Theor Chem Acc (2007) 117: 253–258
DOI 10.1007/s00214-006-0139-8

REGULAR ARTICLE

Slawomir Orlowski · Wieslaw Nowak

Oxygen diffusion in minihemoglobin from *Cerebratulus lacteus*: a locally enhanced sampling study

Received: 26 June 2005 / Accepted: 6 October 2005 / Published online: 5 July 2006

Abstract Heme proteins serve as a source of oxygen in nervous tissue during anoxia. The functional routes of a dioxygen (O_2) diffusion in a novel structure of a minihemoglobin (CerHb) molecule present in worm *Cerebratulus lacteus* are not known. In this paper, the results of 1 ns molecular dynamics simulations of this process are presented. The locally enhanced sampling method (LES) and CHARMM force field were used for simulations of CerHb with 1–15 copies of O_2. It was found, that several alternative routes are possible. The dominant path consists of two steps. Firstly, ligands move from the heme pocket to a different cavity through the barrier defined by the residues Phe10 and Tyr48. Secondly, ligands leave the protein passing through the more complex barrier situated between the E/F loop and the H helix. We note that the number of paths observed depends on a number of LES copies of (O_2).

Keywords Minihemoglobin · Locally enhanced sampling · Molecular dynamics

1 Introduction

Globins are small respiratory proteins (140–160 residues) containing heme prosthetic group, namely the Fe-protoporphyrin IX ring. Heme is hosted in a protein pocket defined by distal E- and proximal F-helices. Gaseous ligands (e.g. dioxygen, O_2) are bound reversibly by the heme-Fe ion on distal side. On the proximal side, the Fe atom is coordinated by a characteristic HisF8 residue. Despite the large differences in their primary and quaternary structures, the globins present a well-conserved tertiary structure organized in a three-on-three α-helical sandwich, the so-called globin fold [1, 2] (Fig. 1a). Most globins serve either as O_2 and CO_2 transport proteins (hemoglobins) or as oxygen storage in muscles (myoglobins).

S. Orlowski · W. Nowak (✉)
Theoretical Molecular Biophysics Group, Institute of Physics,
N. Copernicus University, 87-100 Torun, Poland
E-mail: wiesiek@phys.uni.torun.pl

Recent research extended the list of vertebrate globins by two new members: Neuroglobin (Ngb) and cytoglobin (Cgb) [3, 4]. Both molecules belong to a class of hexa-coordinated globins. Ngb is expressed in nervous tissue, whereas Cgb is expressed at low concentration in all tissues.

Here, we describe ligand diffusion in nervous tissue minihemoglobin from Milky ribbon-worm (*Cerebratulus lacteus* – CerHb) – another member of the neural globin family [5, 6]. Its sequence contains only 109 amino acids, thus this molecule is a very attractive model system (Fig. 2b). The deletion of the A-helix and a shortening of the H-helix in comparison with the "standard" myoglobins may form the structural basis for the difference in affinity of CerHb to gaseous ligands and that measured for the other known globins. The structure of CerHb was discovered by Pesce et al. [5]. These authors indicated a wide channel leading from the heme cavity to the solvent but there is no useful information about a detailed route of the ligand from the cytosol to the heme pocket. The approximate channel has been determined using SURFNET code [5] applied to the static X-ray structure of CerHb. While CO diffusion in myoglobins was intensively studied using computer modeling [7–11], no molecular dynamics simulations of ligand diffusion in this interesting system have yet been performed. Since it is important to know what residues affect ligand diffusion, in this work we have initiated extensive, nanosecond scale MD studies of O_2 diffusion in the CerHb matrix using CHARMM27 force field. Molecular dynamics data support hypothesis presented by Pesce et al. [6] that Thr-E11 regulates O_2 affinity in CerHb. Moreover, our results may encourage further kinetic experiments on CerHb variants and suggest what model of ligand diffusion – "ballistic" or "sponque" is more appropriate for minihemoglobins [7].

2 Methods

Structure of *C. lacteus* minihemoglobin was obtained from Protein Data Bank (PDB [12] code of the oxy form is 1KR7). Simulations of the minihemoglobin using the locally enhanced sampling (LES) [7,13] algorithm were carried out

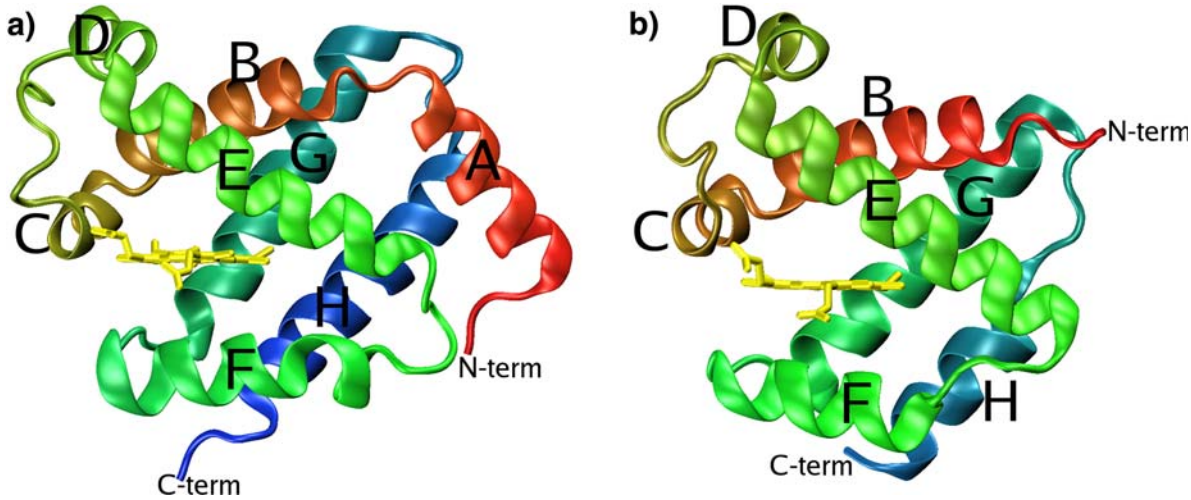

Fig. 1 **a** The classical globin fold (human hemoglobin 1A3N). **b** Minihemoglobin from *Cerebratulus lacteus*. The helices are labeled from N-terminus to C-terminus (A, B, . . . , H). Heme group is shown in *yellow*

using the NAMD program [14] with CHARMM27 force field. The LES method is briefly described in the Appendix. Four simulations, each 1 ns long, with 1, 5, 10, and 15 copies of O_2 ligand were obtained. All simulations were performed in a model water box (TIP3, $49.5\times49.0\times50.0\,\text{Å}^3$, at least 7 Å distance from the protein atoms to the border was kept). The cutoffs for van der Waals and electrostatic interactions were 12 Å. The switch distance was 8 Å, and 1–4 scaling was used. The protein structures were initially frozen, thus the equilibration of water molecules was performed for 50 ps. The heating from 0 to 300 K was performed during the next 100 ps of dynamics. During the main simulations the temperature was held at 300 K by Langevin dynamics (Langevin damping factor was 5). Final structures used in these simulations had over 11,000 atoms. The analysis was performed using the VMD code [15]. Each trajectory was calculated on a cluster of 4×PentiumIV 2.0 GHz, 512 MB RAM running Linux Fedora 2.0 operation system. The Figs.7,8,9,10,11 were prepared using trajectory path script from the VMD web site (authors: Andrew Dalke and Axel Kohlmeyer). Aminoacid composition of ligands diffusion pathways were obtained using a script written by S.O. It was assumed that protein–ligand contacts (collisions) occurred when the distance between the ligand and an atom from hemoglobin was lower than 2.5 Å.

The idea of the LES method is rather simple [7] (see Appendix). The whole system is divided into two subsystems: a big one and a small one. For example, in our case: The protein (CerHb – 1,674 atoms)+water box (over 9,000 atoms) compose the big subsystem and the ligand (dioxygen – 2 atoms) — the small one. Several non-interacting copies of the small subsystem (copies of the ligand) were built. These atoms are affected by a non-bonding (van der Waals and electrostatic) potential from the protein. On the other hand, the big system (protein+water) experiences an average potential from all copies of the ligand. In this way, enhanced atoms can occupy the same space and sampling of the conformational space is more efficient than in the case of just one small ligand and the barriers are reduced [16]. This approach not only increases the sampling intensity, but also the transition rates for ligands. The LES trajectories with ten copies may in principle give information equivalent to data obtained from one order of magnitude longer classical MD simulations.

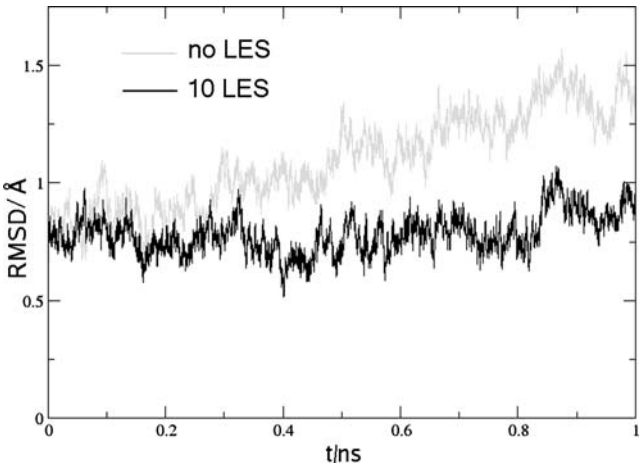

Fig. 2 RMSD (Cα atoms) for trajectory with one and ten copies of dioxygen

3 Results and discussion

In all trajectories Cα RMS (root-mean-square distance) with respect to the pdb protein structure did not exceed 1.5 Å, thus we consider these model as reasonably stable (see Fig. 2).

The lack of the A helix, a shorter H helix and the presence of large, wide tunnel indicate that a novel diffusion pathways in CerHb may be expected in comparison to the standard myoglobin picture [7,17,18]. 1 ns classical MD trajectory (1 copy, no LES) was too short to observe the dioxygen

escape to the exterior of the protein matrix. The ligand sampled heme cavity and occasionally visited the other cavity — CavI, somehow close to the so called pocket Xe4 in Mb [17–19] — see Fig. 3a. The model system with five LES copies of O_2 did not give any exit paths either (Fig. 3b). Only for LES trajectories with 10 and 15 copies we found successful 2 and 4 exit paths, respectively. Thus, six different routes were found in total.

Analysis of all MD trajectories and O_2 collisions shows that at least two pockets accessible to O_2 are present inside CerHb except for the heme distal pocket (CavH, Fig. 3). The main heme pocket is composed of Tyr11, Leu14, Phe15, Phe25, Gln44, Thr48 and Phe10 residues. The first pocket is rather small and is composed of Tyr11, Phe27, Leu35, Ala40, Tyr41 (CavI). The second pocket (CavII) is composed of Ile52, Ile56, Asp61, Leu65, Leu86, Leu98, Ile102, Ala101.

There are two alternative routes for trajectory with ten copies of the dioxygen:

T1. The ligand left heme pocket, just after 15 ps of dynamics, through a cleft between Gln44 and Lys47 (Fig. 4a). This path is an analog to the classical ligand exit path postulated for myoglobins, where E7 residue gate controls release and uptake of the O_2 [5]. Interestingly, Pesce et al. [5] considered this path in CerHb to be rather impossible, since substantial conformational transitions of X-ray determined location of side chain of Gln 44 would be required. Results of LES10 simulations have shown that such a path is possible. The shortest distance responsible for a "bottleneck" in **T1** path is that for OE1 Gln44-CD Lys44 pair of atoms. The 1 ns average value of this distance is 5.2 Å, minimum is 3.6 Å, maximum – 7.1 Å, but a value of 6.2 Å is large enough to observe the passage of O_2 through this gate. Such an "open" distance was observed only during 3% of the simulation time. Thus, despite the fact that 50% of all O_2 collisions noticed in the first 100 ps of **T1** trajectory are with Gln44 and/or Lys47 residues (other collisions are with heme, Thr48, Thr43, Tyr11, Phe10) the "classical" E7 exit path in CerHb is of rather low probability.

T2. Three ligands moved from the heme pocket through the "gate" at Tyr11 into CavI (Fig. 4b). Then O_2 copies left CerHb through a passage composed of Tyr41, Asp8, Ala45, Cys42, Val7, Ala4, Val49 (between B and E helices). The close contacts between O_2 and Ala45 (2.45 Å) and Asp8 (3 Å) were observed. The process of O_2 diffusion from heme pocket to the solvent required long times of 720 and 800 ps to be completed.

In LES10 trajectory none of the ligands escaped through the main channel postulated on the basis of the X-ray data [1], however the trajectory with 15 copies gave such events. In LES15 the following four possible paths for ligand diffusion were obtained:

T3. The ligand left the heme pocket at 380 ps directly to the solvent through the barrier at Leu14, Asn81; a final phase of this O_2 copy diffusion is not far from His18 (Fig. 5a). Out of the 2,916 collisions of this O_2 ligand with aminoacids the following statistics is obtained: Phe25–269, Leu14–265, Lys88–245, Cys85–244, Tyr11–240, Phe10–230, Heme–215, Thr48–192, Gln44–169 and Asn81–120. There are also much less frequent collisions with 14 other aminoacids. **T3** path clearly represents a variant of "classical" path where a ligand finds a way out almost directly from CavH to the solvent.

T4. Is yet another representative of such path. A ligand left the heme pocket (320 ps) through the barrier at Lys47 and Tyr51 (Fig. 5b). The critical distance between heme CMA–Tyr51 CD1 atoms fluctuates within 3.5–4.5 Å, but some five episodes (<10% of simulation time) of large distances of 6 Å were found in our 1 ns LES15 trajectory. Our "successful" ligand passes through the heme-Tyr51 during the first of these events.

Finally, **T5** and **T6** routes are located within the large and wide tunnel present in CerHb [5]. The corresponding ligands moved from the heme pocket to the second pocket (Cav II) passing by Phe10, Tyr48 and then Ile52, Leu86. This looks like a sort of double gate. The following collisions (number in parentheses) with the hydrophobic residues which form the channel are observed: Val7 (145), Phe10 (215), Ile52 (508), Ala55 (256), Leu86 (278), Leu98 (316), Ala101 (244), Ile102 (146), Ile105 (128), Thr48 (217), Tyr51 (164). From there the ligands had left protein (at 750 ps) close to Ile56, Leu98 and Ala101 (Fig. 6a). In **T6**, like in the route **T5** the ligand moved from the heme pocket to the CavII pocket, but then left the protein matrix through a slightly different neighborhood of Ala62, Ala101, Asp104 (Fig. 6b). Statistics for collisions in the channel is similar to that of **T5**. It was rather suprising to see that our standard 1 ns MD simulations was not long enough to allow for O_2 diffusion through the wide escape channel present in CerHb. Even in LES calculations over 50% of escape routes were through more or less direct paths to the solvent. The main residues blocking the access to the channel are Phe10 and Thr48. It seems that the structure of minihemoglobins, or at least CerHb allows for alternate pathways for dioxygen uptake and release.

Presented data, showing multiple cavities in CerHb, are consistent with the recent results obtained by Pesce et al. [6]. Low temperature derived spectroscopy (3 –160 K) indicates that recombination of photolyzed CO ligands occurs from some distant C and D sites. According to these authors, the hydrogen bond between Thr–E11 OG1 atom (or Thr48) and Tyr–B10 (or Tyr11) side chain plays a critical role in regulating O_2 affinity of CerHb. In all the calculated MD trajectories the distances d between Thr–E11 OG1 and Tyr–B10 OH atoms are indeed within the hydrogen bonding range, for example, $d = 2.86 \pm 0.16$ Å in 1 ns standard MD trajectory, $d = 2.85 \pm 0.16$ Å in LES10 trajectory (the X-ray value of d is 2.59 Å [6]). Thus, this hydrogen bond helps to keep Tyr–B10 away from the O_2 binding site. It was observed that, in ThrE11Val variant the Tyr–B10 hydroxyl group moves towards heme plane and changes O_2 kinetics by a few orders of magnitude [6]. While MD studies of CerHb mutants are still under way in our laboratory, it is interesting to note that the presence of the oxygen ligand in the heme cavity affects

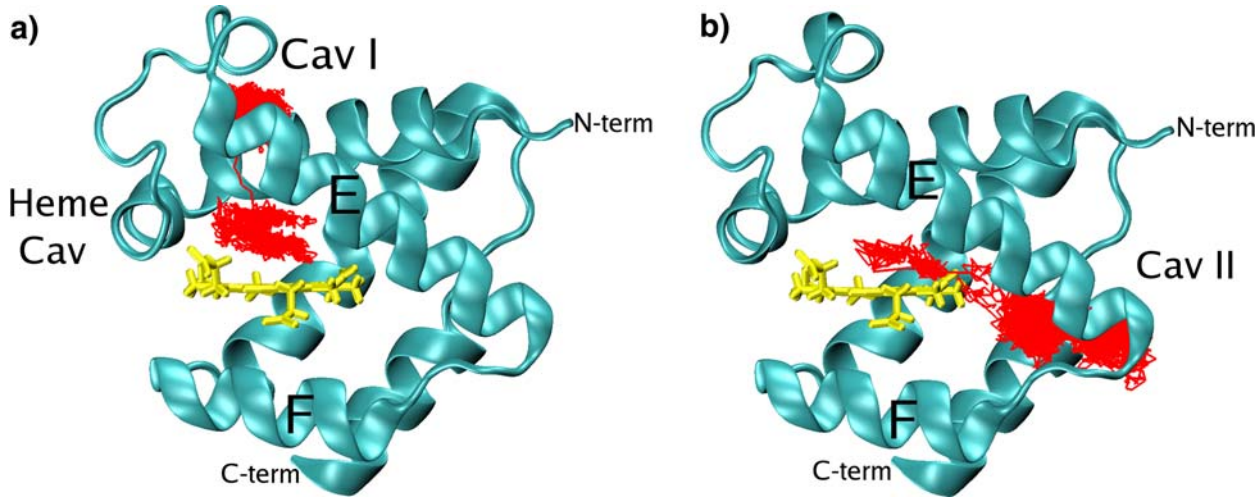

Fig. 3 **a** Result from simulation without LES. The heme cavity an the CavI pocket are shown. **b** Result from LES simulation with five copies of the ligand. The CavII is shown

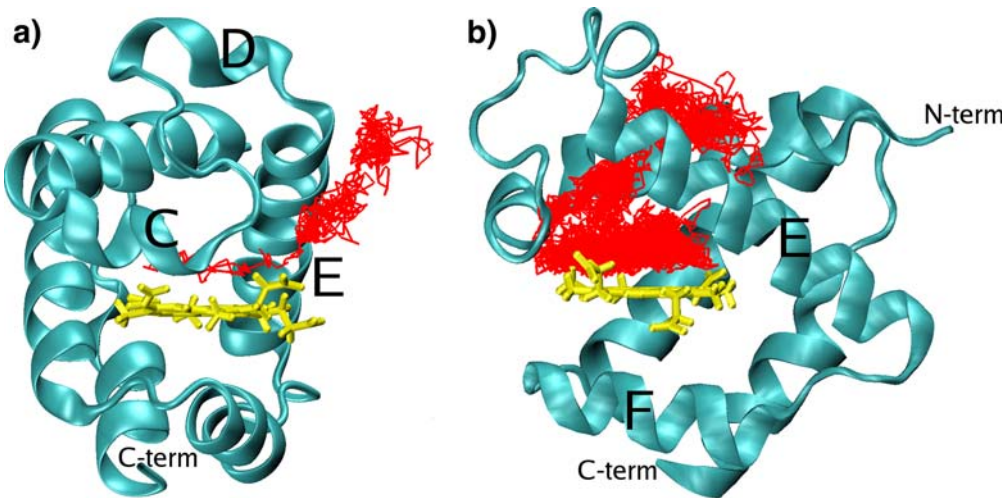

Fig. 4 **a** Routes T1. **b** Routes T2. Description in text

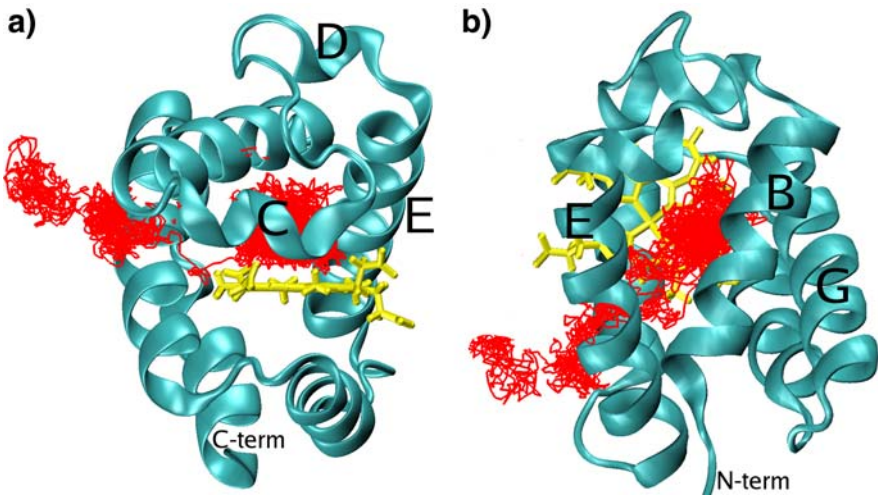

Fig. 5 **a** Routes T3. **b** Routes T4. Description in text

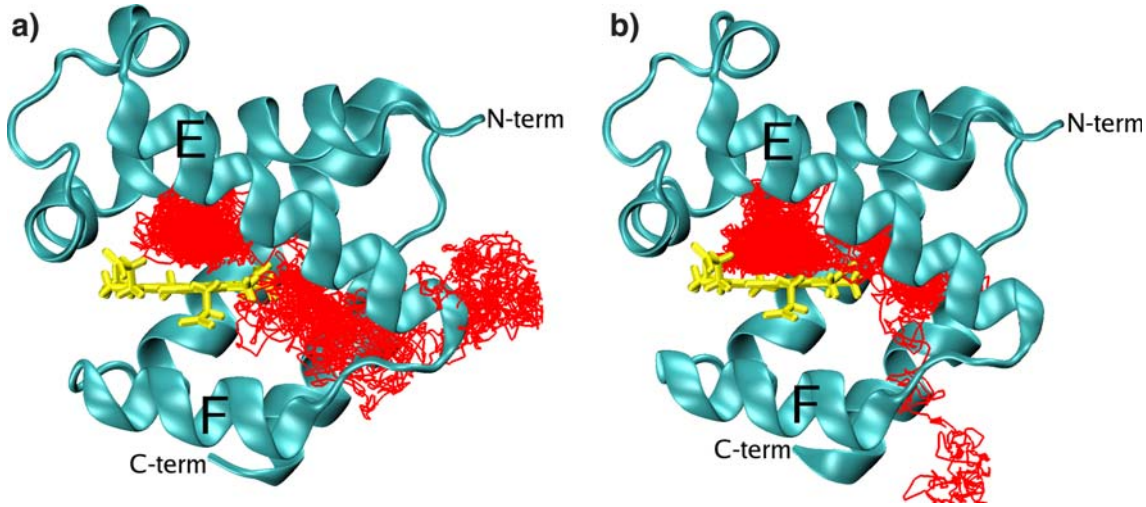

Fig. 6 **a** Routes T5. **b** Routes T6. Description in text

the Fe – Tyr–B10 OH distance in the native CerHb. In the standard (i.e., one O_2 copy) 1 ns MD trajectory, this distance decreases from 6.23 ± 0.34 to 5.53 ± 0.30 Å upon O_2 leaving the heme pocket. This effect indicates a strong Tyr–B10 ligand interactions.

4 Conclusions

In CerHb a very long (>10 ns) simulation will be necessary to obtain a classical diffusion path for a single dioxygen ligand at ambient temperature. The main barriers on crossing from the heme pocket to the next stages of ligand transport through the large CerHb channel are Phe10 and Thr48. The LES method helps to find out possible diffusion paths. Our data (10 and 15 LES ligands) suggest a possibility of multiple diffusion routes in CerHb, as at least distinct six paths were identified. Only two paths were located within the channel, the other paths were variants of a "classical" direct escape of ligands from the heme pocket to solvent. In our opinion MD simulations give a better insight into the quite dynamic phenomenon of ligand diffusion in biopolymer interior than even careful, but static, inspection of the rigid protein X-ray structure.

Acknowledgements UMK grant (2005 444-F, WN) and President of Poland grant "Superpracownia 2002" supported this research. Calculations were performed in part on UCI-UMK cluster.

Appendix

The LES method developed by Elber and Karplus [7,13] is based on the assumption that the phase-space density can be written as a product:

$$\rho(\vec{P}, \vec{Q}, t) = \rho_s(\vec{P}_s, \vec{Q}_S, t)\rho_B(\vec{P}_B, \vec{Q}_B, t),$$

where ρ_S is the density of the ligands (a small subsystem) and ρ_B is the density of the protein (a big subsystem). $\mathbf{R} = (\mathbf{P}, \mathbf{Q})$ is the location of system in the phase space of momenta P and space coordinates Q. It is also assumed that the density of the protein can be written as:

$$\rho_B(\vec{P}_B, \vec{Q}_B, t) = \delta(\vec{P}_B - \vec{P}_{0B}(t), \vec{Q}_B - \vec{Q}_{0B}(t)),$$

and the copied subsystem's density can be written as:

$$\rho_s(\vec{P}_s, \vec{Q}_s, t) = \sum_{k=1}^{N} w_k \delta(\vec{P}_s - \vec{P}_{0s}(t), \vec{Q}_s - \vec{Q}_{0s}(t)),$$

where w is a weight function, and a complete set of Dirac delta functions is used for the expansion of densities. In the LES weights, w is assumed to be $1/N$, where N is number of copies. From these assumptions we obtain the equations describing motions of the bath and the copied system, which are equivalent to the Newton equations of motion [7]:

$$\frac{\partial \vec{Q}_{0Sk}}{\partial t} = \frac{\partial H(\vec{P}_{0Sk}, \vec{Q}_{0Sk}, \vec{P}_{0B}, \vec{Q}_{0B})}{\partial \vec{P}_{0Sk}}$$

$$\frac{\partial \vec{P}_{0Sk}}{\partial t} = -\frac{\partial H(\vec{P}_{0Sk}, \vec{Q}_{0Sk}, \vec{P}_{0B}, \vec{Q}_{0B})}{\partial \vec{Q}_{0Sk}}$$

$$\frac{\partial \vec{Q}_{0B}}{\partial t} = \sum_{k=1}^{N} w_k \frac{\partial H(\vec{P}_{0Sk}, \vec{Q}_{0Sk}, \vec{P}_{0B}, \vec{Q}_{0B})}{\partial \vec{P}_{0B}}$$

$$\frac{\partial \vec{P}_{0B}}{\partial t} = -\sum_{k=1}^{N} w_k \frac{\partial H(\vec{P}_{0Sk}, \vec{Q}_{0Sk}, \vec{P}_{0B}, \vec{Q}_{0B})}{\partial \vec{Q}_{0B}}$$

References

1. Kendrew JC, Dickerson RE, Strandberg BE, Hart RG, Davies DR, Philips DC, Shore VC (1960) Nature 185:422
2. Perutz MF (1979) Annu Rev Biochem 48:327
3. Burmester T, Weich B, Reinhardt S, Hankeln T (2000) Nature 407:520

4. Burmester T, Ebner B, Weich B, Hankeln T (2002) Mol Biol Evol 19:416
5. Pesce A, Nardini M, Dewilde S, Geuens E, Yamauchi K, Ascenzi P, Riggs AF, Moens L, Bolognesi M (2002) Structure 10:725
6. Pesce A, Nardini M, Ascenzi P, Geuens E, Dewilde S, Moens L, Bolognesi M, Riggs AF, Hale A, Deng P, Nienhaus UG, Olson JS, Nienhaus K (2004) J Biol Chem 279:33662
7. Elber R, Karplus M (1990) J Am Chem Soc 112:9161
8. Vitkup D, Petsko GA, Karplus M (1997) Nat Struct Biol 4:202
9. Bossa C, Anselmi M, Roccatno D, Amadei A, Vallone B, Brunori M, Di Nola A (2004) Biophys J 86:3855
10. Nutt DR, Meuwly M (2004) Proc Natl Acad Sci USA 101:5998
11. Hummer G, Schotte F, Anfirud A (2004) Proc Natl Acad Sci USA 101:15330
12. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) Nucleic Acids Res 28:235
13. Verkhivker G, Elber R, Nowak W (1992) J Chem Phys 97:7838
14. Kalr L, Skeel RT, Bhandarkar M, Brunner R, Gursoy A, Krawetz N, Phillips J, Shinozaki A, Varadarajan K, Schulten K (1999) J Comp Phys 151:283
15. Humphrey W, Dalke A, Schulten K, (1996) J Mol Graphics 14.1:33
16. Roitberg A, Elber R (1991) J Chem Phys 95:9277
17. Teeter M (2004) Protein Sci 13:313
18. Scott EE, Gibson QH, Olson JS (2001) J Biol Chem 276:5177
19. Bossa C, Amadei A, Daidone I, Anselmi M, Vallone B, Brunori M, Di Nola A (2005) Biophys J 89:465