Ligand Binding Rate Constants in Heme Proteins Using Markov State Models and Molecular Dynamics Simulations

*Mauro Bringas[1,2], Leandro E. Lombardi[3], F. Javier Luque[4,5], Darío A. Estrin[1,2], Luciana Capece[1,2]**

[1] Departamento de Química Inorgánica, Analítica y Química Física, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, C1428EGA, Buenos Aires, Argentina.

[2] Instituto de Química Física de los Materiales, Medio Ambiente y Energía (INQUIMAE-CONICET), C1428EGA, Buenos Aires, Argentina.

[3] Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, C1428EGA, Buenos Aires, Argentina.

[4] Department of Nutrition, Food Sciences and Gastronomy, Faculty of Pharmacy and Food Sciences, University of Barcelona, Campus Torribera, 08921, Santa Coloma de Gramenet, Spain.

[5] Institute of Biomedicine (IBUB) and Institute of Theoretical and Computational Chemistry (IQTCUB), University of Barcelona, 08028, Barcelona, Spain.

Corresponding Author

*To whom correspondence should be addressed: e-mail: lula@qi.fcen.uba.ar.

Postal Address: Facultad de Ciencias Exactas y Naturales, Intendente Güiraldes 2160 – Ciudad Universitaria Pabellón 2 (C1428EGA) – Ciudad de Buenos Aires – Argentina. Phone number: +54 11 5285 8298

ABSTRACT

Computer simulation studies of the molecular basis for ligand migration in proteins allow the description and quantification of the key events implicated in this process as, such as the transition between docking sites, displacements of existing ligands and solvent molecules, and open/closure of specific "gates", among other factors. In heme proteins, especially in globins, these phenomena are related to the regulation of protein function, since ligand migration from the solvent to the active site preludes ligand binding to the iron in the distal cavity, which in turn triggers the different globin functions. In this work, a combination of molecular dynamics simulations with a Markov-state model of ligand migration is used to the study the migration of O_2 and $^{\cdot}NO$ in two truncated hemoglobins of *Mycobacterium tuberculosis* (truncated hemoglobin N -Mt-TrHbN- and O -Mt-TrHbO). The results indicate that the proposed model provides trends in kinetic association constants in agreement with experimental data. In particular, for Mt-TrHbN, we show that the difference in the association constant in the oxy and deoxy states relies mainly in the displacement of water molecules anchored in the distal cavity by O_2 in the deoxy form, whereas the conformational transition of PheE15 between open and closed states plays a minor role. On the other hand, the results also show the relevant effect played by easily diffusive tunnels, as the ones present in Mt-TrHbN, compared to the more impeded passage in Mt-TrHbO, which contributes to justify the different $^{\cdot}NO$ dioxygenation rates in these proteins. Altogether, the results in this work provide a valuable approach to study ligand migration in globins using molecular dynamics simulations and Markov-state model analysis.

KEYWORDS. Truncated Hemoglobins – ·NO detoxification - Ligand selectivity – Markov state model – Mycobacterium tuberculosis – Molecular Dynamics

Introduction

Ligand migration is a key process in protein function since it involves the transition of the ligand from the protein exterior to the active site, which in turn can trigger the molecular mechanism associated with the specific protein function[1]. In heme proteins, small ligands as CO, NO, O_2 and H_2S can migrate from the solvent to the distal cavity, where they bind to Fe(II) and/or Fe(III) in the heme group, leading to ligand transport and storage, sensing properties, and enzymatic catalysis, among other functions. Globins are a widely expressed family of heme proteins, which share a 3-over-3 helical sandwich fold. Structurally closely related, truncated hemoglobins (TrHb) are smaller members of the globin family, they being characterized by a 2-over-2- helical motif[2].

The characterization of ligand migration through computer simulations usually requires a vast exploration of the free energy surface of the system, which exceeds the possibilities of standard atomistic molecular dynamics (MD) simulations to attain a complete sampling and derive an accurate kinetic and thermodynamic description. Accordingly, one has to resort to enhanced sampling techniques to overcome these difficulties. As an example, free energy profiles for ligand migration through the tunnels present in TrHbs have been obtained by resorting to biased MD techniques, such as the coupling between Multiple Steered MD (MSMD) methods[3–6] and Jarzynski's equality[7]. In this framework, the ligand is forced to migrate along the tunnel using a variable harmonic restraint, which implies the definition of a geometrical coordinate that guides the ligand passage through the tunnel[8]. Although this procedure is computationally efficient, it also presents several drawbacks, especially because the tunnel topology cannot be generally described using a simple geometrical coordinate. Furthermore, the results may be affected by the pulling velocity and the total amount of trajectories, among other factors. Finally, estimating the free energy from the exponential average of the irreversible pulling work along the reaction coordinate for a finite number of trajectories can lead to notable errors[9].

There have also been attempts to tackle these issues using other techniques, such as the implicit ligand sampling methodology[10] or combining free MD simulations with a numerical solution of the Smoluchowski equation[11]. Metadynamics[12] and weighted ensamble[13] simulations have also been used to characterize the passage of ligand through channels. Alternatively, a promising strategy consists of the use of Markov-State Models (MSM), which have been widely applied to protein folding processes[14,15]. These methods discretize the conformational space into distinct "sites", and characterize the transition probabilities from one site to another. A clear advantage is the possibility to characterize those transition rates or probabilities independently by performing multiple short MD trajectories, ensuring the sampling of transitions between contiguous sites.

In the field of ligand migration, a continuous time-discrete space MSM (ctdsMSM) was used by Blumberger and coworkers[16–19], to study the migration of CO, O_2 and H_2 in Fe-Fe and Fe-Ni hydrogenases, and then obtain second-order kinetic rates with a phenomenological kinetic model. This approach is based on the construction of rate matrices, by estimating the number of times the system transitions between two given states and the average lifetime of the system in each site. It is chemically intuitive, considering that the system can transition in a unimolecular fashion from one site to another and transitions depend on the lifetime of a given state. They also explored the possibility of calculating the fluxes through surfaces that separate the initial and final states. To our experience, the need of estimating the average lifetime of the ligand in a certain state turned out to be not easy to be calculated in a reliable way. Furthermore, the uncertainty in cluster definitions may introduce spurious transitions, making it difficult to estimate the lifetimes of the ligand at the different sites.

On the other hand, Meuwly's group studied ligand binding to Myoglobin[20] and *Mycobacterium tuberculosis* (Mt) trHbN[21–23] using a discrete time-discrete space MSM (dtdsMSM) approach. This strategy avoids the calculation of the lifetime of the system in each state. Nevertheless, it is necessary to introduce and choose a temporal parameter (τ) to ensure that the model satisfies the Markov assumption. Meuwly et al. efficiently characterized the transition rates between sites and determined a characteristic time for the decay of the population in each of them. This information is of great relevance, since it allows obtaining a detailed characterization of individual sites.

Here, using that transition network information, we propose to extract additional information by evaluating the collective evolution of the sites as a function of time, starting from an initial proposed population, so that a phenomenological model can be fitted. To this end, we apply a novel combination of MD simulations, MSM and empirical kinetic considerations to characterize the migration of ligands in Mt-TrHbs (Figure 1).

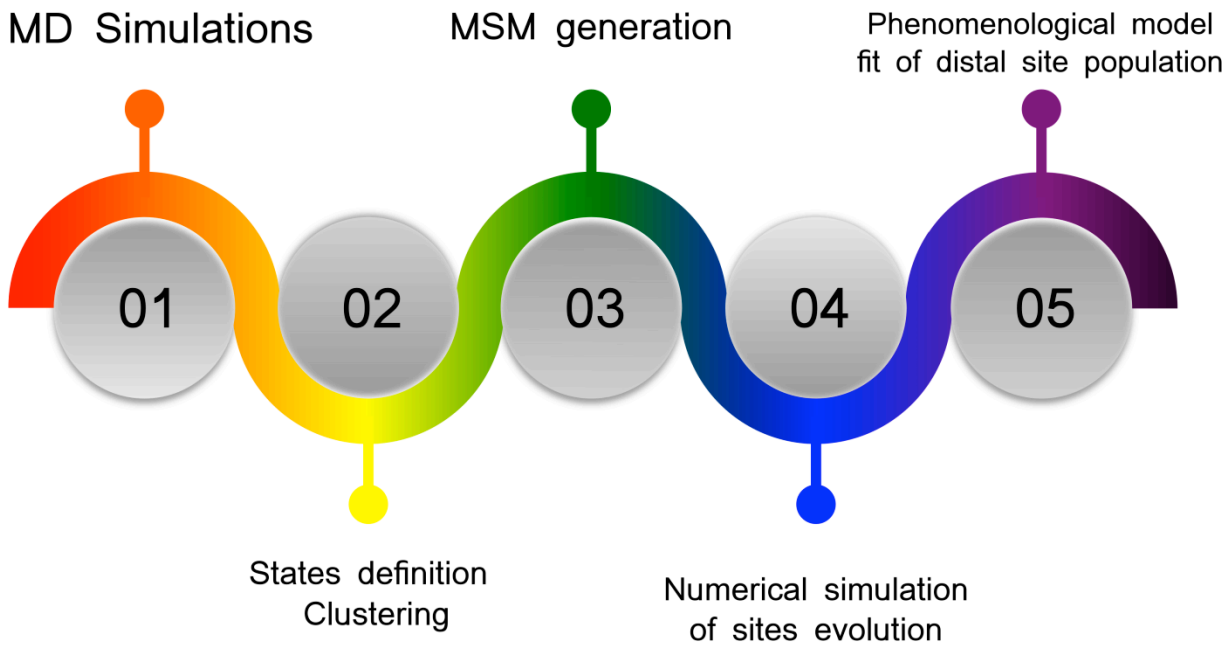

Figure 1. Work pipeline to obtain kinetic rates from MD simulations and Markov State Models. Briefly, the protocol consists in the performance of short MD simulations starting from different ligand positions (1), a clustering analysis of ligand positions (2), the generation of the MSM transition matrix (3), the numerical propagation of this transition matrix along time (4) and finally the fitting to a phenomenological kinetic model (5).

Mt is the pathogen that causes tuberculosis, a disease that produces over three millions deaths each year and affects two billion people around the world[24]. It expresses two members of the TrHb family[25,26], named O (Mt-TrHbO) and N (Mt-TrHbN). This latter protein has a crucial role in ·NO detoxyfication (NOD), a process relevant to warrant the survival of the bacillus[27]. Active in the ferrous form, Mt-TrHbN converts O_2 and ·NO into the harmless species NO_3^-, avoiding the attack from the immune system of the host. To do so, ligands diffuse from the protein exterior to the distal cavity through a complex network of

tunnels (Figure 2) which has been extensively characterized both experimentally[28,29] and computationally[3,21,30,31]. Briefly, three main tunnels were identified in Mt-TrHbN: the long tunnel (LT), located between the B and the E helixes, the short or G8 tunnel (STG8), and the E7-gate tunnel, which is found in many globins, Finally, an additional channel named EH tunnel was also described[27].

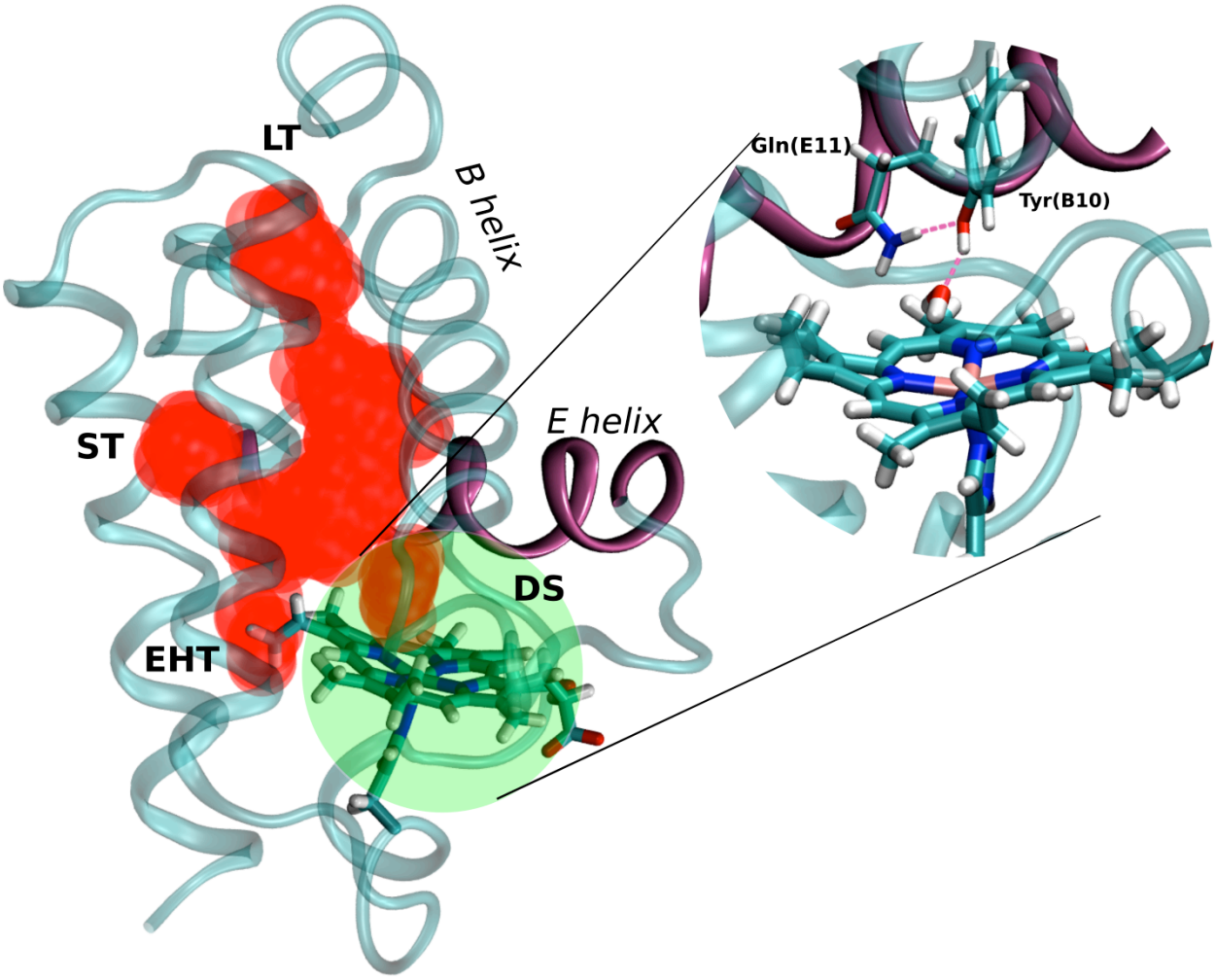

Figure 2. Tubular channels present in Mt-TrHbN and their corresponding entrances: long (LT), short (ST) and EH (EHT). Distal site shaded in green. Detail of polar sidechains relevant in the distal pocket (inset).

The biological function of Mt-TrHbO remains unclear[32]. It oxidizes nitric oxide, decomposes hydrogen peroxide[33] and has been proposed to act as a dioxygen carrier from the cytoplasm towards the cell membrane[34,35] under hypoxia conditions. Its NOD activity is less efficient compared to other heme proteins in the same organism. Ligand migration in Mt-TrHbO is mainly carried out through the LT. The

short channel is blocked by the presence of a tryptophan in the G8 position[36], and the E7 entrance is sealed by electrostatic interactions between positively charged residues and heme propionates[37].

There is a considerable amount of biophysical kinetic data, which is briefly summarized in Table 1. The difference in rate constants for O_2 uptake in deoxy Mt-TrHbN (k_{on}) and ·NO dioxygenation in oxy Mt-TrHbN (k_{ox}) is remarkable, and this selectivity has been a subject of discussion for almost twenty years[3,38]. In this study, for the sake of clarity, we will refer to k_{ox} ·NO as k_{on}, although it is worth noting that it does not correspond to ·NO binding to the deoxy form. Since ligand migration has been previously shown to be the rate limiting step for ·NO oxidation, the entrance kinetic rate is expected to be similar to the overall k_{ox}. Many distal mutants have been expressed and characterized in order to provide a microscopic interpretation[38] to this ligand uptake selectivity. Thus, it was proposed that the deoxy form of the protein presents a conformation that hinders O_2 entrance through the LT into the distal cavity, where the free sixth position of the heme moiety is located. Oxygen binding was proposed to trigger a conformational change that would open the LT, allowing nitric oxide to reach the heme with a higher kinetic rate. This hypothesis was based on the existence of a gate-like phenylalanine residue in the LT, known as the "PheE15 gate". The open conformation was observed to be favored in the oxy form. In this context, the binding of O_2 favors the opening of the E15 gate and consequently a faster migration of the second ligand through the LT[3].

Lately, the lower k_{on} for O_2 relative to NO oxidation was also attributed to the presence of strongly retained water molecules in the vicinity of the ferrous unligated heme iron[38]. Although it is known that water molecules do not bind to the ferrous heme, polar distal residues can act as hydrogen bond donors and therefore interact with water molecules in the distal pocket. In Mt-TrHbN, key distal amino acids are TyrB10 and GlnE11 (Figure 2). It is important to note that while the first ligand (O_2) needs to displace the water molecule to bind to the heme, the second ligand just needs to approach the bound O_2 for the reaction to proceed, avoiding the need to overcome the free energy barrier for water release. The effect of displacing water molecules in the distal cavity on the kinetic association constants was analyzed globally in the truncated hemoglobin family, and was proved to be a key factor in determining ligand migration[39]. Let us note that another mechanism for ·NO dioxygenation where nitric oxide could bind first to the

heme, especially under high ·NO concentrations[40], may be considered. In this case the rate limiting step was proposed to be the ·NO substitution by O_2 in the distal coordination position. Therefore, it is important to remark that the physiologically relevant mechanism will depend on the intracellular concentrations of both gaseous ligands.

In this work, we use a computational scheme outlined above, and summarized in Figure 1, to address a comprehensive analysis of ligand migration in Mt-TrHbN and Mt-TrHbO. The aim is to describe qualitatively and quantitatively the process of ligand migration in these proteins and derive kinetic rate constants from nonguided MD trajectories. For Mt-TrHbN, we aim to obtain a quantitative description of ligand migration towards the distal cavity in the oxy and deoxy states, and disclose the molecular basis for the observed difference in ligand migration for O_2 and ·NO in deoxy and oxy ferrous forms. Specifically, attention will be paid to i) the existence and relevance of the PheE15 gate, and ii) the presence of water molecules in the distal cavity. Finally, we also compare the ·NO oxidation rate between Mt-TrHbN and Mt-TrHbO where, as observed in the constant rate values shown in Table 1, the protein matrix is likely to hinder ·NO migration towards the distal cavity.

Theoretical Methods

In order to estimate the rate constants in TrHbs, we applied a protocol (Figure 1) that consists of i) the generation of multiple short MD simulations, starting with the ligand in different locations of the long and short channels, ii) clustering of the ligand positions sampled in the MD trajectories and definition of the sites in the interior of the protein, iii) construction of the Markov-state transition matrix, iv) propagation of the population vector by application of the transition matrix and determination of the distal cavity population evolution, and v) application of an appropriate phenomenological kinetic model to determine apparent migration rate constants (or rate constant ratios, in our case). In the following we describe the methodological details of these steps.

Molecular dynamics simulations. MD simulations were performed using the Amber 16 package[41], with the amberff99SB forcefield for the protein residues and TIP3P water molecules. Parameters for both oxy and deoxy ferrous heme were generated and validated in our group[42–45]. Diatomic ligands, O_2 and ·NO, were modeled using three-site molecules[46–48], to adequately describe the quadrupole moment of the

ligands. The initial structures for Mt-TrHbN and Mt-TrHbO were taken from the PDB (entries 1IDR[49] and 1NGK[50], respectively). Ionizable side chains were assigned to the predominant form at physiological pH, and histidine protonation state was determined according to the hydrogen bond interactions formed with surrounding residues. The proximal histidine was protonated in the δN, so that it binds to the heme iron through its εN.

For both oxy and deoxy Mt-TrHbN, and oxy Mt-TrHbO systems, a truncated octahedron solvation box was constructed with approximately 9000 TIP3P water molecules. We used a 2 fs time step for the MD integrator and Berendsen thermostat and barostat[51]. A short energy minimization was performed to avoid steric clashes, followed by an 800 ps heating to a final temperature of 300K. The system was simulated for 500 ps using the NPT ensemble in order to achieve a suitable density for the system. Finally, production MD runs were carried out in the NVT scheme. Harmonic restraints were applied to the H-Cα-Cβ-Cγ torsional angle of PheE15 to hold the side chain in the open or closed conformation, using a force constant of 10 kcal mol^{-1} rad^{-2}. Initial structures for independent short MD runs were obtained every 2 ns from long equilibrium production trajectories.

MD trajectory clustering and site definition. The generation of states for the Markov model was performed by geometric clustering of 25 to 30 short (3 ns) MD trajectories with the ligand initially located in the distal site or in any of the previously described channels. Snapshots of these short simulations were recorded every 100 fs. Those MD runs were truncated when the ligand reached the solvent region, to prevent the formation of a large amount of uninformative and poorly sampled states. We found that 12 to 16 clusters represented adequately the distribution of ligand positions inside the protein. The *K*-means algorithm implemented in cpptraj[52] was used. After the unsupervised clustering of ligand positions, in which only the total number of clusters was set but not their positions, the presence of cluster centers close to the reported Xe sites was checked (let us note that the experimentally known Xe sites were not used as an input for the clustering algorithm). Due to the possibility to validate the results from the clustering method with the positions of the experimentally determined docking sites, we did not use any additional clustering methodology. For a detailed comparison of different clustering methodologies applied in this system for a similar analysis, see Cazade et al.[53] Let us note that in this

latter work, the authors indicate that the location and population of the clusters are not significantly affected by the clustering methodology. However, some differences in the kinetic behavior are observed especially when using a more complex method such as the locally scaled diffusion map method.

Construction of the transition matrix and intersite kinetic rate estimation. Each snapshot in the short trajectories was assigned to a previously defined cluster, which are numbered 1 to N. Thus, every trajectory is encoded as the succession of the clusters the system has visited. Let τ be a temporal parameter, known as lag time. A transition from cluster *i* to *j* is considered to occur in a certain frame if the given frame belongs to the cluster *i* but τ frames later it was labeled as *j*.

$c(\tau)_{ij}$ = { # of transitions from cluster i to cluster j in the τ posterior frame}

The method of the sliding window[54] was employed to construct the count matrix. This method evaluates every frame of the trajectory, retrieving non-independent information on the transitions, which are statistically correlated. The other possible scheme is the "independent count" scheme, where the frames inside the interval [x, x+τ] are not used. We preferred the first scheme, as it allows us to extract as much information as possible from our trajectories. However, the "independent count" scheme was applied in order to estimate uncertainties for the calculated rate constants, as explained in the SI.

The transition matrix is constructed as an N dimensional squared matrix in the following way.

$$t_{ij} = \frac{c_{ij}}{\sum_i c_{ij}} \qquad (1)$$

The lag time must be selected in order to capture the characteristic time of the physical process of interest, and was done as suggested by Bowman, Pande and Noé[54]. The choice of a too small value for τ leads to non-Markovian behavior of the resulting MSM. We tested different values of the lag time and observed that when $\tau < 300$ frames, the Markov property was not ensured in the model. We considered that a value of τ=400 frames, which is equivalent to 40 ps in time units, was adequate for this study. As stated before, this lag time indicates that the count matrix construction will be performed comparing every frame of the MD trajectory with the 400th posterior frame.

Numerical simulation of site populations. Fractional populations of different sites can be represented by means of a N-dimensional vector, where each element represents the fraction of ligands inhabiting the corresponding site. The sum of the elements of this vector should always be one, as it was constructed to be fractional. Given an initial vector of populations (in this case, only the clusters at the channel entrance had non-zero population), the temporal evolution of the vector is given as noted in Equation 2.

$$P(n.) = P(0).T(n.\tau) = P(0).T(\tau)^n \quad (2)$$

These equalities are the basis of the Markovian behavior. Plotting the *i*th element of the different P(t) vectors generated as a function of time retrieves the temporal evolution of the *i*th cluster (Figure 3).

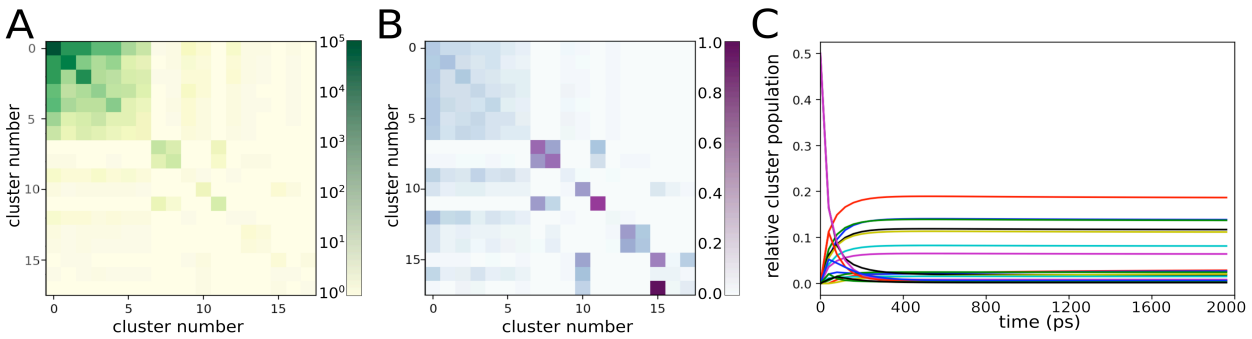

Figure 3. An example of a count matrix (A), which by means of row-normalization turns into a transition matrix (B). (C) Population of the clusters evolution by successive application of the transition operator over an initial distribution vector, using τ=40 ps. The initial population vector concentrated the ligand population in the channel entrances. Each curve corresponds to the temporal evolution the population of a different site. Decreasing traces correspond to initially populated sites, i.e. entrance of the tunnels.

Kinetic analysis and empirical model. It is usual to propose a plausible reaction mechanism and try to fit its rate equation to the kinetic traces determined experimentally. For simple protein-ligand binding and reaction processes, kinetics can be microscopically interpreted in terms of the mechanism presented in Scheme 1. Microscopic kinetic rates corresponding to ligand migration to (k_{in}) and from (k_{out}) the proximity of the reactive center, and the reaction rate (k_r), may be obtained from different types of computer simulations, depending on the physicochemical nature of the involved process. The "reactive" kinetic rate may require a quantum mechanical description of the electronic structure to represent the cleavage or formation of chemical bonds, or it might be related to a process where only non-bonding

interactions are relevant, and there is no actual chemical reaction. In this work, both k_{in} and k_{out} are obtained from the Markov State Model.

$$P + L \overset{k_{bi}\,[L]}{\rightleftharpoons} P + L_{(sup)} \underset{k_{out}}{\overset{k_{in}}{\rightleftharpoons}} \{P \cdots L\} \overset{k_r}{\rightarrow} P - L$$

Scheme 1. General scheme to model ligand binding kinetics in proteins. The first step represents the bimolecular component where the ligand is transferred from the bulk solvent into the proximity of the protein tunnel entrances. The species in brackets represents an encounter complex where the ligand has reached a geminate position to the active center.

The microscopic kinetic rates were estimated by fitting a phenomenological equation (Equation 3) to our numerical simulation, considering that initially the ligand population is located in the tunnel entrance. This model fitting procedure yielded a pseudo first-order kinetic rate (values can be found in Table S2 from the SI).

$$P(t) = \frac{k'_{in}}{k_{out} + k'_{in}} * (1 - e^{-(k_{out} + k'_{in}) * t}) \qquad (3)$$

Since our simulation scheme did not include the estimation of the kinetic rates from the bulk solvent into the channel entrances, the calculation of the absolute second-order kinetic rates is not possible. Nevertheless, we present ratios between those fitted pseudo first-order kinetic rates, which reproduce the experimental trends. Although we intended to minimize possible methodological restrictions to the determination of rate constants, the obtained values are affected by different intrinsic factors, such as the definition of the different states, and the temporal parameter chosen to build the transition matrix, among others. In this context, although only estimations of rate constants can be obtained, this approach is valid to compare the ligand migration through the protein assuming that the bimolecular contribution to the kinetic rate is analogous for the TrHbs in their oxy and deoxy forms.

Tunnel entrance solvent accessible surface (SASA) and cavity volume estimation. Since we do not model the bimolecular diffusional step (Scheme 1), we try to determine how encounter pair formation between a ligand and different proteins will differ by means of analyzing the surface of the tunnel entrances. Solvent accessible area (SASA) was calculated using CPPTRAJ[52] program from the Amber Package, employing default probe radius parameters. To evaluate the tunnel entrances, we calculated the SASA for those protein residues that were more exposed to the solvent. In the case of Mt-TrHbN, the long channel entrance was estimated using Ile19 Ala24 and Ile25. The short channel exposed area was calculated using Ala95, Leu116 and Ile119. In the case of Mt-TrHbO, Met90, Leu50, Leu114 and Leu110 were considered to act as the entrance of the long channel. The volume of internal cavities was estimated using Caver 2.0[55], using an spherical 0.87 Å probe. The cavities calculated that corresponded to any of the described tunnels were summed.

Unimolecular water self-exchange process. As stated before, it is known that water molecules can be retained in the proximity of the ferrous iron by polar distal residues in this family of proteins. To examine this effect, we generated 1 μs MD trajectory of deoxy Mt-TrHbN and evaluated the presence of water molecules in the distal cavity. It was found that a water molecule is constantly anchored to TyrB10 and GlnE11 residues, the distance from the water oxygen atom being in agreement with the formation of hydrogen bond interactions. The residence time of those molecules was measured and an average residence time was calculated.

Results and discussion

Ligand migration in Mt-TrHbs is an interesting test system to calibrate the methodological strategy outlined above, while providing insights into the molecular factors that modulate the passage of gaseous ligands in Mt-TrHbN and Mt-TrHbO, specifically regarding the topological differences in the inner tunnels, the role of gating residues, such as the PheE15 gate in Mt-TrHbN, and the presence of water molecules in the distal cavity in Mt-TrHbN, which altogether determine the differences in ligand migration in two truncated hemoglobins.

·NO and O_2 selectivity in Mt-TrHbN. O_2-bound Mt-TrHbN carries the conversion of nitric oxide to nitrate anion following an almost diffusionally controlled kinetic bimolecular rate, whereas O_2 binding to deoxy Mt-TrHbN is 30–50-fold slower (Table 1). PheE15 may exert a role in this process by regulating the accessibility through the LT, changing its side chain conformation from "closed" to "open" states, and hence allowing a faster migration of the second ligand (·NO) after O_2 binding. Indeed, Figure 4 shows that the PheE15 side chain modifies the ligand positions sampled in MD simulations.

	O_2 k_{on} (10^6 M^{-1} s^{-1})	NO k_{on} (10^6 M^{-1} s^{-1})
deoxy-trHbN	25[32]	–
oxy-trHbN	–	745[27]-1370[56]
deoxy-trHbO	0.11[33]	–
oxy-trHbO	–	0.6[33]

Table 1. Ligand uptake (k_{on}) values for O_2 and ·NO to deoxy and oxy forms, respectively, of Mt-TrHbN and Mt-TrHbO. It should be noticed that O_2 uptake is measured experimentally as the appearance of ferrous oxy Mt-TrHbN spectral bands, and ·NO oxidation is probed by the disappearance of ferrous oxy Mt-TrHbN or the increase of met Mt-TrHbN signals.

In order to test the effect of the PheE15 gating, we estimated k_{in} from the MSM values (Scheme 1) for both ligands in oxy and deoxy Mt-TrHbN, restraining PheE15 to either the open or closed conformation. With this strategy, the aim is to evaluate the ligand migration process to and from the distal site through the protein matrix in the two distinct PheE15 conformational states. As can be observed in Figure S2, the transition between open and closed PheE15 conformations occurs in the time scale of tens of nanosecond. Thus, in the context of the simulations performed here to obtain the transition matrix and the associated rate constants, the sampling of the PheE15 transitions between open and closed states in a totally unbiased scheme would be poor and highly dependent on the initial conformation. It is important to remark, however, that the role of PheE15 side chain has also been studied and described by Diamantis et al.[57], where certain Markovian characteristic loss was observed in Xe migration when performing MD simulations without restraining the PheE15 conformation. Thus, although the computational strategy adopted here presents the limitation of being unable to evaluate directly the effect of the PheE15

conformational change on ligand migration, it allows estimating the kinetic rates for the two conformational states.

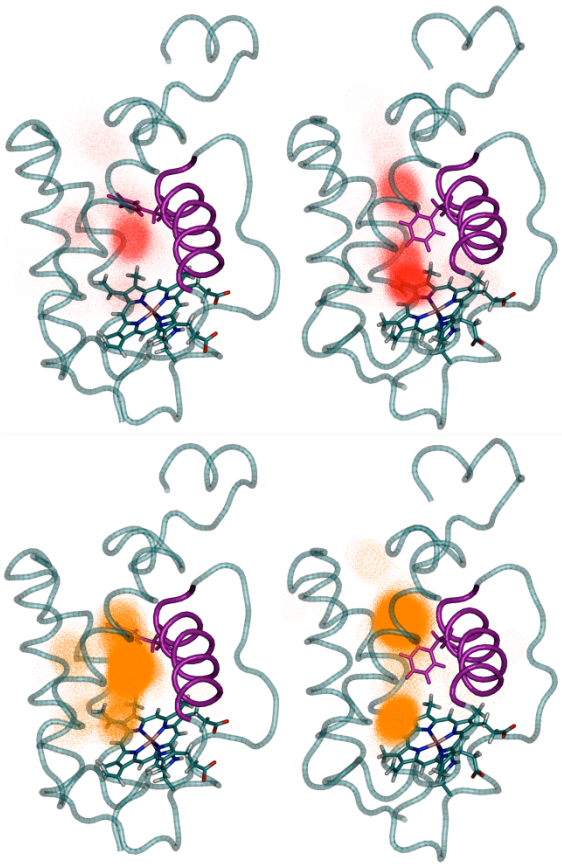

Figure 4. ·NO and O_2 positions in MD simulations for Mt-TrHbN system in the deoxy (above) and oxy (below) forms, with both the PheE15 gate in the closed (right) and open (left) conformations.

In the case of ·NO entrance to the oxy protein, k_{in} were almost equal in the two PheE15 conformations (Table S1). For O_2 migration in the deoxy form, k_{in} was increased 2-fold in the open state relative to the closed one. Additionally, the ratio $k_{in}(O_2)/k_{in}(NO)$ was almost 1.0 for the open state, and 4.2 when PheE15 is in the closed conformation (Table S2). These results indicate that the PheE15 conformation has a mild effect on the ligand migration, which does not suffice to explain the differences in the experimental rates. An alternative factor for ligand differentiation is the displacement of the water molecule anchored close to the sixth coordination position (see Figure 5), already proposed by Guertin et al[35], a process that should affect the O_2 binding in the deoxy state, since the approach of O_2 to the heme Fe(II) is hindered by the

water molecule in the distal cavity. In this scenario, it is possible to propose a unimolecular reaction step that involves water self-exchange (or at least, a release from the hydrogen bond anchoring), a rare event that would act as a rate limiting step in the mechanism proposed in Scheme 2.

$$trHbN^{II}(H_2O) + O_2 \underset{k_{out}}{\overset{k_{in}}{\rightleftharpoons}} trHbN^{II}(H_2O)(O_2) \xrightarrow{k_{exc}} trHbN^{II} - O_2$$

Scheme 2. Kinetic scheme proposed for oxygen binding to deoxy Mt-TrHbN. In this case, the second step is proposed to be rate limiting.

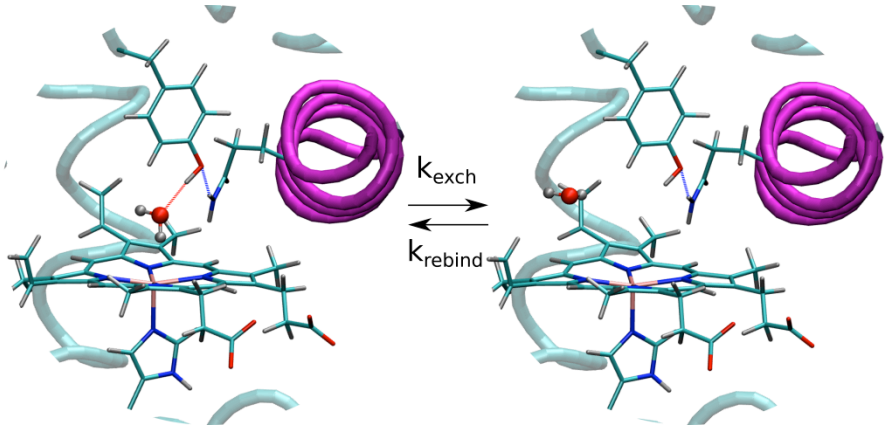

Figure 5. Self-exchange of water molecules from the distal site proximity. Involved polar sidechains are depicted with sticks. E7 helix is depicted in violet for clarity.

Under the assumption that this first-order process is independent of the presence of other molecules in the distal cavity, the associated rate can be estimated as the inverse of the mean lifetime of an anchored water molecule in the distal cavity, as follows

$$k_{exch} = \frac{N}{\sum T_i} \quad (4)$$

where T_i is the lifetime of the *i*th water molecule that has occupied the distal position (Figure 5).

This rate was estimated from the analysis of the residence time for water molecules found in the distal cavity in 1μs unrestrained MD simulation of the deoxy Mt-TrHbN system. Approximately 25 water

exchange events were observed along the trajectory, and the average lifetime was estimated to be about 30 ns.

Since this elementary step of the mechanism is rate limiting, we propose that pseudo preequilibrium conditions are established and the k_{on} could be estimated as

$$k_{on} = \frac{k_{in}}{k_{out}}.k_{exc} = K_{mig}.k_{exc} \qquad (4)$$

where the equilibrium constant K_{mig} was estimated from the ratio between k_{in} and k_{out}, obtained by adjusting the distal site population in an equilibrium relaxation scheme (see SI for more details).

For ·NO oxidation in oxy Mt-TrHbN, the migration through the protein matrix is a kinetically relevant step of the mechanism (Scheme 3). Since O_2 is already bound to the heme iron, there is no need to displace water molecules. Furthermore, the chemical reaction has been reported to occur in a barrierless process[50,51] Considering the reaction mechanism shown in Scheme 3, the k_{on} estimated for NO corresponds to the calculated k_{in} value.

Under these considerations, the ratio $k_{on}(NO)/k_{on}(O_2)$ is estimated to be 15 in the open conformation of PheE15, and 30 in the closed conformation (Table 2). Additionally, if one takes into account the fraction of open and closed states determined from the 1μs free MD of the protein, the $k_{on}(NO)/k_{on}(O_2)$ ratio weighted by the relative population of the two states (closed and open) is estimated to be 26, which is in agreement with the reported experimental data.

$$trHbN^{II} - O_2 + \cdot NO \underset{k_{out}}{\overset{k_{in}}{\rightleftharpoons}} trHbN^{II} - O_2(\cdot NO) \overset{k_{react}}{\rightarrow} trHbN^{III} + NO_3^-$$

Scheme 3. *Kinetic scheme proposed for ·NO oxidation in oxy-Mt-TrHbN. In this case, the second step is known to be fast compared to ligand migration.*

	PheE15 in *open* conformation	PheE15 in *closed* conformation	Weighted-average value (open and closed PheE15 conformations)	Exp.[27,32,56]
$k_{on}(\cdot NO)/ k_{on}(O_2)$	15	30	26	30-55

Table 2. Calculated ·NO and O_2 kinetic rates ratio in open and closed states defined by PheE15 gate conformations (absolute pseudo first-order constants are reported in Table S2). The experimental value corresponds to the ratio between the rates determined for nitric oxide oxidation and oxygen binding.

Overall, the results allow us to estimate in a relative way the influence exerted by the i) PheE15 conformation, and ii) the distal water displacement on the ligand migration in Mt-TrHbN. The analysis of the two molecular assumptions in a phenomenological kinetic scheme allowed us to envision how distal water displacement accounts for most of the ligand selectivity, whereas PheE15 gate conformation provides a minor contribution.

NO dioxygenation in Mt-TrHbO and Mt-TrHbN. To further validate the predictive power of the applied methodology, and the ability to capture the effect of the protein matrix in ligand migration, we analyzed the case of ·NO migration in oxygenated Mt-TrHbO.

In this case, oxygen k_{on}[35] (1.1 10^5 $M^{-1}s^{-1}$) can be explained as well in terms of distal water anchoring, mimicking the case of the Mt-TrHbN. The NO dioxygenation rate (6.105 $M^{-1}s^{-1}$),[35] however, is considerably lower than the almost diffusionally limited rate observed in the Mt-TrHbN case. In our model, the kon for NO can be directly estimated from the obtained kin value, as in the case of TrHbN, since the chemical activation of dioxygen to react with nitric oxide is considered to occur with a low activation barrier, and hence the kinetic efficiency loss for ·NO oxidation is mainly due to protein matrix effects. The calculated value for the pseudo-first order rate constant (*kon NO*) resulted (4±2).10^6 s^{-1} for TrHbO, compared to (1.5±0.7) 10^8 s^{-1} determined for Mt-TrHbN (Table S1).

Although there is qualitative agreement with the experimental data, the calculated ratio is smaller than the experimental one determined from the available rate constants (Table 1). This difference may be attributed, at least in part, to the distinctive topological features of the migration tunnels in the two proteins. First, inspection of the positions occupied by ·NO in the protein interior throughout the MD

simulations, as noted in Figure 6, shows that ligand uptake mainly involves the migration through the long channel[5,38]. Accordingly, the total volume of the internal cavities visited by the ligand in the two proteins is larger in Mt-TrHbN (Table S3), especially when PheE15 is in the open conformation (with a volume of 670 Å^3), than in Mt-TrHbO (with a volume of 460 Å^3), probably affording a larger reservoir for external ligands. Furthermore, the different accessibility to the internal cavities in the two TrHbs may also explain the difference between the respective NOD experimental rates. Indeed, while the solvent accessible surface (SASA) of the entrance to the only channel used for ligand migration in Mt-TrHbO is approximately 20 Å^2, the sum of the SASA the two accessible tunnels in Mt-TrHbN is approximately 175 Å^2.

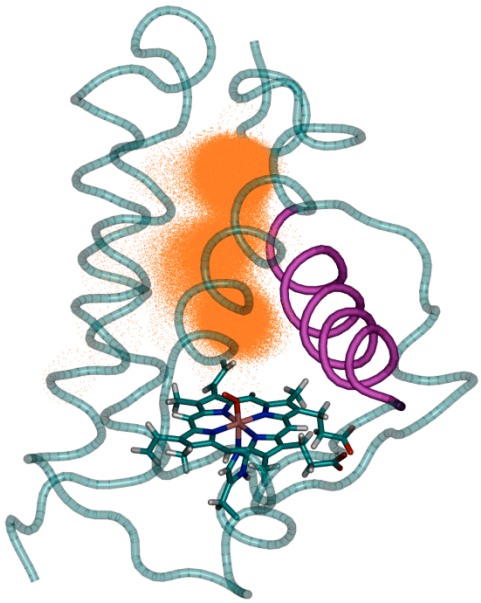

Figure 6. NO positions (orange cloud) in MD simulations for Mt-TrHbO in the oxy form. E helix is depicted in purple for clarity.

These two effects, i.e. the larger probability of the ligand to find the channel entrance from the bulk solvent, and the ability to concentrate ligand molecules inside the channel, are factors not directly accounted for in the MSM scheme. However, they should influence the translational entropy loss due to ligand association to the protein, [58,59] which in turn should facilitate the ligand migration in Mt-TrHbN, explaining the difference between the k_{on}(·NO) values in the two proteins.. These results show the effect

that may be played by easily diffusive tunnels, as the ones present in Mt-TrHbN, compared to the more impeded passage seen in Mt-TrHbO.

Conclusions

Our results attribute the lower rate for O_2 binding to Mt-TrHbN to the displacement of the water molecule anchored by GlnE10 and TyrB11 distal residues, a step required to enable the binding of O_2 to the heme iron. The release of the solvent molecule is the rate limiting step, whereas the PheE15 conformation has a mild effect on the ligand migration. Regarding NO migration, our calculations showed how the protein matrix effects are more important in Mt-TrHbO, where the side chains present in the channels provide a frustrated pathway to reach the distal cavity from the tunnel entrances. Other topological effects that differ between the two proteins may also contribute to the different rates for ·NO migration: i) the different solvent exposure of the tunnel entrance, which is almost 10-fold smaller in Mt-TrHbO relative to trHbN, and ii) a preconcentration process related to the larger volume of the cavity entrance in Mt-TrHbN.

Overall, the combined approach used in thus study, which combines MSM in conjunction with the analysis of simplified systems for the characterization of specific mechanism steps, and a phenomenological kinetic model, allowed us to interpret the differences in the kinetic rate constant determined experimentally for two representative truncated hemoglobins. This computational framework allowed us to discriminate global protein matrix effects from temporally uncoupled molecular rare events, such as water self-exchange from a position with favorable interactions, and the relative importance of the PheE15 gate in ligand migration in Mt-TrHbN.

Supporting Information Description

Additional considerations on the phenomenological models for the estimation of kinetic reaction rates are available. We also present the open and closed conformation fractions obtained from MD, volume of the internal cavities and the fitted pseudo first order kinetic rates.

Acknowledgements

This research was supported by grants from the Universidad de Buenos Aires (UBACYT 20020120300025BA), Agencia Nacional de Promoción Científica y Tecnológica (PICT 2012-2571, PICT 2014-1022, PICT 2015-2761, PICT 2016-0568), Spanish Ministerio de Economía y Competitividad (SAF2017-88107-R, MDM-2017-0767) and Generalitat de Catalunya (2017SGR1746). We acknowledge CECAR and the Barcelona Supercomputer Center (BCV-2016-3-0015) for providing access to computation resources. M.B. holds a CONICET PhD fellowship. L.E.L., D.A.E. and L.C. are members of CONICET.

References

(1) Pang, X.; Zhou, H.-X. Rate Constants and Mechanisms of Protein–Ligand Binding. *Annu. Rev. Biophys.* **2017**, *46* (1), 105–130. https://doi.org/10.1146/annurev-biophys-070816-033639.

(2) Nardini, M.; Pesce, A.; Milani, M.; Bolognesi, M. Protein Fold and Structure in the Truncated (2/2) Globin Family. *Gene* **2007**, *398* (1), 2–11. https://doi.org/10.1016/j.gene.2007.02.045.

(3) Bidon-Chanal, A.; Martí, M. A.; Crespo, A.; Milani, M.; Orozco, M.; Bolognesi, M.; Luque, F. J.; Estrin, D. A. Ligand-Induced Dynamical Regulation of NO Conversion in Mycobacterium Tuberculosis Truncated Hemoglobin-N. *Proteins Struct. Funct. Bioinforma.* **2006**, *64* (2), 457–464. https://doi.org/10.1002/prot.21004.

(4) Boechi, L.; Mañez, P. A.; Luque, F. J.; Marti, M. A.; Estrin, D. A. Unraveling the Molecular Basis for Ligand Binding in Truncated Hemoglobins: The TrHbO Bacillus Subtilis Case. *Proteins Struct. Funct. Bioinforma.* **2010**, *78* (4), 962–970. https://doi.org/10.1002/prot.22620.

(5) Boechi, L.; Martí, M. A.; Milani, M.; Bolognesi, M.; Luque, F. J.; Estrin, D. A. Structural Determinants of Ligand Migration in Mycobacterium Tuberculosis Truncated Hemoglobin O. *Proteins Struct. Funct. Bioinforma.* **2008**, *73* (2), 372–379. https://doi.org/10.1002/prot.22072.

(6) Forti, F.; Boechi, L.; Estrin, D. A.; Marti, M. A. Comparing and Combining Implicit Ligand Sampling with Multiple Steered Molecular Dynamics to Study Ligand Migration Processes in Heme Proteins. *J. Comput. Chem.* **2011**, *32* (10), 2219–2231. https://doi.org/10.1002/jcc.21805.

(7) Jarzynski, C. Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.* **1997**, *78* (14), 2690–2693. https://doi.org/10.1103/PhysRevLett.78.2690.

(8) Bikiel, D. E.; Boechi, L.; Capece, L.; Crespo, A.; Biase, P. M. D.; Lella, S. D.; Lebrero, M. C. G.; Martí, M. A.; Nadra, A. D.; Perissinotti, L. L.; et al. Modeling Heme Proteins Using Atomistic Simulations. *Phys. Chem. Chem. Phys.* **2006**, *8* (48), 5611–5628. https://doi.org/10.1039/B611741B.

(9) Arrar, M.; Boubeta, F. M.; Szretter, M. E.; Sued, M.; Boechi, L.; Rodriguez, D. On the Accurate Estimation of Free Energies Using the Jarzynski Equality: On the Accurate Estimation of Free Energies Using the Jarzynski Equality. *J. Comput. Chem.* **2018**. https://doi.org/10.1002/jcc.25754.

(10) Cohen, J.; Arkhipov, A.; Braun, R.; Schulten, K. Imaging the Migration Pathways for O2, CO, NO, and Xe Inside Myoglobin. *Biophys. J.* **2006**, *91* (5), 1844–1857. https://doi.org/10.1529/biophysj.106.085746.

(11) Banushkina, P.; Meuwly, M. Hierarchical Numerical Solution of Smoluchowski Equations with Rough Potentials. *J. Chem. Theory Comput.* **2005**, *1* (2), 208–214. https://doi.org/10.1021/ct0499480.

(12) Nishihara, Y.; Hayashi, S.; Kato, S. A Search for Ligand Diffusion Pathway in Myoglobin Using a Metadynamics Simulation. *Chem. Phys. Lett.* **2008**, *464* (4), 220–225. https://doi.org/10.1016/j.cplett.2008.09.012.

(13) Zwier, M. C.; Adelman, J. L.; Kaus, J. W.; Pratt, A. J.; Wong, K. F.; Rego, N. B.; Suárez, E.; Lettieri, S.; Wang, D. W.; Grabe, M.; et al. WESTPA: An Interoperable, Highly Scalable Software Package for Weighted Ensemble Simulation and Analysis. *J. Chem. Theory Comput.* **2015**, *11* (2), 800–809. https://doi.org/10.1021/ct5010615.

(14) Lane, T. J.; Shukla, D.; Beauchamp, K. A.; Pande, V. S. To Milliseconds and beyond: Challenges in the Simulation of Protein Folding. *Curr. Opin. Struct. Biol.* **2013**, *23* (1), 58–65. https://doi.org/10.1016/j.sbi.2012.11.002.

(15) Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F. Markov Models of Molecular Kinetics: Generation and Validation. *J. Chem. Phys.* **2011**, *134* (17), 174105. https://doi.org/10.1063/1.3565032.

(16) Greco, C.; Fourmond, V.; Baffert, C.; Wang, P.; Dementin, S.; Bertrand, P.; Bruschi, M.; Blumberger, J.; de Gioia, L.; Léger, C. Combining Experimental and Theoretical Methods to Learn about the Reactivity of Gas-Processing Metalloenzymes. *Energy Env. Sci* **2014**, *7* (11), 3543–3573. https://doi.org/10.1039/C4EE01848F.

(17) Kubas, A.; Orain, C.; De Sancho, D.; Saujet, L.; Sensi, M.; Gauquelin, C.; Meynial-Salles, I.; Soucaille, P.; Bottin, H.; Baffert, C.; et al. Mechanism of O2 Diffusion and Reduction in FeFe Hydrogenases. *Nat. Chem.* **2016**. https://doi.org/10.1038/nchem.2592.

(18) Kubas, A.; De Sancho, D.; Best, R. B.; Blumberger, J. Aerobic Damage to [FeFe]-Hydrogenases: Activation Barriers for the Chemical Attachment of O_2. *Angew. Chem. Int. Ed.* **2014**, *53* (16), 4081–4084. https://doi.org/10.1002/anie.201400534.

(19) Fourmond, V.; Greco, C.; Sybirna, K.; Baffert, C.; Wang, P.-H.; Ezanno, P.; Montefiori, M.; Bruschi, M.; Meynial-Salles, I.; Soucaille, P.; et al. The Oxidative Inactivation of FeFe Hydrogenase Reveals the Flexibility of the H-Cluster. *Nat. Chem.* **2014**, *6* (4), 336–342. https://doi.org/10.1038/nchem.1892.

(20) De Sancho, D.; Kubas, A.; Wang, P.-H.; Blumberger, J.; Best, R. B. Identification of Mutational Hot Spots for Substrate Diffusion: Application to Myoglobin. *J. Chem. Theory Comput.* **2015**, *11* (4), 1919–1927. https://doi.org/10.1021/ct5011455.

(21) Cazade, P.-A.; Meuwly, M. Oxygen Migration Pathways in NO-Bound Truncated Hemoglobin. *ChemPhysChem* **2012**, *13* (18), 4276–4286. https://doi.org/10.1002/cphc.201200608.

(22) Cazade, P.-A.; Berezovska, G.; Meuwly, M. Coupled Protein–Ligand Dynamics in Truncated Hemoglobin N from Atomistic Simulations and Transition Networks. *Biochim. Biophys. Acta BBA - Gen. Subj.* **2015**, *1850* (5), 996–1005. https://doi.org/10.1016/j.bbagen.2014.09.008.

(23) Mishra, S.; Meuwly, M. Quantitative Analysis of Ligand Migration from Transition Networks. *Biophys. J.* **2010**, *99* (12), 3969–3978. https://doi.org/10.1016/j.bpj.2010.09.068.

(24) Sudre, P.; Ten Dam, G.; Kochi, A. Tuberculosis: A Global Overview of the Situation Today. *Bull. World Health Organ.* **1992**, *70* (2), 149–159.

(25) Davidge, K. S.; Dikshit, K. L. Chapter Five - Haemoglobins of Mycobacteria: Structural Features and Biological Functions. In *Advances in Microbial Physiology*; Poole, R. K., Ed.; Microbial Globins -

Status and Opportunities; Academic Press, 2013; Vol. 63, pp 147–194. https://doi.org/10.1016/B978-0-12-407693-8.00005-4.

(26) Milani, M.; Pesce, A.; Nardini, M.; Ouellet, H.; Ouellet, Y.; Dewilde, S.; Bocedi, A.; Ascenzi, P.; Guertin, M.; Moens, L.; et al. Structural Bases for Heme Binding and Diatomic Ligand Recognition in Truncated Hemoglobins. *J. Inorg. Biochem.* **2005**, *99* (1), 97–109. https://doi.org/10.1016/j.jinorgbio.2004.10.035.

(27) Ouellet, H.; Ouellet, Y.; Richard, C.; Labarre, M.; Wittenberg, B.; Wittenberg, J.; Guertin, M. Truncated Hemoglobin HbN Protects Mycobacterium Bovis from Nitric Oxide. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (9), 5902–5907. https://doi.org/10.1073/pnas.092017799.

(28) Milani, M.; Pesce, A.; Ouellet, Y.; Dewilde, S.; Friedman, J.; Ascenzi, P.; Guertin, M.; Bolognesi, M. Heme-Ligand Tunneling in Group I Truncated Hemoglobins. *J. Biol. Chem.* **2004**, *279* (20), 21520–21525. https://doi.org/10.1074/jbc.M401320200.

(29) Boron, I.; Bustamante, J. P.; Davidge, K. S.; Singh, S.; Bowman, L. A.; Tinajero-Trejo, M.; Carballal, S.; Radi, R.; Poole, R. K.; Dikshit, K.; et al. Ligand Uptake in Mycobacterium Tuberculosis Truncated Hemoglobins Is Controlled by Both Internal Tunnels and Active Site Water Molecules. *F1000Research* **2015**. https://doi.org/10.12688/f1000research.5921.2.

(30) Daigle, R.; Guertin, M.; Lagüe, P. Structural Characterization of the Tunnels of Mycobacterium Tuberculosis Truncated Hemoglobin N from Molecular Dynamics Simulations. *Proteins* **2009**, *75* (3), 735–747. https://doi.org/10.1002/prot.22283.

(31) Mishra, S.; Meuwly, M. Nitric Oxide Dynamics in Truncated Hemoglobin: Docking Sites, Migration Pathways, and Vibrational Spectroscopy from Molecular Dynamics Simulations. *Biophys. J.* **2009**, *96* (6), 2105–2118. https://doi.org/10.1016/j.bpj.2008.11.066.

(32) Yeh, S.-R.; Couture, M.; Ouellet, Y.; Guertin, M.; Rousseau, D. L. A Cooperative Oxygen Binding Hemoglobin from Mycobacterium Tuberculosis. *J. Biol. Chem.* **2000**, *275* (3), 1679–1684. https://doi.org/10.1074/jbc.275.3.1679.

(33) Ouellet, H.; Juszczak, L.; Dantsker, D.; Samuni, U.; Ouellet, Y. H.; Savard, P.-Y.; Wittenberg, J. B.; Wittenberg, B. A.; Friedman, J. M.; Guertin, M. Reactions of Mycobacterium Tuberculosis Truncated Hemoglobin O with Ligands Reveal a Novel Ligand-Inclusive Hydrogen Bond Network. *Biochemistry* **2003**, *42* (19), 5764–5774. https://doi.org/10.1021/bi0270337.

(34) Pathania, R.; Navani, N. K.; Rajamohan, G.; Dikshit, K. L. Mycobacterium Tuberculosis Hemoglobin HbO Associates with Membranes and Stimulates Cellular Respiration of RecombinantEscherichia Coli. *J. Biol. Chem.* **2002**, *277* (18), 15293–15302. https://doi.org/10.1074/jbc.M111478200.

(35) Liu, C.; He, Y.; Chang, Z. Truncated Hemoglobin o of Mycobacterium Tuberculosis: The Oligomeric State Change and the Interaction with Membrane Components. *Biochem. Biophys. Res. Commun.* **2004**, *316* (4), 1163–1172. https://doi.org/10.1016/j.bbrc.2004.02.170.

(36) Ouellet, H.; Milani, M.; LaBarre, M.; Bolognesi, M.; Couture, M.; Guertin, M. The Roles of Tyr(CD1) and Trp(G8) in Mycobacterium Tuberculosis Truncated Hemoglobin O in Ligand Binding and on the Heme Distal Site Architecture,. *Biochemistry* **2007**, *46* (41), 11440–11450. https://doi.org/10.1021/bi7010288.

(37) Guallar, V.; Lu, C.; Borrelli, K.; Egawa, T.; Yeh, S.-R. Ligand Migration in the Truncated Hemoglobin-II from Mycobacterium Tuberculosis. *J. Biol. Chem.* **2009**, *284* (5), 3106–3116. https://doi.org/10.1074/jbc.M806183200.

(38) Bustamante, J. P.; Radusky, L.; Boechi, L.; Estrin, D. A.; ten Have, A.; Martí, M. A. Evolutionary and Functional Relationships in the Truncated Hemoglobin Family. *PLOS Comput. Biol.* **2016**, *12* (1), e1004701. https://doi.org/10.1371/journal.pcbi.1004701.

(39) Bustamante, J. P.; Szretter, M. E.; Sued, M.; Martí, M. A.; Estrin, D. A.; Boechi, L. A Quantitative Model for Oxygen Uptake and Release in a Family of Hemeproteins. *Bioinformatics* **2016**, *32* (12), 1805–1813. https://doi.org/10.1093/bioinformatics/btw083.

(40) Das, A. K.; Meuwly, M. Kinetic Analysis and Structural Interpretation of Competitive Ligand Binding for NO Dioxygenation in Truncated Hemoglobin N. *Angew. Chem. Int. Ed.* **2018**, *57* (13), 3509–3513. https://doi.org/10.1002/anie.201711445.

(41) Salomon□Ferrer, R.; Case, D. A.; Walker, R. C. An Overview of the Amber Biomolecular Simulation Package. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2013**, *3* (2), 198–210. https://doi.org/10.1002/wcms.1121.

(42) Nadra, A. D.; Martí, M. A.; Pesce, A.; Bolognesi, M.; Estrin, D. A. Exploring the Molecular Basis of Heme Coordination in Human Neuroglobin. *Proteins Struct. Funct. Bioinforma.* **2008**, *71* (2), 695–705. https://doi.org/10.1002/prot.21814.

(43) Capece, L.; Estrin, D. A.; Marti, M. A. Dynamical Characterization of the Heme NO Oxygen Binding (HNOX) Domain. Insight into Soluble Guanylate Cyclase Allosteric Transition †. *Biochemistry* **2008**, *47* (36), 9416–9427. https://doi.org/10.1021/bi800682k.

(44) Arroyo-Mañez, P.; Bikiel, D. E.; Boechi, L.; Capece, L.; Di Lella, S.; Estrin, D. A.; Martí, M. A.; Moreno, D. M.; Nadra, A. D.; Petruk, A. A. Protein Dynamics and Ligand Migration Interplay as Studied by Computer Simulation. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2011**, *1814* (8), 1054–1064. https://doi.org/10.1016/j.bbapap.2010.08.005.

(45) Capece, L.; Boechi, L.; Perissinotti, L. L.; Arroyo-Mañez, P.; Bikiel, D. E.; Smulevich, G.; Marti, M. A.; Estrin, D. A. Small Ligand–Globin Interactions: Reviewing Lessons Derived from Computer Simulation. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2013**, *1834* (9), 1722–1738. https://doi.org/10.1016/j.bbapap.2013.02.038.

(46) Nutt, D. R.; Meuwly, M. Theoretical Investigation of Infrared Spectra and Pocket Dynamics of Photodissociated Carbonmonoxy Myoglobin. *Biophys. J.* **2003**, *85* (6), 3612–3623.

(47) Meuwly, M.; Becker, O. M.; Stote, R.; Karplus, M. NO Rebinding to Myoglobin: A Reactive Molecular Dynamics Study. *Biophys. Chem.* **2002**, *98* (1–2), 183–207.

(48) Straub, J. E.; Karplus, M. Molecular Dynamics Study of the Photodissociation of Carbon Monoxide from Myoglobin: Ligand Dynamics in the First 10 Ps. *Chem. Phys.* **1991**, *158* (2), 221–248. https://doi.org/10.1016/0301-0104(91)87068-7.

(49) Milani, M.; Pesce, A.; Ouellet, Y.; Ascenzi, P.; Guertin, M.; Bolognesi, M. Mycobacterium Tuberculosis Hemoglobin N Displays a Protein Tunnel Suited for O2 Diffusion to the Heme. *EMBO J.* **2001**, *20* (15), 3902–3909. https://doi.org/10.1093/emboj/20.15.3902.

(50) Milani, M.; Savard, P.-Y.; Ouellet, H.; Ascenzi, P.; Guertin, M.; Bolognesi, M. A TyrCD1/TrpG8 Hydrogen Bond Network and a TyrB10—TyrCD1 Covalent Link Shape the Heme Distal Site of Mycobacterium Tuberculosis Hemoglobin O. *Proc. Natl. Acad. Sci.* **2003**, *100* (10), 5766–5771. https://doi.org/10.1073/pnas.1037676100.

(51) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81* (8), 3684–3690. https://doi.org/10.1063/1.448118.

(52) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9* (7), 3084–3095. https://doi.org/10.1021/ct400341p.

(53) Cazade, P.-A.; Zheng, W.; Prada-Gracia, D.; Berezovska, G.; Rao, F.; Clementi, C.; Meuwly, M. A Comparative Analysis of Clustering Algorithms: O2 Migration in Truncated Hemoglobin I from Transition Networks. *J. Chem. Phys.* **2015**, *142* (2), 025103. https://doi.org/10.1063/1.4904431.

(54) *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*; Bowman, G. R., Pande, V. S., Noé, F., Eds.; Advances in Experimental Medicine and Biology; Springer Netherlands, 2014.

(55) Jurcik, A.; Bednar, D.; Byska, J.; Marques, S. M.; Furmanova, K.; Daniel, L.; Kokkonen, P.; Brezovsky, J.; Strnad, O.; Stourac, J.; et al. CAVER Analyst 2.0: Analysis and Visualization of Channels and Tunnels in Protein Structures and Molecular Dynamics Trajectories. *Bioinformatics* **2018**, *34* (20), 3586–3588. https://doi.org/10.1093/bioinformatics/bty386.

(56) Koebke, K. J.; Waletzko, M. T.; Pacheco, A. A. Direct Monitoring of the Reaction between Photochemically Generated Nitric Oxide and Mycobacterium Tuberculosis Truncated Hemoglobin N Wild Type and Variant Forms: An Assessment of Computational Mechanistic Predictions. *Biochemistry* **2016**, *55* (4), 686–696. https://doi.org/10.1021/acs.biochem.5b01145.

(57) Diamantis, P.; Unke, O. T.; Meuwly, M. Migration of Small Ligands in Globins: Xe Diffusion in Truncated Hemoglobin N. *PLOS Comput. Biol.* **2017**, *13* (3), e1005450. https://doi.org/10.1371/journal.pcbi.1005450.

(58) Amzel, L. M. Loss of Translational Entropy in Binding, Folding, and Catalysis. *Proteins Struct. Funct. Bioinforma.* **1997**, *28* (2), 144–149. https://doi.org/10.1002/(SICI)1097-0134(199706)28:2<144::AID-PROT2>3.0.CO;2-F.

(59) Murphy, K. P.; Xie, D.; Thompson, K. S.; Amzel, L. M.; Freire, E. Entropy in Biological Binding Processes: Estimation of Translational Entropy Loss. *Proteins Struct. Funct. Genet.* **1994**, *18* (1), 63–67. https://doi.org/10.1002/prot.340180108.

Table of Contents Image