Biochimica et Biophysica Acta 1834 (2013) 1739–1743

Contents lists available at SciVerse ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbapap

Single vs. multiple ligand pathways in globins ☆ A computational view

M. Fátima Lucas [a], Víctor Guallar [a,b,*]

[a] Joint BSC-IRB Research Program in Computational Biology, Barcelona Supercomputing Center, Jordi Girona 29, 08034 Barcelona, Spain
[b] Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys 23, 08010 Barcelona, Spain

ARTICLE INFO

Article history:
Received 16 December 2012
Received in revised form 25 January 2013
Accepted 28 January 2013
Available online 4 February 2013

Keywords:
Globin
Ligand migration
PELE
Myoglobin
Cerebratulus lacteus

ABSTRACT

Diatomic ligand migration in globins has been the subject of numerous studies. Still, a consensus picture for the ligand entrance is not clear, with a growing concern among experimental researchers that computational simulations always show multiple pathways for any globin. Modeling non-biased ligand entrance from conventional molecular dynamics techniques, however, has shown to be difficult (and expensive). Here we use our Monte Carlo methodology, capable of freely mapping ligand diffusion and the description of rare events, to two well-studied systems: myoglobin and the mini-hemoglobin from the sea worm *Cerebratulus lacteus*. Our results clearly show that the simulations are specific to the system providing a different trend in the entrance pathway, as expected from experiments. While Mb presents multiple entrance pathways, populating the well-known xenon cavities, in CerHb the ligand enters the protein only by one apolar channel. Most of the trajectories (64%) visiting myoglobin's active site though, are gated by the distal histidine. Such detailed information, accessible through the state of the art algorithms in PELE, is computationally inexpensive and available to all non-profit researchers. This article is part of a Special Issue entitled: Oxygen Binding and Sensing Proteins.

1. Introduction

Due to its implication in oxygen binding, globins have been widely studied, hemoglobin (Hb) and myoglobin (Mb) being the best known family members. Mb, the first protein to have its three dimensional structure solved, presents a reduced size and has been taken as a model for other similar but more complex systems [1–3]. Despite its apparent simplicity, understanding how ligands reach the binding pocket is not a trivial task. Observation of Mb's structure led Perutz et al. to propose that O_2 molecules should enter/exit through a short pathway, gated by the distal E7 histidine [4]. This hypothetical pathway, requiring the histidine side chain rotation, was later supported by experiments [5–8] and is now regularly designated as distal path. Studies with xenon (Xe) gas showed the presence of four internal cavities: Xe1–Xe4 and put forward the discussion for alternative passages of small ligands in and out of the protein [9]. Digression of photo-dissociated ligands into these cavities has been confirmed both experimentally and computationally [6,10–13]. For example, Cohen et al. who have estimated the gas migration pathways through a free-energy perturbation approach show multiple low energy passages through the protein matrix [14]. Simulation methods, however, systematically concur with the assumption of diverse passages [9,15–22] while many experimentalists argue that just one main passage exists [6,23].

More recently, research studies have centered on a mini-hemoglobin from the sea worm *Cerebratulus lacteus* (CerHb) which stores oxygen in order to maintain neural activity under hypoxic conditions. Divergent to Mb, it appears to deliver oxygen through a passage other than the distal path [24–26]. In Fig. 1 the main differences between these two proteins are depicted. CerHb's structure has a large apolar channel starting in the vicinity of Ala55, between the E and H helices, and directed toward the distal pocket [24]. This tunnel is expected to be the result of an evolutive adaptation where myoglobin's equivalent N-terminal A helix has been removed (on the left side of Fig. 1 the missing helix of Mb is shown in a darker shade of green). Additionally, a glutamine residue in the active site replaces the distal histidine present in Mb (also depicted in Fig. 1). Finally, other structural differences between the two globins include the shortening of the CD loop and the reduction of the organized secondary structure in the F helix.

Fourier transform infrared-temperature derivative spectroscopy (FTIR-TDS), commonly used to map protein–ligand interactions in globins [27], has evidenced the possibility of several intermediate docking sites along the tunnel in CerHb [28]. The primary docking site B has been assigned above the C pyrrole ring and beneath Phe10 and Tyr11. Two secondary docking sites, C and D, were assigned to a cavity on the pathway for ligand migration to the apolar channel and to a position between Tyr11 and Gln44, respectively. Simple theoretical approaches, however, identified several different passages opposing the experimental evidence of a main

☆ This article is part of a Special Issue entitled: Oxygen Binding and Sensing Proteins.

* Corresponding author at: Life Science Program, Barcelona Supercomputing Center, Jordi Girona, 29, Barcelona 08028, Spain. Tel.: +34 934137727; fax: +34 934137721.
E-mail address: victor.guallar@bsc.es (V. Guallar).

1570-9639/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.bbapap.2013.01.035

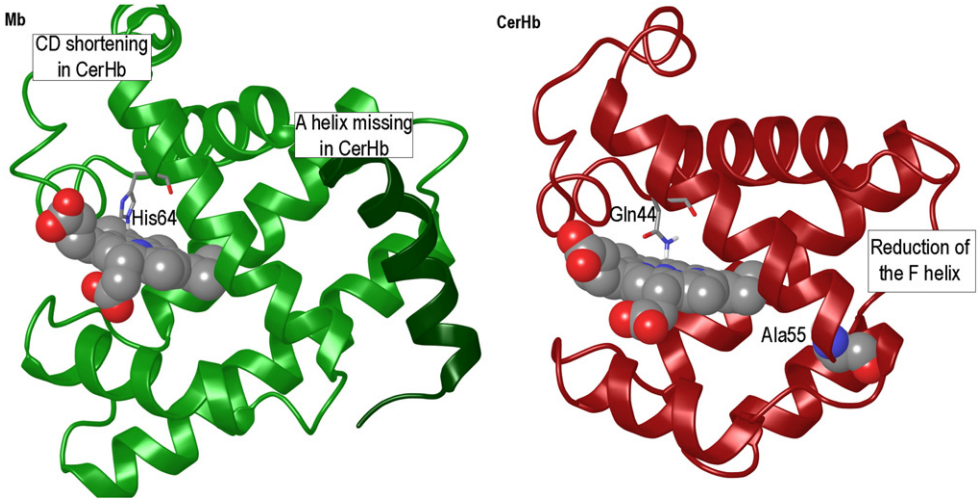

Fig. 1. Crystallographic structures for MB (left panel) and CerHb (right panel).

passage by Ala55 [29]. Biased molecular dynamics simulations have also been performed, showing that the apolar channel is a low energy lying pathway [30].

Despite all the work done, we are still missing a detailed consensus mechanism for the ligand entrance in globins. Molecular dynamic (MD) simulations may provide such atomic detailed information, although it requires several microsecond simulations. State of the art MD simulations, using specialized purpose machines or accelerated hardware (graphic processing units, for example), are capable of producing trajectories in the microsecond timescale. In any case, producing 50–100 microsecond MD trajectories, which would be necessary to observe few entrance events, is computationally prohibitive or accessible only to few research groups. These limitations have reduced most of the studies to escape pathways and steered molecular dynamics. Time saving alternatives such as using high ligand concentrations might be used, rising the probability of observing the rare event of ligands' entrance. Ruscio et al., for example, used this approach in Mb, obtaining several entrance and exit passages [20]. These and other results using approximate methods [29], however, have extended the concern among experimental researchers that computational simulations would always show multiple pathways (regardless of the system).

We have recently developed a Monte Carlo ligand dynamics code, PELE (Protein Energy Landscape Exploration), which is a powerful alternative to MD. PELE has been shown to be capable of describing ligand escape routes of globins such as: truncated hemoglobin, human hemoglobin and myoglobin [21,31,32]. Furthermore, PELE reproduces microsecond molecular dynamics at a fraction of its computational cost. Here, we set to explore the entrance gates of molecular oxygen in Mb and CerHb. For each system we have produced 95 unbiased simulations where the ligand is initially placed in random positions near the protein's surface. PELE's Monte Carlo random walk diffuses the ligand freely, with no restrains, and in about 10–20% of cases the ligand enters the protein matrix. Our results show that the simulations are specific to the system providing a clear different trend in the entrance pathway, as expected from experiments [25,26,30]. Mb presents multiple entrance pathways. Most of the trajectories (64%) visiting the active site, however, are gated by the distal histidine. For CerHb the ligand enters the protein by the apolar channel only, constituting the first comprehensive simulation of ligand entrance in CerHb. Such detailed information, accessible through the state of the art algorithms in PELE, is computationally inexpensive and available to all non-profit researchers through our server (www.pele.bsc.es).

2. Methods

2.1. Mb and CerHb setup

Initial deoxy models for CerHb and Mb were obtained from the pdb structures 1KR7 [24] and 1A6G [33], respectively. A hybrid quantum mechanical/molecular mechanical deoxy heme template was added following previous studies [32] and hydrogen atoms were added with Schrodinger's Protein Preparation [34]. This utility optimizes the hydrogen bond network by sampling side chain dihedral angles in hydroxyl and amine groups, together with water molecules and histidine protonation state. In the case of the (gate) distal histidine 64, in myoglobin, it has been modeled in the epsilon protonation state.

2.2. PELE

Long timescale protein dynamics are computationally feasible with PELE thanks to a successful combination of Monte Carlo techniques with protein structure prediction methods [21,35]. The methodology has been applied to ligand dynamics in systems such as cytochrome P450 and myoglobin [21], truncated hemoglobin [31], human hemoglobin [32] and aryl-alcohol oxidase [36], showing to provide reliable results. The program uses an OPLS (optimized potentials for liquid simulations) all-atom force field [37] with an improved implicit surface-generalized Born (SGB) continuum solvent model [38]. Three main steps define each Monte Carlo cycle: a protein backbone and ligand perturbation, specific side-chain sampling and a final minimization [39,40]. In the case of ligands the perturbations refer to a random rotation and translation of the center of mass of the ligand where the limits are user defined. In the case of the protein the perturbations are based on the displacement of α-carbon according to an anisotropic network model (ANM) [41]. The side-chain sampling step involves arranging all side chains present in a set of conditions: a) adjacent to the ligand within a predefined distance; b) side chains with the overall largest changes in energy identified in the ANM step; c) residues included by the user in the input file (for example x random side chains can be always included but a specific residue can also be added). The last stage involves the minimization of a region including, at least, all residues local to the atoms involved in the perturbation and side-chain steps. Finally, the new structure is accepted or rejected based on a Metropolis test for a given temperature.