PLOS | COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Migration of small ligands in globins: Xe diffusion in truncated hemoglobin N

Polydefkis Diamantis, Oliver T. Unke, Markus Meuwly*

Department of Chemistry, University of Basel, Basel, Switzerland

* m.meuwly@unibas.ch

Check for updates

OPEN ACCESS

Citation: Diamantis P, Unke OT, Meuwly M (2017) Migration of small ligands in globins: Xe diffusion in truncated hemoglobin N. PLoS Comput Biol 13(3): e1005450. https://doi.org/10.1371/journal.pcbi.1005450

Editor: Alexander MacKerell, Baltimore, UNITED STATES

Received: October 6, 2016

Accepted: March 13, 2017

Published: March 30, 2017

Copyright: © 2017 Diamantis et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by the Swiss National Science Foundation (http://www.snf.ch/en/Pages/default.aspx) grant 200020 169079 (to MM), the NCCR MUST, http://www.nccr-must.ch/home.html (to MM), and the University of Basel, https://www.unibas.ch/en.html. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

In heme proteins, the efficient transport of ligands such as NO or O_2 to the binding site is achieved via ligand migration networks. A quantitative assessment of ligand diffusion in these networks is thus essential for a better understanding of the function of these proteins. For this, Xe migration in truncated hemoglobin N (trHbN) of *Mycobacterium Tuberculosis* was studied using molecular dynamics simulations. Transitions between pockets of the migration network and intra-pocket relaxation occur on similar time scales (10 ps and 20 ps), consistent with low free energy barriers (1-2 kcal/mol). Depending on the pocket from where Xe enters a particular transition, the conformation of the side chains lining the transition region differs which highlights the coupling between ligand and protein degrees of freedom. Furthermore, comparison of transition probabilities shows that Xe migration in trHbN is a non-Markovian process. Memory effects arise due to protein rearrangements and coupled dynamics as Xe moves through it.

Author summary

Binding and transport of ligands in proteins is essential, in particular in globular proteins which often exhibit internal cavities. In truncated Hemoglobin N (trHbN) these cavities are arranged as a network with particular connectivities. The present work supports the notion that ligand diffusion in trHbN is an active process and coupled to the protein dynamics. Furthermore, transition probabilities between neighboring pockets depend on the location from where the ligand entered the transition, which is typical for non-Markovian processes. Hence, ligand migration in trHbN exhibits memory effects due to dynamical coupling between the protein and ligand motion.

Introduction

The interplay between protein and ligand degrees of freedom is of great interest in understanding protein function including catalysis, signaling or transport. Both, ligand binding and transport between internal localization sites often leads to small changes in the functionally relevant protein structure and fluctuations around it. Changes in structure and fluctuations constitute a way to store information in proteins, for example through the concept of

Competing interests: The authors have declared that no competing interests exist.

configurational entropy. [1, 2] The transfer of such information between spatially separated regions is essential for transducing signals in chemical and biological systems. One concrete manifestation of this is allosteric communication in proteins whereby information is transferred over distances of the size of a protein. [3–7] Allostery involves a cascade of information exchange processes which can be initiated by uptake of an exogenous ligand. This raises the question whether and if yes, in which way and on what time scale ligands in proteins influence "information flow" in proteins. Starting from a "system-bath" perspective it appears natural that ligands in proteins are active entities beyond their mere chemical role in chemical transformations.

Evolution has developed a range of mechanisms to transmit information across a protein following specific triggering events such as binding of a ligand or absorption of a photon. A recent review of the current state of understanding of allostery pointed out that the "new view" of allostery, which is based on population shift, is actually reminiscent of the 1965 model put forward by Monod, Wyman and Changeux (MWC). [3] This view of allostery underlines the importance of equilibrium fluctuations for the function of allosteric proteins. In a protein-ligand context these equilibrium fluctuations may be subject to localization of a ligand in different pockets and following different pathways. Allosteric control involves several key steps which eventually lead to biological function. Often, the first step is a local structural change, induced by ligand binding. Next, this information is propagated through the protein network. The time scale for this ranges from picoseconds to microseconds, highlighting the large variation in temporal and spatial scales involved in allostery. In a later phase, structural changes lead to affinity changes which complete the chain of events. Hence it is of genuine interest to better understand the mechanistic basis of local structural changes induced by ligand localization and migration, and the time scales on which they occur and relate them to the fate of the system on longer time scales. [8–11]

The binding and reaction of small gas phase ligands like oxygen (O_2) or nitrogen monoxide (NO) with heme proteins is involved in processes that are essential for the physiological function of living organisms. The ligand binding sites of heme proteins are buried. Therefore, ligand migration from the outside towards the active site involves ligand diffusion, often occurring through a network of connected, internal sites. Extensive research has shown that such networks of internal cavities (pockets, packing defects) exist particularly in globular proteins, and has been confirmed for heme containing proteins including myoglobin (Mb) [12–14], cytochrome bα3 oxidase [15, 16], dimeric hemoglobin [8, 14], or truncated hemoglobins N and O [17–19].

The physico-chemical (thermodynamic and kinetic) properties of ligand migration and reaction with heme proteins have been extensively investigated. Besides facilitating ligand diffusion towards the binding site of the protein, ligand migration networks have also been suggested to detain ligand molecules in their pockets, either for storage [13, 16], or for subsequent detoxification [18]. Nonetheless, further investigation on the structural and temporal nature of ligand motion in the protein is required to better understand whether and, if so, how protein and ligand motion are coupled.

In Mb, it has long been suggested that the binding of small ligands to specific pockets affects the internal motions and conformational substates adopted by the protein [12]. This should also apply to other heme proteins exhibiting a similar ligand migration network. In fact, recent computational work on i) NO and O_2 migration in truncated hemoglobin N of Mycobacterium Tuberculosis [20–22], ii) CO migration on myoglobin [23] or iii) Benzamidine in the trypsin inhibitor [24] point in a similar direction: the conformational dynamics of the protein and the ligand (probe) are coupled and, depending on the localization of the ligand within the protein (or at its surface), the conformational dynamics of the protein changes.

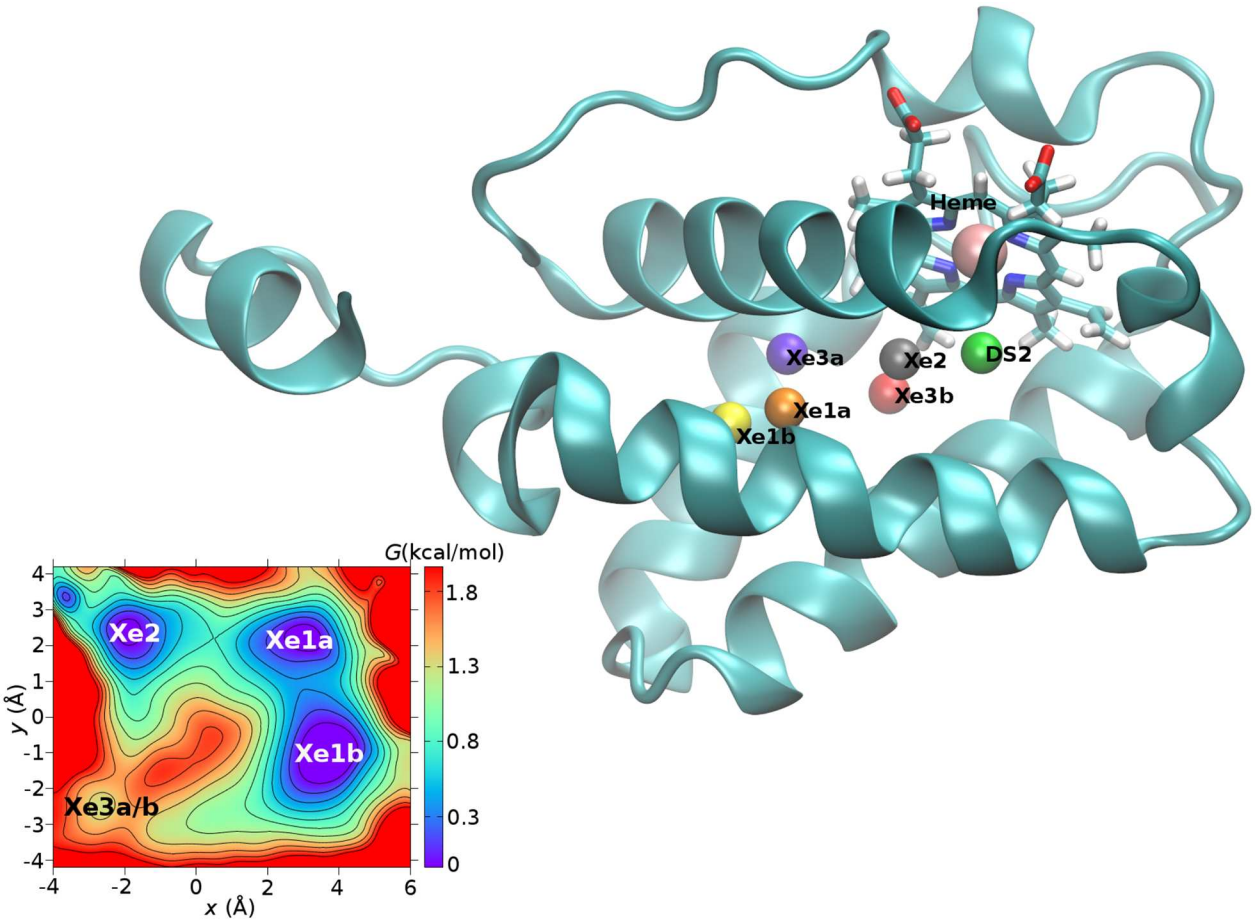

Fig 1. TrHbN and important pockets. Main Figure: The crystal structure of trHbN, with the heme group in licorice, Xe1a/b, Xe2, Xe3a/b and DS2 pockets as spheres. Inset: 2-dimensional cut through the free energy surface along the plane defined by Xe2, the saddle points between Xe1a and Xe1b, and between Xe3a and Xe3b, respectively.

https://doi.org/10.1371/journal.pcbi.1005450.g001

In this work, the nature of ligand migration in truncated hemoglobin N (trHbN, see Fig 1) of Mycobacterium Tuberculosis has been investigated using molecular dynamics (MD) simulations. The migration pathways for NO and O_2 in trHbN have been extensively characterized experimentally [17, 18] and computationally [20–22, 25–29]. TrHbN is an ideal protein for such an investigation as earlier studies [20, 21, 25–27, 29] have revealed that the free energy barriers between different pockets are low (1 to 2 kcal/mol) which allows to extensively sample transitions between them. Xenon (Xe) has been selected as the ligand for this study as it has been used in the experimental determination of binding pockets within trHbN [18] (see Fig 1) and cytochrome bα3 oxidase [16]. In addition, Xe has been suggested to be a good mimic of O_2 [16], a physiologically relevant ligand of trHbN and other heme proteins. Furthermore, evidence from theoretical studies on myoglobin [13] suggests that Xe migration networks in heme proteins share the same pockets and pathways with natural ligands, such as O_2, NO and CO.

In the present work ligand migration and the interplay between ligand and protein dynamics is investigated from a ligand perspective and at atomic resolution. In particular, the question whether, how and on what time scale the motion of a diffusing Xenon atom influences protein dynamics is quantitatively assessed. Furthermore, we analyze potential memory effects

in the transition dynamics between neighboring pockets depending on where the Xenon atom entered a transition path of interest.

Methods

Molecular dynamics simulations

All simulations were performed using the CHARMM suite of programs [30] with the CHARMM22 force field [31] and the van der Waals radius of Xe was R_{min} = 2.25 Å which is close to a radius of 2.24 Å used in previous work [13]. For water, the TIP3P model was used [32]. All simulations were carried out with a timestep of 1 fs and a cutoff of 14 Å was used for the non-bonded interactions, and SHAKE was used for constraining the bonds involving hydrogen atoms [33, 34]. Long-range electrostatic interactions were treated with Ewald summation [35, 36], employing 96, 64 and 64 grid points along the *x*–, *y*– and *z*–axis, respectively.

The B subunit of a crystallized structure of the trHbN dimer (Protein Data Bank (PDB) Index: 1S56 [18]) was used as the starting structure with one Xe atom initially placed in either the Xe1, Xe2 or Xe3 pocket, which are the focus of the present work. The monomer was solvated in a pre-equilibrated periodic water box (71Å, 56Å, 56Å). The system was then heated and equilibrated at 300 K. Two different equilibration procedures were considered. In the first, Xe was weakly restrained with a force constant of 5 kcal/mol/Å^2 to its initial pocket (Xe1, Xe2 or Xe3) which helps to better sample the region of interest (Xe1, Xe2, Xe3). Simulations based on a second equilibration scheme (with no restraints imposed on Xe) confirmed that restraints do not lead to artifacts. In both cases, frames were extracted every 2.5 ps and used as starting points for the production phase (*NVE*) simulations. In total 725.5 ns of simulations were run and analyzed.

Free energy surface for Xe migration

An effective 3-dimensional free energy surface for Xe migration in trHbN was constructed using the Xe coordinates from ≈ 200 ns of simulation. A regular grid with spacing of 1 Å along the *x*–, *y*– and *z*–direction was employed for this. The Helmholtz free energy $G(x, y, z)$ at each grid point was computed according to $G(x, y, z) = -RTln(P(x, y, z))$ where $P(x, y, z)$ is the probability of observing Xe in point (x, y, z), R is the universal gas constant and T is the temperature in K. A smooth free energy surface was then constructed by interpolating the discrete points of $G(x, y, z)$ using 3-dimensional cubic B-splines [37].

Clustering

***k*–means clustering.** In *k*–means clustering [38] a collection of data points is decomposed into disjoint sets with respect to the distance of each point from a number of geometrical centers, each defining a cluster. Every point is assigned to the cluster center closest to it. In the system investigated here, the centers of mass of 9 main pockets (Xe1a, Xe2, Xe3, Xe4, Xe1b, DS2, ENT, IS1, PDS) were provided as an initial input. For each pocket, the coordinates of its center of mass were determined from the amino acids forming it [29]. The algorithm was used using two cut-off distances r_{min} = 1.75 Å and r_{min} = 6.20 Å, similar to a previous analysis of ligand migration in trHbN with *k*–means [21]. The inner cut-off defines a sphere within which Xe positions should be used for re-calculating the centers of the clusters during the analysis, whereas the outer cut-off decides which frames should not be assigned to any of the 9 clusters. These points (outliers) were assigned to a tenth cluster, named "Else".

Robust Growing Neural Gas algorithm. Xe coordinates were also clustered using the Robust Growing Neural Gas (RGNG) algorithm [39]. Contrary to *k*–means, RGNG does not

require an initial guess for the centers of the clusters. The algorithm is very robust due to (i) its insensitivity to the presence of outliers, and (ii) its ability to auto-determine the optimal number of cluster centers for a given dataset based on a minimum description length criterion [40].

Final clustering. Based on the free energy surface, the position of each pocket center was refined using conjugate gradient descent starting from the pocket center coordinates obtained by the clustering techniques (*k*–means, RGNG). In addition, the minimum energy paths (MEPs) for relevant transitions within the (Xe1, Xe2, Xe3) states including their substates were also determined. For each transition, the corresponding MEP was obtained using conjugate peak refinement [41]. The pockets (defined by pocket centers) and the transition regions (defined by saddle points) determined on the free energy surface of trHbN were used for the final clustering of the Xe coordinates.

Results/discussion

Before discussing the transition dynamics between neighboring pockets and their dependence on the trajectory history, the relevant states for the present work are discussed.

Characterization of the states

Table 1 provides a summary of the percentage occupation of each state (i.e. the fraction of total simulation time spent by Xe in a particular pocket), according to *k*–means clustering, RGNG, and the final clustering (clustering with respect to both, pocket centers and transition regions).

According to Table 1, based on *k*–means clustering, the most occupied pocket is Xe1b, followed by DS2, Xe1a, Xe3b, Xe2, and Xe3a. In addition, Table 1 shows that over the course of the MD simulations, the region comprising pockets Xe1a/b, Xe2, and Xe3a/b are highly populated with an aggregated population of 64.4% of the total simulation time. The region containing Xe3 is particularly interesting as previous work showed that it is a hub of the ligand migration network of trHbN [22, 27–29]. The high occupation of the (Xe1, Xe2, Xe3) region was expected, given that the majority of MD simulations were started with Xe in any of these

Table 1. Clustering of Xe position.

Pocket	*k*–means (%)	RGNG (%)	Final Clustering (%)
Xe1a	13.1	12.9	11.5
Xe1b	31.5	33.0	33.4
Xe2	6.9	6.9	6.4
Xe3a	5.9	6.1	5.5
Xe3b	7.0	7.2	7.5
Xe4	1.5	—	0.5
DS2	15.2	15.1	15.2
ENT	2.9	—	0.9
IS1	1.5	—	1.1
PDS	0.6	—	0.0
Else	13.9	18.8	13.6
Transition States	—	—	4.9

The fraction of total simulation time that Xe spent in each pocket (occupation) according to *k*–means clustering, the RGNG algorithm, and the final clustering. In RGNG, "Else" corresponds to *k*–means "Else" + Xe4, ENT, IS1 and PDS, which are not found by RGNG. In the final clustering, "Else" is used to cluster all points not belonging to any of Xe1a/b, Xe2, Xe3a/b, Xe4, DS2, ENT, IS1, PDS pockets or transition states. The occupation of PDS for the final clustering is 0.003%

https://doi.org/10.1371/journal.pcbi.1005450.t001

three pockets to increase sampling probability of the relevant transitions involving these pockets. Hence, the occupations in Table 1 are probably not representative of an equilibrium population which, however, is of no concern for the purpose of the present work. Ligand diffusion in trHbN has previously been investigated for O_2 and was analyzed with different clustering methods, including *k*–means, kinetics-based Markov Clustering (MCL) and the locally scaled diffusion map (LSDMap) [21]. The O_2–occupation of trHbN differs significantly from the one found in the present work for Xe a) because Xe and O_2 interact differently with the protein, and b) because the simulations started with Xe in either Xe1, Xe2 or Xe3. Hence, for O_2 exploration of the network started from an equilibrium distribution which was not the case here.

As an independent validation of *k*–means, pocket analysis was carried out with the RGNG algorithm. The RGNG clustering yielded six clusters which correspond to the most frequently visited Xe pockets according to *k*–means clustering, see Table 1. Compared to *k*–means, the centers of pockets Xe1a, Xe1b, Xe2 and DS2 shift by less than 1 Å. The finding of 2 clusters that both belong to the Xe3 pocket agrees with previous indications of this pocket being large and diffuse [29]. In hindsight the separation of Xe3 into Xe3a and Xe3b could already have been anticipated in an earlier investigation of O_2 diffusion in trHbN [21]. In the present work, RGNG allows to explicitly identify two separate sub-states within the Xe3 pocket. This is similar to the situation in Mb where the Xe4 pocket is also large but contains two substates which are separated by a small but nonzero free energy barrier [42]. To distinguish these two pockets, they are referred to as Xe3a and Xe3b, respectively. These two states are separated by a 1.1 kcal/mol and 0.9 kcal/mol forward and reverse barrier, respectively. Finally, for Xe3a and Xe3b, the respective occupations are 6.1% and 7.2% in RGNG. For the (Xe1, Xe2, Xe3) region the total occupation fraction is 66.1%, which supports the 64.4% fraction obtained from *k*–means.

In order to validate the assignment of Xe positions to particular pockets, explicit time series along the Xenon *x*–, *y*–, and *z*–coordinates are compared with the discrete time series from the clustering. It is found (see Fig 2) that pocket assignments can be difficult at boundaries between regions because of the non-spherical shape of the cavities. As a consequence, an event may be assigned to the wrong state if Xe is in a transition region between neighboring states. This was indeed confirmed when visualizing *k*–means- and RGNG-clustered points in 3 dimensions. Further analysis of the clustered points would thus lead to non-realistic estimates of life times, transition times, and transition probabilities. As a consequence, additional states corresponding to transition regions were introduced to alleviate this problem (see top panel of Fig 3 for more detail).

The pockets (defined by pocket centers) and transition regions (defined by saddle points) determined on the free energy surface of trHbN were used for the final clustering of Xe coordinates for the subsequent analysis. The occupation of each pocket following this procedure is presented in the fourth column of Table 1. When transition states are included in the clustering, the total occupation of the (Xe1, Xe2, Xe3) region is 64.3% of the total simulation time, consistent with *k*–means and RGNG clustering. In addition, Xe is found in transition regions during 4.9% of the total simulation time.

Dwell and transition time distributions

After clustering with respect to both pockets and transition regions, 19252 transitions were identified from more than 700 ns of simulation. The transitions between neighboring pockets are summarized in Table 2 and graphically represented in Fig 3. Once "states" are defined, transitions between them can be analyzed quantitatively by means of residence ("dwell") times τ, their distributions $p(\tau)$ and transition times. The residence time is defined by the time Xe

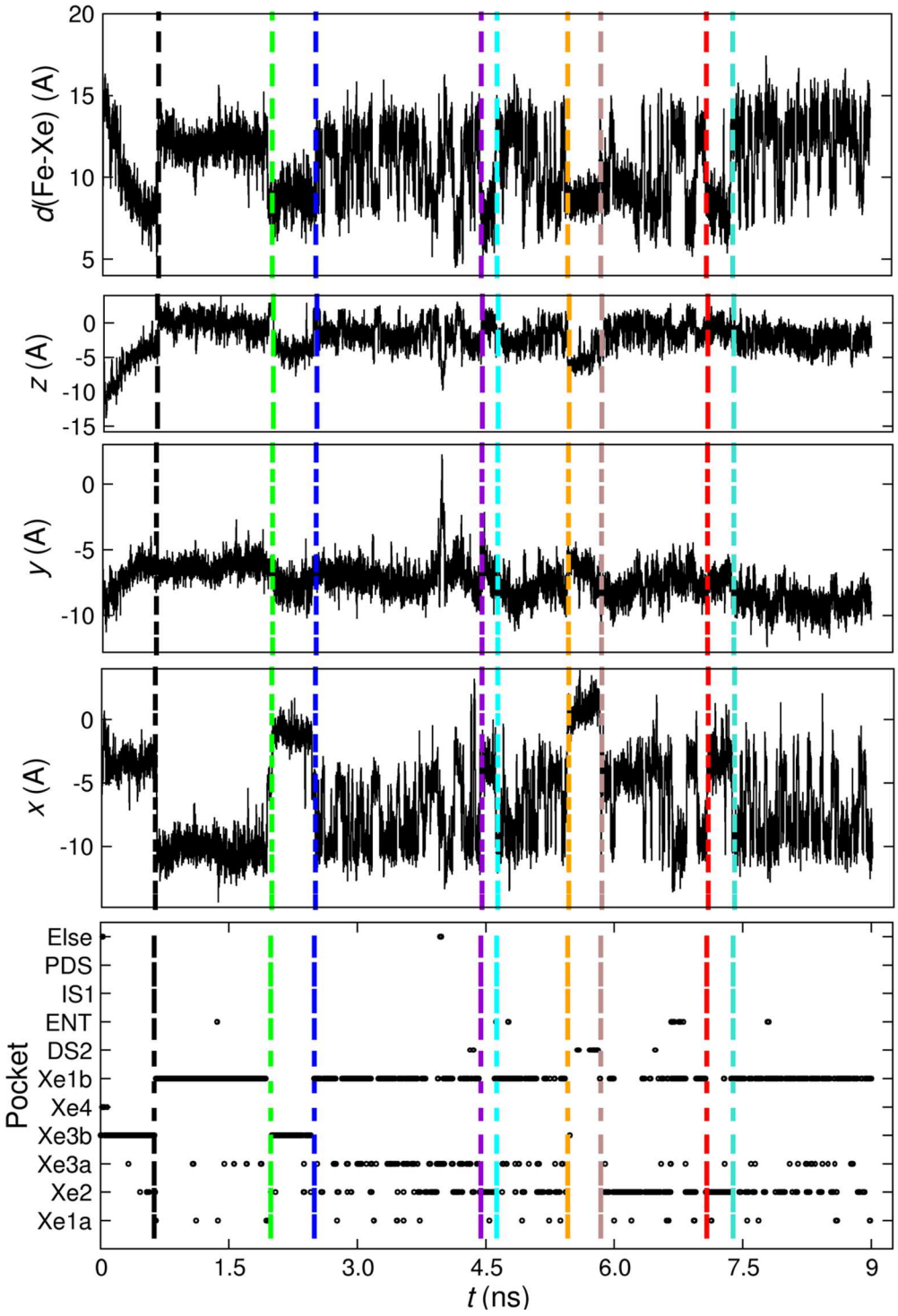

Fig 2. Illustration of transitions in a 9 ns trajectory. Identification of Xe transitions by tracking the evolution of the Fe-Xe distance and *x*−, *y*−, and *z*−distances from origin (Fe atom), over 9 ns of explicit MD. The selected transitions are shown via vertical dashed lines. For clarity, points belonging to transition regions (index 12) are not displayed in the bottom panel of the figure.

https://doi.org/10.1371/journal.pcbi.1005450.g002

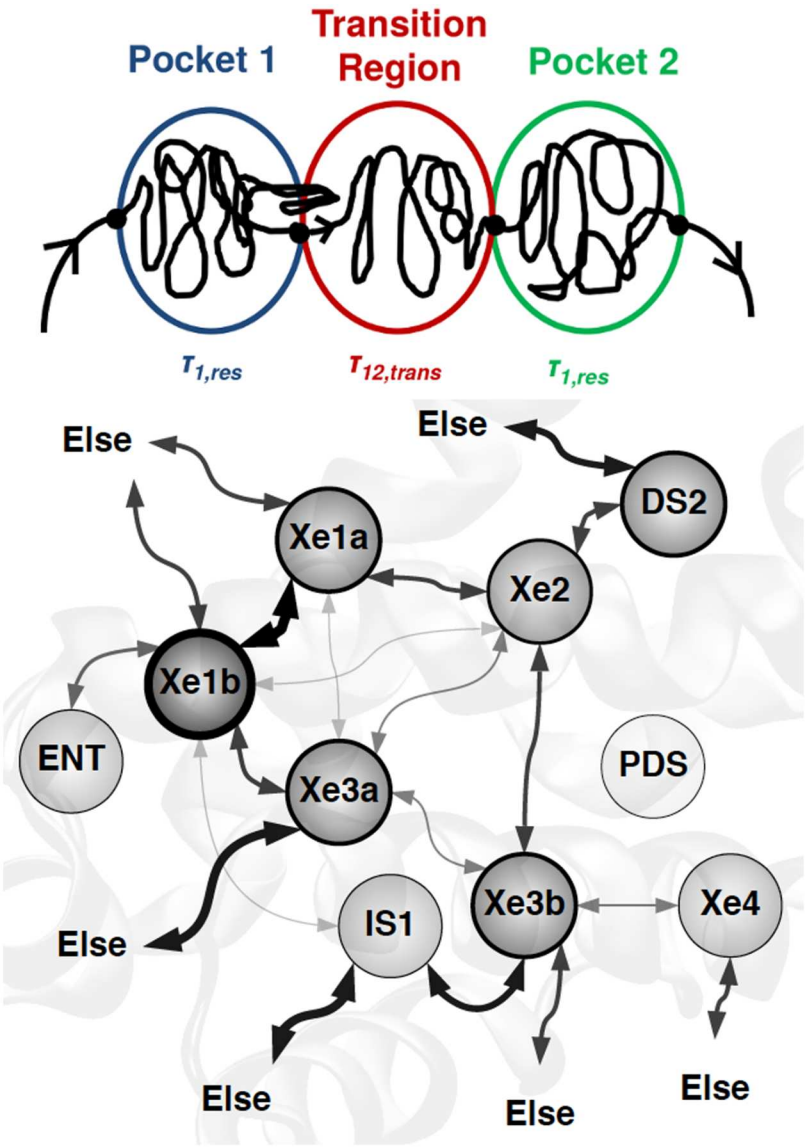

Fig 3. The network of Xe migration in trHbN. The top part of the figure illustrates the definition of residence times $\tau_{1,res}$ and $\tau_{2,res}$ in pockets 1 (blue) and 2 (green) and the definition of transition time $\tau_{12,trans}$ between the two pockets. For the network, the thicker the arrow is, the more transitions are observed between the two pockets it connects. The width of the circular borders represents the occupation probability of the respective pocket.

https://doi.org/10.1371/journal.pcbi.1005450.g003

remains at a given basin, prior to its transition to another basin via the corresponding transition state, and the transition time is defined as the time spent by Xe in a transition region as it moves between different pocket basins, see Fig 3.

Table 2 and Fig 3 highlight the role of the entire Xe3 pocket as a hub in the Xe-migration network in trHbN. Xe3a and Xe3b are connected to all other pockets, except for ENT. It is noted that the transition matrix is nearly symmetric. Furthermore, previous work for O_2 migration in trHbN found barriers ranging from 0.5 kcal/mol to 1.5 kcal/mol between neighboring sites when treating O_2 with a fluctuating charge model. [29] Given that the present simulations were initiated with Xe being in one of the Xe1a/b, Xe2 or Xe3a/b pockets, the

Table 2. Table of transitions.

	Xe1a	Xe2	Xe3a	Xe3b	Xe4	Xe1b	DS2	ENT	IS1	PDS	Else
Xe1a	0	391	61	4	1	1467	0	0	0	0	418
Xe2	396	0	110	383	0	101	348	0	8	1	3
Xe3a	57	103	0	129	1	326	1	1	11	1	1208
Xe3b	5	428	128	0	171	13	6	1	636	1	454
Xe4	0	0	0	175	0	0	1	0	0	0	505
Xe1b	1505	94	316	11	3	0	0	230	35	0	561
DS2	0	356	1	9	1	1	0	0	0	0	998
ENT	0	0	2	2	0	224	0	0	0	0	28
IS1	0	10	11	610	0	28	0	0	0	0	1216
PDS	0	0	1	0	0	0	1	0	0	0	2
Else	417	3	1201	462	505	543	994	28	1188	1	0

The transition matrix between initial (row) and final (column) pocket as obtained from the explicit MD simulations. A total of 19252 transitions was found.

https://doi.org/10.1371/journal.pcbi.1005450.t002

populations of pockets outside this region do in general not correspond to an equilibrium distribution. This is, however, of no concern because the focus here is on ligand diffusion within the (Xe1a/b, Xe2, Xe3a/b) region. The stability of O_2 in these 5 pockets is within 0.5 kcal/mol. [22] Given the total simulation time of 725.5 ns, the low barriers between the pockets and the large number of transitions observed (almost 20000), the Xe-distributions for the relevant Xe1a/b, Xe2 and Xe3a/b states are expected to be unbiased by the initial conditions.

The residence time distributions were analyzed for the core network (Xe1a, Xe1b, Xe2, Xe3a, Xe3b, and DS2), and is shown in Fig 4. For all states, 70% or more of the residence times are shorter than 25 ps, see inset of Fig 4. On the other hand, a small fraction of dwell times is on the 1 ns time scale or longer.

The time scale within which transitions occur was also determined. For later analysis, transition times between two particular pockets A and B were separately determined depending on the pocket C the Xe atom originated from. In other words, for a transition A→B the transition time was separately determined for C1→A→B to C*n*→A→B, where C1 to C*n* runs over all *n* pockets connected to pocket A. A→B transitions within the (Xe1a/b, Xe2, Xe3a/b) region were all found to occur on the 10 ps time scale, independently of where Xe came from before. This analysis for five different transitions is reported in Table 3. For each event, (C→)A→B, the probability was obtained by normalizing with respect to the total number of transitions from A to any other pocket, with Xe having arrived to A from C. For instance, taking Xe2 as state A and Xe1a as C*n*, (Xe1a→Xe2→B), the probabilities are 0.070 for Xe1a→Xe2→Xe1b and 0.720 for Xe1a→Xe2→Xe1a. The remaining probabilities for Xe1a→Xe2→(Xe3a,Xe3b,DS2,IS1) are 0.060, 0.078, 0.067 and 0.005 respectively, which are not shown in Table 3.

Table 3 shows that transition probabilities for A→B→A transitions are considerably more likely to occur than C→B→A transitions. This suggests that the physical process of ligand migration in trHbN involves appreciable dynamical coupling between ligand and protein degrees of freedom. This finding underlines the importance of trHbN fluctuations in Xe migration.

Analysis of the Xe1a↔Xe2 transition

In a next step the dynamical coupling between ligand and protein motion is further analyzed. For this, the Xe1a↔Xe2 transition is considered in more detail. Fig 5 shows pockets Xe1a and

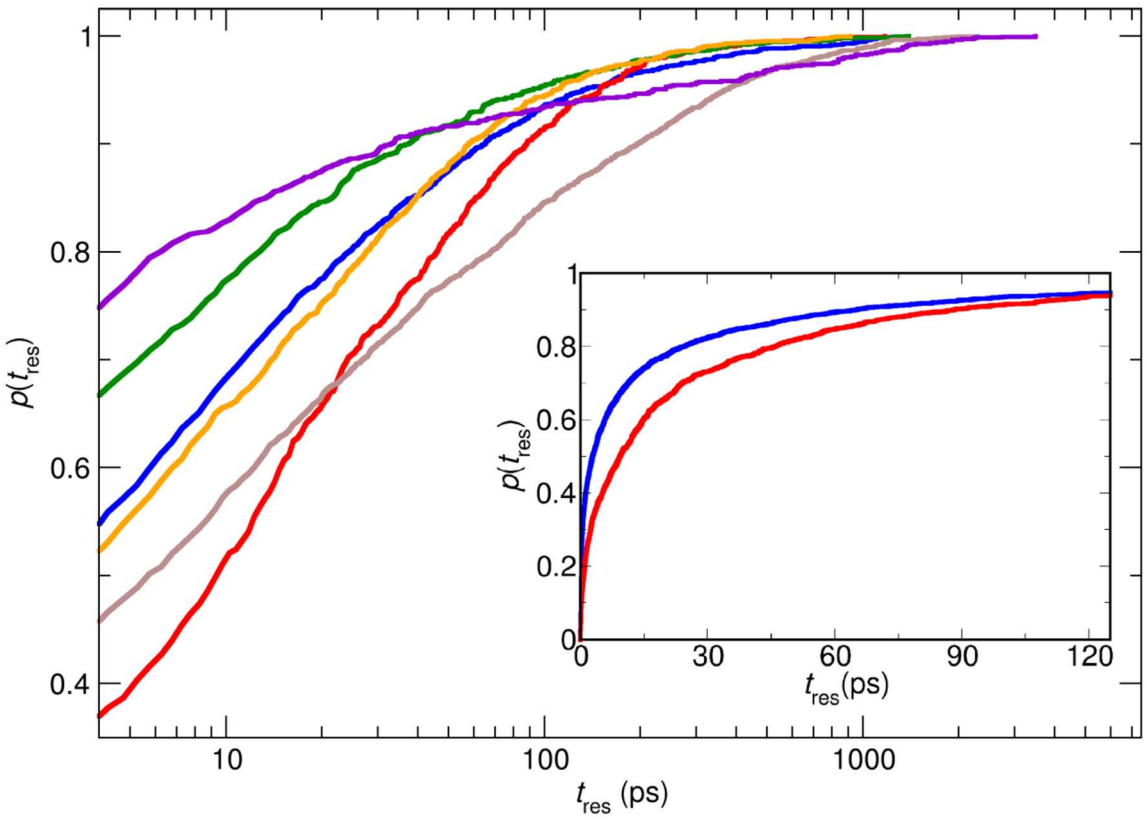

Fig 4. Residence time distributions. Cumulative distributions, $p(t_{res})$ of Xe residence times, t_{res}, in states Xe1a (blue), Xe2 (red), Xe3a (green), Xe3b (orange), Xe1b (brown) and DS2 (magenta) (double-logarithmic representation). While in all cases, up to at least 70% of the residence times are shorter than 25 ps, the logarithmic scale allows for a better illustration of the difference in the statistics of infrequently sampled long residence times for each pocket. The inset reports dwell times using a linear time axis for pockets Xe1a (blue) and Xe2 (red); 5% of the dwell times are longer than 150 ps.

https://doi.org/10.1371/journal.pcbi.1005450.g004

Xe2 along with the five amino acids that according to the MEP are involved in the Xe1a↔Xe2 transition. This transition was selected because a) it is extensively sampled, b) both Xe1a and Xe2 are connected with several other pockets, in particular, Xe1a is connected to Xe2, Xe3a and Xe1b, while Xe2 is linked to Xe1a, Xe3a, Xe3b, Xe1b and DS2, and c) Xe1a↔Xe2 is a representative transition of the (Xe1,Xe2,Xe3) core of trHbN ligand migration network.

The transition and dwell times for the Xe1a↔Xe2 transition depending on from where Xe arrived in either Xe1a or Xe2 are summarized in Tables 2 and 3. The analysis indicates that A→B→A transitions (e.g. Xe2→Xe1a→Xe2) are considerably more probable than C→B→A (e.g. Xe1b→Xe1a→Xe2) (0.698 vs. 0.071). Furthermore, the probability for a particular transition also depends on where Xe originally came from. However, transition times for Xe1a↔Xe2 do not differ significantly regardless of where the ligand came from. In a broader perspective *all* transition times are on the several picosecond time scale (3 to 8 ps) as can be seen in Table 2. Similarly, typical dwell times are 5 to 10 ps for all states, see Fig 4, although a small but potentially interesting fraction resides for up to 3.5 ns in individual states.

Pockets Xe1a and Xe2 are connected through a channel which involves amino acids Phe^{B9}, Phe^{E15}, Val^{B6}, Gln^{E11} and Leu^{G12} (see Fig 5). The conformational space sampled by all residues involved in the channel is analyzed in the following. Only amino acid- and Xe-coordinates during the time Xe spends in the transition region (see Fig 5) are considered and analyzed. For

Table 3. Overview of selected Xe transitions.

C	A→B	Probability	Time (ps)	Counts
Xe1a	Xe2→Xe1b	0.070	6.6	27
Xe3a		0.098	6.8	10
Xe1b		0.564	4.5	53
DS2		0.014	4.8	5
Xe1a	Xe3a→Xe1b	0.138	5.9	8
Xe3a		0.211	2.4	23
Xe1b		0.142	4.8	18
DS2		0.789	4.2	247
Xe1a	Xe3a→Xe2	0.155	4.1	9
Xe3a		0.532	5.0	58
Xe1b		0.063	8.3	8
DS2		0.064	3.6	20
Xe2	Xe1a→Xe2	0.698	6.9	270
Xe3a		0.268	4.2	15
Xe1b		0.071	8.6	104
Xe1a	Xe2→Xe1a	0.720	5.3	278
Xe3a		0.186	8.9	19
Xe3b		0.116	3.2	49
Xe1b		0.191	7.2	18
DS2		0.086	5.9	30

Transition probabilities, average transition times, and number of events observed (counts) for selected transitions, obtained from the explicit MD simulations.

https://doi.org/10.1371/journal.pcbi.1005450.t003

this analysis, the protein backbone atoms, excluding the flexible N-terminus region of trHbN (first 13 amino acids) were reoriented with respect to the crystal structure of trHbN.

The Cα-Cβ-Cγ-Cδ2 dihedral angle distribution $p(\phi)$ of Phe^{E15} is considered first. All distributions are normalized with respect to their individual flux as indicated in Fig 6B and Table 3 and a bin size of 5˚ was employed. The results show that depending on where Xe came from, $p(\phi)$ is different, see Fig 6A, and all distributions differ from the equilibrium distribution (orange).

The equilibrium distribution is characterized by two peaks, at around −50˚ and 150˚ respectively. On the contrary, in distributions corresponding to C→Xe1a→Xe2 transitions, these peaks are shifted. In particular, the relative populations of the two favorable orientations differ whether Xe accesses the Xe1a→Xe2 coming from Xe2 (blue) or Xe3a (red). Also, the maxima of $p(\phi)$ for these two cases differ by up to −100˚. Moreover, besides the two peaks at −100˚ and 100˚, the distribution for Xe2→Xe1a→Xe2 events is characterized by a significant presence of points in the intermediate region, contrary to the distributions of Xe3a/Xe1b→Xe1a→Xe2 transitions. This outcome implies a greater rotational flexibility for the Phe^{E15} ring when Xe originates from Xe2, which would indeed facilitate Xe2→Xe1a→Xe2 transitions.

In addition, we also considered the distribution of Xe distances from the initial pocket (Xe1a pocket in Xe1a→Xe2 transitions, and Xe2 pocket in Xe2→Xe1a transitions). The results for Xe1a→Xe2 transitions are illustrated in Fig 6B, while those for Xe2→Xe1a transitions are depicted in Figure S1B in S1 Text. Depending on the pocket Xe was before the transition, the magnitude and shape of $p(d_{\mathrm{Xe}\rightarrow\mathrm{Xe1a}})$ differs. For Xe3a/Xe1b→Xe1a→Xe2 $p(d)$ extends to

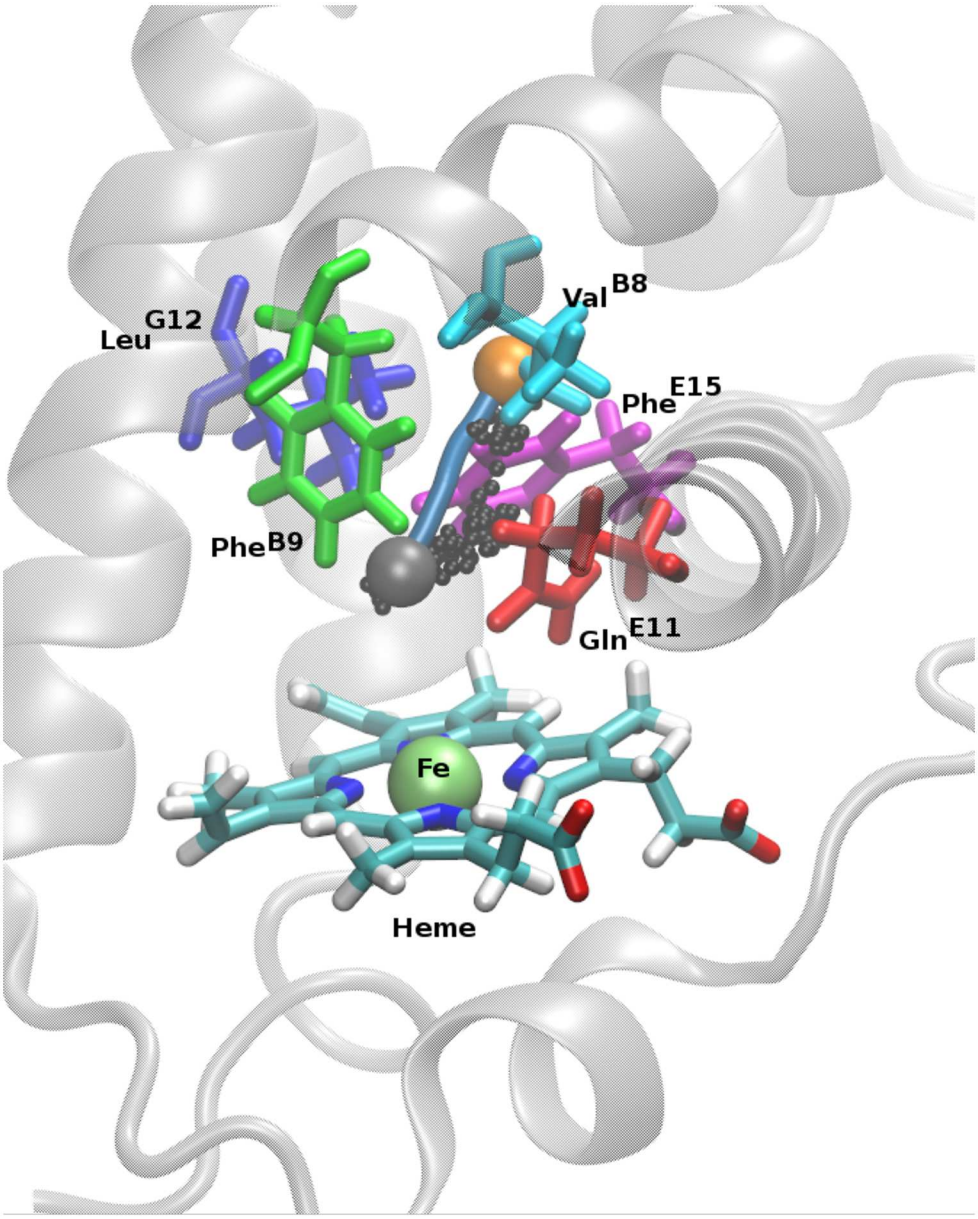

Fig 5. Xe1a↔Xe2 transition region. Xe1a and Xe2 pockets (orange and gray sphere, respectively), together with surrounding residues Val^{B6} (cyan), Phe^{B9} (green), Phe^{E15} (magenta), Gln^{E11} (red), and Leu^{G12}(blue). The minimum energy transition path (in blue cylinder) and one particular transition path (small black spheres) are shown between the two pockets. In grey (background) the ribbon structure of the protein and in licorice the heme-unit.

https://doi.org/10.1371/journal.pcbi.1005450.g005

larger distances between Xe and the center from the initial pocket compared to the other two transitions.

Next, the interplay between the Xe motion and the surrounding amino acids (Phe^{B9}, Phe^{E15}, Val^{B6}, Gln^{E11} and Leu^{G12}) in the transition region was analyzed. For that, a plane orthogonal to the tangent along the minimum energy path for the Xe1a↔Xe2 transition containing the saddle point as the origin was defined, see Fig 7B. All positions of the Xe and amino acid atoms were projected onto this plane and their densities are shown as isocontours

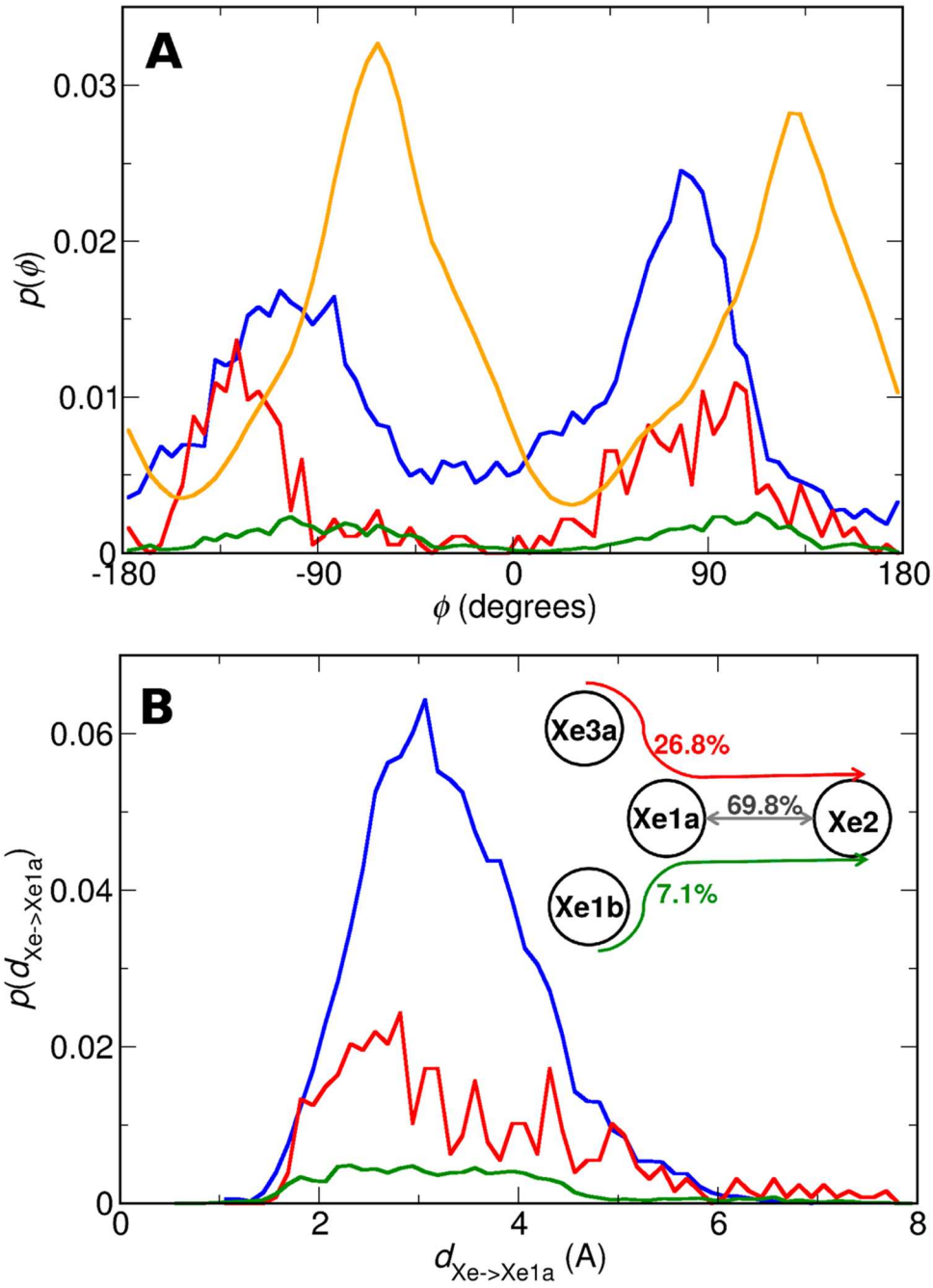

Fig 6. Analysis for C→Xe1a→Xe2 transitions. Distribution of the (A) Cα-Cβ-Cγ-Cδ2 dihedral angle of Phe[61] depending on state C in the transition C→Xe1a→Xe2. The equilibrium distribution (orange) is shown together with $p(\phi)$ for C = Xe2 (blue), C = Xe3a (red), C = Xe1b (green). (B) Xe distance from center of the Xe1a state for the Xe1a→Xe2 transitions, from explicit MD simulations. Depending on where Xe entered the Xe1a→Xe2 transition the distributions differ.

https://doi.org/10.1371/journal.pcbi.1005450.g006

in Fig 7. For clarity, filled contours and non-filled contours are used for Xe and the amino acids, respectively. The analysis finds that depending on state C before the C→Xe1a→Xe2 transition the orientation of the amino acids is different, i.e. the side chain and ligand dynamics are coupled and influence each other.

Fig 7A and 7C establish that for Xe2→Xe1a→Xe2 transitions the side chain of Phe^{E15} is oriented predominantly perpendicular to the transition plane. Contrary to that, for Xe3a2→Xe1a→Xe2 transitions it is oriented parallel to that plane. This can be clearly seen in Fig 7A where individual atom projections of the side chain ring atoms (in color) are shown.

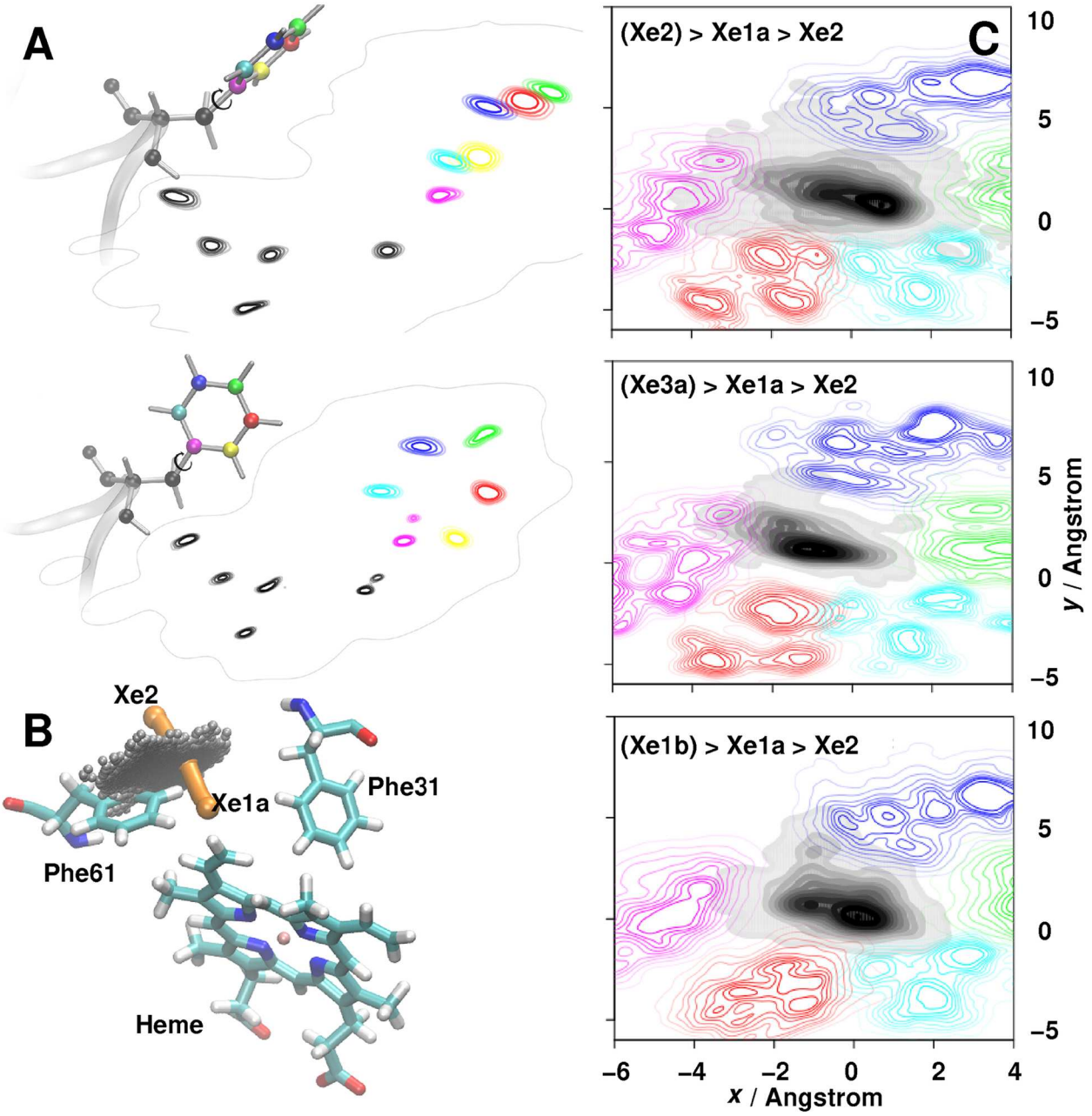

Fig 7. Difference in Phe^{E15} ring orientation and projection of Xe and amino acid atom positions upon C→Xe1a→Xe2 transitions. Panel A: The average orientations of the benzene ring of Phe^{61} for the two transitions (top) Xe2→Xe1a→Xe2 and (bottom) Xe3a→Xe1a→Xe2 together with the projection of each atom onto the transition plane (colored isocontours). Panel B: The projection of Xenon-positions onto the transition plane (grey)—and hence the transition plane itself-together with the Xe2→Xe1a transition (orange). Panel C: Projection of the position of Xe (greyscale) and ValB6 (cyan), Phe^{B9} (green), Gln^{E11} (red), Phe^{E15} (magenta) and Leu^{G12} (blue) atoms onto the plane containing the transition points for the Xe1a→Xe2 transition. Only transition points are included. The integrated density corresponds to 100%. Moving from inner to outer contours, each of them contains an additional 10% of the total number of points. The overall orientations of all side chains are the same throughout the Figure.

https://doi.org/10.1371/journal.pcbi.1005450.g007

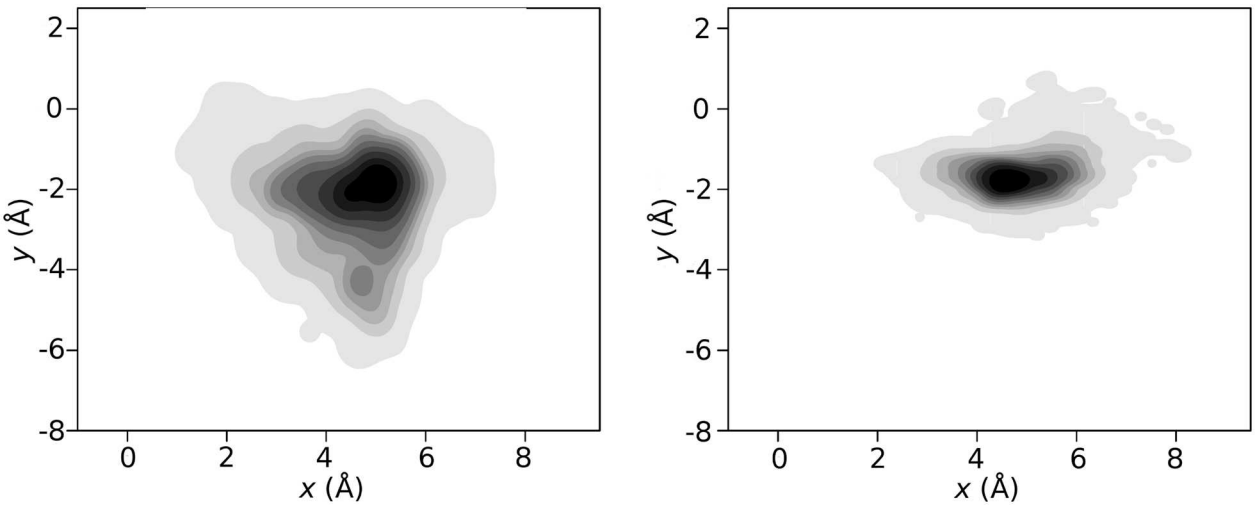

Fig 8. Xe position probability distributions, short vs long residence times. Comparison of probability distributions $P(x, y)$ from many short dwell times (left, 5 ps) with $P(x, y)$ of a simulation with a 1.2 ns dwell time (right) in the Xe1a state. The two distributions differ in shape and size because for short dwell times, Xe samples the available volume with little protein adaptation taking place whereas for long dwell times the protein is able to pack more closely around the Xe atom which typically leads to reduction of the volume of the sampled space.

https://doi.org/10.1371/journal.pcbi.1005450.g008

The probability distribution functions in Fig 7C underline the importance of Phe^{61} in this transition but other residues also show differing orientations depending on the transition considered. It is also noteworthy that the overlap of the Xe distributions differs for the three transitions considered. However, the present illustration may be exaggerated as all snapshots were reoriented with respect to the same protein structure (X-ray structure) which brings out the differences rather than the common features of individual transition paths for the same transition.

Depending on the time spent by the Xe probe in a particular pocket its shape can change in different ways. Fig 8 shows that the probability distribution of Xe positions in Xe1a from many short ($\approx$ 5 ps) dwell times differs considerably from the space sampled during one long (continuous 1.2 ns) occupation trajectory of the same pocket. This suggests that for longer dwell times the pocket has time to adapt its shape whereas for short (picosecond) dwell times the pocket does not have sufficient time to adapt.

Finally, the orientation of the Phe^{E15} side chain as defined by the (χ_1, χ_2) angles is analyzed for Xea1↔Xe2 transitions depending on state C. The equilibrium distribution (see Fig 9A) exhibits pronounced maxima at $\chi_1 = -90^\circ$ and $\chi_2 = (145, -45)^\circ$. Due to the small number of events for certain transitions (see Table 3) the probability distribution functions $p(\chi_1, \chi_2)$ are not converged. Nevertheless, they show rather characteristic structures. For the Xe2→Xe1a→Xe2 transition (270 transitions, well sampled) $p(\chi_1, \chi_2)$ exhibits a clear shift of the maximum occupation to $(\chi_1 = -145^\circ, \chi_2 = 145^\circ)$. Hence, no new state is found compared to the equilibrium distribution, but the occupation of the states changes. Conversely, for Xe1b→Xe1a→Xe2 transition (104 transitions, well sampled) the distribution is more reminiscent of the equilibrium distribution. Finally, for Xe3a→Xe1a→Xe2 (15 transitions, undersampled) the majority of the distribution is along $\chi_1 = -145^\circ$ which differs from the equilibrium distribution.

For the reverse transition Xe2→Xe1a similar observations can be made. The most sampled transition is Xe1a→Xe2→Xe1a (278 transitions) for which $p(\chi_1, \chi_2)$ confirms the finding for the Xe2→Xe1a→Xe2 transition. All other transitions are sampled less than 50 times and

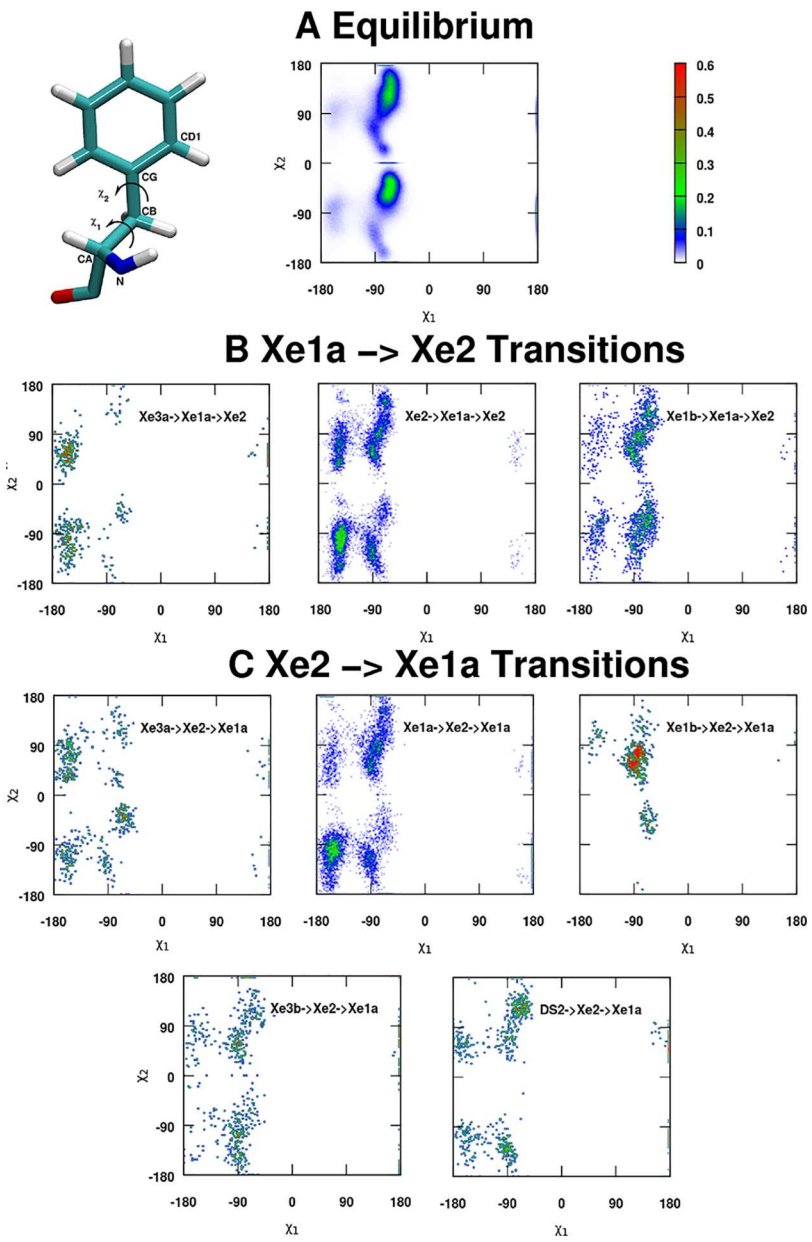

Fig 9. Dihedral probability distributions for Phe^{E15}. 2-dimensional probability distributions $p(\chi_1, \chi_2)$ of χ_1 and χ_2 dihedral angles upon Xe1a→Xe2 transitions, compared to the equilibrium distribution. A single Phe is reported in the left upper corner, along with the definition of angles χ_1 and χ_2. $p(\chi_1, \chi_2)$ in panels B and C are different from the equilibrium distribution. Distributions also differ depending on where Xe came from before the transition. Characteristic examples are Xe2→Xe1a→Xe2 vs Xe1b→Xe1a→Xe2 events from panel B and Xe1a→Xe2→Xe1a vs Xe1b→Xe2→Xe1a events from panel C.

https://doi.org/10.1371/journal.pcbi.1005450.g009

hence not particularly well converged. It is found that for the Xe3a→Xe2→Xe1a, the distribution resembles that for Xe3a→Xe1a→Xe2, whereas the remaining transitions are more similar to the equilibrium distribution. Hence, overall it is found that for the best sampled transitions the distribution functions $p(\chi_1, \chi_2)$ differ in a characteristic fashion from the equilibrium distribution, whereas for other transitions the distributions can be more similar to p_{eq}.

These observations establish that depending on from where Xe enters the Xe1a↔Xe2 transition region the channel-forming amino acids are arranged differently. Ultimately, this strongly supports the notion that protein and ligand motion are coupled for Xe migration through trHbN. Here, this was explicitly demonstrated for amino acids involved in the Xe1a↔Xe2 transition, and especially for the Phe^{E15} side chain. It is known from recent work [22, 24, 43–45] that Markov State Models (MSMs) containing protein degrees of freedom can be constructed. However, the purpose of the present work is not to construct a MSM, but to characterize the nature of Xe migration itself.

Discussion

The present work suggests that for Xe migration through the network formed by the Xe-pockets (which are internal packing defects) in trHbN, the protein and ligand degrees of freedom are coupled. The analysis of the MD simulations demonstrated that even the diffusion of the weakly interacting Xe atom is coupled to the continuous structural rearrangement of trHbN as the ligand migrates through it. In fact, trHbN directs Xe migration to the extent that the probability of observing a specific transition (between states A and B) at a given time depends on where the ligand was before (state C). This was explicitly illustrated for Xe1a↔Xe2 events, where the amino acids composing the Xe1a↔Xe2 transition channel adopted different conformations depending on where Xe arrived from before the transition took place. The explicit MD simulations imply the presence of protein-induced memory effects through coupling between ligand and protein degrees of freedom, which greatly affect the migration of Xe. In other words, for a given transition, A→B, depending on where Xe comes to A from, trHbN rearranges differently.

The observation of memory effects raises the question whether, and if so, on what time scale ligand migration is a Markovian process. Based on time scale considerations [46, 47] a process is expected to be Markovian when one of its major sub-processes occurs on a considerably longer time scale with respect to other major sub-processes. For ligand migration, this would translate in transition times and intra-pocket relaxation occurring on different time scales. This should be possible when the free energy barriers separating the different pockets of the network are sufficiently high to ensure dwell times that are significantly longer than transition times, which is not the case in the present work.

While the assumption of Markovianity is commonly employed when describing chemical and biological processes, it has been known from several cases that it is not always valid [46, 48]. Characteristic examples of processes exhibiting non-Markovian behaviour are enzymatic reactions [49, 50] or the glass transition in polymers [51]. The true nature of the dynamics of such processes remains hidden unless appropriate reaction coordinates [46] or analysis techniques [48] are used for their investigation.

Ultimately, the preferable approach to probe the Markov-assumption is to address the problem in terms of probabilities. A ligand migration network is only Markovian if the Markov property holds for the transition probabilities between two pockets, A and B: $P_{(C\rightarrow)A\rightarrow B} = P_{(D\rightarrow)A\rightarrow B}$. In this equation, C and D correspond to any two different pockets of the migration network from which the ligand can arrive to A. From a physical point of view, this equation states that the probability of observing A→B transitions is the same no matter from where the ligand arrived to A. That is, memory effects do not impact ligand diffusion through the protein.

However, as repeatedly shown in the present work, this is not found for Xe migration in trHbN. Rather, depending on the origin of the ligand (state C), the probability for the transition and the ensuing dynamics and conformations of the protein differ. The comparison of

transition probabilities revealed that back-and forth events (B→A→B) are far more common than transitions C→A→B. Hence Xe migration in trHbN belongs to an expanding group of chemical and biological processes [46, 49–51] for which such an unbiased description reveals that the Markovianity assumption is not valid. The present study finds that Xe migration in trHbN is non-Markovian and occurs on a time scale of 10 to 20 ps. This time scale is identical to the one determined on a theoretical investigation of enantioselective reactions [52]. Moreover, the present work provides solid evidence that memory effects are operative in protein motion [53]. More precisely, for a specific event (transition), this translates to a different rearrangement of the amino acids involved in the transition region, depending on where Xe came from before the transition.

A question that arises is whether the present findings also apply to the migration of trHbN's natural ligands, O_2 and NO. Compared to Xe, both of these ligands have a different van der Waals shape, as well as the potential of interacting with the side chains of the amino acids that are involved in trHbN's ligand migration network. Earlier computational investigations of O_2, NO, CO and Xe migration in myoglobin [13] demonstrated that all four ligands localize in the same pockets and migrate along the same channels, but the free energy barriers separating these pockets are different for each ligand. For trHbN, Xe, NO and O_2 migrate through the same pockets and transition channels, and the free energy barriers separating the pockets in NO and O_2 migration [20, 21, 25–27, 29] (1-2 kcal/mol) are on the same order of magnitude as those for Xe migration determined in the present work. This suggests that the nature of Xe migration in trHbN is similar for NO and O_2 migration in trHbN. Hence, it is expected that for the physiologically relevant NO and O_2 ligands, migration is also non-Markovian but further investigation of this point is warranted to substantiate this.

Finally, it is also of interest to briefly touch on the question whether and how such effects could be observed experimentally. Potential experiments for the problem at hand require a spatial resolution of a few Å and a time resolution on the picosecond time scale. In addition, it will be advantageous to control the number of possible migration pathways through suitable point mutations as has already been successfully done for Myoglobin. [54] Time-resolved Laue diffraction allows characterization of structural changes on the sub-Å scale with a time resolution of 100 ps which is, however, probably too slow for the present purposes. [55] On the other hand, using different ligands (NO, CO) the dynamics may be slowed down due to increased barriers for ligand migration. X-ray free electron lasers (XFELs) offer new promising avenues to characterize the short-time dynamics of biological systems. It is expected that time-resolved serial femtosecond crystallography (SFX) and wide angle X-ray scattering (WAXS) at XFELs allow to investigate ultrafast protein structural dynamics on the femtosecond to picosecond time-scale. [56, 57] Alternatively, femtosecond X-ray solution scattering has been used to characterize the ultrafast increase of the radius of gyration of Mb on the 1 ps time scale. Such studies suggest that with ultrafast laser pulses the intrinsic motions in proteins can be characterized in the nonequilibrium (protein quake) regime. [58]

In summary, the migration of Xe in trHbN of Mycobacterium Tuberculosis was investigated using classical MD simulations. The analysis shows that Xe migration in trHbN occurs on the 10 to 20 ps time scale and is non-Markovian. In addition, the results demonstrate that memory effects are in operation during Xe motion through the internal protein pockets. For a specific transition, this translates into different rearrangements of the amino acids lining the transition region, depending on where Xe came from before. Whether or not this observation is physiologically relevant requires simulations with chemically active ligands such as NO or O_2.

Supporting information

S1 Text. In this document, additional analysis of the simulations is reported.
(PDF)

Author Contributions

Conceptualization: PD OTU MM.

Data curation: PD OTU MM.

Formal analysis: PD OTU.

Funding acquisition: MM.

Investigation: PD OTU.

Methodology: PD OTU MM.

Project administration: MM.

Resources: MM.

Software: OTU.

Supervision: MM.

Validation: PD OTU MM.

Visualization: PD OTU MM.

Writing – original draft: PD OTU MM.

Writing – review & editing: PD OTU MM.

References

1. Karplus M, McCammon J. Molecular dynamics simulations of biomolecules. Nat Struct Biol. 2002; 9 (9):646–652. https://doi.org/10.1038/nsb0902-646 PMID: 12198485
2. Kamberaj H, van der Vaart A. Extracting the Causality of Correlated Motions from Molecular Dynamics Simulations. Biophys J. 2009; 97(6):1747–1755. https://doi.org/10.1016/j.bpj.2009.07.019 PMID: 19751680
3. Cui Q, Karplus M. Allostery and cooperativity revisited. Prot Sci. 2008; 17(8):1295–1307. https://doi.org/10.1110/ps.03259908
4. Suel GM, Lockless SW, Wall MA, Ranganathan R. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. Nat Struct Biol. 2003; 10(1):59–69. https://doi.org/10.1038/nsb881 PMID: 12483203
5. del Sol A, Arauzo-Bravo MJ, Amoros D, Nussinov R. Modular architecture of protein structures and allosteric communications: potential implications for signaling proteins and regulatory linkages. Gen Biol. 2007; 8(5):1–11. https://doi.org/10.1186/gb-2007-8-5-r92
6. Martin C, Berridge G, Higgins CF, Mistry P, Charlton P, Callaghan R. Communication between multiple drug binding sites on P-glycoprotein. Mol Pharmar. 2000; 58(3):624–632.
7. Kalescky R, Zhou H, Liu J, Tao P. Rigid Residue Scan Simulations Systematically Reveal Residue Entropic Roles in Protein Allostery. PLoS Comp Biol. 2016; 12(4). https://doi.org/10.1371/journal.pcbi.1004893
8. Nienhaus K, Knapp JE, Palladino P, W E Royer J, Nienhaus U. Ligand Migration and Binding in the Dimeric Hemoglobin of Scapharca inaequivalvis. Biochemistry. 2007; 46(49):14018–14031. https://doi.org/10.1021/bi7016798 PMID: 18001141
9. Sharp K, Skinner JJ. Pump-probe molecular dynamics as a tool for studying protein motion and long range coupling. Proteins-Struct Funct Bioinf. 2006; 65(2):347–361. https://doi.org/10.1002/prot.21146

10. Miyashita O, Onuchic JN, Wolynes PG. Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins. Proc Natl Acad Sci USA. 2003; 100(22):12570–12575. https://doi.org/10.1073/pnas.2135471100 PMID: 14566052

11. Buchli B, Waldauer SA, Walser R, Donten ML, Pfister R, Bloechliger N, et al. Kinetic response of a photoperturbed allosteric protein. Proc Natl Acad Sci USA. 2013; 110(29):11725–11730. https://doi.org/10.1073/pnas.1306323110 PMID: 23818626

12. Tilton RF, Kuntz ID, Petsko GA. Cavities in Proteins—Structure of a Metmyoglobin-Xenon Complex Solved to 1.9-A. Biochemistry. 1984; 23(13):2849–2857. https://doi.org/10.1021/bi00308a002 PMID: 6466620

13. Cohen J, Arkhipov A, Braun R, Schulten K. Imaging the migration pathways for O-2, CO, NO, and Xe inside myoglobin. Biophys J. 2006; 91(5):1844–1857. https://doi.org/10.1529/biophysj.106.085746 PMID: 16751246

14. Nienhaus K, Nienhaus GU. Comparison of ligand migration and binding in heme proteins of the globin family. Chin Phys B. 2015; 24(12):article no: 128705. https://doi.org/10.1088/1674-1056/24/12/128705

15. Luna VM, Chen Y, Fee JA, Stout CD. Crystallographic Studies of Xe and Kr Binding within the Large Internal Cavity of Cytochrome ba3 from Thermus thermophilus: Structural Analysis and Role of Oxygen Transport Channels in the Hemeâ??Cu Oxidases,. Biochemistry. 2008; 47(16):4657–4665. https://doi.org/10.1021/bi800045y PMID: 18376849

16. Luna VM, Fee JA, Deniz AA, Stout CD. Mobility of Xe Atoms within the Oxygen Diffusion Channel of Cytochrome ba3 Oxidase. Biochemistry. 2012; 51(23):4669–4676. https://doi.org/10.1021/bi3003988 PMID: 22607023

17. Milani M, Pesce A, Ouellet Y, Ascenzi P, Guertin M, Bolognesi M. Mycobacterium tuberculosis hemoglobin N displays a protein tunnel suited for O-2 diffusion to the heme. EMBO Journal. 2001; 20 (15):3902–3909. https://doi.org/10.1093/emboj/20.15.3902 PMID: 11483493

18. Milani M, Pesce A, Ouellet Y, Dewilde S, Friedman J, Ascenzi P, et al. Heme-ligand tunneling in group I truncated hemoglobins. J Biol Chem. 2004; 279(20):21520–21525. https://doi.org/10.1074/jbc.M401320200 PMID: 15016811

19. Heroux MS, Mohan AD, Olsen KW. Ligand Migration in the Truncated Hemoglobin of Mycobacterium tuberculosis. IUBMB Life. 2011; 63(3, SI):214–220. https://doi.org/10.1002/iub.438 PMID: 21445853

20. Bidon-Chanal A, Marti MA, Estrin DA, Luque FJ. Exploring the Nitric Oxide Detoxification Mechanism of Mycobacterium tuberculosis Truncated Haemoglobin N. In: Russo N and Antonchenko, V Y and Kryachko, E S, editor. Self-Organization of Molecular Systems: From Molecules and Clusters to Nanotubes and Proteins. NATO Science for Peace and Security Series A-Chemistry and Biology. NATO; 2009. p. 33–47.

21. Cazade PA, Zheng W, Prada-Gracia D, Berezovska G, Rao F, Clementi C, et al. A comparative analysis of clustering algorithms: O-2 migration in truncated hemoglobin I from transition networks. J Chem Phys. 2015; 142(2):article no: 025103. https://doi.org/10.1063/1.4904431

22. Cazade PA, Berezovska G, Meuwly M. Coupled protein-ligand dynamics in truncated hemoglobin N from atomistic simulations and transition networks. Biochim Biophys Acta-General Subjects. 2015; 1850(5, SI):996–1005. https://doi.org/10.1016/j.bbagen.2014.09.008

23. Plattner N, Meuwly M. Quantifying the Importance of Protein Conformation on Ligand Migration in Myoglobin. Biophys J. 2012; 102(2):333–341. https://doi.org/10.1016/j.bpj.2011.10.058 PMID: 22339870

24. Plattner N, Noe F. Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models. Nature Commun. 2015; 6:article no. 7653. https://doi.org/10.1038/ncomms8653

25. Bidon-Chanal A, Marti MA, Crespo A, Milani M, Orozco M, Bolognesi M, et al. Ligand-induced dynamical regulation of NO conversion in Mycobacterium tuberculosis truncated hemoglobin-N. Proteins-Struct Funct Bioinf. 2006; 64(2):457–464. https://doi.org/10.1002/prot.21004

26. Boron I, Bustamante JP, Davidge KS, Singh S, Bowman LAH, Tinajero-Trejo M, et al. Ligand uptake in Mycobacterium tuberculosis truncated hemoglobins is controlled by both internal tunnels and active site water molecules. F1000Research. 2015; 4(22). https://doi.org/10.12688/f1000research.5921.1 PMID: 26478812

27. 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 PMID: 19289037

28. 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 PMID: 21156139

29. Cazade PA, Meuwly M. Oxygen Migration Pathways in NO-bound Truncated Hemoglobin. ChemPhysChem. 2012; 13(18):4276–4286. https://doi.org/10.1002/cphc.201200608 PMID: 23161831

30. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM—A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. J Comput Chem. 1983; 4 (2):187–217. https://doi.org/10.1002/jcc.540040211

31. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B. 1998; 102 (18):3586–3616. https://doi.org/10.1021/jp973084f PMID: 24889800

32. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of Simple Potential Functions for Simulating Liquid Water. J Chem Phys. 1983; 79(2):926–935. https://doi.org/10.1063/1.445869

33. VanGunsteren WF, Berendsen HJC. Algorithms for Macromolecular Dynamics and Constraint Dynamics. Mol Phys. 1977; 34(5):1311–1327. https://doi.org/10.1080/00268977700102571

34. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical-Integration of Cartesian Equations of Motion of a System With Constraints—Molecular-Dynamics of N-Alkanes. J Comput Phys. 1977; 23(3):327–341. https://doi.org/10.1016/0021-9991(77)90098-5

35. Darden T, York D, Pedersen L. Particle Mesh Ewald—An N.log(N) Method for Ewald Sums In Large Systems. J Chem Phys. 1993; 98(12):10089–10092. https://doi.org/10.1063/1.464397

36. Toukmaji AY, Board JA. Ewald summation techniques in perspective: A survey. Comp Phys Commun. 1996; 95(2–3):73–92. https://doi.org/10.1016/0010-4655(96)00016-1

37. Abramowitz M, Stegun IA. Handbook of mathematical functions with formulas, graphs, and mathematical tables. Dover, New York; 1970.

38. Hartigan JA, Wong MA. Algorithm AS 136: A K-Means Clustering Algorithm. J Royal Stat Soc Series C (Applied Statistics). 1979; 28(1):100–108.

39. Qin AK, Suganthan PN. Robust growing neural gas algorithm with application in cluster analysis. Neur Net. 2004; 17(8–9):1135–1148. https://doi.org/10.1016/S0893-6080(04)00166-2

40. Rissanen J. Modeling by shortest data description. Automatica. 1978; 14(5):465–471. https://doi.org/10.1016/0005-1098(78)90005-5

41. Fischer S, Karplus M. Comjugate Peak Refinement—An Algorithm for Finding Reaction Paths and Accurate Transition-States in Systems with Many Degrees of Freedom. Chem Phys Lett. 1992; 194 (3):252–261. https://doi.org/10.1016/0009-2614(92)85543-J

42. Banushkina P, Meuwly M. Diffusive dynamics on multidimensional rough free energy surfaces. J Chem Phys. 2007; 127(13). https://doi.org/10.1063/1.2775444 PMID: 17919054

43. Silva DA, Bowman GR, Sosa-Peinado A, Huang X. A Role for Both Conformational Selection and Induced Fit in Ligand Binding by the LAO Protein. PLoS Comp Biol. 2011; 7(5):article no. e1002054. https://doi.org/10.1371/journal.pcbi.1002054

44. Noe F, Wu H, Prinz JH, Plattner N. Projected and hidden Markov models for calculating kinetics and metastable states of complex molecules. J Chem Phys. 2013; 139(18):article no. 184114. https://doi.org/10.1063/1.4828816

45. Gu S, Silva DA, Meng L, Yue A, Huang X. Quantitatively Characterizing the Ligand Binding Mechanisms of Choline Binding Protein Using Markov State Model Analysis. PLoS Comp Biol. 2014; 10(8): article no. e1003767. https://doi.org/10.1371/journal.pcbi.1003767

46. Borek JA, Perakis F, Hamm P. Testing for memory-free spectroscopic coordinates by 3D IR exchange spectroscopy. Proc Natl Acad Sci USA. 2014; 111(29):10462–10467. https://doi.org/10.1073/pnas.1406967111 PMID: 25002483

47. Hou Z, Filar JA, Chen A. Markov Processes and Controlled Markov Chains. Berlin, Heidelberg: Springer (Kluwer Academic Publishers); 2002.

48. Chiarugi D, Falaschi M, Hermith D, Olarte C, Torella L. Modelling non-Markovian dynamics in biochemical reactions. BMC Sys Biol. 2015; 9(3).

49. Min W, English BP, Luo GB, Cherayil BJ, Kou SC, Xie XS. Fluctuating enzymes: Lessons from single-molecule studies. Acc Chem Res. 2005; 38(12):923–931. https://doi.org/10.1021/ar040133f PMID: 16359164

50. English BP, Min W, van Oijen AM, Lee KT, Luo GB, Sun HY, et al. Ever-fluctuating single enzyme molecules: Michaelis-Menten equation revisited. Nature Chem Biol. 2006; 2(2):87–94. https://doi.org/10.1038/nchembio759

51. Schmidt-Rohr K, Spiess HW. Nature of Nonexponential Loss of Correlation above the Glass-Transition Investigated by Multidimensional NMR. Phys Rev Lett. 1991; 66(23):3020–3023. https://doi.org/10.1103/PhysRevLett.66.3020 PMID: 10043678

52. Carpenter BK, Harvey JN, Glowacki DR. Prediction of enhanced solvent-induced enantioselectivity for a ring opening with a bifurcating reaction path. Phys Chem Chem Phys. 2015; 17(13):8372–8381. https://doi.org/10.1039/C4CP05078A PMID: 25521804

53. Zheng W, Best RB. Reduction of All-Atom Protein Folding Dynamics to One-Dimensional Diffusion. J Phus Chem B. 2015; 119(49):15247–15255. https://doi.org/10.1021/acs.jpcb.5b09741

54. Ostermann A, Waschipky R, Parak FG, Nienhaus GU. Nature. 2000; 404:205. https://doi.org/10.1038/35004622 PMID: 10724176

55. Bourgeois D, Schotte F, Brunori M, Vallone B. Time-resolved methods in biophysics. 6. Time-resolved Laue crystallography as a tool to investigate photo-activated protein dynamics. Photochem Photobiol Sci. 2007; 6(10):1047–1056. https://doi.org/10.1039/b704249c PMID: 17914477

56. Arnlund D, Johansson LC, Wickstrand C, Barty A, Williams GJ, Malmerberg E, et al. Visualizing a protein quake with time-resolved X-ray scattering at a free-electron laser. Nat Meth. 2014; 11(9):923–926. https://doi.org/10.1038/nmeth.3067

57. Neutze R. Opportunities and challenges for time-resolved studies of protein structural dynamics at X-ray free-electron lasers. Phil Trans Royal Soc B-Biol Sci. 2014; 369 (1647). https://doi.org/10.1098/rstb.2013.0318

58. Levantino M, Schiro G, Lemke HT, Cottone G, Glownia JM, Zhu D, et al. Ultrafast myoglobin structural dynamics observed with an X-ray free-electron laser. Nat Comm. 2015; 6. https://doi.org/10.1038/ncomms7772