Biochimica et Biophysica Acta xxx (2010) xxx–xxx

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

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

Review

Protein dynamics and ligand migration interplay as studied by computer simulation

Pau Arroyo-Mañez [a,b], Damián E. Bikiel [a], Leonardo Boechi [a], Luciana Capece [a], Santiago Di Lella [a], Darío A. Estrin [a,*], Marcelo A. Martí [a,c,*], Diego M. Moreno [a], Alejandro D. Nadra [c,d], Ariel A. Petruk [a,e]

[a] Departamento de Química Inorgánica, Analítica y Química-Física (INQUIMAE-CONICET), Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad de Buenos Aires, Argentina
[b] Departamento de Química Orgánica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad de Buenos Aires, Argentina
[c] Departamento de Química Biológica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad de Buenos Aires, Argentina
[d] Departamento de Fisiología, Biología Molecular y Celular, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad de Buenos Aires, Argentina
[e] Instituto Superior de Investigaciones Biológicas (INSIBIO-CONICET), Facultad de Bioquímica, Química y Farmacia, Universidad Nacional de Tucumán, San Miguel de Tucumán, Argentina

ARTICLE INFO

Article history:
Received 11 June 2010
Received in revised form 12 August 2010
Accepted 13 August 2010
Available online xxxx

Keywords:
Ligand migration
Protein dynamics
Molecular dynamics

ABSTRACT

Since proteins are dynamic systems in living organisms, the employment of methodologies contemplating this crucial characteristic results fundamental to allow revealing several aspects of their function. In this work, we present results obtained using classical mechanical atomistic simulation tools applied to understand the connection between protein dynamics and ligand migration. Firstly, we will present a review of the different sampling schemes used in the last years to obtain both ligand migration pathways and the thermodynamic information associated with the process. Secondly, we will focus on representative examples in which the schemes previously presented are employed, concerning the following: i) ligand migration, tunnels, and cavities in myoglobin and neuroglobin; ii) ligand migration in truncated hemoglobin members; iii) NO escape and conformational changes in nitrophorins; iv) ligand selectivity in catalase and hydrogenase; and v) larger ligand migration: the P450 and haloalkane dehalogenase cases.

1. Introduction

1.1. Protein dynamics and protein function

Proteins are intrinsic dynamic systems, and for that reason, this characteristic has been recognized to be of crucial importance in many molecular processes. The most successful theory that allows us to understand and describe protein dynamics is the protein energy landscape theory, which postulates that proteins do not exist in a unique conformation, but in a number of conformational organized hierarchically in the energy landscape, with valleys within valleys [1–7]. Each substrate has different structure and therefore different reactive properties. An allosteric change or a large conformational change correspond to changes from one big valley to another, while thermal fluctuations representing side chain and small backbone movements correspond to movements across smaller valleys [1–7].

One of the most important ways in which dynamics affects the protein function is by regulating the ligand migration process, either allowing faster/slower substrate access to the active site [1,3,8–10], promoting product release in enzymes [11], or by subtly regulating ligand affinity in a sensor or transporter [12]. Both large- and small-scale motions may affect the process in a different way. Larger motions are relevant in the case of allosteric proteins (i.e., hemoglobin) [13] or in the open/closed movement of the HIV flaps that allow substrate to enter the protein [6,14]. Small-scale faster thermal motions are essential for ligand migration inside the proteins, since a rigid structure would inhibit ligand migration due to sterical clashes. This intimate relationship between ligand migration and dynamics has promoted intense research in the field, to understand how protein dynamics affect ligand migration at the atomic/molecular level.

1.2. Ligand migration as the determinants of protein function

The protein acting as an enzyme, a transporter, and a sensor must precisely regulate its ligand (or substrate) affinity to fulfill its function. This is specially the case of metalloproteins which use small molecules (O_2, CO, NO,....) as ligands [15]. Experimentally, the ligand affinity is characterized by the equilibrium constant K_d, determined by the ratio between the association and dissociation kinetic constants (k_{on}/k_{off}). In metalloproteins, the ligand association kinetic constant k_{on} depends on two processes: ligand migration from the solvent to the protein active site and the ligand coordination to metal [15]. In the case of protein without coordinated ligand, only the migration step is operative. The ligand migration to the active site is determined by the presence of internal cavities or tunnels involved in the migration

* Corresponding authors. Departamento de Química Inorgánica, Analítica y Química-Física (INQUIMAE-CONICET), Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad de Buenos Aires, Argentina. Tel.: +54 11 45763378x105; fax: +54 11 45763348.
E-mail addresses: dario@qi.fcen.uba.ar (D.A. Estrin), marcelo@qi.fcen.uba.ar (M.A. Martí).

1570-9639/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.bbapap.2010.08.005

process [16–20] and the presence of specific residues acting as "gates" promoting or blocking the entry process [16]. All these phenomena are intrinsically dynamic and are determined and controlled by the protein motions. The second step (coordination) is determined by the intrinsic reactivity between the ligand and the metal and it is not usually subject of dynamic control [15].

Typically, the migration process is the key issue in determining the overall affinity, especially when comparing the same ligand and active site metal. Values of association rate constants span a range of many orders of magnitude, starting from $10^3\ M^{-1}\ s^{-1}$ in proteins where the entry is blocked to values of $10^9\ M^{-1}\ s^{-1}$ when the association rate is mainly controlled by the diffusion, as observed for isolated prosthetic groups such as porphyrins [15,16,21].

The dissociation rate (k_{off}) is also determined by two processes, thermal breaking of the protein–ligand interactions, and ligand escape from the active site to the bulk of the solvent. Both processes depending on the protein and ligand may be rate-limiting and are strongly controlled by protein dynamics [12,15,16,22]. The escape process is similar to the ligand entry described above and depends on the presence of tunnels and gates. On the other hand, in some cases, protein–ligand interactions may determine the presence of more than one conformation significantly populated [23,24].

1.3. Experimental and theoretical methods for the study of ligand migration

Determining the ligand association and dissociation constants for wild type (wt) and mutant proteins, combined with available structural data for the protein, can yield useful insights into the role of the protein structure and dynamics, and the presence of cavities, tunnels, and gates in relation with the ligand migration process [16,18–20]. However, understanding microscopic details and evaluating the dynamics of the system require a methodology capable of probing molecular structure with the ligands inside the protein. Several experimental techniques such as crystallization under high Xe pressure [18–20,25] and time-resolved X-ray crystallography [26,27] allow the determination of ligand migration paths. On the other hand, computer simulation techniques provide an effective tool for the investigation of the ligand migration mechanism within the protein matrix because these can provide a combined atomistic detailed structural and thermodynamic picture of the phenomena that include protein dynamics explicitly [9,12,22,28–33]. However, obtaining a complete picture of the ligand migration process by means of standard molecular dynamics (MD) involves extremely long simulations with an associated high computational cost. To overcome this limitation, several methods to enhance sampling at an affordable computational cost were developed in the last years [9,11,12,31–39].

In this work, we will review the different computational methods developed to study the ligand migration process, paying particular attention into the protein dynamics role. In the second part, we will briefly describe selected examples in which these methods have been successfully applied to the ligand migration process study.

2. Computational methods

2.1. Molecular dynamics and ligand migration

Atomistic simulations of complex systems are a common task of today research. However, while it is feasible to perform molecular dynamics simulations along a time scale of about hundreds of nanoseconds, most of the interesting phenomena usually occur within time scales orders of magnitude beyond usual MD simulation times. Therefore, they cannot be observed spontaneously in a typical MD simulation. To overcome this difficulty and to obtain precise estimates of thermodynamic properties of the process (i.e., the associated free energy profiles), several techniques and algorithms of enhanced sampling were developed [9,11,12,31–39]. In the next section, we will describe the basic ideas of the most widely used techniques to explore migration of ligands in proteins.

The first group of methods treats explicitly the interaction between ligand and protein and can be classified as explicit ligand sampling (ELS) methods, allowing for the elucidation of ligand migration pathways. In some cases, but upon the payment of a considerable computational cost, free energy profiles associated with the migration process can also be obtained. The second group of methods is called implicit ligand sampling and is based on performing the molecular dynamics simulation of the protein without the ligand and treating its presence as a weak perturbation added afterwards.

Both ELS and ILS methods typically employ mean field force fields, and so neglect effects related to charge polarization, which in some cases may be significant.

2.2. Multiple steered molecular dynamics (MSMD)

In the MSMD method, the original potential force field is modified by the addition of a new time-dependent term $V_{add}(t)$, which is set to drive the system through an arbitrary reaction path coordinate. The idea is to force the system to visit energetically less-probable configurations, similar to the well-known "umbrella sampling method" [40] but using a time-dependent restrain (Eq. (1)), committing the system to explore the desired phenomenon.

$$V_{add}(t) = \left(\tfrac{1}{2}\right)k[x - x_0(t)]^2 \tag{1}$$

In Eq. (1), k is an arbitrary constant, x is the actual reaction coordinate (RC) value of the system, and $x_0(t)$ is the time-dependent restrain expressed as a moving equilibrium value of the RC, as described by Eq. (2),

$$x_0 = x_i + vt \tag{2}$$

where t is the time in the MD simulation and v is the velocity at which the harmonic restrain potential equilibrium value is guided. The selection of this velocity results critical when performing MSMD. When treating small ligand migration through internal tunnels of proteins, the usual chosen RC is the distance between the ligand and the active site, and the external force given by the new modified potential, pushes (or pulls) the ligand to enter (or escape) from the bulk of the solvent into the protein active site (or from the active site to the bulk solvent) [9,11,22,32,33].

Interestingly, the external (nonequilibrium) work (W) necessary to move the system (i.e., the ligand) along the chosen RC can be computed by integrating over the external force. Starting from different initial configurations, we can compute many different nonequilibrium works, for the same reaction path. Using the Jarzynski's equality (Eq. (3)) [41], which relates these irreversible works with free energy changes (ΔG), it is possible to obtain the free energy change of the process under study.

$$e^{\frac{-\Delta G}{k_B T}} = \langle e^{\frac{-W}{k_B T}} \rangle \tag{3}$$

In Eq. (3), k_B is the Boltzman constant, T is the temperature, and W is the calculated work for each independent nonequilibrium process. MSMD refers to the idea that multiple MD simulations of the process are needed in order to obtain accurate free energy profiles. Despite the considerable computational cost, this methodology provides accurate free energy profiles for small ligand migration across protein tunnels, as will be shown in the examples section [9,11,22,32,33]. Noteworthy, each performed SMD simulation represents a possible "real" trajectory of the system. Moreover, the closer the computed W to the final free energy profile, the more likely that this trajectory actually happens.

Please cite this article as: P. Arroyo-Mañez, et al., Protein dynamics and ligand migration interplay as studied by computer simulation, Biochim. Biophys. Acta (2010), doi:10.1016/j.bbapap.2010.08.005

2.3. Metadynamics

Similarly to the MSMD approach described above, metadynamics also allows the determination of ligand migration trajectories and the associated free energy profile. The metadynamics technique requires a preliminary identification of a set of collective variables (or reaction coordinates) which are assumed to be able to describe the process of interest. If such process is ligand migration, the center of mass of the ligand and its relative position to the protein is a wise first choice. The method has the ability to reconstruct the free energy profile associated with the process and accelerate the occurrence of rare events. To accomplish this, the dynamics along the chosen reaction coordinate is enhanced by a time-dependent potential that modifies the system energy along it. This potential is constructed as a sum of three-dimensional Gaussian functions (GF) centered at different points along the reaction coordinate, which are different points to iteratively reconstruct an estimation of the free energy [35,39,42].

As a descriptive example, imagine a ligand freely moving inside the protein matrix. In a standard MD simulation, the position of its center of mass will explore accessible regions (free energy valleys), and after some time, it will define an equilibrium position with some dispersion. However, other stable space regions (or energy valleys) will not be accessible to the ligand possibly due to the presence of a high-energy barrier separating them. In metadynamics, positive energy GF potentials are added to the system, centered at the starting equilibrium position of the ligand. This bias forces the ligand to move from its well towards a new one. Now, the dynamics of the system in the modified surface will define a new equilibrium position, which is subsequently modified by adding additional GF, forcing a new exploration and so on. Once the ligand explores with the same probability all the desired reaction coordinate range (i.e., there are no more energy wells), by keeping track of all the added GF, the corresponding free energy profile can be reconstructed as the sum of the introduced GFs. This approach can be refined using both repulsive and attractive functions to assure the equilibrium sampling [35,39,42]. Compared with MSMD the computational cost of metadynamics is similar, and also the accuracy of the obtained free energy profile which depends, if statistically converged, on the force field.

2.4. Protein energy landscape exploration (PELE)

An interesting methodology, which is costly but extremely useful for studying larger ligands and/or conformational changes due to the migration process, is the PELE method [43]. This approach maps the energy landscape for large conformational rearrangements in proteins, combining all-atom energy force field functions with specialized sampling algorithms for side chain and loop prediction [44]. The methodology comprises several steps, including Monte Carlo moves, rotamer library side chain optimizations, truncated Newton minimizations, and Metropolis acceptance tests. The method generates and propagates changes in a system by means of a series of structurally similar local minima, which are combined into a trajectory.

The sampling procedure is initiated by the generation of a local perturbation on the ligand. If the ligand can be treated as a rigid body, only the rotational and translational degrees of freedom are required. The next step filters if there exist steric contacts between the ligand and the protein backbone. If a contact is found, the perturbation is rejected. A set of perturbations are generated, and the best one, based in a score function, is selected. Once the perturbation is selected, the algorithm proceeds to optimally arrange all the side chains perturbed, using a rotamer library. The rotamer algorithm includes a steric filtering and a rotamer clustering used to decrease the number of rotamers to be minimized. The last step is the minimization of the region including, at least, all residues close to the atoms involved in the two previous steps. The objective of the minimization is to generate the backbone response to the initial local perturbation. The method is based on the assumption that the side chains act as sensors responding to ligand motion. These three steps described are defined as a "move," which is accepted or rejected via a Metropolis criterion. The conformational changes are propagated by means of the diffusion of the local perturbation. The direction of the perturbation is kept constant for a set number of steps making up a steering cycle. The translation is described by a translational vector, a variable describing the translational range and a random number. After the defined set of steering steps, the direction can be randomly updated using a Metropolis criterion and a user-defined overall displacement. The result of this procedure is a series of local minima which are structurally highly correlated. However, large structural changes can be observed during a trajectory. Atomic-scale details can be observed, and possible mechanisms can be explored. One of the advantages of this method is that by modifying the steering and acceptance parameters, different tasks can be done such as equilibration procedures and exploration runs [43].

The results obtained by this methodology comprise both pathways for ligand migration, with the associated changes in protein structure, and the potential energy profile including the barriers. The method, however, does not include any entropic contribution, so the free energy associated with the process is not obtained.

2.5. Accelerated molecular dynamics (AMD)

Another method which allows faster sampling, and the determination of the potential energy associated to the ligand migration path is called accelerated molecular dynamics (AMD). In this method, a bias potential $\Delta V(\mathrm{r})$ is included allowing the system to overcome the barriers and thus accelerating the sampling [45,46]. The real potential $V(\mathrm{r})$ is only modified near the minima, but remains unaffected in the barriers zones. The modified potential $V^*(\mathrm{r})$ depends on an arbitrary value E, which determines when the bias potential will be applied. If the chosen E value is lower than the global minimum of the potential energy function, the simulation will be not accelerated at any point. However, if the E value is higher than the barriers, the potential will becomes flat in all the profile, and the shape of the potential energy surface will be lost.

There are different ways to define the bias potential $\Delta V(\mathrm{r})$. Rahman et al. [47] and Hamelberg et al. [48] introduced an accurate way to define $\Delta V(\mathrm{r})$, which is derivable at all points along the potential energy function, and it is able to maintain the shape of the potential surface. As in metadynamics, it is possible to reconstruct the real potential energy surface along the migration path using the bias at every point of the accelerated simulation.

2.6. Activated molecular dynamics (ActMD)

In the activated molecular dynamics method (ActMD), the protein and the ligand present different temperatures. The idea is to increase the kinetic energy of the ligand, allowing the diffusion along the pathway [49]. Two independent Berendsen's thermostats are employed for the protein and for the activated ligand. This interesting scheme was used by Kalko et al. [49] in a water diffusion study in catalase. The temperatures used were 300 K for the protein and 10,000 K for internal water molecules. The results revealed the importance of an accurate selection of a suitable relaxation time for the activated ligand to maintain the high temperature along the simulation. Coordinates and velocities of a standard MD snapshot are taken as starting point. Once the ligand is chosen (if there exists more than one, as in the case of water molecules), its velocity is scaled to 10,000 K and its direction is randomly selected. After 5–10 ps of constant temperature (with both thermostats), the ligand can be reselected and its velocity reassigned. The steps are repeated for a user-defined number of times. By means of this procedure, complete diffusion of the ligands can be obtained in a nanosecond time scale, giving the potential pathways for ligand migration from the active site towards the solvent [49].

Please cite this article as: P. Arroyo-Mañez, et al., Protein dynamics and ligand migration interplay as studied by computer simulation, Biochim. Biophys. Acta (2010), doi:10.1016/j.bbapap.2010.08.005

2.7. Random activated molecular dynamics (RAMD)

In the RAMD method, an additional force is applied to the center of mass of the ligand in a randomly chosen direction. After an arbitrary number of steps, the distance traveled by the ligand is compared to a threshold parameter. If the ligand does not reach the threshold distance, a new randomly chosen direction is given; otherwise, the force direction is maintained. This approach allows to observe the potential pathways for the ligand escape, without any thermodynamic information [50].

2.8. Locally enhanced sampling (LES)

The methods described above are based in the use of classical MD with the addition of a biasing potential or force that drives the ligand migration process. Another possibility to increase sampling efficiency is to introduce multiple copies of the ligand [51,52]. In this method called ligand-enhanced sampling, the system is divided in two subsystems (i.e., the protein and the ligand). The point is to simulate hundreds of copies of one of the subsystems (i.e., the ligand) at the same time with one single copy of the other subsystem (i.e., the protein). Each copy of ligand does not interact with the other copies and as a consequence of that, they can occupy the same space at the same time. The protein "feels" the ligand as an average of the mean force generated by these copies. This approach increases the sampling and reduces the transition barriers, allowing the system to explore different regions of the protein at lower computational cost. The method, however, does not allow to obtain thermodynamic information.

2.9. Implicit ligand sampling

The ILS approach is based on the potential of mean force (PMF) calculation, corresponding to the placement of a small ligand inside the whole protein matrix [37,38,53,54]. The corresponding PMF is associated to the free energy cost of incorporating the ligand at each particular position and therefore describes the accessible regions of the protein. The methodology relies on the fact that small ligands are small and interact weakly with the protein matrix and therefore the interaction can be computed for protein structures obtained without the presence of the ligand. By performing the molecular dynamics simulation of the protein without the ligand and treating its presence as a weak perturbation added afterwards, sampling of the ligand can be performed on the whole protein matrix simultaneously, from just one sufficiently long plain MD simulation of about 50 ns (unpublished results).

From a computational point of view, the averages are obtained from a finite set of snapshots extracted from molecular dynamic simulations. If M different states of the protein were used, the PMF is modified to:

$$W(r) = -k_B T \ln \sum_{m=1}^{M} \sum_{k=1}^{C} \frac{e^{-\beta \Delta E(q_m, r, \Omega_k)}}{MC} \quad (4)$$

where C different orientations for the ligand were taken for each protein state.

In practice, the computation of the PMF is carried out over a regularly spaced grid, in which several orientations for the ligand are computed. The ligand interaction energy $\Delta E(q,r)$ is usually computed using a Lennard–Jones potential. After performing a MD simulation of the protein system, the PMF is computed for each snapshot in the trajectory. For a more detailed derivation, see the works by Schulten et al. [37,38,53,54].

2.10. Another ILS method: the pathfinder algorithm

Another ILS method based on the use of a previously obtained MD trajectory is the pathfinder algorithm developed by Ruscio et al. [36]. The method is based on the search of cavities inside the protein matrix which are capable to contain the ligand. The input needed to start the algorithm is an MD trajectory, with a large number of statistically representative snapshots of the protein. The algorithm treats both the protein atoms and the solvent molecules as hard spheres. The first step of the algorithm is to find the points in the space in which the ligand does not intersect any atom of the protein. This is done by placing the ligand in each space region and looking for overlap between protein and probe atoms treated as hard spheres. Once all the points that do not intersect the protein are found, an internal cavity is defined as a bounded, maximally connected subset of the corresponding points. The algorithm is based on a grid computation in order to define the internal points where the probe is placed. If the procedure is repeated for each of the n snapshots, an integration of the cavities over the time can be performed. By marking which points of the grid have been found to belong to a particular cavity in at least k snapshots, the point is kept. Then, after the analysis of the n snapshots, the structure of the time-integrated cavities is obtained. The method allows identifying the protein cavities and tunnels but does not allow obtaining proper free energy profiles [36].

2.11. Another ILS method: the GRID–MD method

GRID–MD is a recently developed method based on precomputed grid of forces of proteins [34]. Using a conformational ensemble previously calculated, it can detect feasible channels with a very low computational cost. Approximately, trajectories of 10 ns are sufficient to sample the equilibrium conformations of proteins. By means of classical molecular interaction potentials (CMIP), the van der Waals and Poisson–Boltzmann electrostatic potentials are calculated to generate a regularly spaced grid points (the usual value is 0.5 Å). After the generation of the grid, a probe (usually corresponding to a carbon atom) is placed in the active site of the protein and its velocity is randomly selected. The individual GRID–MD trajectories are simulated during 0.1 to 1 μs, depending on the time that needs the probe to reach the exterior of the protein. Several independent simulations are needed in order to achieve convergence (ca., 10^3).

The interaction between a protein and a polyatomic rigid probe is classically determined by the addition of van der Waals and Coulombic contributions and the force experimented by the probe due to the protein can be easily computed using Newton's second law. If several proteins' conformations are used, each snapshot contribution to the energy/force are Boltzmann averaged. After the derivation of the averages forces, Brownian dynamics are used to explore the motion of the probe inside the protein matrix.

The method is very efficient due to the fact of the precomputed usage of data. On the other hand, the implementation of the Boltzmann average of different conformations, guarantees that the mean forces at each grid point is mainly determined by the best conformation that accepts the probe at that point. The method is then able to detect rare events. The limitation of the technique is obviously the precomputed data: the only conformations tested are the microstates visited during the original trajectory and no proper thermodynamic quantity is obtained [34].

2.12. Summary of the presented methods

The methods presented in this section are summarized and classified in Table 1. For this purpose, the different methods have been classified according to their computational cost (high, middle, low), the kind of information provided (ligand migration path (LMP), potential energy profile (PEP) , free energy profile (FEP)), their applicability depending on the size of the ligand and their approximations (besides the approximations inherent to the utilization of a classical force field).

Please cite this article as: P. Arroyo-Mañez, et al., Protein dynamics and ligand migration interplay as studied by computer simulation, Biochim. Biophys. Acta (2010), doi:10.1016/j.bbapap.2010.08.005

Table 1
Comparison of the different methods used to study ligand migration in proteins.

Method	Computational cost	Results	Type of ligand	Approximations
MSMD	High	FEP, LMP	Small, might be applied to large	None
Metadynamics	High	FEP, LMP	Small, might be applied to large	None
PELE	Middle	PEP, LMP	Specially large	None
AMD	Middle	PEP, LMP	Small	None
ActMD	Middle	LMP	Small	None
RAMD	Middle	LMP	Small	None
LES	Middle	LMP	Small	Diffused protein ligand interaction
ILS	Low	FEP, LMP	Small	Neglect of ligand–protein interactions
Pathfinder	Low	LMP	Small	Neglect of ligand–protein interaction
Grid-MD	Low	LMP	Small	Diffused ligand–protein interaction

LMP, PEP, and FEP stand for ligand migration path, potential energy profile, and free energy profile.

Having described different methods used for study ligand migration in protein and its relation to protein dynamics, we will present now several examples.

3. Small ligand migration

3.1. Dynamics of protein cavities and ligand migration: the myoglobin and neuroglobin cases

Protein cavities were first described in myoglobin (Mb) [25,55,56] and were originally termed as "packing defects" underestimating their importance in protein function [57]. There are many works in which the role of these cavities in ligand migration has been studied both experimentally and theoretically [8,16,19,25–30,35–38,55,58,59]. We will focus on the later ones, using the widely studied Mb and the related neuroglobin (Ngb) as illustrative examples.

Mb is possibly the most studied protein by biophysical methods, specifically concerning the small ligand migration process [1,3,16,56,60]. Ngb is an ancient heme protein that belongs to the globin superfamily, present in the nervous system of vertebrates, with a particular high level of expression in highly oxygen-consuming tissues such as the retina. Since its discovery 10 years ago, it was the subject of study of many groups, although its function in vivo is still unclear [61–65]. Interestingly, its crystal structure presents an endogenous hexacoordinated heme, with HisE7 as the sixth ligand. In this context, for the small ligands to bind to the heme, HisE7 needs to dissociate from the iron, involving an important structural reorganization process. In addition, in the CD region, there are two cysteine residues (Cys46 and Cys55) that are able to form a disulfide bond, modifying the kinetic and thermodynamic properties of the ligand binding process [61–69]. Thus, a key element for determining Ngb function concerns understanding the process of ligand migration in the different coordination/redox states. In Mb, four small hydrophobic ligand-docking sites were described, named Xe1, Xe2, Xe3, and Xe4, as schematically shown in Fig. 1 [25]. Comparatively, in Ngb, the X-ray structure displays the striking presence of a huge cavity (around 300 $Å^3$), connecting and expanding the cavities found in Mb. Since the existence of this huge cavity is supposed to involve a very high energy cost, many efforts have been devoted to understand its role [69–73].

Theoretically, small ligand migration across Mb secondary docking sites has been studied using almost all the above described methods. The most relevant cases are the studies of Di Nola et al., in which a MD simulation of 90 ns of free CO inside Mb was performed [28,29], the works of Meuwly using MD to study rebinding properties [30,72,74], the works by Cohen et al. [37] using and presenting the ILS methodology and those by Ruscio et al. [36] combining extensive MD with the pathfinder algorithm. All the works showed that ligands move across the experimentally found sites, and that barriers that separate them are small.

In Ngb, although its relatively recent discovery, many works have been devoted to study protein dynamics and its effect on ligand migration has also been performed in Ngb. Given its resting state equilibrium between a hexacoordinated (6c) and pentacoordinated (5c) state, and since ligands can only bind to the 5c state, understanding how the equilibrium is regulated is the first key issue in determining ligand migration and binding to Ngb. To understand the role of protein dynamics in the 6c/5c equilibrium, including the effect of S–S bond formation and oxygenation, Nadra et al. [68] performed MD simulations of the process, showing that the CD loop dynamics is the key element that controls the equilibrium and therefore ligand accessibility. Furthermore, in a second work, using high-pressure conditions in the MD simulations, it has been shown that the loss of flexibility associated to this environmental condition disfavors the 5c state, forcing the protein to stay hexacoordinated [66].

Concerning the specific study of the small ligand migration process Lutz et al. [72], explored CO migration using MD, finding eight possible docking sites, three of which were also identified experimentally. The free energy barriers between the eight docking sites and their relative stabilization of the ligand in the site were calculated by means of umbrella sampling calculations. The results showed that the majority of the docking sites are separated by barriers sufficiently low to be overpassed fast at room temperature, denoting the fact that the ligand is able to explore the different cavities easily. By evaluating mutants and low-temperature simulations, they mapped the ligand migration network and proposed that Phe28 and Pro52 might be two key residues in regulating the migration process. PheB(10)28, was already identified as a gate in Mb, separating the primary docking site located at the distal cavity from the other docking sites also observed in Mb. Pro52, on the other hand, is involved in the transition between sites located further away from the active site. The authors also suggest that ligand migration is more accessible in Ngb compared to Mb [72].

The existence of multiple docking sites for small gaseous ligands is relevant for the proposed NO dioxygenase activity of globins in which two substrates are required to react sequentially within the heme, as will be further discussed below. To study docking preferences for different ligands, Orlowski and Nowak [73] used ILS for sampling CO, O_2, and NO migration in human Ngb using a short MD simulation of 5 ns.

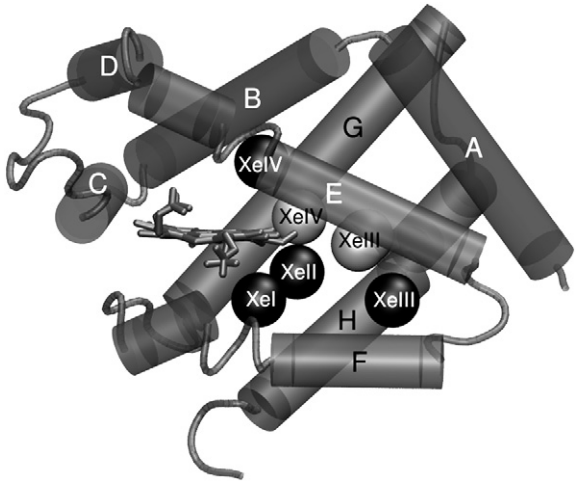

Fig. 1. Comparison of the Xe binding sites in Ngb (gray) and Mb (black). Cartoon representation of Ngb displaying Xe atom positions inside the cavity, represented as spheres. alpha-Helices are labeled according to the globin fold nomenclature and heme is depicted as sticks.

Please cite this article as: P. Arroyo-Mañez, et al., Protein dynamics and ligand migration interplay as studied by computer simulation, Biochim. Biophys. Acta (2010), doi:10.1016/j.bbapap.2010.08.005

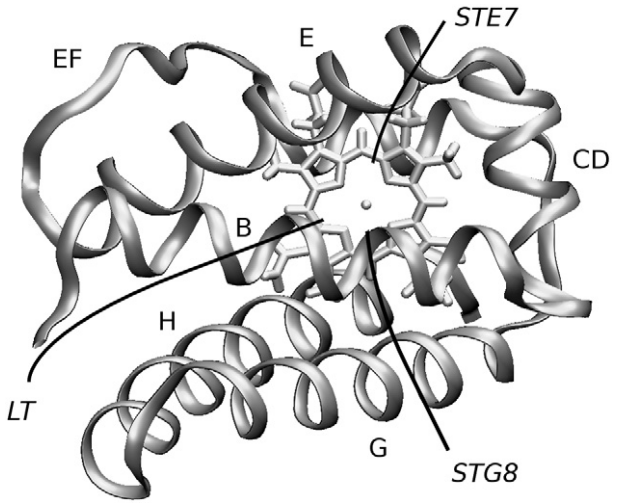

Fig. 2. Representation of possible ligand migration tunnels in a schematic model of a truncated hemoglobin: long (LT), short E7 (STE7), and short G8 (STG8).

Their results point out that there are certain regions of the protein that might be occupied by O_2 and NO, but not by CO. However, due to the intrinsic approximations of the ILS method and the short MD time, caution should be taken in interpreting the observed differences between these similar ligands. In another theoretical work, Bocahut et al. [71] went deeper in this kind of analysis, taking into account not only these three ligands (O_2, NO, and CO) but also Ngb conformational state induced by HisE7 coordination and the redox state affecting the disulfide bond in the CD region. They performed MD simulations combined with coarse-grain Brownian dynamics to explore the phase space in conjunction with metadynamics simulations to obtain free energy profiles for the migration of the different small ligands. This work shows how the reshaping of the cavities due to protein conformational freedom affects ligand migration. The authors concluded that the coordination of the heme iron as well as the redox state of the disulfide bond significantly affects the volume of the different cavities [71].

In summary, the theoretical and experimental work devoted to study the cavity network present in Mb and Ngb shows that these proteins display a complex network of docking sites separated by small free energy barriers which allow the different ligands to move from one to the other. In addition, in Ngb, the coordination state of the iron in the heme and the redox state of the disulfide bond present in the CD region produces important conformational changes that modify the distribution and shape of the different cavities [66,68].

3.2. Ligand migration through internal "tunnels" in the truncated hemoglobin family

In the last decade, a subfamily of globins called truncated hemoglobins (trHb) was discovered. Its members are widely distributed in bacteria, fungi, unicellular eukaryotes, and plants. Phylogenetic analysis allowed to classify the members of this subfamily into three groups named N, O, and P. Regarding its function, very different proposals have been made, suggesting they can work as NO detoxification proteins, O_2 transporters to peroxidases, among others [17,18,20]. One of the most striking structural feature of the trHb N family consists in the presence of a set of conserved tunnels, connecting the heme active site with the solvent (Fig. 2). Considerable effort has been made to identify the role of possible tunnels and cavities in the ligand migration process in truncated hemoglobins [9–11,17,18,20,32,33,52].

The most intensively studied case is the truncated hemoglobin N of *Mycobacterium tuberculosis* (Mt-trHbN), which is able to detoxify efficiently NO protecting the bacteria from the macrophage nitrosative attack. The detoxification reaction occurs in the active site of the protein, between the heme-bound O_2 and the incoming NO [75]. In this context, both ligands need to migrate through the complex net of tunnels and reach the active site. Mt-trHbN presents two tunnels for ligand migration, named long tunnel (LT) and short tunnel G8 (STG8) (Fig. 2) [9,10,18,20]. Experimental and theoretical evidence (using both MSMD and ILS method combined with long-range MD simulations) indicate that the first ligand (O_2), enters the active site mainly through the STG8, being the LT partially blocked by a phenylalanine residue in the E15 position (PheE15). Nevertheless, the coordination of O_2 to the iron triggers a conformational change which involves PheE15 side chain rotation and the associated opening the LT. This process allows the second ligand (NO) to migrate through the LT and encounter the active site. In this sense, PheE15 acts as a gate, modulating ligand migration through the internal LT [9,34,52,75,76].

Interestingly, it has been observed that in the O and P groups the presence of a conserved Trp residue (TrpG8) blocks the STG8 tunnel. Consequently, ligands must either use the LT to migrate or an alternative tunnel, located in the same region than the classical myoglobin E7 gate [16], which was named short tunnel E7 (STE7) (Fig. 2) [32]. In addition, it has been observed by means of MSMD calculations performed on different members of the O group that the residue located at the E11 position also plays a key role in the migration through the LT. For

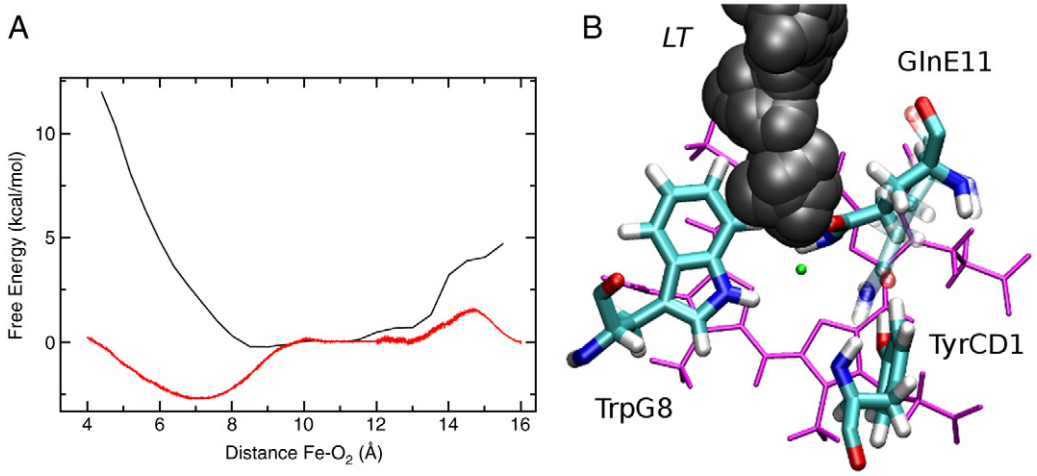

Fig. 3. (A) Free energy profile of ligand migration through LT in Mt-trHbO (black) and Bs-trHbO (red). (B) View of the active site of truncated hemoglobin O of *B. subtilis* (Bs-trHbO). Heme group, relevant residues are depicted. GlnE11 can rotate to form an HB with TyrCD1, opening LT for ligand migration as can be seen by the volume contour plot (black). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Please cite this article as: P. Arroyo-Mañez, et al., Protein dynamics and ligand migration interplay as studied by computer simulation, Biochim. Biophys. Acta (2010), doi:10.1016/j.bbapap.2010.08.005

instance, in *Bacillus subtilis* truncated hemoglobin O (*Bs*-trHb), where a Gln residue is found in the E11 position, the migration occurs almost barrier-less [33]. On the other hand, *M. tuberculosis* truncated hemoglobin O (*Mt*-trHbO), which displays a Leu residue in the E11 position, presents a barrier larger than 10 kcal/mol for the ligand to reach the active site through the LT (Fig. 3A) [32]. This result has been ascribed to the fact that GlnE11 can rotate to establish an hydrogen bond (HB) with TyrB10, opening the LT for ligand migration (Fig. 3B). The blocking of both STG8 and LT in *Mt*-trHbO leaves the STE7 as the only available path for ligand migration in this protein. In addition, the relevance of the residues G8, E11, and E7 was confirmed studying several *in silico* mutants, showing that the ligand migration through internal tunnels in truncated hemoglobins is governed by those residues [32,33]. Another ligand migration investigation in M.T. TrHbO was performed by Guallar et al. [77], combining experimental with a simulation approach. In that work, the authors analyzed different internal cavities and barriers for ligand migration through the internal tunnels using the PELE approach, obtaining a picture fully consistent with the experimental results. Interestingly, they found a similar barrier to that reported by Boechi et al. for the LT, but a higher barrier for the STE7 was found, a fact that was discussed in the work.

3.3. ILS study of Hansenula polymorpha copper-containing amine oxidase oxygen migration

Another interesting example where ligand migration has been studied by ILS methods concerns *Hansenula polymorpha* copper-containing amine oxidase (HPAO) which is able to oxidize primary amines to aldehydes reducing molecular oxygen to hydrogen peroxide [78]. HPAO has been intensively studied experimentally, showing an extremely fast reduction of O_2 by Cu(I) to give superoxide [79–81]. To study ligand migration paths, the protein was crystallized under high xenon pressure. The results showed four major Xe binding sites per HPAO monomer. The location of Xe1 resulted especially intriguing, since this region seemed inaccessible to the solvent. These data, in combination with results obtained for other copper amine oxidases [82], revealed the existence of a chain of Xe atoms occupying the hydrophobic core of the catalytic β-sandwich (Fig. 4). This chain of Xe atoms indicates the pathway that oxygen molecules may follow to enter the catalytic β-sandwich and then proceed towards the active site.

Starting from these structures, the authors performed MD simulations of 10 ns that were subsequently used as input for ILS analysis, resulting in detailed three-dimensional energy profiles for O_2 placement inside the protein. The energy profile was created by contouring energy isosurfaces representing elevated PMF values of 1.8 and 3.0 kcal/mol. The results showed two major regions connecting the solvent boundary with the active site. While one of them starts from the known amine entry channel (left channel), the other one originates at the hydrophobic core of the β-sandwich (right channel). Both pathways merge near the active site in an energetically favorable region, which probably corresponds to an O_2 prebinding site. Interestingly, the location of the minimum energy regions obtained in the ILS calculations coincides with the experimentally observed xenon binding sites, validating the ILS technique. In summary, based on the ILS results, the authors proposed that both pathways are almost isoenergetically favored for molecular oxygen migration. Taken together, the experimental and computational results of HPAO support the idea that small ligand migration occurs through specific pathways specifically designed for that purpose [79,80].

3.4. Dynamical regulation of ligand escape: the nitrophorins

An interesting case of the regulation of ligand migration process by protein dynamics is presented by the NO carriers nitrophorins (NPs). Most blood-sucking insects possess salivary proteins which, upon injection into the victim's tissue, help them to improve their feeding. To fulfill this task, the NPs take advantage of the vasodilator properties of NO. The NPs are heme proteins that store and transport NO in a pH dependent allosterically controlled manner where NO binds tightly at low pH (around 5.6) and is released in the victims´ tissue, at pH of approximately 7.4. The X-ray structures of several NPs show a conserved structure consisting of a eight-stranded antiparallel β-barrel typical of the lipicalin family with the heme group buried in the barrel and capped by the four loops AB, CD, EF, and GH. Experimental data show that pH-dependent NO differential affinity is mainly controlled by NO release [83–89].

To understand the way protein structure and dynamics affect NO release, the process was separated into two steps: the breaking of the Fe–NO bond and the subsequent ligand diffusion from the protein active site to the bulk solvent. Quantum mechanics/molecular mechanics (QM/MM) calculations were employed to show that both low and high pH structures need similar energy for breaking the Fe–NO bond and therefore NO escape must be regulated at the migration step. The corresponding ligand escape process at both pH values was

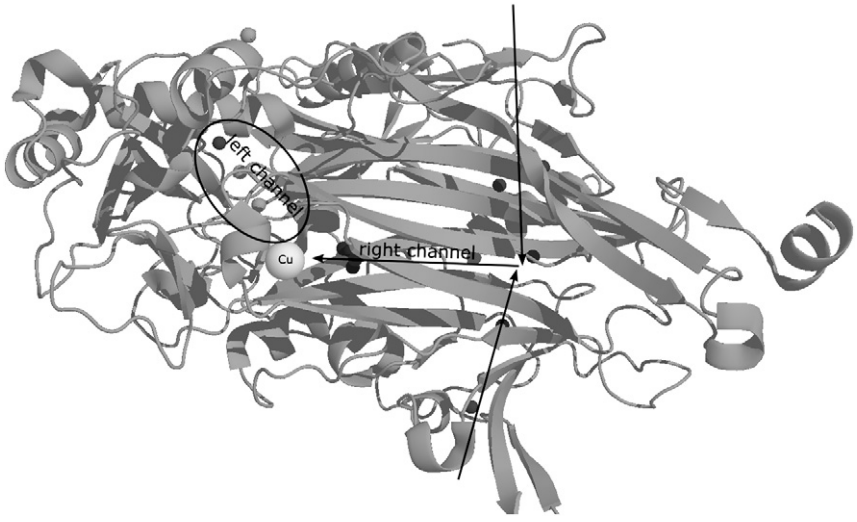

Fig. 4. *Hansenula polymorpha* amine oxidase (HPAO)D3 β-sandwich domain. Spheres represent Xe sites obtained from the X-ray structures of different copper amine oxidases. Black spheres correspond to the Xe sites form the HPAO bound to Xe (PDB ID: 2OQE) [78], while grey ones correspond to the selected from other copper amine oxidases: *Arthrobacter globiformis* (AGAO, PDB ID: 1RJO) [82], *Pichia pastoris* (PPLO, PDB ID: 1RKY) [82], and *Pisum sativum* (PSAO, PDB ID: 1W2Z) [82]. Combined Xe sites forming the Xe chain which corresponds to the left channel (ellipse) and the new O_2 channel proposed by Johnson et al. [78] (black arrows).

Please cite this article as: P. Arroyo-Mañez, et al., Protein dynamics and ligand migration interplay as studied by computer simulation, Biochim. Biophys. Acta (2010), doi:10.1016/j.bbapap.2010.08.005

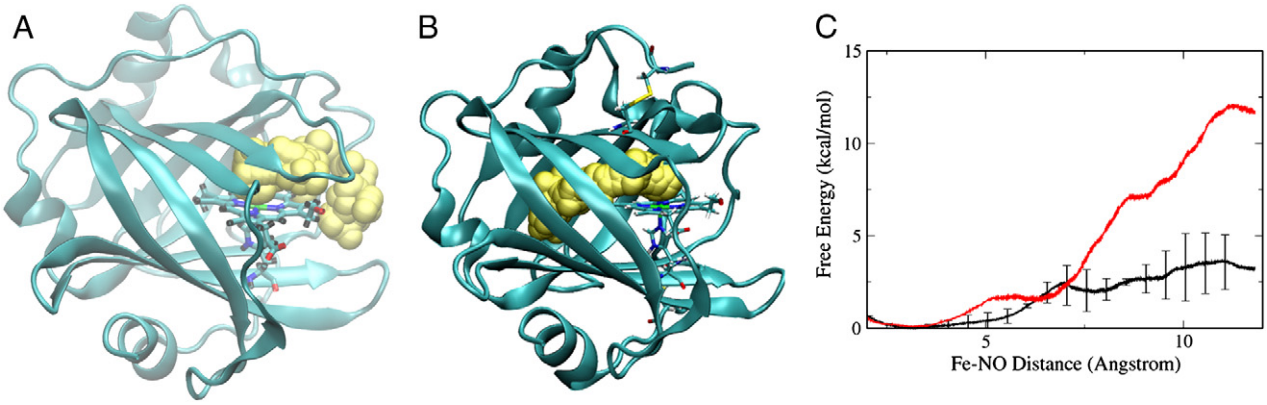

Fig. 5. (A and B) NO location in NP4 along a selected SMD run in high-pH (A) and low-pH (B) conditions. NO positions are shown as yellow spheres; protein is shown as blue ribbons, and the heme group and proximal histidine. (C) Free energy profiles for NO escape from NP4. Results for the high-pH and low-pH structures are depicted using black and red lines, respectively. In the high-pH profile, the data correspond to mean (SD) of two independent sets of 10 MSMD calculations. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

characterized by performing MD simulations with a free NO molecule inside the heme cavity in each case. As expected, in the high pH conditions, NO moves away from the iron and escapes from the protein in three out of three simulations, in between 10 and 20 ns, while at low pH, the NO remains restricted in the active site and no NO escape event is observed during three different MD simulations of 50 ns [12].

Finally, the NO escape mechanism at both pH conditions was studied using MSMD, pulling the NO out of NP4 for each protein conformation. Consistently with the above presented MD results, in the 95% of the MSMD performed at the high-pH condition, the NO escapes yielding a free energy profile with a barrier of ~3 kcal/mol. As expected, the results obtained for the SMD runs at low pH conditions are completely different. The analysis of the work vs RC profiles suggest that the barrier is too high for NO to escape, but instead for about half of the simulations, it migrates deep into the protein matrix. The results are summarized in Fig. 5 where the NO escape, and deeper migration paths and the corresponding free energy profiles are shown [12].

The slightly different NP2 has also been investigated. The results showed the same trend, with the dynamic NO escape process as the responsible for regulating ligand affinity. The main difference between both proteins concerns the amount of pH-dependent structural change which is much smaller (involving only the titratable residue Asp30) in NP2 [22]. In summary, NO escape from NPs presents an interesting example of how to use MD and MSMD to study ligand migration process across the protein matrix and explicitly taking into account protein dynamics and conformational substrates. The obtained results, which perfectly match the experimental observations, make a strong point for the use of this type of methodology [12,22].

3.5. Ligand selectivity: catalase and hydrogenase

Many proteins require for function the entry/exit of significantly different ligands. In these cases, one of the most interesting issues is whether ligand migration paths or tunnels are selective. A few selected interesting cases are reviewed below.

3.5.1. Combining LES and umbrella sampling to study H_2O_2 and oxygen migration in catalase

Catalases are proteins that participate in the dismutation of hydrogen peroxide (H_2O_2) in water and molecular oxygen (O_2). The structural differences between both ligands raise a key issue regarding the entry of H_2O_2 and O_2 through the same pathway. Amara et al. [90] have elegantly elucidated the role of channels and cavities present in the catalase from *Proteus mirabilis* (PMC) using the LES method. Initial cavity calculations performed with the CAVEnv program [91] showed a main channel starting at the surface and leading to the active site where a small cavity is visible. The channel is interrupted at a region called "gate" which blocks the way to the heme. In addition, there are two lateral channels which could participate in the exit of products (Fig. 6).

Using the LES protocol, 20 H_2O_2 molecules were placed at the entrance of the hydrophobic tunnel and were allowed to diffuse freely. However, the results showed that only 2% of the trajectories brought the ligand spontaneously to the active site. This observation led the authors to study migration of H_2O_2 towards the active site employing umbrella sampling to overcome the observed barrier due to the gate residues (Met53, Val95, and Phe132). Results demonstrated that, although H_2O_2 did not diffuse beyond the gate in the plain simulations, the free energy

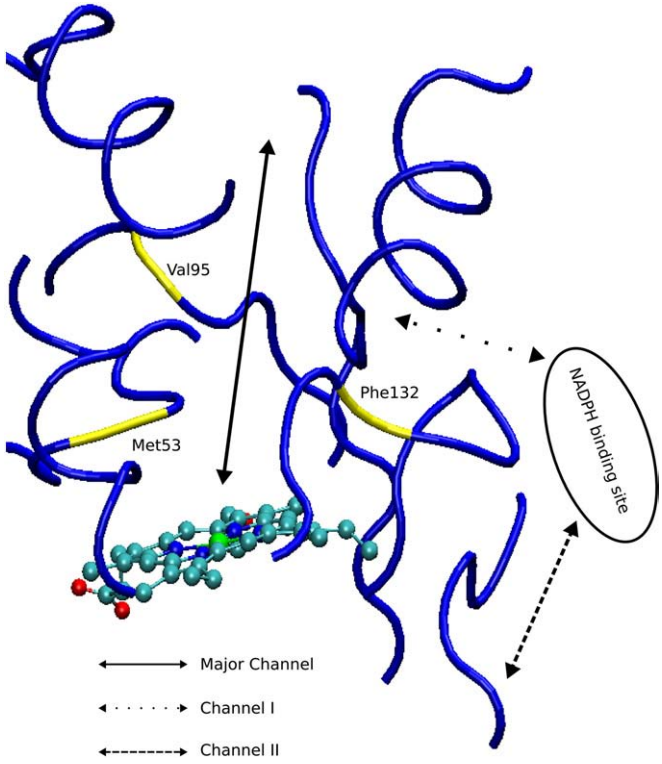

Fig. 6. Scheme of the different diffusion pathways to the active site. The residues Met53, Val95, and Phe132 (in yellow) form the "gate" which blocks the way to the heme. The major and the two lateral channels are described with different types of arrows in the figure. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Please cite this article as: P. Arroyo-Mañez, et al., Protein dynamics and ligand migration interplay as studied by computer simulation, Biochim. Biophys. Acta (2010), doi:10.1016/j.bbapap.2010.08.005

profiles showed that the barrier to reach the active site is small (less than 4.8 kcal/mol), confirming the role of these amino acids in controlling substrate entry.

To reveal the escape pathway, the authors started with H_2O and O_2 placed above the heme in the active site cavity. Once again, a small fraction of runs employing MD plain simulations let the ligands leave the active site through the main channel. The corresponding free energy profiles predicted barriers of 2.2 kcal/mol for O_2 and 3.6 kcal/mol for H_2O, indicating that the main channel favors the escape of products (O_2 and H_2O) rather than substrate (H_2O_2). Finally, the authors evaluated the barrier for O_2 and H_2O escape through both proposed alternative channels. Results showed that for channel I case barriers are 7.9 and 5.3 kcal/mol for O_2 and H_2O, respectively, suggesting that exit by this route may be possible, although it is less likely than in the main channel. On the other hand, channel II did not experiment any opening event during MD simulations, which may prevent any product release using this path. Based on all these observations, the authors concluded that dynamical fluctuations of the protein matrix are essential for the connectivity of tunnel cavities [90].

3.5.2. O_2 vs. H_2 migration in hydrogenase

Another interesting example is the diffusion pathway taken by H_2 and O_2 across CpI [FeFe] hydrogenase from *Clostridium pasteuriarum* studied by Cohen et al. using temperature-controlled locally enhanced sampling (TLES) simulations. Hydrogenases are enzymes that catalyze the reversible oxidation of molecular hydrogen (H_2). Such reaction is achieved by the reduction of H^+ ions from the external solution by electrons provided by a reduced carrier such as ferredoxin [92,93]. In addition, CpI must allow the exit of the product H_2, but it also may allow small molecules such as O_2 and CO to enter the active site. Those molecules bind to Fe and inactivate the enzyme. To understand the possible role of the protein in ligand discrimination, the authors performed different simulations for both H_2 and O_2, to determine the pathways taken for those gases in their way from the active site of the protein to the solution. In addition, they used a "heavy dihydrogen" (hH_2), with molecular mass set equal to that of O_2 to specifically address the role played by ligand size [38,53,54].

The results showed that starting with hH2 in the active site, the ligand escaped through two mayor diffusion pathways, the first one (pathway A) having been previously proposed as an H_2 channel candidate, and the another one referred as pathway B. In the O_2 simulations, the molecules of gas were placed at the active site and allowed to diffuse. In this case, from a total of five simulations, only one provided the O_2 leaving the central cavity through the newly discovered pathway B. In all others, O_2 remained in the central cavity near the binding site. To confirm that O_2 is not able to diffuse through pathway A, possibly due to the narrow tunnel size, the authors performed simulations with O_2 placed inside the tunnel. The results for this case showed that O_2 either diffuses along that channel (avoiding it to leave the protein), or it goes inward to the central cavity and back.

In summary, for hydrogenase both O_2 and hH_2 may diffuse across the protein and exit through two common pathways (A and B), but due to its smaller size, H_2 penetrates in a broader region of the protein and on shorter time scales, allowing partial discrimination between both ligands [38,53,54].

4. Large ligand migration

The migration pathway of ligands other than small molecules represents also a vast field of study, although due to the longer time scales of the process, it represents a tougher task compared with the previous examples. Typical examples include permeation of ligands across membrane, carried out by transmembrane protein channels which are key elements for the cell metabolism [94–98]. For nonchannel proteins, like enzymes, usually binding of large molecules mostly occurs on the surface. However, there are several cases where large ligands must reach a deeply buried active sites, and several possible access or exit channels and their open/close mechanisms have been investigated for a variety of proteins [95,96,99–106].

A powerful technique for this type of studies is as already mentioned; the metadynamics methodology which was successfully applied for the determination of free energy profiles for enter/exit of substrates in several proteins which in all cases showed good agreement with the experimental binding values and also correctly predicting the equilibrium structure of the complex. The main conclusion of these studies is that the flexibility of both the protein and the ligand appears to be crucial in the migration processes, highlighting the impact of protein dynamics in the process [39,42].

Another used technique for the study of large ligand migration is PELE; with this methodology, for example, the microscopic details of palmitate diffusion in the intestinal fatty acid-binding protein were determined, showing, interestingly, that many polar residues in the protein interior were found to be crucial in the migration process, since they induced a 180° turn on the fatty acid in relation to its bound crystal structure, resulting in an orientation where the carboxylic head of the ligand faces the solution, facilitating its exit pathway [43].

An interesting case of regulation of large ligand migrations is presented by the P450 family of proteins. P450s are a ubiquitous protein family with functions including the synthesis and degradation of physiologically important compounds including steroids and prostaglandins and of many xenobiotics such as drugs and procarcinogens. Several studies of ligand migration were performed on this system, using different methodologies including PELE, RAMD, and others [43]. Interestingly, and despite having the same overall fold, and a topologically identical access channel, all P450s studied were proved to utilize very different mechanisms for substrate passage, specifically appropriate for the physicochemical nature of their specific substrate [43,107]. For example, while in P450cam, a natural breathing mode of the enzyme results in the opening of the channel allowing camphor migration, P450eryF uses both a natural breathing motion and an induced fit mechanism to permit 6-deoxyerythronolide B passage. In contrast, the fatty acid (palmitoleic acid) substrate in P450BM-3 is tethered via a salt link to an arginine; the passage then requires the folding of the substrate into a "U" shape as it enters or exits the hydrophobic site. Clearly, for this family of proteins, protein dynamic regulation migration is a key element for determining substrate specificity [107].

Some other works tackling selectivity of large ligand migration in specific proteins by MD simulations can be found in recent literature [108,109]. Elegantly, Pongprayoon et al studied the mechanism of phosphate permeation and selectivity in an outer membrane protein from *Pseudomonas aeruginosa,* using different simulation techniques.

The last presented example concerns the release pathways for products in haloalkane dehalogenase (DhaA) [110]. This enzyme catalyzes hydrolytic dehalogenation of various halogenated aliphatic hydrocarbons to the corresponding alcohol and a halide derivative. In the case studied, 1,2,3-trichloropropane converts into chloride anion and 2,3-dichloropropane-1-ol. The interesting fact about this family of enzymes is that the reaction always takes place in a hydrophobic active site cavity deep inside the enzyme. Although the hydrophobic reactant reaches the cavity easily driven by hydrophobic interactions, the release of the halide and alcohol products occurs through specific tunnels. Upon the reaction, the release of the chloride anion proceeds through a unique channel accompanied by solvation of the negatively charged ion by water molecules, stabilizing the halide charge by asparagine, tryptophan, and lysine residues. On the other hand, the release of the dichloropropanol proceeds through a variety of pathways, favored by the high mobility of the protein backbone and progressive rotations of the amino acid side chains leading to the formation of transient tunnels. Klvana et al. [110] studied DhaA from *Rhodococcus* sp., using RAMD simulations. Their results showed the presence of five pathways connecting the active site to the bulk

Please cite this article as: P. Arroyo-Mañez, et al., Protein dynamics and ligand migration interplay as studied by computer simulation, Biochim. Biophys. Acta (2010), doi:10.1016/j.bbapap.2010.08.005

solvent and displaying a dynamical open/close process. Although two of five described pathways could have previously been observed in the X-ray crystal structure, all the other three were identified and characterized solely by the MD simulations [110].

5. Conclusions

The results presented in this review make apparent the contribution of computer simulation schemes to the exploration of the connection of protein dynamics with ligand migration. In particular, it has been shown how structural and conformational features and ligand migration paths can be investigated with classical MD simulations in combination with advanced sampling tools, such as MSMD, metadynamics, and implicit ligand sampling, to yield information about the migration free energy landscape and possible docking sites. The consistency with the experimental data, when available, constitutes a stringent test to the reliability of these approaches and in this context we have included kinetic and structural information. Computational simulation turns out to be a precious tool because it offers a microscopic view, sometimes resolved in real time. Besides these molecular interactions insight, computer simulations offer a critical tool for applied protein–ligand interactions studies, such as inhibitor design. With the refinement of the methodology and the hardware, computer simulation schemes will be more and more at the center of the stage in structural and molecular biology, likely to become an independent tool that complements nicely experimental evidence.

Acknowledgments

This work was partially supported by the University of Buenos Aires, Agencia Nacional de Promoción Científica y Tecnológica (project PICT Raíces 157), CONICET (PIP 02508), and European Union Project FP7-NOstress. L.B., L.C., S.D.L., D.M.M., and A.A.P. are fellows from CONICET. D. A.E., M.A.M., and A.D.N. are members of CONICET.

References

[1] H. Frauenfelder, G. Chen, J. Berendzen, P.W. Fenimore, H. Jansson, B.H. McMahon, I.R. Stroe, J. Swenson, R.D. Young, Proc. Natl Acad. Sci. USA 106 (2009) 5129–5134.
[2] H. Frauenfelder, P.W. Fenimore, R.D. Young, IUBMB Life 59 (2007) 506–512.
[3] H. Frauenfelder, B.H. McMahon, R.H. Austin, K. Chu, J.T. Groves, Proc. Natl Acad. Sci. USA 98 (2001) 2370–2374.
[4] H. Frauenfelder, F. Parak, R.D. Young, Annu. Rev. Biophys. Biophys. Chem. 17 (1988) 451–479.
[5] O. Jardetzky, Prog. Biophys. Mol. Biol. 65 (1996) 171–219.
[6] A. Bakan, I. Bahar, Proc. Natl Acad. Sci. USA 106 (2009) 14349–14354.
[7] D. Kern, E.R.P. Zuiderweg, Curr. Opin. Struct. Biol. 13 (2003) 748–757.
[8] A. Ostermann, R. Waschipky, F.G. Parak, G.U. Nienhaus, Nature 404 (2000) 205–208.
[9] A. Bidon-Chanal, M.A. Marti, A. Crespo, M. Milani, M. Orozco, M. Bolognesi, F.J. Luque, D.A. Estrin, Proteins: Struct., Funct., Bioinf. 64 (2006) 457–464.
[10] A. Bidon-Chanal, M.A. Marti, D.A. Estrin, F.J. Luque, J. Am. Chem. Soc. 129 (2007) 6782–6788.
[11] M.A. Marti, A. Bidon-Chanal, A. Crespo, S.R. Yeh, V. Guallar, F.J. Luque, D.A. Estrin, J. Am. Chem. Soc. 130 (2008) 1688–1693.
[12] M.A. Marti, M.C.G. Lebrero, A.E. Roitberg, D.A. Estrin, J. Am. Chem. Soc. 130 (2008) 1611–1618.
[13] W.A. Eaton, E.R. Henry, J. Hofrichter, A. Mozzarelli, Nat. Struct. Biol. 6 (1999) 351–358.
[14] F. Pietrucci, F. Marinelli, P. Carloni, A. Laio, J. Am. Chem. Soc. 131 (2009) 11811–11818.
[15] A. Ghosh, The Smallest Biomolecules: Diatomics and Their Interactions with Heme Proteins, vol. 1, Elsevier, 2007.
[16] E.E. Scott, Q.H. Gibson, J.S. Olson, J. Biol. Chem. 276 (2001) 5177–5188.
[17] J.B. Wittenberg, M. Bolognesi, B.A. Wittenberg, M. Guertin, J. Biol. Chem. 277 (2002) 871–874.
[18] A. Pesce, M. Milani, M. Nardini, M. Bolognesi, Meth. Enzymol. 436 (2008) 303–315.
[19] D. de Sanctis, S. Dewilde, A. Pesce, L. Moens, P. Ascenzi, T. Hankeln, T. Burmester, M. Bolognesi, Biochem. Biophys. Res. Commun. 316 (2004) 1217–1221.
[20] M. Milani, A. Pesce, Y. Ouellet, S. Dewilde, J. Friedman, P. Ascenzi, M. Guertin, M. Bolognesi, J. Biol. Chem. 279 (2004) 21520–21525.
[21] L.E. Laverman, P.C. Ford, J. Am. Chem. Soc. 123 (2001) 11614–11622.
[22] J.M. Swails, Y. Meng, F.A. Walker, M.A. Marti, D.A. Estrin, A.E. Roitberg, J. Phys. Chem. B 113 (2009) 1192–1201.
[23] M.A. Marti, D.E. Bikiel, A. Crespo, M. Nardini, M. Bolognesi, D.A. Estrin, Proteins: Struct., Funct., Genet. 62 (2006) 641–648.
[24] M.A. Marti, L. Capece, D.E. Bikiel, B. Falcone, D.A. Estrin, Proteins: Struct., Funct., Bioinf. 68 (2007) 480–487.
[25] R.F. Tilton, I.D. Kuntz, G.A. Petsko, Biochemistry 23 (1984) 2849–2857.
[26] M. Schmidt, K. Nienhaus, R. Pahl, A. Krasselt, S. Anderson, F. Parak, G.U. Nienhaus, V. Åraje, Proc. Natl Acad. Sci. USA 102 (2005) 11704–11709.
[27] F. Schotte, M. Lim, T.A. Jackson, A.V. Smirnov, J. Soman, J.S. Olson, G.N. Phillips Jr., M. Wulff, P.A. Anfinrud, Science 300 (2003) 1944–1947.
[28] C. Bossa, A. Amadei, I. Daidone, M. Anselmi, B. Vallone, M. Brunori, A. Di Nola, Biophys. J. 89 (2005) 465–474.
[29] C. Bossa, M. Anselmi, D. Roccatano, A. Amadei, B. Vallone, M. Brunori, A. Di Nola, Biophys. J. 86 (2004) 3855–3862.
[30] D.R. Nutt, M. Meuwly, Proc. Natl Acad. Sci. USA 101 (2004) 5998–6002.
[31] R. Baron, C. Riley, P. Chenprakhon, K. Thotsaporn, R.T. Winter, A. Alfieri, F. Forneris, W.J.H. van Berkel, P. Chaiyen, M.W. Fraaije, A. Mattevi, J.A. McCammon, Proc. Natl. Acad. Sci. 106 (2009) 10603–10608.
[32] L. Boechi, M.A. Marti, M. Milani, M. Bolognesi, F.J. Luque, D.A. Estrin, Proteins 73 (2008) 372–379.
[33] L. Boechi, P. Arroyo Mañez, F.J. Luque, M.A. Marti, D.A. Estrin, Proteins 78 (2009) 962–970.
[34] O. Carrillo, M. Orozco, Proteins: Struct., Funct., Genet. 70 (2008) 892–899.
[35] Y. Nishihara, S. Hayashi, S. Kato, Chem. Phys. Lett. 464 (2008) 220–225.
[36] J.Z. Ruscio, D. Kumar, M. Shukla, M.G. Prisant, T.M. Murali, A.V. Onufriev, Proc. Natl Acad. Sci. USA 105 (2008) 9204–9209.
[37] J. Cohen, A. Arkhipov, R. Braun, K. Schulten, Biophys. J. 91 (2006) 1844–1857.
[38] J. Cohen, K.W. Olsen, K. Schulten, in: R.K. Poole (Ed.), Methods in Enzymology, vol. 437, 2008, pp. 439–457.
[39] A. Laio, F.L. Gervasio, Rep. Prog. Phys. 71 (2008).
[40] A.R. Leach, Pearson Education EMA, Second Edition, 2001.
[41] C. Jarzynski, Phys. Rev. E 56 (1997) 5018.
[42] F.L. Gervasio, A. Laio, M. Parrinello, J. Am. Chem. Soc. 127 (2004) 2600–2607.
[43] K.W. Borrelli, A. Vitalis, R. Alcantara, V. Guallar, J. Chem. Theory Comput. 1 (2005) 1304–1311.
[44] M.P. Jacobson, D.L. Pincus, C.S. Rapp, T.J.F. Day, B. Honig, D.E. Shaw, R.A. Friesner, Proteins: Struct., Funct., Genet. 55 (2004) 351–367.
[45] A.F. Voter, Phys. Rev. Lett. 78 (1997) 3908.
[46] A.F. Voter, J. Chem. Phys. 106 (1997) 4665–4677.
[47] J.A. Rahman, J.C. Tully, J. Chem. Phys. 116 (2002) 8750–8760.
[48] D. Hamelberg, J. Mongan, J.A. McCammon, J. Chem. Phys. 120 (2004) (1929) 11919–11929.
[49] S.G. Kalko, J. Lluis Test, J. Am. Chem. Soc. 123 (2001) 9665–9672.
[50] S.K. Ludemann, V. Lounnas, R.C. Wade, J. Mol. Biol. 303 (2000) 797–811.
[51] L. Xu, X. Liu, W. Zhao, X. Wang, J. Phys. Chem. B 113 (2009) 13596–13603.
[52] S.D. Golden, K.W. Olsen, Methods Enzymol. 437 (2008) 459–475.
[53] J. Cohen, K. Kim, P. King, M. Seibert, K. Schulten, Structure 13 (2005) 1321–1329.
[54] J. Cohen, K. Kim, M. Posewitz, M.L. Ghirardi, Biochem. Soc. Trans. 33 (2005) 80–82.
[55] M. Brunori, Q.H. Gibson, EMBO Rep. 2 (2001) 674–679.
[56] H. Frauenfelder, B.H. McMahon, P.W. Fenimore, Proc. Natl. Acad. Sci. 100 (2003) 8615–8617.
[57] B. Vallone, M. Brunori, Ital. J. Biochem. 53 (2004) 46–52.
[58] I. Schlichting, J. Berendzen, G.N. Phillips, R.M. Sweet, Nature 371 (1994) 808–812.
[59] H. Hartmann, S. Zinser, P. Komninos, R.T. Schneider, G.U. Nienhaus, F. Parak, Proc. Natl Acad. Sci. USA 93 (1996) 7013–7016.
[60] J.B. Wittenberg, B.A. Wittenberg, J. Exp. Biol. 206 (2003) 2011–2020.
[61] M. Brunori, B. Vallone, Cell. Mol. Life Sci. 64 (2007) 1259–1268.
[62] A. Pesce, M. Bolognesi, A. Bocedi, P. Ascenzi, S. Dewilde, L. Moens, T. Hankeln, T. Burmester, EMBO Rep. 3 (2002) 1146–1151.
[63] A. Pesce, S. Dewilde, M. Nardini, L. Moens, P. Ascenzi, T. Hankeln, T. Burmester, M. Bolognesi, Structure 35 (2004) 63–65.
[64] J.T. Trent, R.A. Watts, M.S. Hargrove, J. Biol. Chem. 276 (2001) 30106–30110.
[65] B. Vallone, K. Nienhaus, M. Brunori, G.U. Nienhaus, Proteins: Struct., Funct., Bioinf. 56 (2004) 85–92.
[66] L. Capece, M.A. Marti, A. Bidon-Chanal, A. Nadra, F.J. Luque, D.A. Estrin, Proteins: Struct., Funct., Bioinf. 75 (2008) 885–894.
[67] D. Hamdane, L. Kiger, S. Dewilde, B.N. Green, A. Pesce, J. Uzan, T. Burmester, T. Hankeln, M. Bolognesi, L. Moens, M.C. Marden, J. Biol. Chem. 278 (2003) 51713–51721.
[68] A.D. Nadra, M.A. Martí, A. Pesce, M. Bolognesi, D.A. Estrin, Proteins: Struct., Funct., Bioinf. 71 (2007) 695–705.
[69] B. Vallone, K. Nienhaus, A. Matthes, M. Brunori, G.U. Nienhaus, Proc. Natl Acad. Sci. USA 101 (2004) 17351–17356.
[70] S. Abbruzzetti, S. Faggiano, S. Bruno, F. Spyrakis, A. Mozzarelli, S. Dewilde, L. Moens, C. Viappiani, Proc. Natl. Acad. Sci. 106 (2009) 18984–18989.
[71] A. Bocahut, S. Bernad, P. Sebban, S. Sacquin-Mora, J. Phys. Chem. B 113 (2009) 16257–16267.
[72] S. Lutz, K. Nienhaus, G.U. Nienhaus, M. Meuwly, J. Phys. Chem. B 113 (2009) 15334–15343.
[73] S. Orlowski, W. Nowak, Biosystems 94 (2008) 263–266.
[74] P. Banushkina, M. Meuwly, J. Phys. Chem. B 109 (2005) 16911–16917.
[75] A. Crespo, M.A. MartÃ-, S.G. Kalko, A. Morreale, M. Orozco, J.L. Gelpi, F.J. Luque, D.A. Estrin, J. Am. Chem. Soc. 127 (2005) 4433–4444.
[76] A. Lama, S. Pawaria, A. Bidon-Chanal, A. Anand, J.L. Gelpi, S. Arya, M. Marti, D.A. Estrin, F.J. Luque, K.L. Dikshit, J. Biol. Chem. 284 (2009) 14457–14468.

Please cite this article as: P. Arroyo-Mañez, et al., Protein dynamics and ligand migration interplay as studied by computer simulation, Biochim. Biophys. Acta (2010), doi:10.1016/j.bbapap.2010.08.005

[77] V. Guallar, C. Lu, K. Borrelli, T. Egawa, S.-R. Yeh, J. Biol. Chem. 284 (2009) 3106–3116.
[78] B.J. Johnson, J. Cohen, R.W. Welford, A.R. Pearson, K. Schulten, J.P. Klinman, C.M. Wilmot, J. Biol. Chem. 282 (2007) 17767–17776.
[79] Y. Goto, J.P. Klinman, Biochemistry 41 (2002) 13637–13643.
[80] S.A. Mills, Y. Goto, Q. Su, J. Plastino, J.P. Klinman, Biochemistry 41 (2002) 10577–10584.
[81] D. Cai, J.P. Klinman, Biochemistry 33 (1994) 7647–7653.
[82] A.P. Duff, D.M. Trambaiolo, A.E. Cohen, P.J. Ellis, G.A. Juda, E.M. Shepard, D.B. Langley, D.M. Dooley, H.C. Freeman, J.M. Guss, J. Mol. Biol. 344 (2004) 599–607.
[83] J.F. Andersen, W.R. Montfort, J. Biol. Chem. 275 (2000) 30496–30503.
[84] J.F. Andersen, X.D. Ding, C. Balfour, T.K. Shokhireva, D.E. Champagne, F.A. Walker, W.R. Montfort, Biochemistry 39 (2000) 10118–10131.
[85] J.F. Andersen, A. Weichsel, C.A. Balfour, D.E. Champagne, W.R. Montfort, Structure 6 (1998) 1315–1327.
[86] A. Weichsel, J.F. Andersen, D.E. Champagne, F.A. Walker, W.R. Montfort, Nat. Struct. Biol. 5 (1998) 304–309.
[87] E.M. Maes, A. Weichsel, J.F. Andersen, D. Shepley, W.R. Montfort, Biochemistry 43 (2004) 6679–6690.
[88] D.A. Kondrashov, S.A. Roberts, A. Weichsel, W.R. Montfort, Biochemistry 43 (2004) 13637–13647.
[89] D.K. Menyhárd, G.M. Keserü, FEBS Lett. 579 (2005) 5392–5398.
[90] P. Amara, P. Andreoletti, H.M. Jouve Test, Protein Sci. 10 (2001) 1927–1935.
[91] A. Volbeda, Ecoles Phys. Chim. Vivant 1 (1999) 47–52.
[92] J.W. Peters, Curr. Opin. Struct. Biol. 9 (1999) 670–676.
[93] Y. Nicolet, C. Cavazza, J.C. Fontecilla-Camps, J. Inorg. Biochem. 91 (2002) 1–8.
[94] M.O. Jensen, Y. Yin, E. Tajkhorshid, K. Schulten, Biophys. J. 93 (2007) 92–102.
[95] T. Wang, Y. Duan, J. Am. Chem. Soc. 129 (2007) 6970–6971.
[96] Y. Wang, K. Schulten, E. Tajkhorshid, Structure 13 (2005) 1107–1118.
[97] K. Tai, P. Fowler, Y. Mokrab, P. Stansfeld, M.S. Sansom, Meth. Cell Biol. 90 (2008) 233–265.
[98] M.Ø. Jensen, D.W. Borhani, K. Lindorff-Larsen, P. Maragakis, V. Jogini, M.P. Eastwood, R.O. Dror, D.E. Shaw, Proc. Natl. Acad. Sci. 107 (2010) 5833–5838.
[99] P. Carlsson, S. Burendahl, L. Nilsson, Biophys. J. 91 (2006).
[100] M. Perakyla, Eur. Biophys. J. 38 (2009) 185–198.
[101] L. Martinez, M.T. Sonoda, P. Webb, J.D. Baxter, M.S. Skaf, I. Polikarpov, Biophys. J. 89 (2005) 2011–2023.
[102] L. Martinez, P. Webb, I. Polikarpov, M.S. Skaf, J. Med. Chem. 49 (2006) 23–26.
[103] D. Genest, G. N., A. Arrault, C. Marot, L. Morin-Allory, M. Genest, Eur. Biophys. J. 37 (2008) 369–379.
[104] A. Aird, J. Wrachtrup, K. Schulten, C. Tietz, Biophys. J. 92 (2007) 23–33.
[105] S.-H. Sheu, T. Kaya, D.J. Waxman, S. Vajda, Biochemistry 44 (2005) 1193–1209.
[106] S. Burendahl, C. Danciulescu, L. Nilsson, Proteins: Struct., Funct., Bioinf. 77 (2009) 842–856.
[107] P.J. Winn, S.K. Ludemann, R. Gauges, V. Lounnas, R.C. Wade, Proc. Natl. Acad. Sci. 99 (2002) 5361–5366.
[108] P. Pongprayoon, O. Beckstein, C.L. Wee, M.S. Sansom, Proc. Natl. Acad. Sci. 106 (2009) 21614–21618.
[109] M. Masetti, A. Cavalli, M. Recanatini, F.L. Gervasio, J. Phys. Chem. B 113 (2009) 4807–4816.
[110] M. Klvana, M. Pavlova, T. Koudelakova, R. Chaloupkova, P. Dvorak, Z. Prokop, A. Stsiapanava, M. Kuty, I. Kuta-Smatanova, J. Dohnalek, P. Kulhanek, R.C. Wade, J. Damborsky, J. Mol. Biol. 392 (2009) 1339–1356.

Please cite this article as: P. Arroyo-Mañez, et al., Protein dynamics and ligand migration interplay as studied by computer simulation, Biochim. Biophys. Acta (2010), doi:10.1016/j.bbapap.2010.08.005