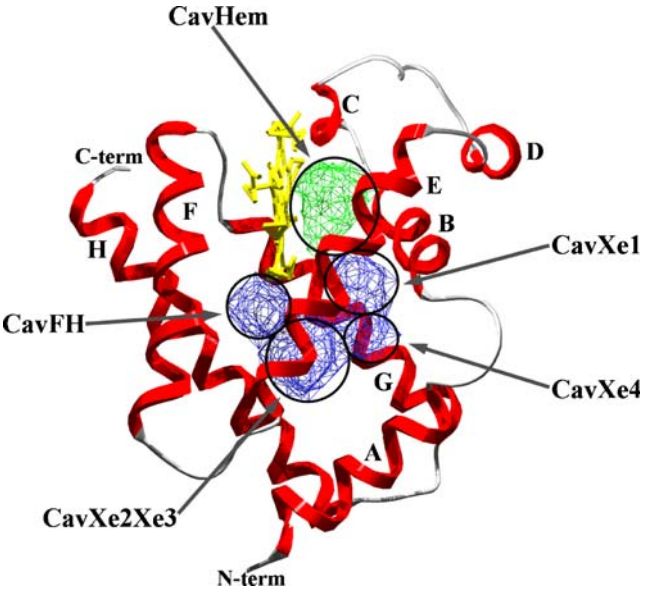

Fig. 2 DeepView drawing of the human Cyg cavities. The helices (*red*) are labeled from N-terminus to C-terminus (*A,B,...,H*). Heme group is shown in *yellow*

The computed cavities (using DeepView code [36]) in randomly selected Cyg structure (370 ps of trajectory LES10) which are schematically presented in Fig. 2 correspond well to the observed Xe occupancy sites [22]. A special overlap using a tool RMSD [align from the VMD code [35] of the X-ray structure of Cyg (1UX9 with docked

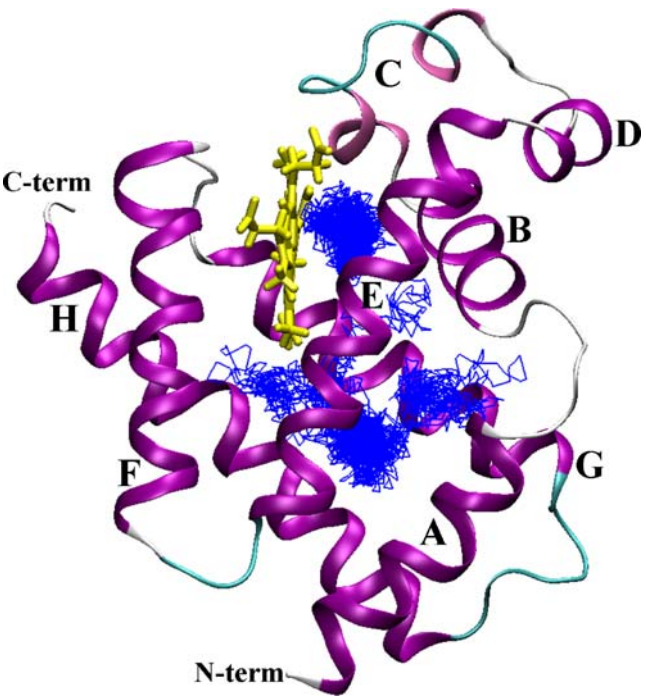

Fig. 3 Trace of the mass center of the randomly chosen ligand during MD simulation. All cavities from Fig. 1 are presented here

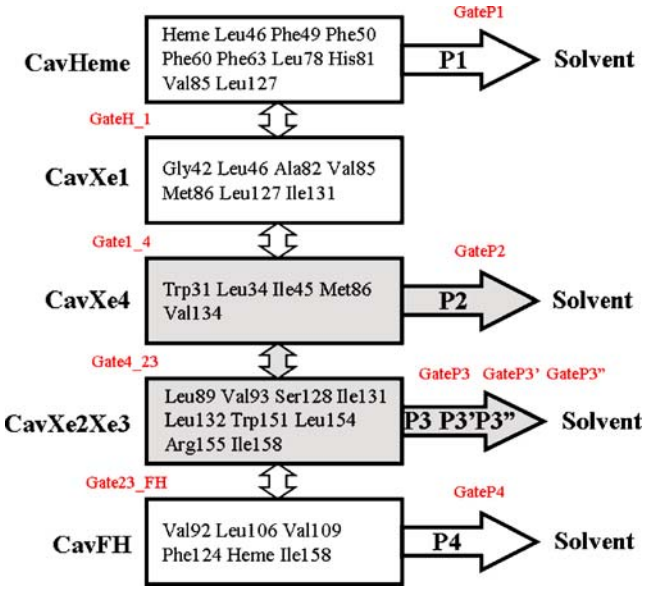

Fig. 4 Diagram of the cavities and gates in human Cyg. Wide hydrophobic tunnel is shown in *light gray*. The exit paths (*P1–P4*) to solvent are labeled. Table 1 contains descriptions of the gates

Xe atoms [22]) and our Cyg model with calculated O_2 trajectory] was performed to find any correspondence between cavities observed in simulations and Xe occupancy sites. The C_α, N, Fe atoms from the following residues were used to make this alignment: Glu21, His81, His113 and the heme group. Then, distances between O_2 copies and all Xe atoms were monitored on representative frames from the trajectory. Regions with frequent presence of O_2 ligand were interpreted as cavities in Cyg. Some cavities were related to the previously noted ones [22] based on the criterion of the shortest distances between O_2 ligands and Xe atoms in the overlapped structures.

Detailed information on amino acids involved in formation of these cavities may be found in Figs. S10, S11, S12, S13, S14 (Electronic Supplementary Material). One should note, however, that in addition to the previously observed Xe1–X4 pockets two new predicted cavities are also indicated. The heme cavity (CavHem, Fig. 2) and a cavity present in vicinity of H and F helices (CavFH, Fig. 2) were not occupied by Xe atoms in the structure studied by de Sanctis et al. [22]. This observation indicates that studies of protein dynamics may reveal novel features important for biological activity.

Using the VMD software [35], traces of every O_2 ligand were registered in all calculated trajectories. Out of 1+5+10+15 ligands observed only 11 copies left the protein on a 2 ns timescale. Such events were possible only in LES10 and LES15 trajectories. A typical trajectory is presented in Fig. 3.

Our dynamical data provide new data in comparison to the static analysis of one X-ray structure. Fluctuations of side groups of amino acids often open transient gates and passages allowing for ligand diffusion. In Cyg, at least three distinct paths (P1–P3) were observed (see scheme in Fig. 4). These routes, illustrated using a protein structure at 360 ps and relevant oxygen positions data from 2 ns trajectory LES10, are presented in Figs. 5, 6, 7 and 8 (and their detailed analysis will be easier using a scheme shown in Fig. 4).

In the standard MD LES1 simulations, O_2 does not leave the Cyg interior. The ligand travels between cavities Hem and Xe2Xe3, visiting on the way Xe1 and Xe4 cavities (see Fig. 2). Approximately 30% of simulation time O_2 is observed in CavHem, 35% in Xe1, 30% in Xe2Xe3, but only 5% in the Xe4 pocket.

A similar picture of ligand diffusion emerges from LES5 simulations. The dioxygen copies penetrate the same cavities as in LES1. Multiple crossings between cavities are observed but none of the copies finds a way into the solvent. The weighted percentage of time spent by copies of oxygen at Heme, Xe1 and Xe2Xe3 cavities is 70%, 10%

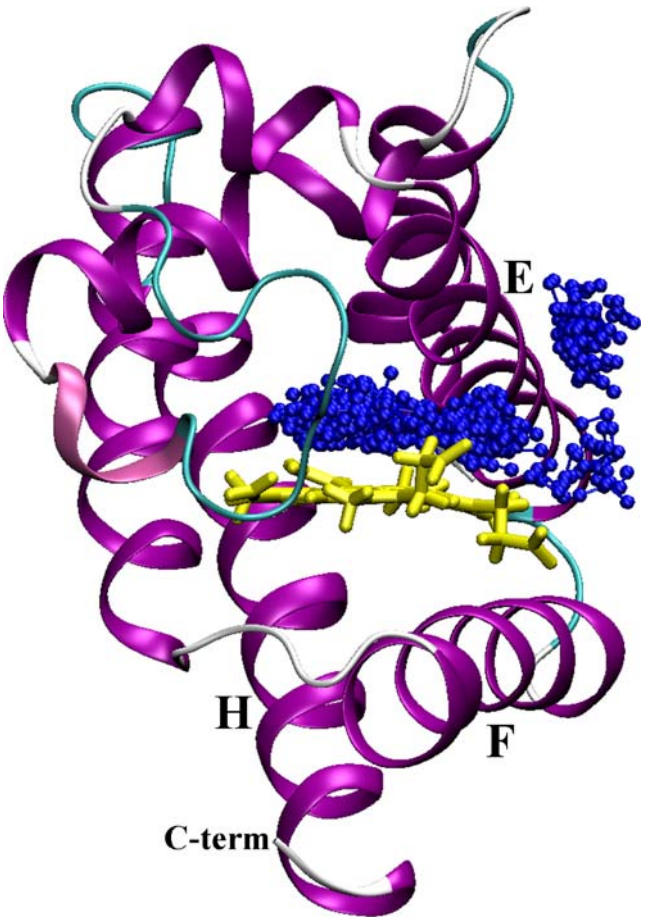

Fig. 5 Pathway P1. Ligand (dioxygen) is shown in *blue*. This route leads directly from HemeCav to solvent through gate GateP1 (see Table 1). The amino acids building HemeCav cavity are listed in Fig. 4. A picture of the HemeCav itself is available in Electronic Supplementary Materials (Fig. S10)

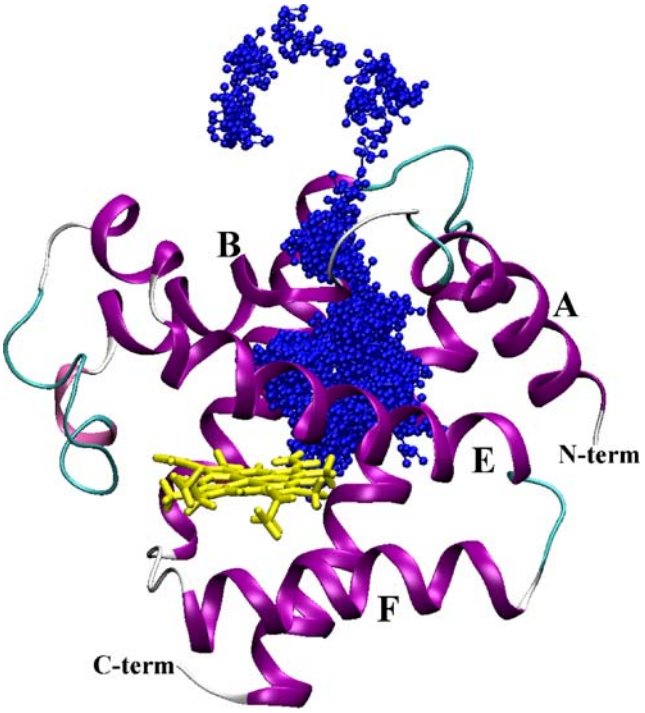

Fig. 6 Pathway P2. Ligand (*blue*) moved from CavXe4 cavity and left Cyg through gate GateP2 (see Table 1). A picture of the CavXe4 itself is available in Electronic Supplementary Materials (Fig. S12). CavXe4 is a part of wide hydrophobic tunnel

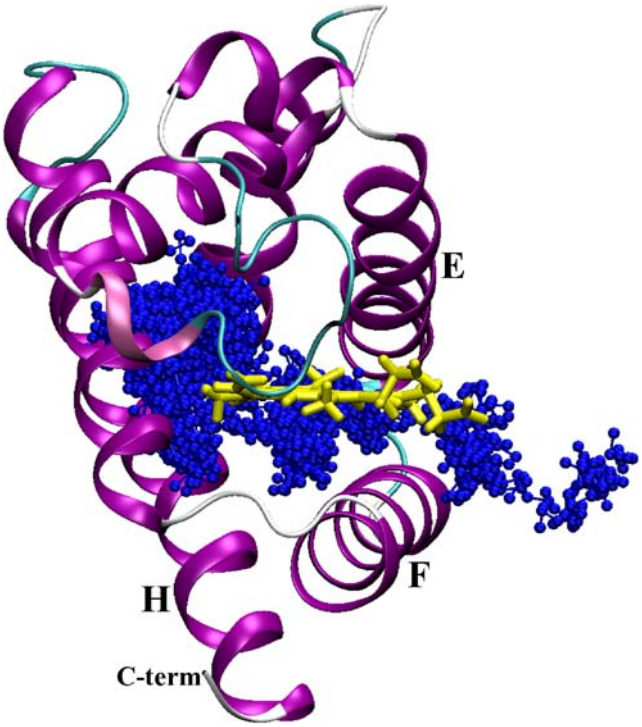

Fig. 8 Pathway P4. Ligand located in the cavity CavFH left protein through gateP4. For more details about the gate and amino acids building the CavFH see Table 1 and Fig. S14 (in Electronic Supplementary Materials)

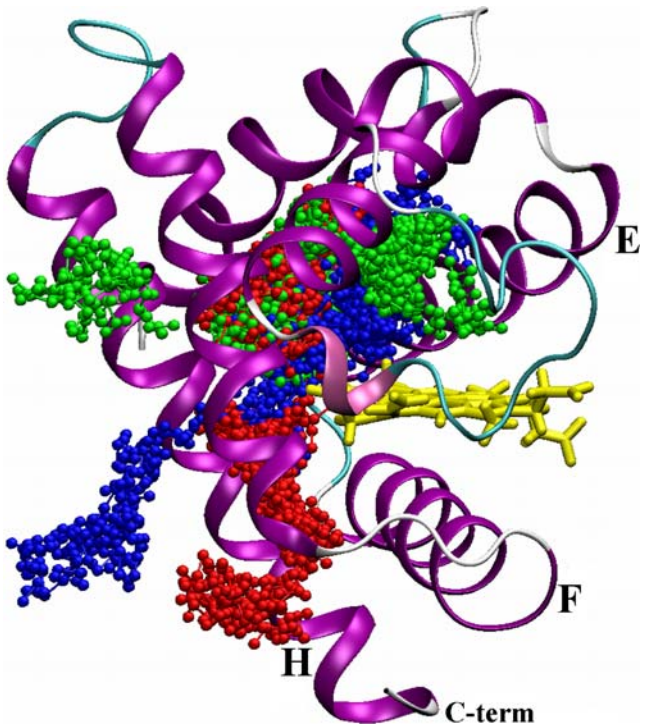

Fig. 7 Pathways P3 (*green*), P3′ (*red*) and P3″ (*blue*). These pathways are located between G and H helices. Ligand moved from CavXe2Xe3 cavity to solvent using gateP3, gateP3′ or gateP3″. A detailed picture of the CavXe2Xe3 is available in Electronic Supplementary Materials (Fig. S13). CavXe2Xe3 constitute the main part of wide hydrophobic tunnel

and 20%, respectively. The time spent by dioxygen molecules in Xe4 is negligible in LES5 trajectory.

In both LES10 and LES15, many copies of the ligand escape to the solvent (6 and 5, respectively). Schematic routes and all determined gates are represented in Fig. 4. Since steric barriers are lowered in the LES approach, ligand copies may even reach a distant cavity, CavFH shown in Fig. 3. De Sanctis et al. [22] have already discussed the possibility of such a cavity in CYGB*, but in crystals filled by xenon gas under pressure this cavity was not observed. Here, we can confirm that this part of the protein is rather rarely visited by small gaseous ligands. However, one should note that this cavity may play some physiological role because at least one escape route (P4) goes through this region. Qualitative picture of oxygen distribution in a real Cyg may be obtained by simultaneous visual inspection of Figs. 5, 6 and 7. In these figures, representative escape paths calculated in the LES10 trajectory are shown.

Possible escape/entry routes (Fig. 4) are described below. These data may help to design engineered variants of human Cyg and to check experimentally the nature of dioxygen diffusion in this globin.

The simplest path, P1, leads directly from the heme distal pocket to the solvent (Fig. 5). The ligand passes "over" the heme group and "below" HisE7(81) and ArgE10 (84). This path corresponds to distal histidine swing often

observed in myoglobins [39]). Such an event has occurred only once in the LES10 trajectory after 128 ps of simulations, thus we expect that it is not the most common exit path for Cyg. No particular sudden jumps in torsional angles of residues involved in this gating event were observed. The main obstacle on this passage appears to be the ArgE10(84) residue.

The remaining exit paths, i.e. P2, P3 and P4, are much longer (see Fig. 4). The ligands travel first to CavXe1, passing several amino acids-GateH_1 (Table 1). We shall call the passage region "gate" although any real opening/closing motions are not necessarily observed. Criteria for selecting gates are presented in Methods. Then, the ligand moves into CavXe4. It remains in this cavity for a long time. In the P2 path (Fig. 6), a copy of the ligand leaves the protein at 930 ps. The GateP2 is located close to LeuG15 (34), ValG15(134) and IleB9(45) (AB loop and helix G). This path seems to be quite important because nearly 50% of exit events in the current simulations were observed in this region. It is a natural path, since it is a part of wide hydrophobic channel already indicated in the X-ray structure [21]. Some ligands trapped in CaveXe4 find the way to the large CavXe2Xe3 through the Gate4_23 (Fig. 4). From this cavity, three exit events (240 ps, 650 ps, 1.3 ns) were observed inbetween the G and H helices (P3, P3′, P3″, respectively). All P3 paths are shown in Fig. 7 and details on locations of these exits are presented in Table 1. All P3 paths are located within the large channel; GateP3′ is just on the opposite side from gateP2 (gray area in Figs. 4 and 7).

Interestingly, in LES10 and LES15 simulations (equivalent to 20 ns or 30 ns standard simulations) a cavity close to the proximal side of heme and the helix H is populated by O_2 copies (CavFH, Figs. 2 and 3). After 1.15 ns in LES10 and 1.9 ns in LES15 single copies of ligands leave this cavity into solvent through gateP4 (see Fig. 8). The exit is located near AlaE14(88), ValG7(105) and ValG11(109).

Table 1 Description of the protein-solvent and protein-protein Cyg gates

Gate name	Residues
Protein–solvent gates	
GateP1	Heme His81 Arg84
GateP2	Leu34 Val134 Ile45
GateP3	Trp151 Leu132 Arg155 Ala152
GateP3′	Thr163 Phe124 Pro121 Thr159
GateP3″	Tyr159 Arg155 Lys125
GateP4	Ala88 Val105 Val109
Protein–protein gates	
GateH_1	Val85 Leu46
Gate1_4	Met86 Ile131
Gate4_23	Ile131 Trp151 Trp31
Gate23_FH	Heme Ile158

Protein–solvent gates lead from protein matrix to solvent. Protein–protein gates are located between cavities inside protein

The ultimate understanding of ligand diffusion in protein matrix is based on time-resolved X-ray crystalography [40] and/or extensive kinetic studies on selected mutants of a given globin [41, 42]. In order to facilitate design of variants of human Cyg suitable for further experimental studies, collision counts for paths P1, P2 and P3 observed in the LES10 trajectory are presented in Fig. 9. The diagrams show that perhaps ArgE10(84), HisE7(81) (P1), IleG12(131) (P2), ThrG20(163), ProG2(121), LeuB10(46) (P3) affect ligand diffusion the most. The statistics of collisions may be improved, but these dynamical data provide for the first time this information on Cyg properties.

It is interesting to compare present findings with the results of MD simulations of ligands diffusion in Mb [23–25, 27, 29]. In their important paper, Elber and Karplus have shown that the LES method predicts multiple diffusion paths for CO in sperm whale (SW) Mb [28]. Out of 60 copies of CO ligand, 30 molecules left SW Mb interior within 100 ps of LES dynamics. Similarly to our results for Cyg, there were five major pathways in Mb. On the most common pathway, CO went from the heme distal pocket to the AB/G cavity (equivalent to Xe4 in xenon-filled Mb [43] or our CavXe1 in Cyg), made a second jump to the EF/G cavity (Xe2 in Mb), and escaped near Trp14 [28]. The first phase of this Mb path is the same as initial parts of all paths observed in Cyg. In present simulations, ligands reached regions of CavXe2Xe3 which are located close to the EF/G cavity observed in [24], but this Cyg cavity is different (see diagram in Fig. 4). The last part of the major CO escape route in Mb is oriented in the direction opposite to paths determined in the present study. In Mb, exits are located close to Trp14, but in Cyg, the majority of exits are close to Arg 155 (see gates P3, P3′, P3″ in Table 1). Out of 12 amino acids which most frequently collide with CO in Mb, 10 corresponding amino acids were also most often participating in O_2 collisions with Cyg (in parenthesis): Leu29 (Leu46), Val68 (Val85), Leu69 (Met86), Ile107 (Leu127), Ile111 (Ile131), Ile28 (Ile45), Leu72 (Val93), Leu135 (Arg155), Phe138 (Ile158), Leu104 (Leu124). Careful analysis of Fig. 4 shows that 19 residues, out of 35 participating in formation of cavities in Cyg, are conserved in Mb sequence. Thus, amino acids making pockets are in Cyg are better conserved (55%) than the average 25%.

This observation is in agreement with analysis of high resolution alternate A1 and A3 Mb crystal structures (higher than 1.5 Å) performed by Teeter [26]. Paths leading from the distal to the proximal side of heme were analyzed. Clusters of cavities were found to be a fundamental part of myoglobin dynamics. While a detailed comparison of cavity composition in Mb and in our Cyg model is not

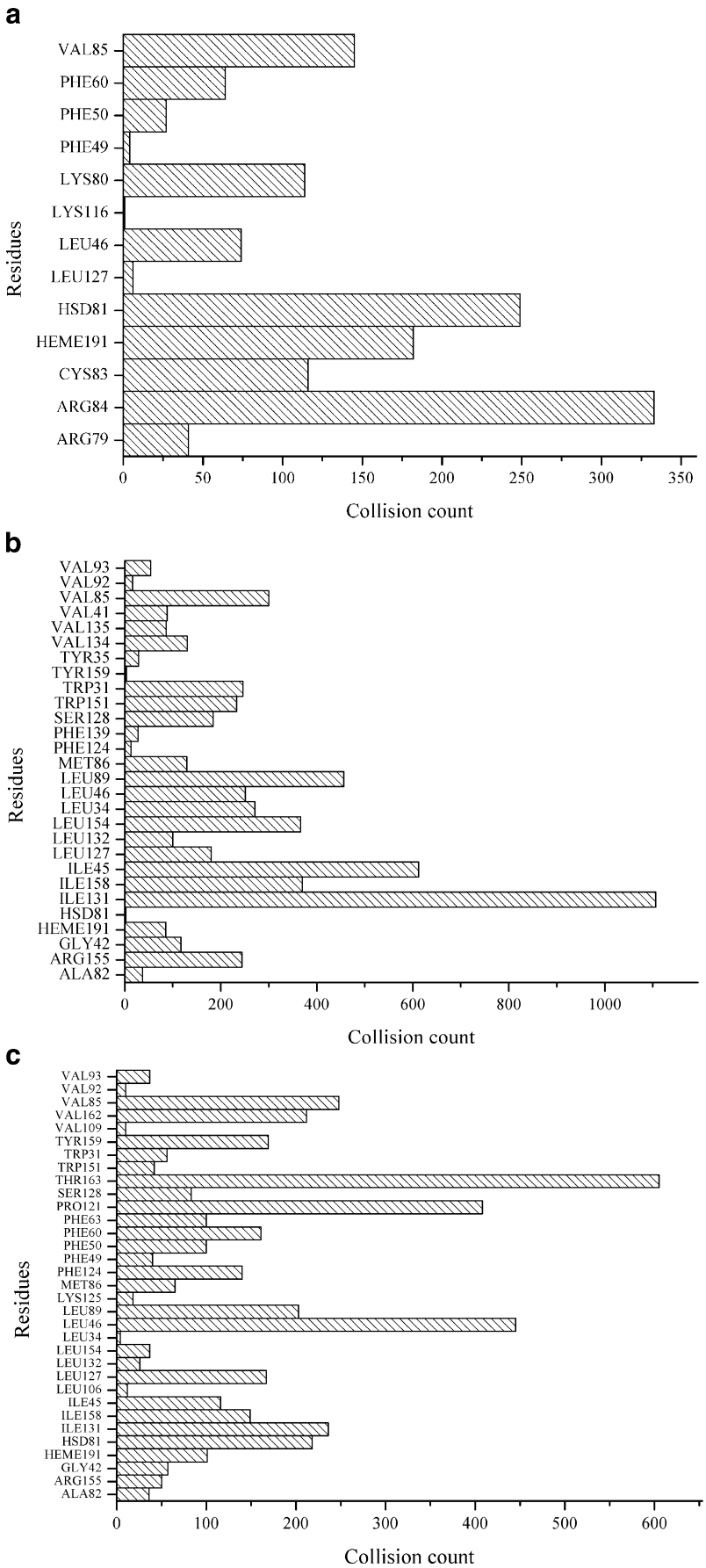

Fig. 9 Statistics of collisions between selected ligands and residues of the Cyg. Plot (**a**) corresponds to P1 path, (**b**) corresponds to P2 path and (**c**) to P3′

possible, due to a lack in the paper [26] of a list of amino acids which form cavities, visual inspection of cavities presented in Fig. 1 of [26] and our data (Fig. 2) gives similar pictures. Correspondence is as follows (Mb names in parentheses): CavHem (Xe5), CavXe1 (Xe4), CavXe2Xe3 (Xe2), CavXe4 (purple transient state), CavFH (~additional cavity in A3 conformation of Mb). We did not observe any population of O_2 on the proximal side of the heme, so the cavity Xe1 of Mb discussed by Teeter [26] was not considered in the present study. One may expect that, on a longer simulation timescale, such a cavity in Cyg may be visible. It is possible that, in contrast to Mb, in Cyg a small gaseous ligand would rather escape to the solvent than pay a visit in the proximal side. Teeter [26] noted a special role of Phe138 in controlling CO diffusion in Mb. We have found that, in human Cyg, Ile158 also controls the passage from cavity CavXe2Xe3 to CavFH (see Table 1 and Fig. 4). Note that in the alignment presented in Fig. 1 these residues are on the same position.

Nutt and Meuwly [23] calculated 24 short (100 ps each) trajectories of photodissociated CO in L29F mutant of SW Mb. On such a short timescale the authors observed CO mostly in the heme cavity. However, to some extent, a cavity corresponding to our Xe1 (Fig. 2) cavity was also populated. Rare presence of CO ligands in the region between Phe29 and His64 (numbering of Mb L29F is used here, see Fig. 1) was noted. In our Cyg simulations, a population of O_2 in this region was also observed, but in the present analysis it has simply been added to the heme cavity statistics. We have found that collisions of O_2 with Leu46 of Cyg (corresponding to position 29 in Mb [23]) are indeed frequent (see Fig. 9a). Thus, this particular amino acid plays a functional role in Cyg as well as in Mb. No ligand exits were observed in the reported simulations [23].

Bossa et al. analyzed much longer (close to 100 ns) simulations of CO diffusion in SW Mb WT [24] and its YQR triple mutant [25]. For WT SW Mb, a classical (one ligand) trajectory was calculated [24] and only one migration path was observed. The ligand spent 10% of simulation time in the heme pocket, then it moved to a cavity corresponding to our Xe1 pocket (8% of time) and through so-called phantom cavities until it had finally reached a cavity on the proximal heme side. CO remained there for over 50% of the simulation time. Some re-crossings to neighboring cavities were observed but no ligand exit to solvent had occurred. In Cyg simulation, no O_2 ligands in a cavity corresponding to this proximal cavity were observed. Possibly, the distinct protein architecture prevents O_2 from visiting the proximal side in Cyg and this feature has appeared rather recently in evolution. Another explanation for this difference between Cyg and Mb may be related to the presence of a large channel leading to Cyg exterior. Present data show that O_2 ligands prefer to escape via this channel instead of visiting the heme proximal cavity. The closest region to the heme proximal side reached by LES ligands in Cyg was the cavity named CavFH (see Fig. 2).

The most recent and comprehensive study by Cohen et al. brings a qualitative picture of O_2, CO, NO and Xe migration in Mb matrix [27]. Using a newly-developed implicit ligand sampling method the authors computed potential of mean force (PMF) corresponding to the placement of small gaseous ligands everywhere inside Mb. Low PMF regions of Xe and CO transport correlate very well with experimental Xe cavities in Mb. Based on 5 ns MD trajectory in addition to a "classical" short distal exit pathway (gated by HisE7(64), in Cyg the corresponding path is named P1) two additional sets of exits from Mb with low energy barriers and three with high barriers were discovered. These findings are in good agreement with our data showing multiple exit paths in Cyg. O_2 ligands in Cyg also reached the same regions distant from the heme group (CavXe2Xe in Fig. 2 or Xe3 in Fig. 4 [27]). However, the detailed route of escaping ligands are different: our paths P2–P4 compare well up to CavXe2Xe3 (see Fig. 4), but the second parts lead to exits located in opposite regions from Mb exits. Our paths P2, P3, P3′ and P3″ are connected with the Cyg hydrophobic channel absent in SW Mb.

A physiological role of Cyg is not yet recognized [14, 18]. Oxygen (or other small gasous ligands such as NO or CO) perhaps penetrate Cyg matrix in biological settings. Our MD LES simulations show that there are at least five cavities in which ligands may remain for prolonged periods (see Fig. 4). However, from four of them there are direct exits to the solvent possible. Only one cavity—CavXe1—is pretty isolated from the exterior. Thus, we put forward a hypothesis that this cavity, located close to heme distal cavity, is a "molecular waiting room". If a small ligand has to react with iron, it has to wait until, normally, six-coordinated Fe^{3+} loses its bound to the distal HisE7(81). Thus, binding of ligand is biphasic [12]. A similar secondary docking site for Mb was discussed by Brunori [44]. When high resolution X-ray structures on Cyg are one day available, present MD data may serve for an interior ligand pathway determination, similar to that performed by Teeter for Mb [26].

Conclusions

Molecular dynamics simulations of a realistic model of newly-discovered ancient human cytoglobin were performed. Increased sampling of conformational space achieved by multiplication of O_2 ligands and appropriate scaling of ligand–protein interactions by the LES method [28, 37] d′g, allowed for determination of five cavities accessible to ligands. It has been found that, apart from a direct exit from the heme distal cavity to the solvent, at

least three additional major routes are possible. Simulation data confirm that the most probable exits are located within a long channel visible in the static X-ray structure of human Cyg [22]. However, many additional details, including the role of a new cavity located on the proximal side close to the H helix, have been revealed by long timescale simulations. It is predicted that the following residues should have the most pronounced effects on ligand transport kinetics in Cyg: LeuB10(46), HisE7(81), ArgE10(84), ProG2(121), IleG12 (131), ThrG20(163). One cavity, lying close to the heme distal side, does not exhibit direct connection with the solvent. Probably small gaseous ligands remain in this isolated space until reaction with the iron centre is possible.

Acknowledgements This work was supported by grant "Krok w przyszlosc-stypendia dla doktorantow" (S.O.), President of Poland grant "Superpracownia 2002" and in part by MEiN grant 2P04A 07229 (W.N.). We thank CI TASK for computer time and Albert Rutkowski for his assistance in computations.

References

1. Kendrew JC (1963) Science 139:1259–1266
2. Perutz MF (1979) Annu Rev Biochem 48:327–386
3. Wittenberg JB (1970) Physiol Rev 50:559–636
4. Elber R, Karplus M (1987) Science 235:318–321
5. Wittenberg BA, Wittenberg JB (1989) Annu Rev Physiol 51:857–878
6. Perutz MF (1990) Annu Rev Physiol 52:1–25
7. Springer BA, Siligar SG, Olson JS, Philips GN (1994) Chem Rev 94:699–714
8. Tilton RF, Kuntz ID Jr, Petsko GA (1984) Biochem 23:2849–2857
9. Meller J, Elber R (1998) Biophys J 74:789–802
10. Burmester T, Weich B, Reinhardt S, Hankeln T (2000) Nature 407:520–523
11. Kawada N, Kristensen DB, Asahina K, Nakatani K, Minamiyama Y, Seki S, Yoshizato K (2001) J Biol Chem 276:25318–25323
12. Pesce A, Bolognesi M, Bocedi A, Ascenzi P, Dewilde S, Moens L, Hankeln T, Burmester T (2002) EMBO Rep 12:1146–1151
13. Burmester T, Ebner B, Weich B, Hankeln T (2002) Mol Biol Evol 19:416–421
14. Riggs A, Gorr T (2006) PNAS 103:2469–2470
15. Fordel E, Thijs L, Martinet W, Lenjou M, Laufs T, van Bockstaele D, Moens L, Dewilde S (2006) Neurosci Lett 10:146–151
16. Fago A, Hundahl C, Dewilde S, Gilany K, Moens L, Weber RE (2004) J Biol Chem 279:44417–44426
17. Schmid M, Gerlach F, Avivi A, Laufs T, Wystub S, Simpson JC, Nevoll E, Saaler-Reinhardt S, Reuss S, Hankeln T, Burmester T (2004) J Biol Chem 279:8063–8069
18. Hankeln T, Ebner B, Fuchs C, Gerlach F, Haberkamp M, Laufs TL, Roesner A, Schmidt M, Weich B, Wystub S, Saaler-Reinhardt S, Reuss S, Bolognesi M, De Sanctis D, Marden MC, Kiger L, Moens L, Dewilde S, Nevo E, Avivi A, Weber RE, Fago A, Burmester T (2005) J Inorg Biochem 99:110–119
19. Nakatani K, Okuyama H, Shimahara Y, Saeki S, Kim DH, Nakajima Y, Seki S, Kawada N, Yoshizato K (2004) Lab Invest 84:91
20. Dewilde S, Kieger L, Burmester T, Hankeln T, Baudin-Creuza V, Aerts T, Marden MC, Caubergs R, Moens L (2001) J Biol Chem 276:38949–38955
21. De Sanctis D, Dewilde S, Pesce A, Moens L, Ascenzi P, Hankeln T, Burmester T, Bolognesi M (2004) J Mol Biol 336:917–927
22. De Sanctis D, Dewilde S, Pesce A, Moens L, Ascenzi P, Hankeln T, Burmester T, Bolognesi M (2004) Biochem Biophys Res Commun 316:1217–1221
23. Nutt DR, Meuwly M (2004) Proc Natl Acad Sci USA 101:5998–6002
24. Bossa C, Anselmi M, Roccatano D, Amadei A, Vallone B, Brunori M, Di Nola A (2004) Biophys J 86:3855–3862
25. Bossa C, Amadei A, Daidone I, Anselmi M, Vallone B, Brunori M, Di Nola A (2005) Biophys J 89:465–474
26. Teeter MM (2004) Protein Sci 13:313–318
27. Cohen J, Arkhipov A, Braun R, Schulten K (2006) Biophys J 91:1844–1857
28. Elber R, Karplus MJ (1990) J Amer Chem Soc 112:9161–9175
29. Roitberg A, Elber R (1991) J Chem Phys 95:9277–9287
30. Verkhivker G, Elber R, Nowak W (1992) J Chem Phys 97:7838–7841
31. Nowak W, Czerminski R, Elber R (1991) J Amer Chem Soc 113:5627–5637
32. Sugimoto H, Makino M, Sawai H, Kawada N, Yoshizato K, Shiro Y (2004) J Mol Biol 339:873–885
33. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) Nucleic Acids Res 28:235–242
34. Kalr L, Skeel R, Bhandarkar M, Brunner R, Gursoy A, Krawetz N, Phillips J, Shinozaki A, Varadarajan K, Schulten K (1999) J Comp Phys 151:283–312
35. Humphrey W, Dalke A, Schulten K (1996) J Molec Graphic 14:33–38
36. Guex N, Peitsch MC (1997) Electrophoresis 18:2714–2723
37. Orlowski S, Nowak W (2007) Theor Chem Acc 117:253–258
38. Karplus M, Straub J (1991) J Chem Phys 94:6737–6739
39. Scott EE, Gibson QH, Olson JS (2001) J Biol Chem 276:5177–5188
40. Schotte F, Lim M, Jackson TA, Smirnov AV, Soman J, Olson JS, Phillips GN, Wulff M, Anfinrud PA (2003) Science 300:1944–1947
41. Trent JT, Hargrove MS (2002) J Biol Chem 227:19538–19545