

Systematic computational analysis of myoglobins conformational dynamics and allosteric mechanism

Ali Razzak and Markus Meuwly*

*Department of Chemistry, University of Basel, Klingelbergstrasse 80 , CH-4056 Basel,
Switzerland.*

E-mail: m.meuwly@unibas.ch

June 17, 2020

Abstract

Proteins possess an intricate intramolecular network that can be modulated through ligand interaction and alter the function of the protein. Myoglobin (Mb) presents a model system for investigating the effects of ligands on protein allostery. Experimental data together with theoretical studies have resolved that ligand interaction perturbs the conformational behaviour and energetic barriers (ΔG) for ligand migration within Mb. This research explores whether xenon (Xe) occupation alters Xe migration, and if so, identifies which regions facilitate Mb allostery. We show how different Xe configurations influence the ΔG for a specific transition. We also locate how structural elements, such as the intramolecular cavity size and residue sidechain geometry respond to various Xe occupancy and migration events. Finally, we demonstrate how a thorough analytical process can resolve residues critical in the allosteric mechanism of Mb. We find that Xe presence and migration induces specific changes in the structural dynamics and intramolecular behaviour of Mb, altering the proteins function.

Introduction

The research of globin family proteins has been instrumental in understanding the structure and dynamics of proteins.^{1–4} Biochemical and computational investigations of globin proteins have revealed complex interactions between proteins and ligands.^{5,6} Computational techniques can identify how ligands migrate through globin proteins and elucidate their physiological function.^{7–9} Structural studies of globin proteins have demonstrated that globin proteins possess an intricate network of channels which manage ligand movement.^{10,11}

Of interest are the molecular dynamics of globin proteins and how they are perturbed in response to ligands.^{3,12} It has been shown that the dynamics and ligand migration routes of globin proteins are sensitive to whether CO, NO, or O₂ are bound.¹³ Furthermore, that globin

protein behaviour and ligand control is impacted by mutations of specific residues.^{2,14,15} Mb research has indicated that these proteins possess an array of cavities which are instrumental in its allosteric mechanism.^{7,16} When ligands interact with globin molecules the protein responds by shifting its conformational behaviour in a manner specific to the ligand to facilitate ligand transport.^{17,18} Mb proteins exhibit a myriad of different intramolecular matrices which ligands diffuse through, thereby a mapping of protein allosteric networks requires thorough investigation.^{13,19}

Through the application of computational molecular dynamic techniques an understanding of the allosteric mechanism of Mb can be identified.^{17,18,20} Analysis of the thermodynamic and chemical properties of Mb as atoms diffuse through the proteins intramolecular environment can uncover the discrete structural behaviour which influences protein allostery.^{19,21,22} It has been shown that computationally studying the interplay between ligand migration and a proteins conformational dynamics can yield valuable allosteric insight.^{23–25} This is especially true in the case of Mb where seminal work on i) O₂ migration,^{26,27} ii) various ligand migration pathways,¹³ and iii) conformational response to NO^{22,28} has been conducted. The aforementioned research has been fundamental in resolving allosteric questions concerning proteins, yet the specifics of the long range electrostatic forces involved are unclear. An inquiry into an inert atom and protein interaction would supplement science relating to allostery, endows a fundamental knowledge of protein and ligand allostery along with postulations of the physiological role of Mb.^{29–31}

The research outlined below investigates Mb allostery in response to presence and migration of Xe.^{5,13,27,32} Mb is among the most extensively studied proteins concerning questions of protein allostery and conformational behaviour.^{33–35} The intramolecular matrix which constitutes Mb is well characterised^{7,36–38} however the migration routes, and how they are modulated are less definitively mapped.^{2,39} What has inhibited this research is that migra-

tion with the diatomic ligands which Mb physiologically binds to rarely exhibit transitions events.^{40,41} In light of this Xe presents a suitable mimic ligand candidate because it is a single atom with a comparable properties to physiological ligands like O₂,²⁵ and theoretical studies suggest it shares migration pathways in heme proteins with physiological ligands.^{13,25,42} Xe resides in its given pocket for a short duration before it rapidly migrates to other pockets.⁴⁰ This makes it difficult to carefully examine Mb dynamics as Xe migrates using conventional free molecular dynamics (MD). However, by retaining the dynamics of Xe at positions across channels which connect pockets within Mb the thermodynamic and kinetic behaviour along Xe migration routes can be carefully interrogated.³⁷ This yields theoretical calculations of the ΔG of Mb at points along the Xe migration route, allowing for the calculation of the activation barrier and the transition state (TS) of Mb as Xe passes between pockets as seen in Figure 1.^{17,27,42,43}

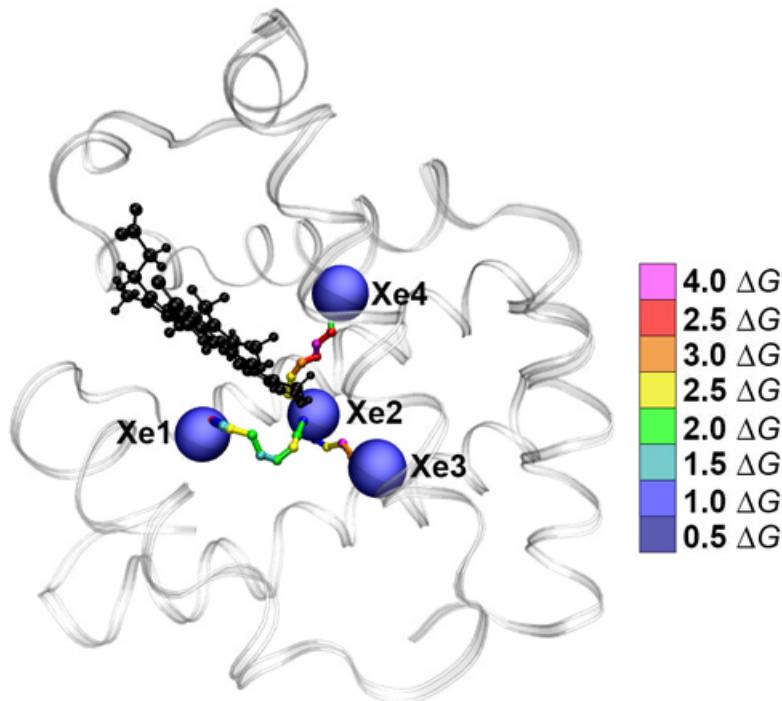


Figure 1: Visualisation of Mb as transparent white, heme group in black licorice, X1, Xe2, Xe3, and Xe4 pockets as annotated blue spheres, and migration routes towards Xe2 highlighted according to the energetic barriers at different points along the migration route for single Xe bound systems.

In this work we computationally investigate transition events at femtosecond (fs) timescales to expose minute kinetic changes that can be coupled with Xe position along the migration pathway, aiding identification of residues critical in the allosteric behaviour of Mb.^{7,41,44} The data from Xe migration free dynamics can also be useful in deriving a free energy profile (FEP) through statistical mechanics.¹⁹ This is achieved by calculating the probability distribution of Xe residing at specific points along the reaction pathway. By investigating Xe effect on the dynamics of Mb we can postulate whether allosteric mechanisms can be triggered exclusively through short range interactions.^{33,38} Additionally, what kind of effects does Xe, an electrostatically inert atom, have on Mb and whether a allosteric network can be conceived from the interaction. Through inspecting the foundations of those conformational shifts the allosteric function, and how allostery relates to the physiological role of Mb can be considered. Together this data would support a detailed understanding of how atoms diffuse through Mb.

Methods

The simulations were performed using the CHARMM⁴⁵ suite with the CHARMM22⁴⁶ all atom force-field and the TIP3P water model.^{47–49} The van der Waals (VDW) radius of Xe was $R_{\min} = 2.25 \text{ \AA}$ which is similar to what was used in previous work^{13,25} The coordinates of the start structure were taken from the X-ray structure of recombinant Sperm whale, PDB code 2W6W (1.99 Å resolution).⁴² The solvent box size has dimensions 65.2 Å x 59.0 Å x 46.6 Å. 16 different Xe occupied systems were setup encompassing all the possible Mb occupation states with regards to myoglobin cavities Xe1, Xe2, Xe3, and Xe4.⁵⁰ The definitions for the cavities were taken from prior characterisations of the Xe pockets as shown in Table 1.^{7,37,51–53}

A binary system is used to identify the Xe occupied systems. For example, 0100 represents

Table 1: Name of each cavity within Mb along with molecules and residues which line the pocket as defined by previous research^{2,17}

Cavity	Molecular definition
Xe1	Leu89, His93, Leu104, Phe138, Ile142, Tyr146 and HEME
Xe2	Leu72, Ile107, Ser108, Leu135, Phe138, Arg139 and HEME
Xe3	Trp7, Thr75, Leu76, Gly80, His82, Ala134, Leu137, Phe138
Xe4	Gly25, Ile28, Leu29, Gly65, Val68, Leu69, Leu72, Ile107, Ile111
Distal pocket (DP)	Leu29, Phe46, Hse64, Val68, Ile107

a Mb system with Xe present in Xe2, while 1011 represents a system where Xe1, Xe3, and Xe4 host a Xe atom.

For all simulations long-range electrostatic interactions were treated with particle mesh Ewald (PME)^{54,55} with a relative tolerance of 10^{-6} , a cutoff of 14 Å, a 12 Å switching cut-off for the Lennard Jones interactions, and the Verlet integrator with a time step of 1 fs was used as the time integrator.^{47,56} The SHAKE algorithm⁵⁷ was used for constraining bonds involving H-atoms. Each system was heated for 100 ps from 0 K to 300 K in increments of 10 K over 100 ps, followed by *NVT* ensemble at 300 K for 200 ps. The unbiased simulations the systems were extended for 1 ns under *NVE*, this was repeated 300 times for each system. Mb rotational and translational movement was restrained and the molecule was fixed to the center of mass of Mb with a harmonic restraint of 1 kcal/mol for every simulation. In total 2400 ns of unbiased simulations were run and analysed.

For the biased Xe migration simulations umbrella sampling (US) was used to calculate the FEP during Xe migration.^{37,52,53} The free energy of Xe along a reaction coordinate between pockets were determined from the US. The umbrella potential has the form

$$U(\delta) = K(\delta - \delta_0)^2 \quad (\text{Eq. 1})$$

where K is the strength of the biasing potential, and δ_0 the center of the umbrella window with reference to the reaction coordinate δ as defined in Table 2. Each umbrella was composed of 51 windows spaced 0.2 Å across the 10 Å reaction coordinate with $K = 50$ kcal/mol. The NVT simulations were carried out at 300 K, with an equilibration time of 300 ps for every window, after which data accumulation was carried out for 250 ps per window. The free energy change along the Xe migration path is the potential of mean force $W(\delta)$ which is defined as

$$W(\delta) = -k_B T \ln[P(\delta)] - U(\delta) + C_i \quad (\text{Eq. 2})$$

where $P(\delta)$ is the equilibrium distribution in the presence of the biasing potential and C_i is a different constant for each window. The Weighted Histogram Analysis Method (WHAM)⁵⁸ was used to determine C_i values. The Xe atoms which were not sampled along the reaction coordinate were restrained to their respective pockets using a harmonic restraint of 1000 kcal/mol/Å. in total 204 ns of US simulations were run and analysed.

Table 2: Description of the Xe migration pathway for each system along with its reaction coordinate definition for US between different pockets. The reaction coordinate was the distance between the center of mass (COM) of the atoms which line the pocket as defined in Table 1.

Pathway	Xe systems	Reaction coordinate (δ)
$\text{Xe2} \rightarrow \text{Xe1}$	0100, 0110, 0101, 0111	COM of Xe2 \leftrightarrow COM of Xe1
$\text{Xe1} \rightarrow \text{Xe2}$	1000, 1010, 1001, 1011	COM of Xe1 \leftrightarrow COM of Xe2

Downhill simulations were produced by selecting windows which were nearby the TS of Xe between two pockets as determined by the US FEP for a specific system. After which, coordinates for seeding unbiased simulations were taken from those windows which initialised 300 20 picoseconds (ps) unbiased NVT simulations for each coordinate. The final position of Xe was used to map how Xe would distribute depending on its initial coordinates.^{22,44,59} A k-means clustering was used to sort the final position of Xe into different events depending

on their position relative to geometrical centers of Xe pockets, and each pocket center was used to initialise a cluster.⁶⁰ The COM of each cavity represented the four main Mb pockets and DP, as described in Table 2, and were used to initialise the clustering. A minimum cutoff of $r_{\min} = 1.5 \text{ \AA}$ was used to define which Xe should be included and to refine the clusters. A maximum cutoff of $r_{\max} = 7 \text{ \AA}$ was used to define which Xe should be excluded from all clusters and were assigned as “Out”.^{21,25,61} The possibility for 11 events were initiated describing Xe1, Xe2, Xe3, Xe4, DP, between Xe2 and Xe1, between Xe2 and Xe3, between Xe2 and Xe4, between Xe4 and DP, and outside the protein.

The SURFNET package⁶² was used for detecting Mb cavities Xe1, Xe2, Xe3, and Xe4 and calculating their volumes. SURFNET searches for gap regions within Mb, once found the gap region was built up by fitting spheres within the gap to form a gap region. This program returns a list of different gap regions each with their associated volume and location of their center if the gap region was within specific dimensions, in our case: ($r_{\min} = 1.2 \text{ \AA}$ and $r_{\max} = 3.5 \text{ \AA}$).^{2,37} The list of gap regions was then parsed to determine which of the gaps was closest to the subject Mb pocket as defined by Table 1⁶³ and suitable for calculating the cavity volume. The average and standard deviation of the volumes for each US window was then reported.

The sidechain dihedral angles were calculated for Leu89, Leu104, Gln105, and Phe138. The dihedral angle χ_1 was measured, the selection was composed the N, C α , C β , and C γ atoms. The average χ_1 for each US window was taken to represent the χ_1 for that window.

The root mean square deviation (RMSD) for Mb at each window across the US systems was calculated as follows

$$RMSD(a, b) = \sqrt{\left(\frac{\sum_{i=1}^N r_i^2}{N}\right)} \quad (\text{Eq. 1})$$

where N are atoms in Mb, r_i is the distance between the absolute coordinates of Mb atoms

i in sets a and in b , taken from the computational simulation and X-ray structure (*PDB: 2W6W*).⁴² The domain RMSD (dRMSD) for each of the eight Mb α helical domains was then calculated where N are all atoms in a given domain (as defined in Table 3), r_i is the distance between the Mb domain i in sets a from computational simulations and X-ray structure (*PDB: 2W6W*).⁴² $RMSD(a, b)$ was then averaged for each US window.

The sidechain RMSD (sRMSD) was calculated in reference to the smallest RMSD for each protein residue in Mb across the 51 US windows. The equation to calculate the sRMSD, $sRMSD^r(a, b)$ is as follows

$$sRMSD(a, b)^r = sRMSD(a, b) - sRMSD(a, b)_{\min} \quad (\text{Eq. 2})$$

where $sRMSD(a, b)$ is the $RMSD(a, b)$ as in equation 1 except only the sidechain atoms of each residue a and in b , taken from the US systems and X-ray structure (*PDB: 2W6W*)⁴² respectively, and $sRMSD(a, b)_{\min}$ represents the minimum $sRMSD(a, b)$ within the US window. The sRMSD is averaged per US window.

The torsional energy contributions were calculated using CHARMM for each protein residue in reference to Xe free Mb for each US window across the US reaction coordinate. The formula used to calculate a given residues torsional energy is as follows

$$U_{tors} = k(1 + \cos(n(\psi) - \phi)) \quad (\text{Eq. 3})$$

where U_{tors} is a residues torsional energy, k is the dihedral force constant, n is the dihedral periodicity, ψ is the dihedral angle and ϕ is the phase. All of which are specific for a residues type and dihedral angle as specified in the force-field.⁴⁶ The U_{tors} was then referenced to free Mb using the following equation

$$U_{tors}^r = U_{tors} - U_{tors}^f \quad (\text{Eq. 4})$$

where U_{tors}^f represents the average U_{tors} (as calculated in Eq 3.) over a 10 ns *NVT* Xe free Mb system. The U_{tors}^r was then averaged for each position along the reaction coordinate.

The VDW contributions were calculated using CHARMM for each protein residue in reference to Xe free Mb for each US window across the US reaction coordinate. The formula used to calculate the VDW energy for each residue is as follows

$$U_{LJ} = \varepsilon_{ij} \left[\left(\frac{r_{ij}^{min}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{ij}^{min}}{r_{ij}} \right)^6 \right] \quad (\text{Eq. 5})$$

where U_{LJ} is a residues Lennard-Jones energy, ε_{ij} is the Lennard-Jones well depth, r^{min} is the distance at the Lennard-Jones minimum, and r_{ij} is the distance between atoms i and j .⁴⁶ The U_{LJ} was then referenced to free Mb using the following

$$U_{LJ}^r = U_{LJ} - U_{LJ}^f \quad (\text{Eq. 6})$$

where U_{LJ}^f represents the average U_{LJ} (as calculated in Eq 5.) over a 10 ns *NVT* Xe free Mb system. The U_{LJ}^r was then averaged for each position along the reaction coordinate.

The residue electrostatic contributions were calculated using CHARMM between each protein residue and the system in reference to Xe free Mb for each US window across the US reaction coordinate. The formula used to calculate the electrostatic energy for each residue is as follows

$$U_{elec} = \frac{q_i q_j}{\epsilon_l r_{ij}} \quad (\text{Eq. 7})$$

where U_{elec} is the electrostatic energy between a residue and the system, q_i is the partial

atomic charge, ϵ_l is the effective dielectric constant, and r_{ij} is the distance between atoms i and j .⁴⁶ The U_{elec} was then referenced to free Mb using the following

$$U_{elec}^r = U_{elec} - U_{elec}^f \quad (\text{Eq. 8})$$

where U_{elec}^f represents the average U_{elec} (as calculated in Eq 7.) over a 10 ns *NVT* Xe free Mb system. The U_{elec}^r was then averaged for each position along the reaction coordinate.

The FEP was calculated by sampling the transition times of 70 1 ns *NVT* dynamics as formulated below

$$G(\delta) = -k_B T \ln[P(\delta)] + C \quad (\text{Eq. 7})$$

where $G(\delta)$ is the free energy for the distribution δ of Xe along reaction coordinate, C is a constant, T is the temperature, and k_B is the Boltzmann constant.^{19,35}

The secondary structure for each residue was assigned according to Sanders Dictionary of secondary structure protein (DSSP). For each window along the US reaction coordinate each residue was assigned a secondary structure. This was done for all the Mb systems.

The intramolecular hydrogen bond (H-bond) analysis was conducted by searching for all sidechain residue H-bonds using CHARMM. The number of H-bonds per residue sidechain were summed to give a total number of sidechain H-bonds. The average number of sidechain H-bonds for each residue for each window was then reported for Mb as Xe is transitioning along the US reaction coordinate. This analysis was repeated for every system.

The dynamics cross-correlation matrix (DCCM) and principal component analysis (PCA) was calculated using the Bio3d engine in R.⁶⁴ The US sampling windows were divided into reactant (initial pocket), transition (migration channel), and product (final pocket) windows

to represent different stages during Xe migration as informed by the US FEP. The DCCM was then calculated for the aggregated windows and the motions were evaluated using a Pearson's correlation coefficient. The PCA was determined for each US system by aggregating all the windows into a single trajectory and then calculating the principal components for the entire migration event for each system.

The domain dihedral analysis was conducted by taking the dihedral angles between constituent α domains within Mb. The α domains definitions are taken from X-ray crystallography studies⁴² and are listed in Table 3.

Table 3: Identity of each of each α helical domain in Mb along with the residues which compose each domain.

α helices	α 1	α 2	α 3	α 4
Domain definition	Glu4 to Val17	Val21 to Ser35	Pro37 to Lys42	Glu52 to Met55
α helices	α 5	α 6	α 7	α 8
Domain definition	Asp60 to Leu76	Leu86 to His93	Ile101 to Arg118	Asp126 to Lys147

Results

Xe migration between Xe1 and Xe2

To investigate how Xe occupancy influences Xe migration between Xe1 and Xe2 pockets and the conformational dynamics of Mb the unbiased and US FEP were calculated. The ΔG for Xe to migrate from Xe2 to Xe1 was modulated by Xe occupancy in Xe3 and Xe4 while for migration from Xe1 to Xe2 the occupancy of Xe3 affected the barrier. For system 0100 the ΔG was 2 kcal/mol compared with 5.3 kcal/mol and 4.4 kcal/mol for the 0110 or 0101 systems as seen in Table S1. Figure 1 row one demonstrates that the ΔG between Xe2 and Xe1 was reduced by 2.4 kcal/mol to 3.4 kcal/mol when Xe was absent from Xe2 and Xe3.

Xe presence in either Xe3 or Xe4 increased the ΔG from approximately 2.4 kcal/mol to 3.4 kcal/mol for systems 0110, 0101, and 0111. The magnitude of ΔG was in agreement with those reported for previous Mb Xe studies and the knowledge that Xe more readily diffuses through Mb than the physiological diatomic atoms which interact with Mb.^{19,22,23,32,53}

Figure 1 row one showed that there were discrepancies in the ΔG and position of the TS when comparing the unbiased FEP and US FEP. The TS points between the unbiased and US FEP differed by 0.2 Å while the ΔG was 0.6 kcal/mol to 1.5 kcal/mol higher for the US FEP as shown in Table S1. For system 0100 the unbiased FEP ΔG was 0.5 kcal/mol except when Xe was present in Xe3 or Xe4 where the barrier height rose to 3 kcal/mol. The global minima for Xe during the migration for all systems resides in Xe1 and the overall trend between US and FEP systems remained the same. The unbiased FEP for the Xe1 to Xe2 transition, as shown in Figure 1 row one, were more representative of the US determined FEP but exhibited smaller activation barriers. For the Xe2 to Xe1 migration of Xe the ΔG was altered by Xe occupancy in pockets Xe3 and Xe4. Deviations between theoretically calculated ΔG , particularly US overestimation, and shifts in the profile of unbiased and US FEP for Mb are known to occur.^{13,19,22,65}

To probe Xe allosteric effects on the structure of Mb the volume of Xe1, Xe2, Xe3, and Xe4 was calculated for different Xe positions along the reaction coordinate.²⁷ For US and FEP systems 0100 and 0111 Figure 1 row two presented that the cavity size of Xe1 and Xe4 increased by 25 Å³ while Xe3 increased by 55 Å³ when Xe was nearby the TS. The change in pocket volume correlated with Xe presence in the cleft in the US and unbiased FEP. This showed there was structural reorganisation among the pockets in response to Xe migration from Xe2 to Xe1 and that the Mb pockets undergo conformational adaptions to facilitate Xe diffusion through the protein lattice. Additionally, when Xe was either present or absent from Xe3 and Xe4 the cleft played a functional role in the conformational reorganisation of

Mb through triggering the expansion of Xe1, Xe3 and Xe4.^{2,66} The trend between the pockets US and FEP pockets were similar but the unbiased FEP pocket volumes were smaller and had a greater standard deviation, this difference was attributed to the variation in reaction coordinate across the unbiased simulations sampled, as seen in Figure S2. The pocket sizes observed for unbiased and US systems is nearby to those observed in prior studies of Mb pocket volume.^{2,37}

Shifts in the average dihedral χ_1 for residues which line the Xe1 to Xe2 channel were inspected χ_1 to see localised allosteric effects in response to Xe migration. The changes in ΔG were a function of Mb structural changes in response to Xe presence as demonstrated by Figure 1 row three and four. To determine which residues along the migration route may have been responsible for this shift the average χ_1 for residues Leu89, Leu104, Glu105, and Phe138 which line the migration route, were measured for each reaction coordinate for unbiased and US systems.^{13,51} Across all systems there was a sidechain flip represented by the change of the Leu89 χ_1 from about -90° to -150° when Xe was nearby the TS. The regions where Leu89 changed conformation correlate with where the migration barrier was initiated from Xe2 and once Xe arrived at Xe1. Additionally Leu104 showed a change in χ_1 from about -170° to -80° when Xe passed the TS for the Xe2 to Xe1 Xe migration. As Leu89 and Leu104 are located along the migration pathway between the two pockets they are involved in regulating the migration of Xe between Xe1 and Xe2.^{10,13} The differences in the ΔG for Xe migration from Xe2 to Xe1 was related to ligand occupying different pockets in Mb, contributing to the two fold changes of ΔG seen in Figure 1 row one.

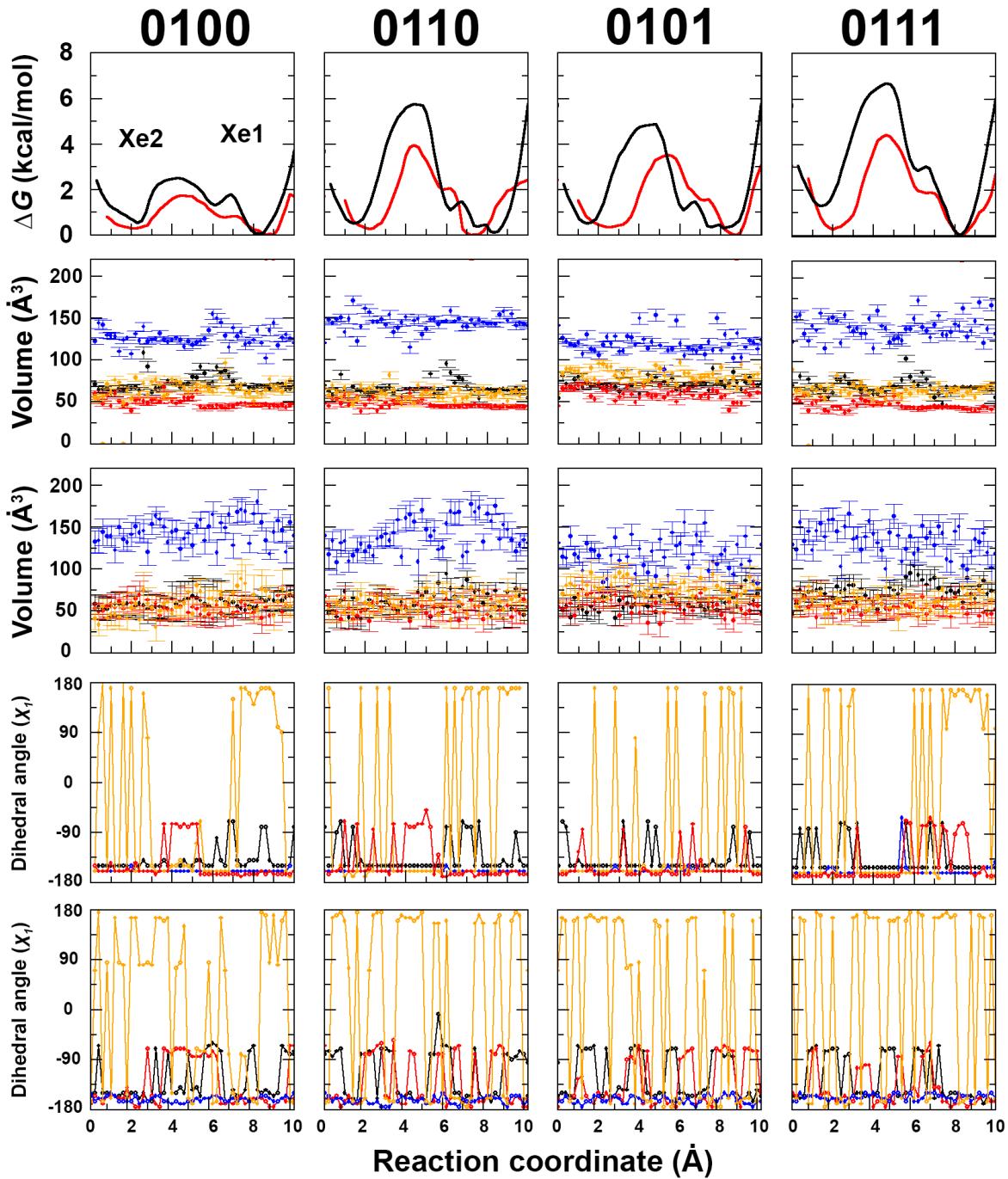


Figure 1: FEP, volume, and χ_1 dihedral angle analysis for different states for the Xe2 to Xe1 transition as defined in Table 2. Row one shows the FEP from US (black) and from unbiased sampling (red). Row two presents the average volume and standard deviation of pockets Xe1 (black), Xe2 (red), Xe3 (blue), and Xe4 (orange) and standard deviations for each position along the US reaction coordinate. Row three presents the average χ_1 dihedral angle for residues Leu89 (black), Leu104 (red), Glu105 (blue), and Phe138 (orange) which are local to the migration route for each position along the US reaction coordinate.

An assessment of sRMSD changes for residues as Xe migrates from Xe2 to Xe1 highlights specific residues important in the allosteric mechanism of Mb.^{10,67} The left panel of Figure 2 showed that the sRMSD shifted by 0.3 Å for Leu104 and Phe138 for all systems when Xe was nearby the TS. Additionally, Leu104 demonstrated a 0.3 Å sRMSD shift at 0 Å to 2 Å along the reaction coordinate. These residues line the Xe1 to Xe2 channel, indicating that they were shifted in response to Xe moving through the channel and are involved in controlling Xe movement across the Xe1 to Xe2 channel. Ser35, Glu38 and Leu49, Lys96 of system 0101 showed sRMSD changes of up to 0.25 Å when Xe was leaving Xe2 however all were distal from the pocket indicating a long range mechanism triggered by Xe presence in Xe3. Gly121 and Gln128 are nearby the DP and presented sRMSD changes up to 0.25 Å most prominently seen in systems 0110 and 0111 nearby the TS, suggesting a role in facilitating Xe exit when Xe3 was occupied. Leu9, Glu59, Lys140, and Met131 were located away from the migration channel and exhibited sRMSD changes of up to 0.2 Å nearby the TS for systems 0101 and 0111, suggesting these residues respond to Xe when it was migrating between Xe1 and Xe2 and when Xe4 is occupied. Gln8 and Leu29 were located on the surface of Mb and showed sRMSD changes and were triggered when Xe was either entering or leaving the Xe1 to Xe2 migration channel. Lys47 is located on the Mb surface and demonstrated migration sensitive sRMSD most prominently for system 0111, changing 0.2 Å. Specific residues sRMSD at different sites on Mb contribute to Mb dynamics during Xe migration from Xe2 to Xe1 and are perturbed in response to Xe migration from Xe2 to Xe1.

As an extension of inspecting the structural effects Xe has on residue sidechains the U_{tors}^r for each residue was calculated. The right panel of Figure 2 shows that majority of shifts in the U_{tors}^r occur for residues on the Mb surface when Xe is at various points along the reaction coordinate. Surface residues Trp4, Pro37 and Lys62 exhibited a U_{tors}^r change of 2 kcal/mol to 4 kcal/mol when Xe was nearby its Xe2, while when Xe was near the TS His12, Asp27 Pro37, Lys56, Asp60, Pro100, His113, Pro120, Ala127, and Gly129 demonstrated 3 kcal/mol

to 6 kcal/mol change in U_{tors}^r , and when Xe was nearby Xe1 Ile30 demonstrated a 3 kcal/mol U_{tors}^r change. Pro88 exhibited up to 6 kcal/mol U_{tors}^r changes at different points throughout the Xe Xe2 to Xe1 transition. Residue His93 was local to Xe1 and showed a 3 kcal/mol U_{tors}^r shift when Xe was leaving Xe2, as was Leu104 which exhibited a 4 kcal/mol shift nearby the TS. Nearby to Xe2 was Arg139 which showed a 4 kcal/mol shift when Xe was 1 Å to 9 Å along the reaction pathway. Trp7 was close to Xe3 and exhibited a 4 kcal/mol change when Xe was near the TS. Nearby the exit channel Arg45 which showed a U_{tors}^r shift of 6 kcal/mol in response to Xe at various points during Xe transition from Xe2 to Xe1. Together this has shown that the U_{tors}^r of residues is dependent on Xe position along the Xe2 to Xe1 migration pathway. Furthermore, that the largest U_{tors}^r changes are seen when Xe is nearby the TS, and very few U_{tors}^r are exhibited when Xe was nearby Xe1, helping identify the source of the ΔG seen in Figure 1 and S1.

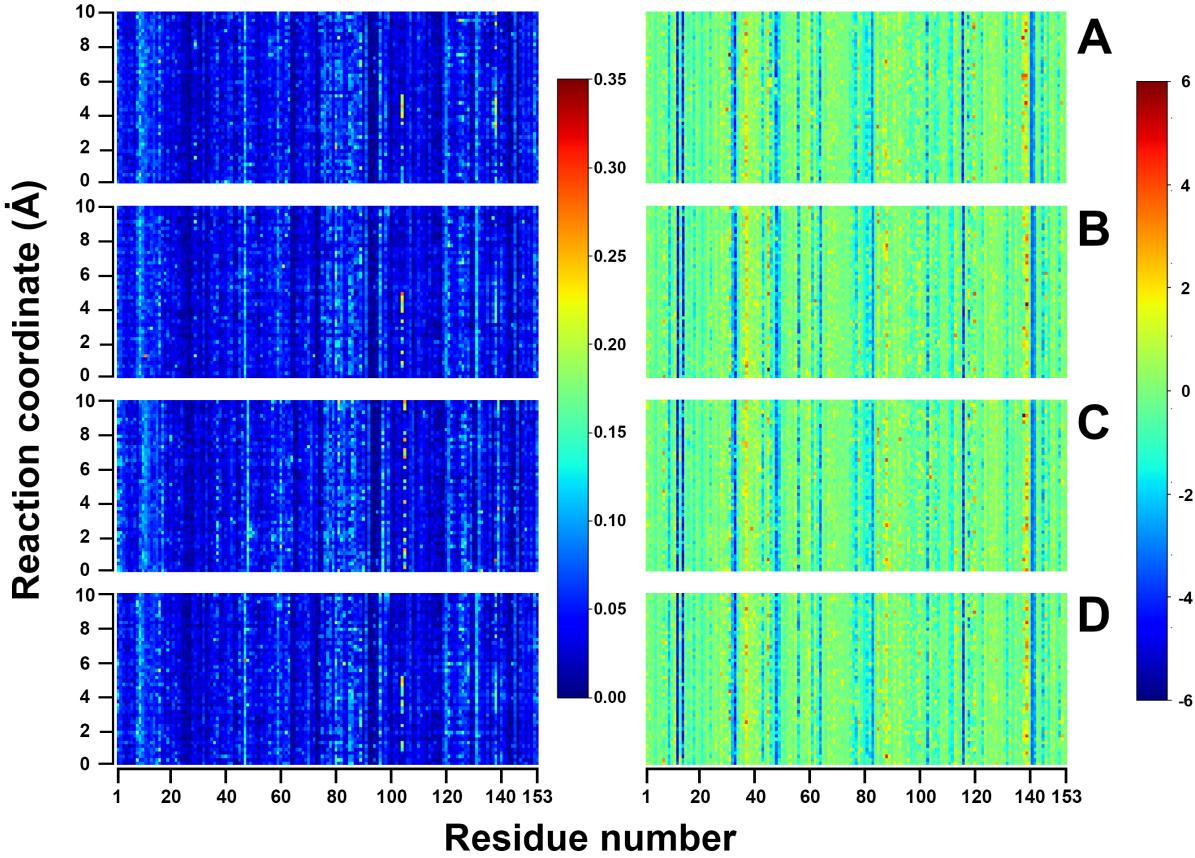


Figure 2: Average sidechain $sRMSD^r$ (a, b) (left) and average U_{tors}^r (right) for each residue at points along the US reaction coordinate from Xe2 during the Xe2 to Xe1 transition. Panels A to D represent systems 0100, 0110, 0101, and 0111 respectively. Legend to the right of each panel are in Å (left), and kcal/mol (right).

To investigate short and long range non-bonded effects Xe position has on the allostery of Mb the U_{LJ}^r term for each residue and U_{LJ}^r between each residue and system was calculated. The surface residue VDW interactions dominate the U_{LJ}^r energies during the Xe2 to Xe1 migration as illustrated in the left panels of Figure 3. Surface residues Trp14, Ile30, Leu32, Glu41, Leu49, Lys62, Pro88, Ile101, Tyr103, Phe123, Lys140, and Tyr151 U_{LJ}^r changes of up to 4 kcal/mol when Xe was leaving Xe2, nearby the TS Ser3, Glu4, Trp14, Asp20, Ile30, Leu32, Thr39, Lys62, Pro88, Tyr103, Leu115, Phe123, Asp126, Met131, Asn132, and Tyr151 exhibited U_{LJ}^r changes of up to 6 kcal/mol. Additionally, surface residues Trp14, Asp20, Ile30, Leu32, Thr39, Tyr103, Phe123, Asp126 and Tyr151 showed U_{LJ}^r changes of up to 3 kcal/mol

to 5 kcal/mol when Xe was nearby Xe1. Leu104 and Leu135 were nearby Xe1 and Xe2 respectively exhibited U_{LJ}^r changes of up to 4 kcal/mol when Xe was near the TS. Trp7 was nearby Xe3 and exhibited U_{LJ}^r changes of up to 4 kcal/mol when Xe was near the TS. Ile28 and Val68 were local to Xe4 and exhibited 4 kcal/mol U_{LJ}^r changes when Xe was leaving Xe2. Nearby to the exit pocket was Arg45 which demonstrated 3 kcal/mol U_{LJ}^r changes at various points along the reaction coordinate. The number and magnitude of U_{LJ}^r change was greatest for the residues at the surface and when in response to Xe near the TS. The right panels of Figure 3 showed that the U_{elec}^r between certain residues and the protein was modulated by Xe2 to Xe1 migration. Surface residues Glu6, Glu18, Asp27, Pro37, Thr51, Thr70, and Asn132 demonstrate U_{elec}^r shifts of up to 40 kcal/mol when Xe nearby Xe2. When Xe was near the TS the U_{elec}^r of Pro37, Lys42, Thr51, Lys77, His97, Arg118, Phe123, and Asn132 changed up to 40 kcal/mol. While Trp14, Asp44, Lys47, Glu54, and Glu85 U_{elec}^r changed by up to 40 kcal/mol when Xe leaves Xe2, transitions between Xe2 and Xe1 and arrived at Xe1. When Xe was migrating from Xe2 to Xe1 the U_{elec}^r between residues and the protein are scarcely effected. The highest and most frequent U_{LJ}^r and U_{elec}^r for residues was observed when Xe was near its TS, followed by changes when Xe was nearby Xe2 and rarely for when Xe was nearby Xe1. Together, this identifies the ΔG as seen in Figure 1 was a culmination of primarily U_{LJ}^r interactions and is dominated by residues located on the surface of Mb.

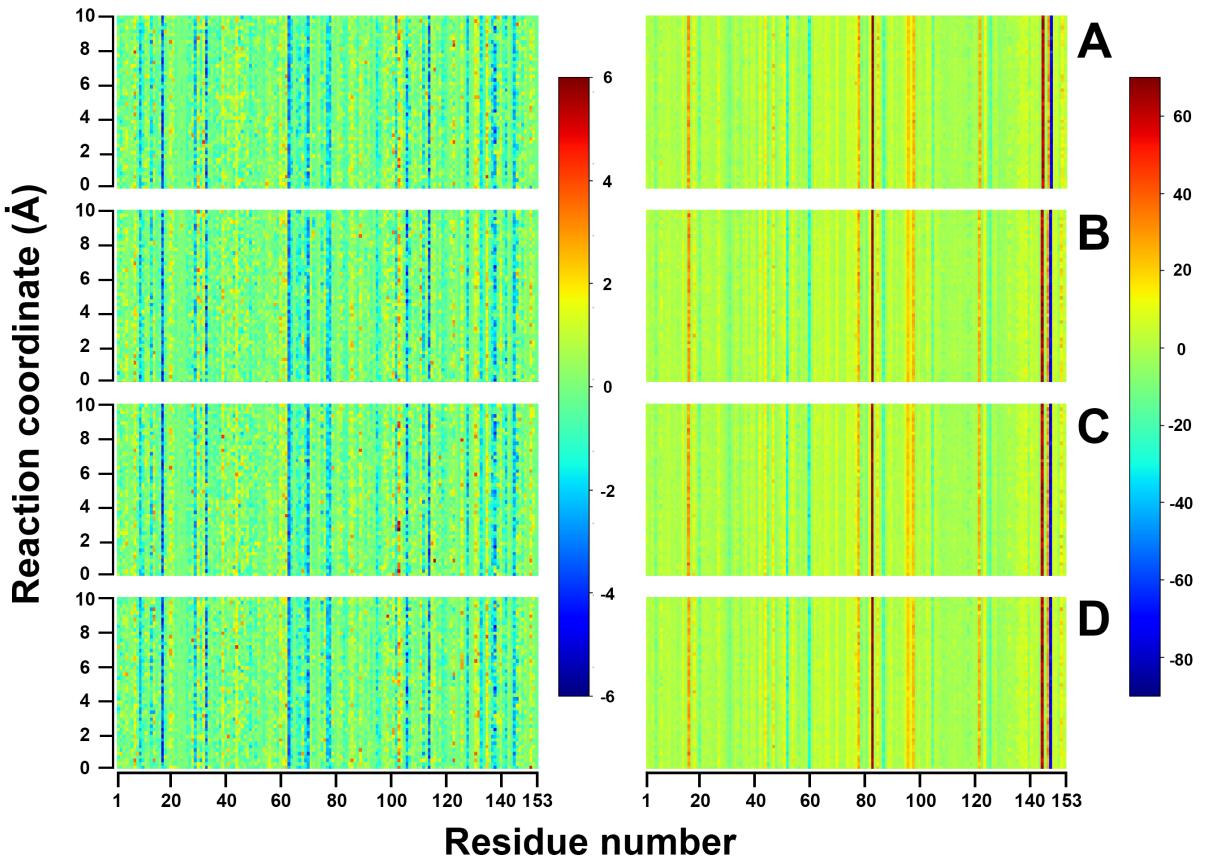


Figure 3: Average U_{LJ} (left) and average U_{elec} (right) for each residue at points along the US reaction coordinate from Xe2 during the Xe2 to Xe1 transition. Panels A to D represent systems 0100, 0110, 0101, and 0111 respectively. Legend to the right of each panel are in Å (left), and kcal/mol (right).

Identification of correlated motion indicates potential allosteric networks that are altered by Xe position along the reaction coordinate.^{10,27,68} Figure 4 shows that for system 0100 the correlation between Glu51 to His48 and Ala74 to Glu83 involved more residues when Xe was nearby Xe2. Correlated motion between His36 to His48 and Lys77 to Glu83 involved fewer residues while Xe was near its TS. The correlated motion between Val1 to Ile30 and His97 to Ile99 was smaller when Xe was nearby to Xe1 while the correlated motion between residues Asp60 to Pro88 and Ala125 to Lys140 was fragmented when Xe was nearby to Xe2, and involved fewer atoms when Xe was near Xe1. For system 0110 the correlated motion between Val1 to Leu32 and His97 to Ile99 and Lys102 to Ile107 was fragmented when Xe

was nearby Xe2, this was also observed when Xe was near Xe1 for system 0101, and when Xe was near its TS for system 0111. Novel correlations were observed for system 0110 when Xe was in the migration channel between Lys77 to Gly80 and Lys133 to Phe138, and while it was nearby Xe1 between Asp60 to Ala74 and Asp141 to Tyr146. System 0101 exhibited novel correlated motions when Xe was nearby Xe2 between Glu54 to Ser58 and Gly124 to Asn132, while when Xe was near its TS unique correlated motion was noted between Asp60 to Thr70 and Pro120 to Asp122. There was distinctive correlated motion for system 0111 depending on Xe position, when Xe was nearby Xe2 this was witnessed between Ile75 to Lys79 and Asn132 to Lys140, and Glu83 to Ser92 and Asn132 to Ile142. Then for when Xe was near its TS for system 0111 between Glu83 to Ala90 and Gly124 and Met131. The correlated motions between groups of residues is effected by Xe position along the Xe2 to Xe1 transition and Xe occupancy however it was not specific to Xe occupancy and Xe position but was dependent on both these elements together.^{17,23,67}

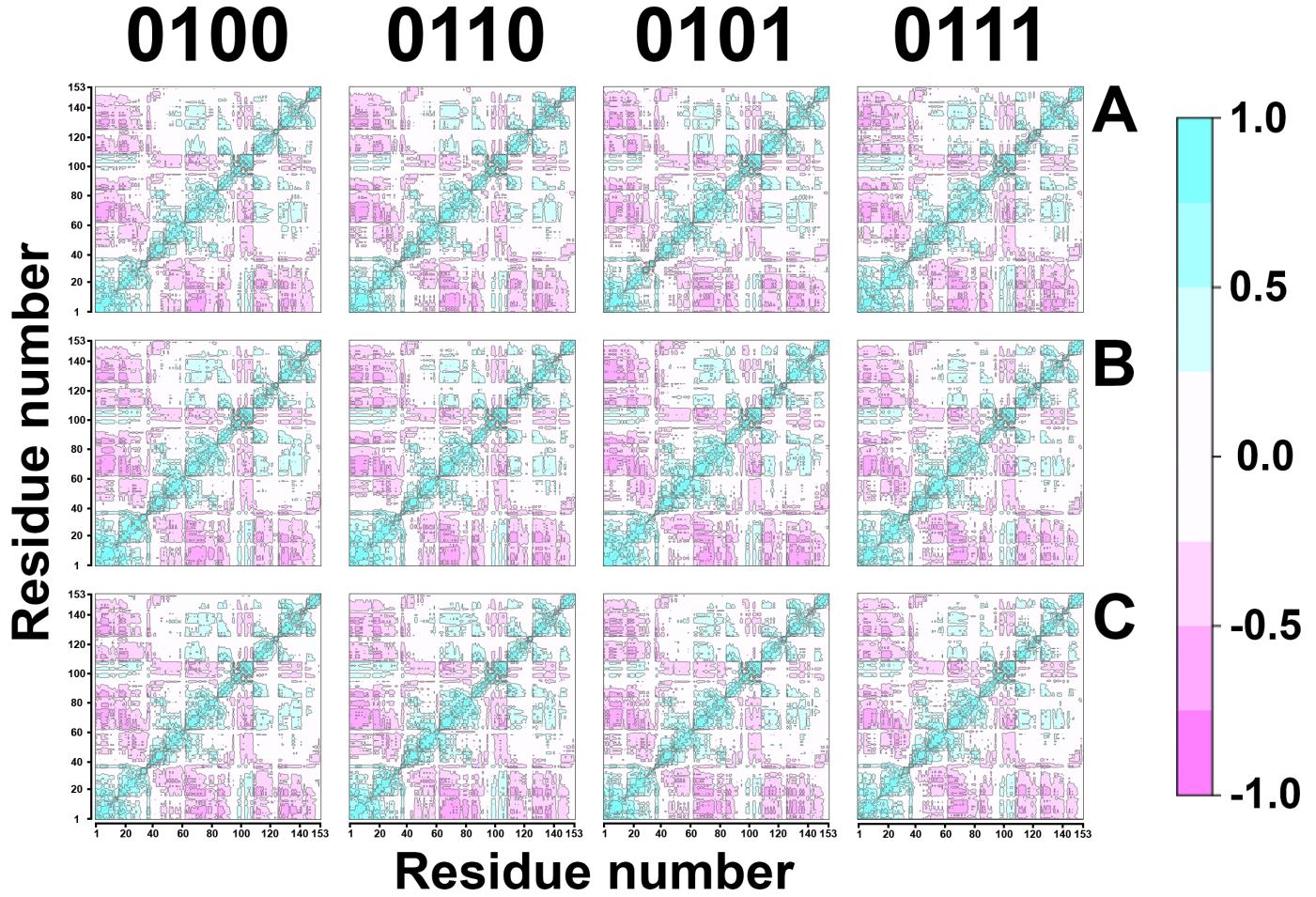


Figure 4: DCCM plots using for US Xe2 to Xe1 systems for reactant (*A*), transition (*B*), and product (*C*) windows exhibited as a Pearson correlation coefficient segmented into 0.25 groups.

Similar to what had been observed in Figure 1, Xe occupancy modulated the ΔG when Xe migrated from Xe1 to Xe2. Figure 5 row one displayed that for US systems 1000 and 1001 Xe migration the ΔG was 3 kcal/mol compared with the 4.5 kcal/mol for 1010 and 1011. This revealed that Xe presence in Xe3 increases the ΔG for by 1.5 kcal/mol while presence in Xe4 had no impact on the ΔG . The ΔG for Xe migrating from Xe1 to Xe2 were smaller than those calculated for when Xe had migrated from Xe2 to Xe1 transition. This suggests a bias in the thermodynamics conferred by the allostery of Mb: that when Xe2 was occupied and Xe migrates from Xe2 to Xe1 the ΔG was modulated to be higher than when Xe1 was

occupied and Xe migrates from Xe1 to Xe2.^{18,22}

The unbiased system ΔG were lower than those calculated for the US and displayed similar Xe migration dynamics and TS position. The ΔG for the unbiased FEPs were between 0.8 kcal/mol to 1.6 kcal/mol smaller than the US FEP calculations while the TS were within 0.2 Å of each other as indicated in Table S1 and Figure 5. The ΔG for unbiased FEP the ΔG was smallest for system 1000 at 1.8 kcal/mol and largest for system 1011 at 3.4 kcal/mol as seen in Table S1. Aside from the magnitude of the ΔG the greatest difference between the profiles was the absence of the cleft located at 6 Å along the reaction coordinate for system 1000. This is due to the relatively low ΔG for this transition, indicating that when Xe was absent from Xe3 and Xe4 the subpocket does impact Xe migration.^{17,20}

Figure 5 rows two and three showed that the volumes of Xe cavities respond to the migration of Xe from Xe1 to Xe2. As Xe migrated from Xe1 to Xe2 system 1010 demonstrated an increase in Xe1 pocket by 50 Å³ when Xe was near the TS, indicating that Xe3 occupancy triggered Xe1 to change volume as Xe passed through the Xe1 to Xe2 channel. A similar migration correlated increase in volume to 125 Å³ was witnessed for Xe2 for system 1010 and 1001, identifying a characteristic reorganisation of the pockets when either Xe3 or Xe4 was occupied.^{22,66} As seen for Figure 1 row 3, the unbiased volume calculations in general exhibited lower volumes and higher deviations but also less distinctive Xe responsive changes. This can be attributed to the variety of different migration routes Xe can take on, as expressed on S1 and S2 and resulting in less distinctive Xe position orientated structural changes.

The differences in the ΔG for Xe migration from Xe1 to Xe2 was related to ligand occupying different pockets in Mb, causing specific structural changes during Xe migration. As an extension of understanding the structural rearrangements of Mb in response to Xe occupancy

the change in χ_1 for Leu89, Leu104, Glu105, and Phe138 was undertaken as shown in Figure 5 rows four and five. There was a characteristic shift in the χ_1 of Leu89 at about the TS for each system, changing from -80° to -170° . Similarly, Leu104 demonstrated a χ_1 shift from -170° to -80° when Xe was nearby Xe2. The χ_1 of Glu105 presented a shift for system 1011 from -160° to -80° at 6 Å along the reaction coordinate. This indicated that when either Xe3 or Xe4 were occupied Glu105 modulated the subpockets interaction.^{13,48} Phe138 exhibited sidechain movements for similar positions across the transition pathway, changing from 170° or -170° to 90° . Together this showed that the χ_1 of Leu89, Leu104, Glu105, and Phe138 were all shifted at different points along the migration route as Xe was migrating from Xe1 to Xe2 and are triggered in response to Xe migration.

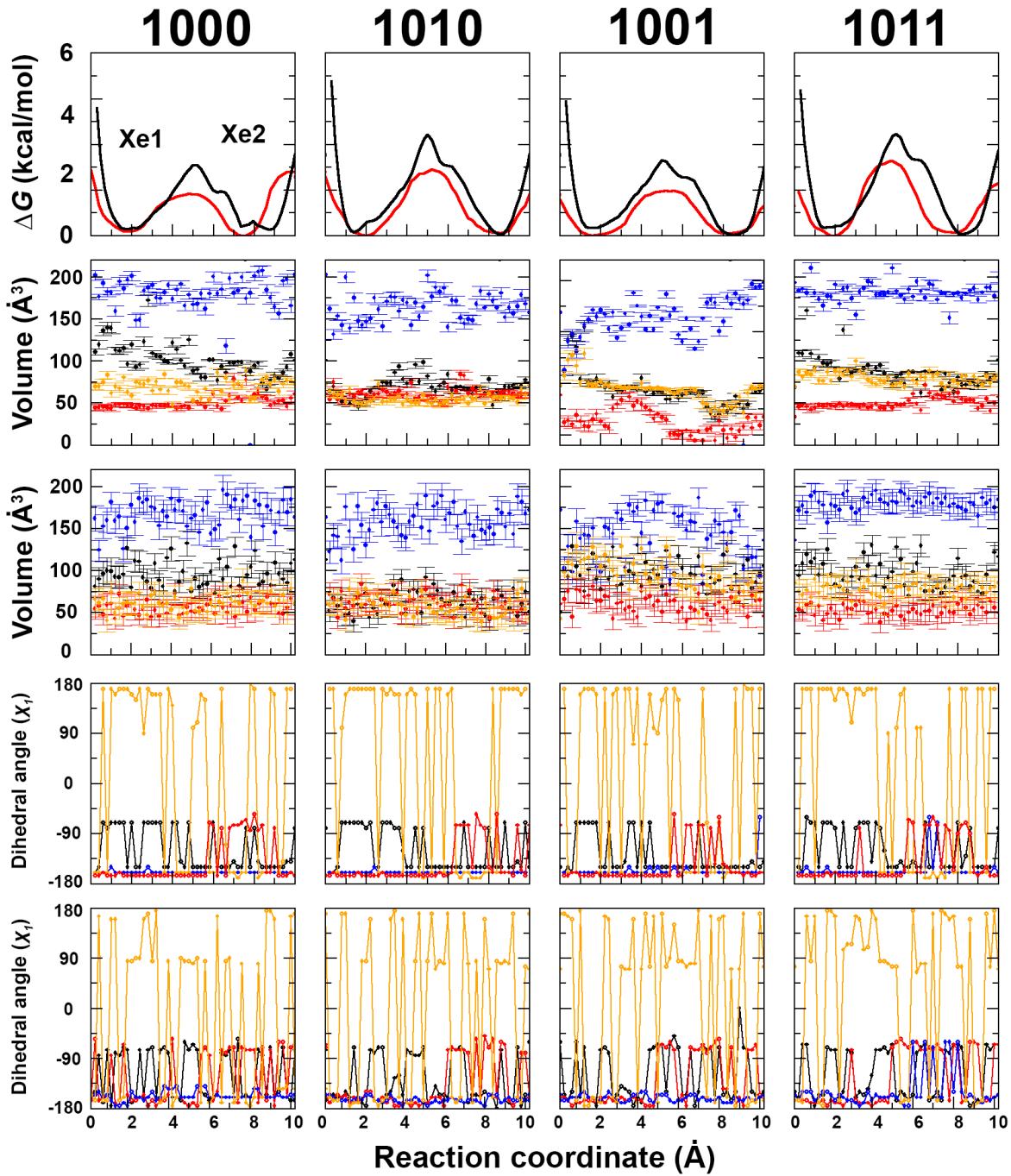


Figure 5: FEP, volume, and χ_1 dihedral angle analysis for different states for the Xe1 to Xe2 transition as defined in Table 2. Row one displayed the FEP from US (black) and from unbiased sampling (red). Row two and three presents the average volume and standard deviation of pockets Xe1 (black), Xe2 (red), Xe3 (blue), and Xe4 (orange) and standard deviations for each position along the US and unbiased reaction coordinate respectively. Row four and five presents the average χ_1 dihedral angle for residues Leu89 (black), Leu104 (red), Glu105 (blue), and Phe138 (orange) which were local to the migration route for each position along the US and unbiased reaction coordinate respectively.

The average sRMSD for specific residues is altered during the Xe1 to Xe2 transition depending on the position of Xe along the reaction coordinate. The left panel of Figure 6 showed Glu85, Leu89, Lys96, and Met131 located on the surface of Mb presented sRMSD shifts of up to 0.25 Å when Xe was nearby Xe1. Lys47, Glu59, and Lys96 which were also located on the surface of Mb demonstrated a shift in sRMSD of 0.2 Å when Xe was nearby Xe1 when both Xe3 and Xe4 are occupied however when Xe4 was not occupied they demonstrated changes when Xe was near the TS, suggesting their involvement in altering Xe1 as Xe leaves Xe1 and migrates towards Xe2. Leu104 and Phe138, which line Xe2, display a sRMSD change of 0.3 Å as Xe was nearby Xe2, revealing their role in changing the shape of Xe2 as Xe approached from Xe1. Leu9 and Lys16 were located on the surface of Mb and influenced Xe migration when Xe passed through the Xe1 to Xe2 channel, their sRMSD shifting by 0.25 Å, indicating a potential long range mechanism. Residues located at both the surface of Mb and along the Xe1 to Xe2 channel exhibited a sRMSD change in response to Xe migration, however fewer in general for the Xe1 to Xe2 when compared to the Xe2 to Xe1 migration.

Changes in the U_{tors}^r as Xe transitions from Xe1 to Xe2 were dominated by residues on the surface of Mb, as presented on the right panels of Figure 6. Surface residues Arg31, His36, Pro37, Glu85, Pro88, and Pro100 showed U_{tors}^r changes of 4 kcal/mol when Xe was leaving Xe2, transitioning between Xe2 and Xe1, and arriving at Xe1 during its Xe2 to Xe1 migration, while nearby the TS Met55 and Asp60 were surface residues which showed 3 kcal/mol changes in U_{tors}^r when Xe was nearby Xe1, Asp27, and Lys40, and Lys62 displayed U_{tors}^r shifts of 3 kcal/mol to 4 kcal/mol when Xe was nearby the TS. Residues nearby Xe1 and were perturbed by Xe migration including His93 and Lys133 exhibiting U_{tors}^r changes up to 4 kcal/mol when Xe was nearby the TS, and His113 and Pro120 throughout the transition pathway with U_{tors}^r changes of between 3 kcal/mol to 6 kcal/mol. Leu104 was near Xe2 and showed U_{tors}^r changes of up to 3 kcal/mol when Xe was near the TS as did Leu139 and was showed U_{tors}^r perturbation for up to 6 kcal/mol throughout the Xe2 to Xe1 transition. Trp7,

Lys78, and Leu135 were nearby Xe3 and exhibited U_{tors}^r changes of 2 kcal/mol to 6 kcal/mol when Xe was near the TS. Local to Xe4 are Ile28, Leu29, and Val66, and Val68 which exhibited U_{tors}^r changes of up to 4 kcal/mol when Xe was leaving Xe1. Arg45 which regulates Xe exit demonstrated U_{tors}^r changes of up to 5 kcal/mol throughout Xe transition from Xe1 to Xe2. Compared to the findings from Figure 2 there were fewer U_{tors}^r identified with smaller U_{tors}^r on average, still triggered when Xe was nearby to the TS, and there was a more equal split between surface and intramolecular residues with notable U_{tors}^r . This information helps locate sources for why the ΔG is smaller for Figure 2 than Figure 1 and the residues which contribute to the ΔG .

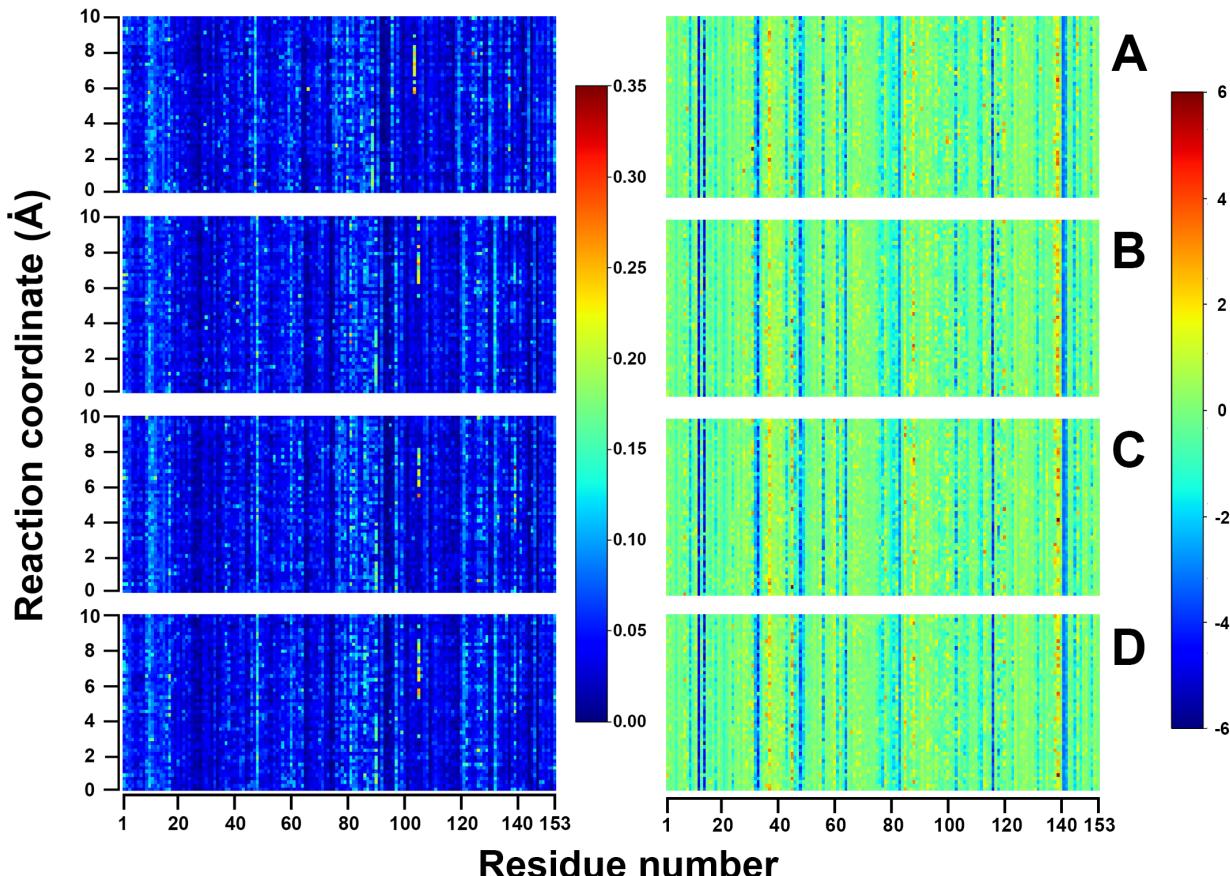


Figure 6: Average $sRMSD^r(a, b)$ (left) and average U_{tors}^r (right) for each residue at points along the US reaction coordinate during the Xe1 to Xe2 transition. Panels A-D represent systems 1000, 1010, 1001, and 1011 respectively. Legend to the right of each panel are in \AA (left), and kcal/mol (right).

The left panel of Figure 7 portrayed the non-bonded energy contributions to the system for each residue as Xe migrated from Xe1 to Xe2. Ser3, Glu4, Asp20, Asp60, Leu76, Asp122, Phe123, and Met131 are surface residues which exhibited 3 kcal/mol to 4 kcal/mol U_{LJ}^r changes when Xe was near Xe1. When Xe was near the TS Ser3, Leu11, Asp20, Arg31, Phe33, Pro37, Glu41, Asp44, Met55, Asp60, Lys62, Glu85, His97, Leu115, Arg118, Asn132, Lys147, Glu148 exhibited 3 kcal/mol to 6 kcal/mol changes in U_{LJ}^r . While when Xe was near Xe2 Met55, Lys62, and Leu76 were surface residues that presented a 4 kcal/mol change in U_{LJ}^r . Surface residues Trp14, Ile30, Thr39, Tyr103, Asp126, and Tyr151 exhibited 4 kcal/mol to 5 kcal/mol changes in their U_{LJ}^r in response to Xe through its transition to Xe2. Leu104 was nearby Xe1 and showed U_{LJ}^r changes of 3 kcal/mol when Xe was near the TS, while Leu135 which is local to Xe2 demonstrated U_{LJ}^r changes of up to 3 kcal/mol in response to Xe when it was leaving Xe1 or near the TS. Trp7 which is nearby Xe3 presented a 4 kcal/mol U_{LJ}^r shift when Xe was near Xe1. Local to Xe4 was Ile111 which showed U_{LJ}^r change of 4 kcal/mol when Xe was nearby its transition state, while Ile28, and Leu32 presented a 4 kcal/mol change in response to Xe being near Xe1, transitioning between Xe1 and Xe2 and nearby to Xe2. As had been seen for Figure 3 the U_{LJ}^r is dominated by surface residues, and the highest contributions were witnessed when Xe was nearby the TS. The U_{elec}^r between a given residue and Mb during the Xe1 to Xe2 transition was calculated on the right panel of Figure 7. Trp14, Asp27, and Glu85 were surface residues which showed a U_{elec}^r 20 kcal/mol shift when Xe was nearby Xe1. Furthermore, Trp14, Glu18, Lys34, Asp44, Lys47, Thr51, Glu85, His113, Phe123, and Asp141 exhibited U_{elec}^r changes of up to 40 kcal/mol shift when Xe was nearby its transition state. Surface residues Glu18, Asp27, Lys47, and Glu85, exhibited 20 kcal/mol shifts when Xe was near Xe2. While Glu4, Glu54, Lys77, Lys78, and Tyr152 exhibited 30 kcal/mol shifts in their U_{elec}^r throughout Xe transition from Xe1 to Xe2. Nearby to Xe2 is Phe138 which presented a 20 kcal/mol U_{elec}^r change when Xe was 1 Å to 7 Å along the reaction coordinate. Aside from Phe138 the notable U_{elec}^r between residues and Mb in response to Xe migration was exclusively witnessed for surface residues. A similar,

but less exaggerated trend had been seen for Figure 3 indicating solvent interactions may play a role in the allosteric mechanism of Mb. Furthermore, that the greatest shifts and most frequent in U_{elec}^r were observed when Xe was near its TS. However, there are fewer residues involved in U_{elec}^r changes than U_{LJ}^r indicating that the U_{LJ}^r was more sensitive to the Xe as it was migrating from Xe1 to Xe2.

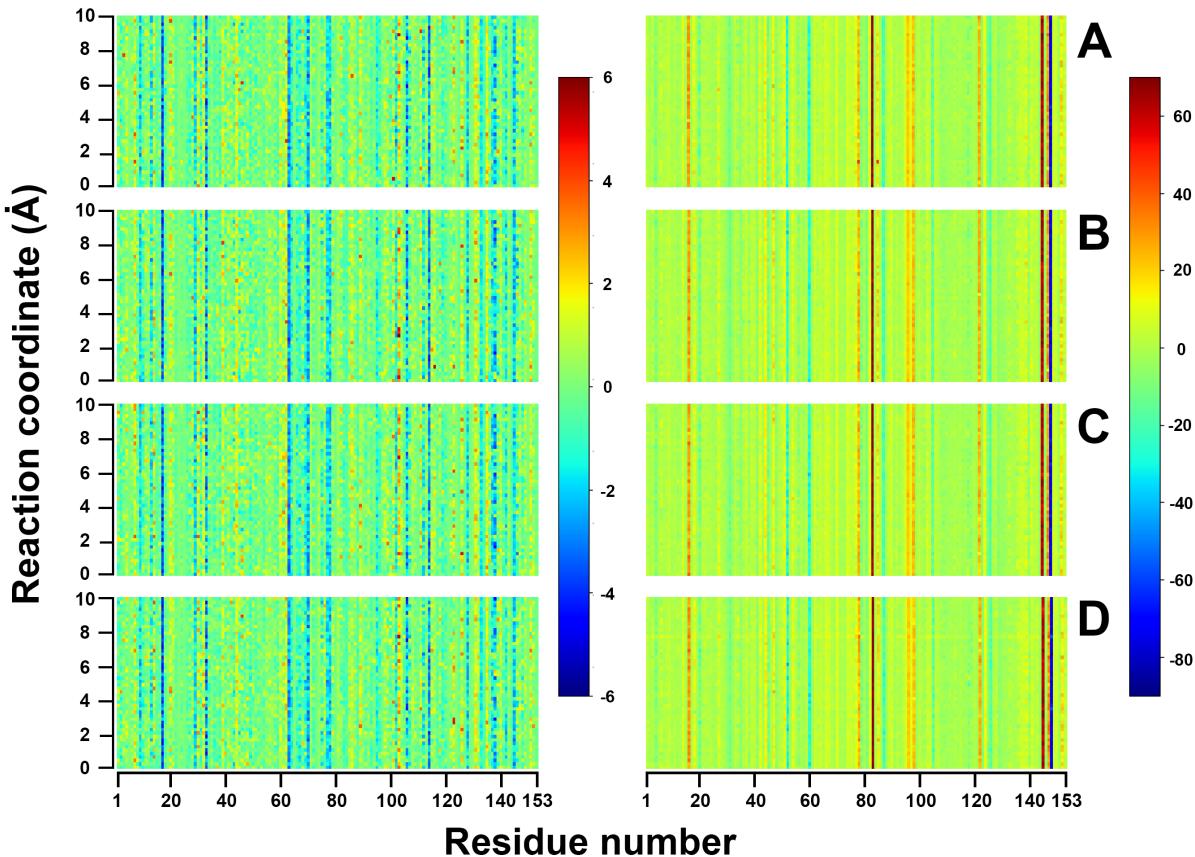


Figure 7: Average U_{LJ}^r (left) and average U_{elec}^r (right) for each residue at points along the US reaction coordinate from Xe2 during the Xe1 to Xe2 transition. Panels A to D represent systems 1000, 1010, 1001, and 1011 respectively. Legend to the right of each panel are in Å (left), and kcal/mol (right).

As shown for the Xe2 to Xe1 transition the correlated motion between residues was dependent on both the Xe occupancy and at which position Xe was along the reaction coordinate. For system 1000 and 1011 it was shown that the correlated motion between Val1 to Leu32

and Lys96 to Ile99 and Lys102 to Ile107 is fragmented when Xe was nearby Xe1 or Xe2, this was also demonstrated for system 1010 and 1001 when Xe was nearby Xe2. System 1000 also showed fragmentation of the Asp60 to Gly80 and Ala125 to Glu148 correlation when Xe was nearby Xe1. System 1001 showed fragmentation of the correlation between Asp60 to Ala74 and Ile142 to Leu149 when Xe was nearby to Xe1. Additionally, the correlation with Asp60 to Gly80 was modulated when Xe was nearby Xe2 as it was for correlation with Ala144 to Glu150 for system 1000. For system 1011 correlation with Ile107 to Leu115 and Lys133 to Ala144 was distinctive depending on Xe position along the reaction coordinate. Similar domains were involved in the correlated motion as had been seen in Figure 4 for the Xe2 to Xe1 transition. This indicates that correlated motion between residues is less sensitive to the direction of the transition but still sensitive to Xe occupancy and position of Xe along the transition pathway.

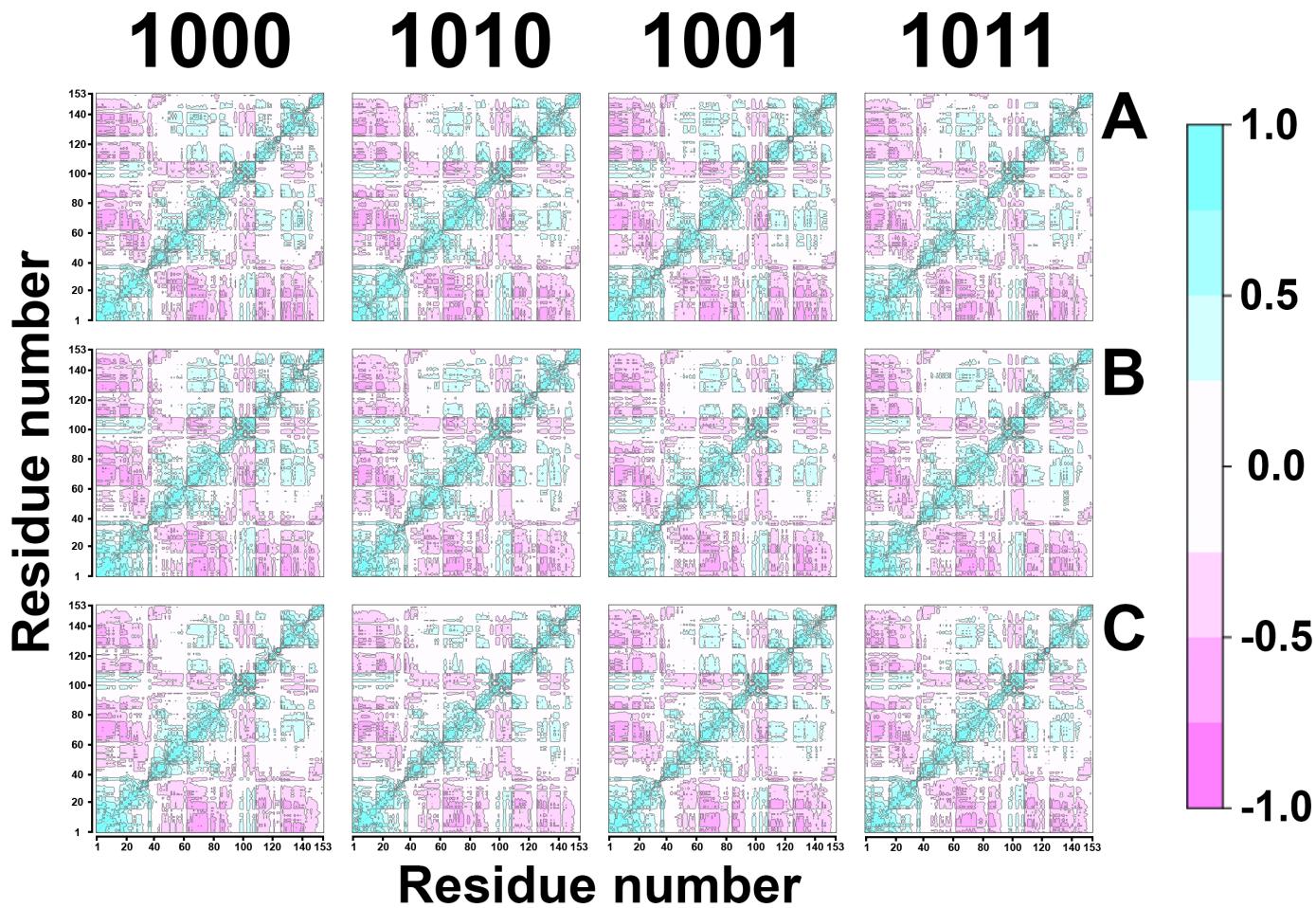


Figure 8: DCCM plots using Mb C α atoms for US Xe1 to Xe2 systems for reactant windows (*A*), transition windows (*B*), and product windows (*C*) exhibited as a Pearson correlation coefficient.

The sidechain intramolecular H-bond network within Mb was responsive to the position of Xe along Xe1 to Xe2. Figure 9 panels A, B,C and D show that when Xe is migrating from Xe2 to Xe1 Glu6, Trp14, Arg45, and Tyr146 present increased H-bonding when Xe was nearby Xe2. The same residues demonstrated influxes in their sidechain H-bonding when Xe was migrating through the Xe1 to Xe2 channel and from Xe2 such as Glu6, Trp14, and Arg45, Glu18, Gln26, and Ser92. Glu6, Glu18, Arg45, and Ser92, Tyr146 also demonstrated changes in their sidechain H-bonding when Xe was nearby Xe1. For figure 9 panels E, F, G, and H some of the same residues were found to rearrange their sidechain H-bonding in

response to Xe migration for the Xe1 to Xe2 migration. When nearby to Xe1 Glu6, Trp14, His36, Arg45, and Tyr146 showed changes in their intramolecular sidechain H-bonding. Additionally Glu6, Trp14, His36, and Arg139 take on changes in their sidechain H-bonding when Xe was nearby the TS. While Trp14, Arg45, Glu54, and Tyr146 also exhibited changes in their sidechain H-bonding network when Xe is nearby Xe2. This has revealed that the intramolecular sidechain H-bonding network is not sensitive to the direction of the Xe migration between Xe1 and Xe2. Furthermore, that the occupancy of Xe has little influence on the intramolecular H-bonding networking, allowing us to highlight residues which are fundamental in the allosteric mechanism of Mb.

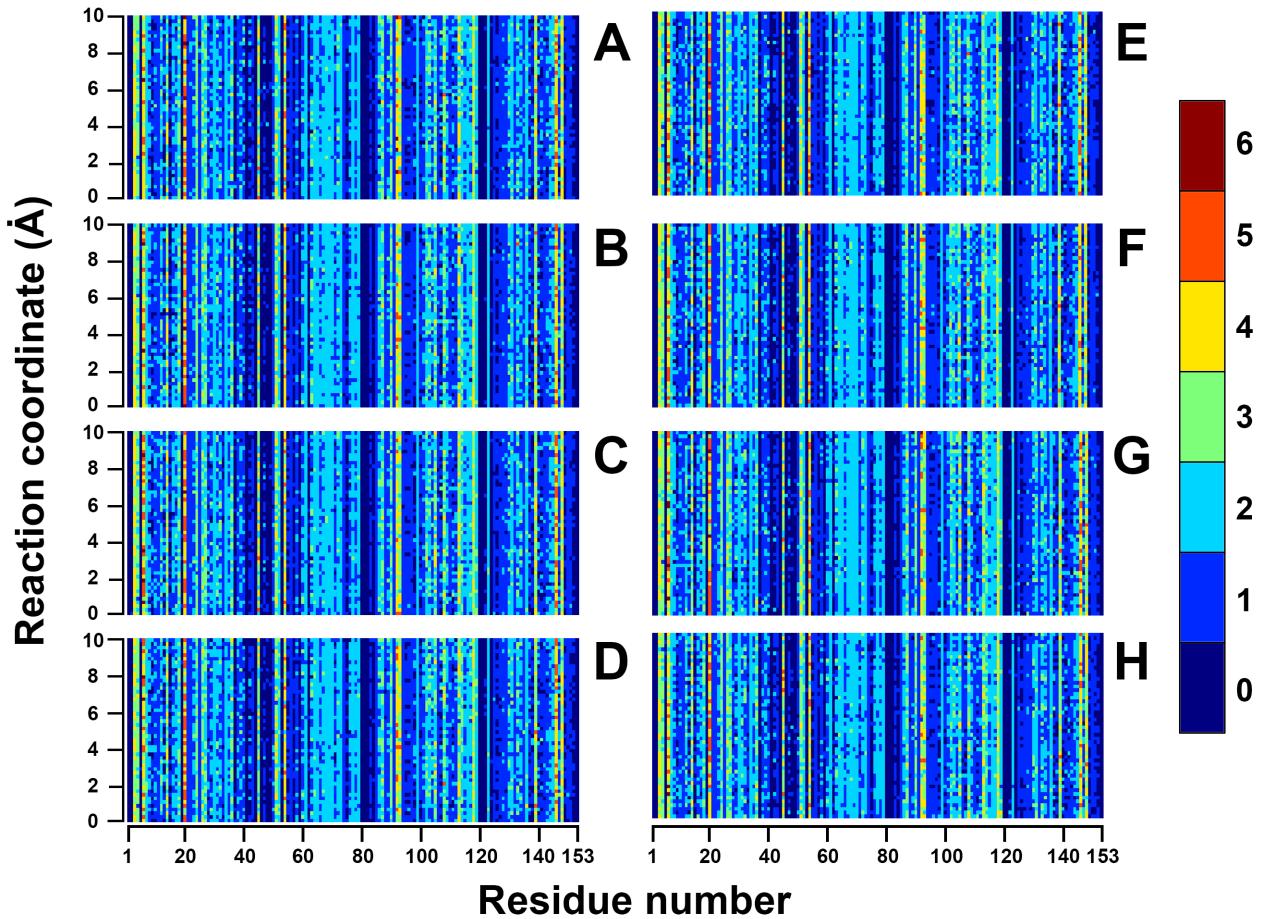


Figure 9: displays the number of intramolecular H-bonds formed for each residues sidechain as Xe migrates from Xe2 to Xe1 for systems 0100 (panel *A*), 0110 (panel *B*), 0101 (panel *C*), and 0111 (panel *D*) and from Xe1 to Xe2 for systems 1000 (panel *E*), 1010 (panel *F*), 1001 (panel *G*), and 1011 (panel *H*) for each point along the reaction coordinate .

Machine Learning

Machine learning models

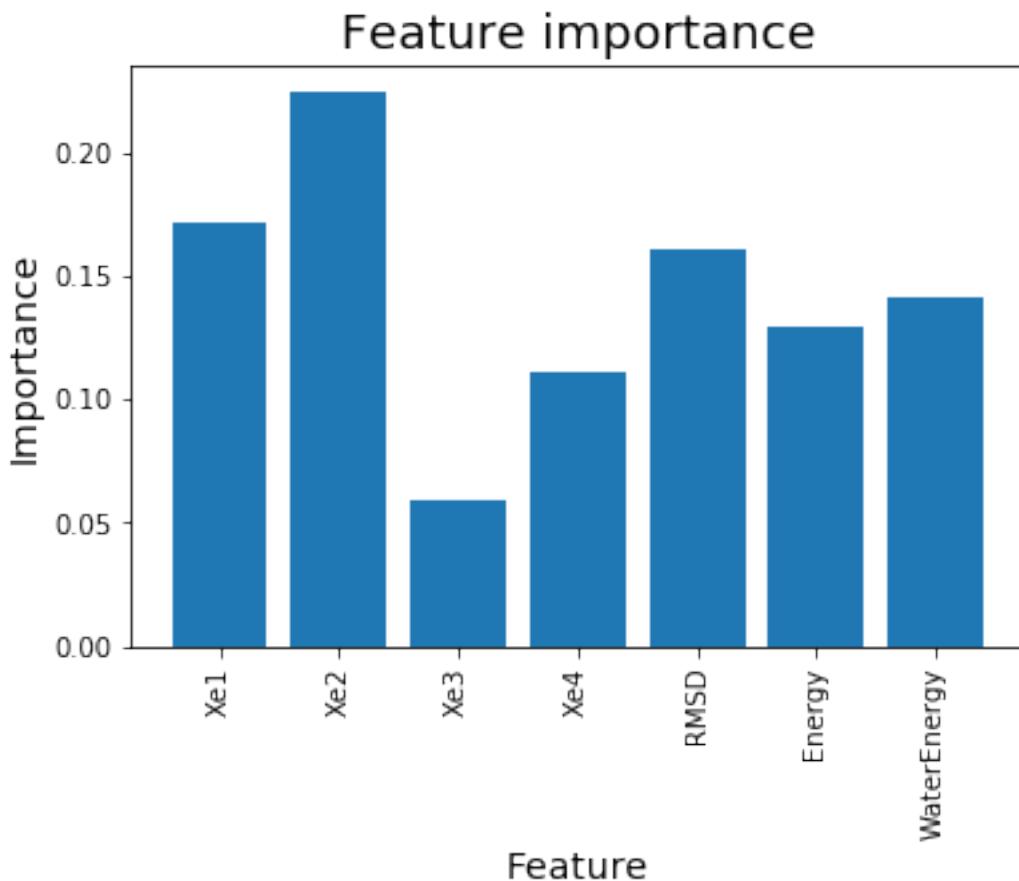


Figure 3: Feature importance weighting of Random Forest algorithm trained on Xe1 volume, Xe2 volume, Xe3 volume, Xe4, volume, protein RMSD, and protein energy contribution data calculated during Xe2 to Xe1 migration for 0100 US system.

Algorithm Comparison

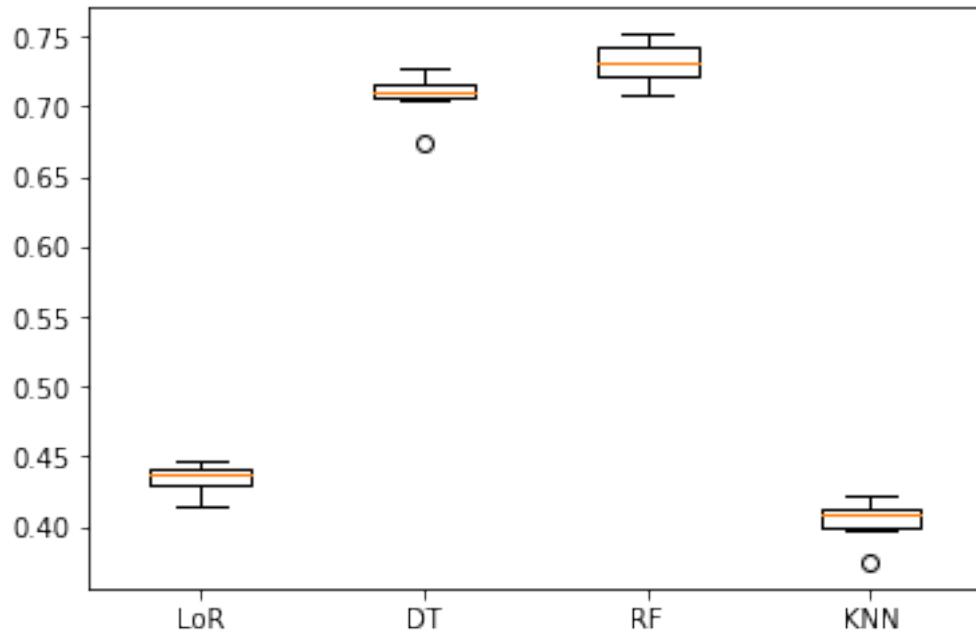


Figure 3: Box and whisker plot accuracy comparison of of Logistic Regression, Decision Tree Random Forest, and K-nearest neighbour machine learning models using K-fold cross validation (K=10) for protein model.

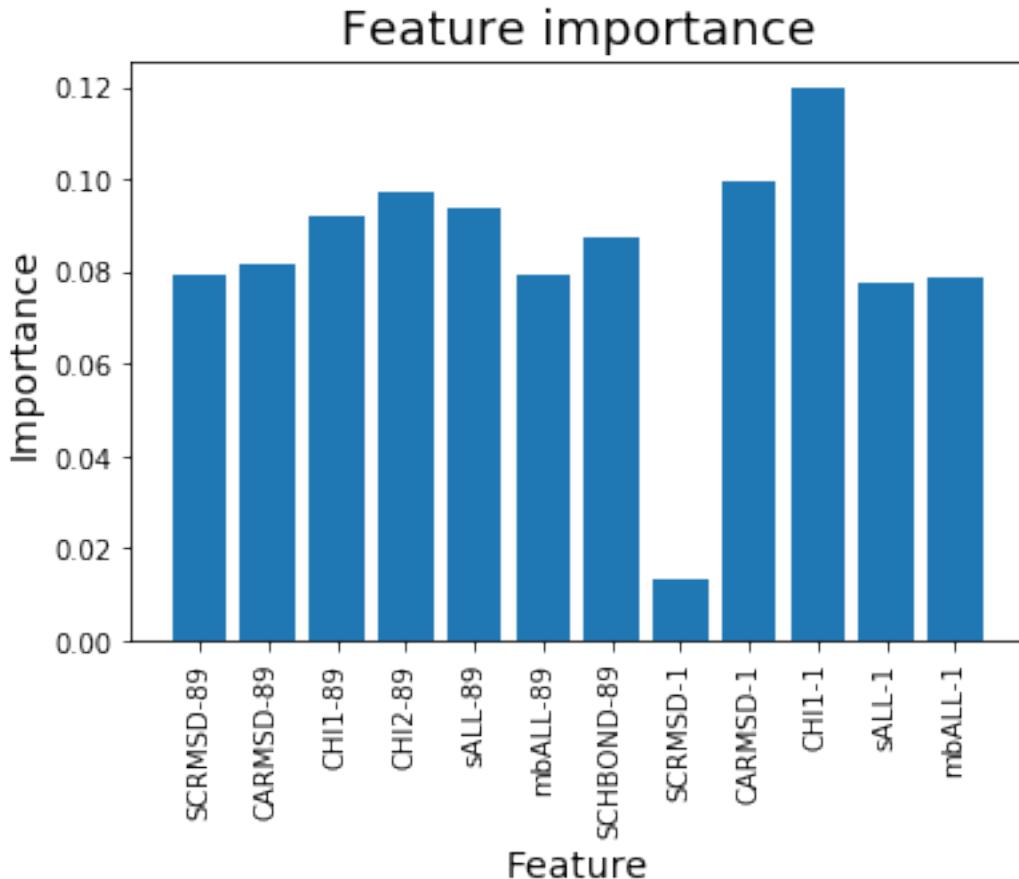


Figure 3: Bar graphs showing how the Random Forest weighted different aspects of Val1 and Leu89 in how the model predicted ligand position across migration route for the Xe2 to Xe1 migration for the 0100 umbrella simulation. These features are sidechain-RMSD (SCRMSD).carbon- α RMSD (CARMSD), χ_1 (CHI1), χ_2 (CHI2), residue torsion energy contributions, all the residues energy terms (SELFALL), All residue energy terms contributions to all other residues (MBALL), number of sidechain H-bonds (SCHBONDS).

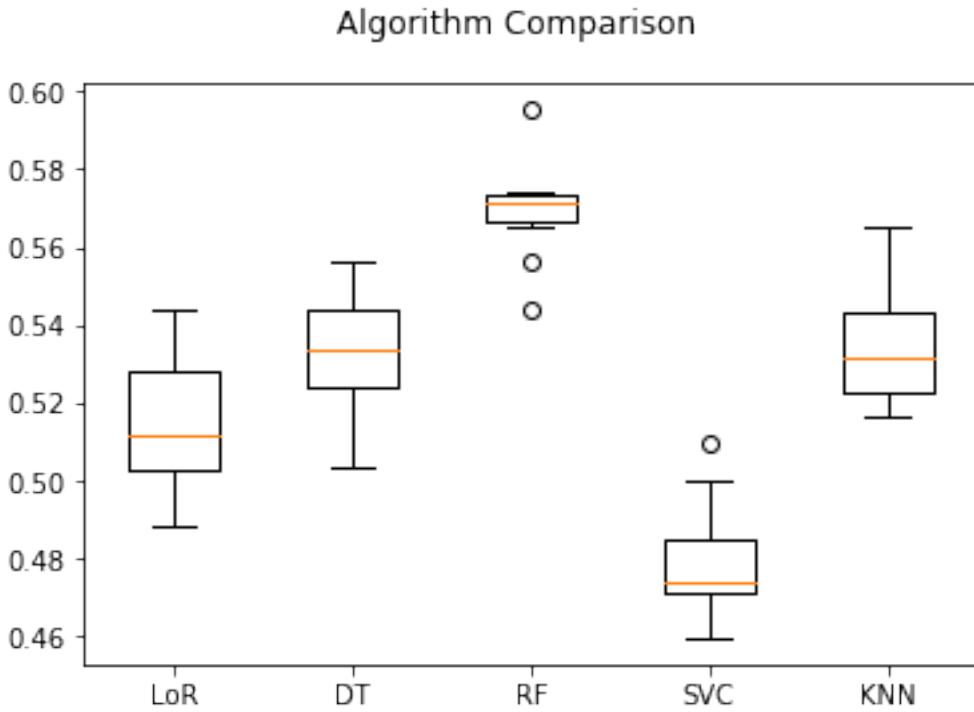


Figure 3: Box and whisker plot accuracy comparison of Logistic Regression, Decision Tree Random Forest, Support Vector Machine, and K-nearest neighbour machine learning models using K-fold cross validation ($K=10$) for residue model.

Conclusions

Discussion

The nature of studying allosteric behaviour in relation to Xe migration introduced a novel process for both computational and allosteric studies. This research inspected how shifting the position of Xe impacts the dynamics of Mb, precipitating a structurally driven understanding of allostery. Several types of computational analytical tools were implemented to probe protein structure and thermodynamics as a means to inspect the allosteric mechanism of Mb in response to Xe migration and occupancy.^{18,20,67} The investigation of Xe migration within Mb benefited from the and extensive kinetic and thermodynamic orientated analysis. Furthermore, that compared to conventional MD analysis, much of the data was represented

in reference to the position of Xe along a explicitly defined reaction coordinate (in the case of US systems),^{, 14,21,37,69} This allowed for the assessment of how the position of Xe altered the structural behaviour and thermodynamics of Mb.

To identify the effects of Mb migration and Xe occupancy the FEP for the Xe migration event was calculated for different Xe bound configurations. Figure 1 and 5 demonstrated that the FEP was modulated depending on the Xe migration direction and the Xe occupancy. The first row of Figure 1 and Figure 5 indicate that depending on which direction Xe migrated the ΔG varies. Table S1 showed that when no Xe was present in Xe3 and Xe4 the ΔG for Xe to migrate from Xe2 to Xe1 was smaller than for Xe to migrate from Xe1 to Xe2. However, for the Xe2 to Xe1 migration the barrier was more sensitive to Xe presence in Xe3 or Xe4, This can be seen as the ΔG for systems 0101, 0110, and 0111 are all higher than those for systems 1010, 1001, and 1011. This revealed that when Xe is present in Xe2, Xe3 and Xe4 are occupied Mb takes on physiognomies that effect the transition channel towards Xe1 differently than when Xe is present in Xe1, Xe3 and Xe4.¹⁸ This informed that depending on the Xe occupancy Mb captured different conformations which altered the ΔG between pockets, demonstrating that pocket occupancy triggers the allosteric mechanism of Mb.^{20,34} Furthermore, there was a consistent difference between the unbiased FEP and US calculated FEP for all systems. These differences were a consequence of the both the selection of the reaction coordinate and the sampling at each point. The US reaction coordinate sampled positions across Xe1 to Xe2 channel that deviate from the innate migration route of Xe, as represented by the unbiased FEP. This lead to Xe encountering resistance from residues lining the pocket, amplifying the energy calculations at those points,^{,13,22} as Xe was left to sample each position along the reaction coordinate for 250 ps during US it will iteratively interact with residues, accentuating the ΔG calculation at that point.^{,19,67,70}

This invited investigation of what changes in Mb contributed to the FEP profile as calcu-

lated by US FEP. For both US FEP and unbiased FEP the smallest ΔG was observed when no Xe was present in Xe3 and Xe4, and the highest was observed for when Xe resides in Xe3 and Xe4. This proposed that Xe presence in the pockets influenced the migration of atoms between channels in Mb. Specifically, that Xe presence in Mb restricted migration between pockets at other sites in the protein however, when Xe was absent those restrictions were relaxed. This suggested that the presence of Xe in other pockets had a regulatory role on Xe storage and diffusion through Mb.^{17,41,71} Furthermore, that there are critical residues along the migration pathway, such as Leu89 and Phe138, which were both conduits in facilitating atom migration between pockets and are triggered by the presence of Xe in Xe3 and Xe4.^{13,23,43} This was illustrated in row four and five of Figure 1 and Figure 5 by the χ_1 dihedral angle of both residues as they take on a distinctive shift when Xe was moving through the channel between Xe1 and Xe2. The distance between the two cavities for all systems was 7 Å. While the distance for X-ray determined distance between cavities was 7.5 Å the difference was likely due to the difference in environmental conditions between the two systems.⁵¹ The MD environment simulates MB when it was enveloped in solvent, causing Xe to localise differently within the pocket and Mb to take on unique conformations compared to those from the experimental X-ray studies.^{42,72}

For the Xe2 to Xe1 and Xe1 to Xe2 transition the volume of Xe3 and Xe4 was modulated by Xe presence configuration as in Table 4. The volume of Xe3 was 30% larger when Xe was present in Xe3 than when it was absent and that the volume of Xe2 and Xe4 was inflated by 25% only when Xe was absent from Xe3 but present in Xe4. Row two and three of Figure 5 illustrated that the size of any given cavity scatters about a consistent volume and only changes in response to the position of Xe along the reaction coordinate.⁶⁶ The extent to which the volumes of each pocket are altered by Xe occupancy as represented in row two and three of Figure 5 was severe. When Xe was present in Xe3 but absent from Xe4 the volume of Xe1 decreased by almost half. Additionally, that when Xe was absent from Xe3 but

present in Xe4 the volume of Xe3 was dramatically reduced but was substantially inflated for Xe1 and Xe2. When only Xe4 was occupied Xe1 and Xe2 possess a similar volume and behaviour as Xe transitions from Xe1 to Xe2. Together this indicates that the volume of Xe cavities was responsive to Xe migrating from Xe1 to Xe2. The reported volumes in Table 4 are consistent with both the experimental and theoretical calculations of cavities Xe1, Xe2, Xe3, and Xe4 for Mb.^{2,37,42,50}

The direction Xe was migrating and Xe occupancy adjust the χ_1 dihedral angles of residues Leu89, Leu104, Glu105, and Phe138 were altered. Figure 1 row four and five shows that Leu89 predominantly possessed a χ_1 angle of -150°, however this shifted to -80° in response to Xe moving through the Xe1 to Xe2 channel. Figure 5 row four and five showed that as Xe approached Xe2 Leu89 a took on different χ_1 of about -80° which shifts to -150° when Xe was present in the channel between Xe1 and Xe2. The χ_1 of Phe138 was also sensitive to the passage of Xe between Xe1 and Xe2. For the Xe1 to Xe2 transition, when Xe was absent from the migration channel Phe138 possessed a χ_1 of -170° or 170° . When Xe was present in the Xe1 to Xe2 migration channel the Phe138 χ_1 changed to 90°. Glu105 χ_1 primarily resided at χ_1 of -160° but for the 1011 system at about the transition point it changed to -50° in response to Xe migration. Together this indicated that Leu89 and Phe138 were critically involved in regulating atom motion between the two channels, effecting the ΔG of the migration. While all four residues reside along the migration channel not all have an effect on Xe migration between Xe1 and Xe2.^{13,51}

Investigations into the volumetric changes of Xe1, Xe2, Xe3, and Xe4 as a function of Xe position along the reaction coordinate demonstrated that there were structural changes which correlated with Xe migration.^{27,53,73} These findings were then supplemented with measurements of the average χ_1 of residues Leu89, Leu104, Glu105, and Phe138 which line the migration route, showing shifts about the TS along with shifts about a cleft along the

transition route specific to the direction Xe was migrating. As Xe is migrating from Xe2 to Xe1 both the US and unbiased FEPs illustrate a cleft at 6.6 Å, as seen in Figure 1 which was indicative of a potential subpocket local to Xe1. Figure 10 illustrated that as Xe approached Xe1 it was trapped behind the sidechains of Leu89, Leu104, and Ile142 at 6.6 Å along the migration pathway. For Xe to migrate to Xe1 the Leu89 sidechain methyl (circled in red) must first swing away from the pathway and into the outer solvent shell of the protein. This change in conformation can be observed in the fourth row of Figure 1 for systems 0100, 0110, and 0111 about 6.6 Å away from Xe2 (before the FEP peak descends) Leu89 takes on a shift in its χ_1 from -160° to -80°. This transient shift in the sidechain conformation allows Xe to pass between the residues and migrate into Xe1 and has been observed for experimental studies of Mb.⁶⁵

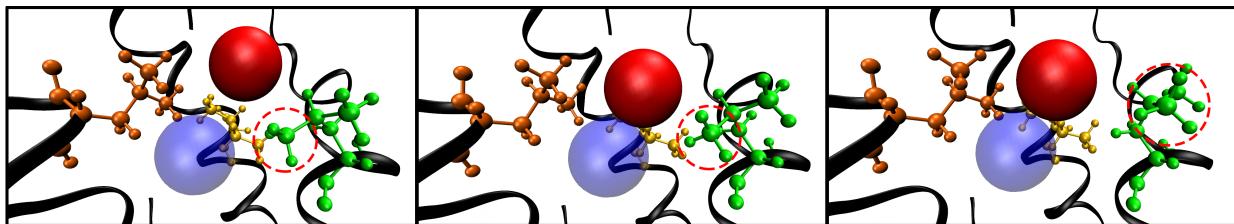


Figure 10: Sequential snapshots of the US dynamics for system 0100 showcasing the gate dynamics as Xe migrates towards Xe1. The methyl group of Leu89 is circled (*red dashed*) highlighting the change in orientation which allows Xe to pass. The sphere represents Xe (*red*), the transparent sphere as Xe1 (*blue*), the ribbon as Mb (*black*), and residues Leu89 (*green*), Leu104 (*orange*), and Ile142 (*yellow*).

Xe was occluded from Xe2 as it migrates from Xe1 to Xe2 as shown in the first row of Figure 5 by a cleft at 6.4 Å. Figure 11 showed that a gate is formed between the heme and the sidechains of Leu72 and Phe138. It was also shown that for Xe to migrate towards Xe2 the Phe138 sidechain (circled in red) needs to flip its orientation. This distinctive change in orientation can be observed in all systems for the χ_1 Phe138 takes on at about 6.6 Å from 170° to -90° in the fourth row of Figure 5.

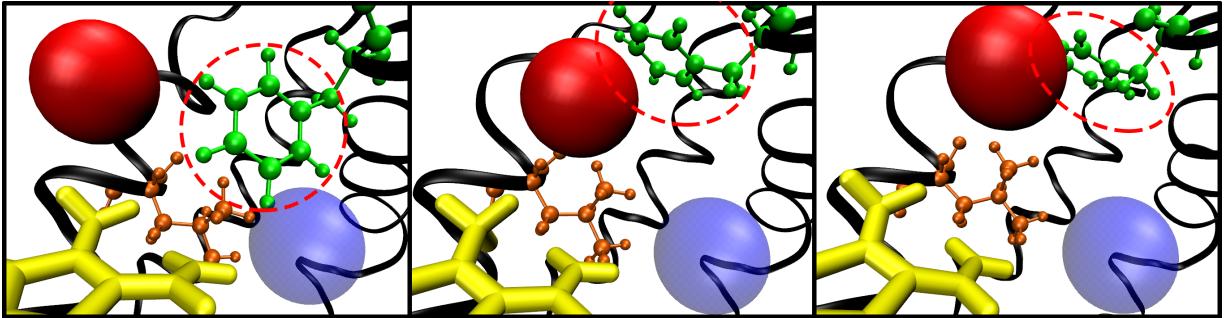


Figure 11: Sequential snapshots of the US dynamics for system 1000 showcasing the gate dynamics as Xe migrates towards Xe2. The phenyl group of Phe138 is circled (*red dashed*) highlighting the change in orientation which allows Xe to pass. The sphere represents Xe (*red*), the transparent sphere as Xe2 (*blue*), the ribbon as Mb (*black*), and residues Leu72 (*orange*), Phe138 (*green*), and Heme molecule (*yellow*).

Analysis of the sRMSD for each residue as a function of time was then calculated, as shown in the left panels of Figures 2 and 6. This analysis highlighted residues which show changes in their sRMSD at different points in response to Xe position along the reaction coordinate.¹⁰ The Xe occupation generally did not effect which residues and when they were involved along the reaction pathway however the magnitude of these changes was consistently greatest about the TS and for when Xe3 or Xe4 were occupied. This was then supplemented with investigation of the each residues U_{tors}^r as this energy term is dominated by sidechain contributions for residues located at the surface of Mb.^{7,19,53,74} Similar to the findings from the sRMSD the magnitude of the changes in U_{tors}^r for all residues was greatest when Xe was about the TS and when more Xe were present in Mb. The U_{LJ}^r and U_{elec}^r were then calculated, as seen in Figure 3 and 7 to broaden the search for residues which may contribute to describing the US FEP calculated and aid pinpointing the involvement of particular residues residues. Like the U_{tors}^r , the U_{LJ}^r was dominated by contributions from residues located on the surface and was also greatest when Xe was about the TS and more Xe occupied Mb pockets. The U_{elec}^r presented limited information regarding specific residue contributions in response to Xe binding, this is likely due to Xe not exhibiting an electrostatic energy.^{7,13,25} Finally, the intramolecular H-bonding analysis was the most discrete of all, identifying key residues throughout all the analysis from the χ_1 , to sRMSD, to U_{tors}^r , to U_{LJ}^r , and to U_{elec}^r . The

Critical residues for different points along the reaction coordinate are that were highlighted to be important by this analysis are listed in the table 4.

Table 4: Residues analytically identified to contribute towards the ΔG as Xe is migrating between Xe1 and Xe2 for different stages during the migration. Superscript *a* denotes those contributing to the ΔG during Xe migration from Xe2 to Xe1, and *b* denotes those contributing to the ΔG during Xe migration from Xe1 to Xe2.

Migration stage	Residue
Reactant	Trp7 ^a , Glu18 ^a , Asp27 ^a , His36 ^b , Pro37 ^a , Arg45 ^{a,b} , Asp60 ^b , Glu85 ^b Asn132 ^a , Lys140 ^a
Transition	Glu6 ^a , Trp7 ^b , Asp27 ^b , Arg31 ^b , His36 ^b , Pro37 ^{a,b} , Arg45 ^a , Lys56 ^b , Lys62 ^b , Pro88 ^a , Glu85 ^b , Leu89 ^b Leu104 ^{a,b} , Ala127 ^a , Met131 ^a , Leu135 ^b , Leu137 ^b , Phe138 ^b , Arg139 ^b
Product	Glu6 ^a , Arg45 ^{a,b}

The residues listed in Table 4 have been identified as critical during Xe migration between Xe1 to Xe2, in that they presented quantifiable structural and thermodynamic changes in response to Xe migration. To qualify as a notable structural change a given residue would have to exhibit a notable sRMSD, intramolecular H-bonding, or χ_1 while to qualify as a thermodynamic change a residue would have to exhibit a notable U_{tors}^r , U_{LJ}^r , and U_{elec}^r - needing at least one of both and three altogether. Some of these residues have also been identified as critical during Mb interaction between O₂ and CO, such as residues Arg45, Lys56, Leu137, Phe138.^{13,27,37} While my analysis residues which structurally and thermodynamically responded to Xe migration, it is not clear where they present a cumulative effect and require a specific network to exhibit their allosteric behaviour. These findings could be better resolved through mutational studies which probe how the thermodynamics of Mb is effected when the highlighted proteins are perturbed.

Not all the analytical tools yielded information which was sensitive to Xe migration they

aided in orientating the analytical process. Figure S4 panel A demonstrates that protein RMSD for all systems within a 0.7 Å to 0.9 Å range when compared to the crystal structure while panels B through to I show that the dRMSD for each α helical domain consistently resided within a given range specific to that domain. Although this data did not provide allosteric insight it demonstrated that while the system maintains the physiological structure of Mb the RMSD at a protein and domain level does show sensitivity to Xe occupancy or migration. A residue level secondary structure analysis of Xe impact was pursued as shown in S5. Similar to S4, this analysis exhibited that the overall structure of Mb did not differ from the crystal structure. Although there were changes in the secondary structure assignment of residues these should be interpreted as stochastic shifts at the terminal ends of α helical domains (as can be seen for helical domains α 5 and α 6) which have been demonstrated to be the most subject to secondary structure change.^{5,10,33}

The dihedral angle between α helical domains was calculated as seen in Figure S6 to identify if there was a geometric reorganisation among the domains in response to Xe occupancy and migration. The dihedral angles c and d did demonstrate a response to Xe migration changing from for all systems in Panel A from -180° to about -45° when Xe was to 2 Å, 4 Å to 6 Å and 8 Å along the reaction coordinate. These positions correlate with Xe leaving Xe2, nearby the TS of the Xe2 to Xe1 migration, and when Xe approached Xe1. This suggested that α 3 and α 4 rearrange in response to Xe migration from Xe2 to Xe1. While panel B showed that for dihedral c only system 1010 demonstrated a dihedral shift at 2 Å, 4 Å to 6 Å and 8 Å along the reaction pathway, while for system 1000 and 1011 a change was seen at 2 Å and 4 Å to 6 Å, while system 1001 only showed a change at 8 Å. Together this showed that Xe migration influences α 3 and α 4 formation when Xe was leaving Xe1 and Xe4 was not occupied. While there was Xe migration correlated behaviour it is not clear how these domains are involved in managing Xe migration at this level and would be difficult to model allosteric action in reference to these domain level changes.

The analysis of the unbiased simulations supported but also complicated the findings from the US systems. Figure S1 and S2 show that the transition state and distribution of transition times for Xe from Xe2 to Xe1 or Xe1 to Xe2 varies widely for unbiased systems. The systems which take longer transition times (0110, 0111, 1010, 1011) do not form normal distributions but this is likely due to the paucity of unbiased simulations as reported in Table S3. The variety of different transition times is indicative of Xe taking different migration routes through the same Xe1 to Xe2 channel. In general, the transition times do present distributions about the average transition time (Table S3) and according to transition state theory.^{75,76} Furthermore, that the differences in the ΔG , pocket volume, and average χ_1 between unbiased and US FEP as seen in rows Figures 1 and 5 are a consequence of both the selection of reaction pathway (for the US systems) and culmination of different reaction pathways (for the unbiased systems).

In light of the differences between the unbiased and US systems downhill simulations were pursued to identify whether the location of the TS for systems of the same Xe occupancy but different simulation type also varied. Table S8 shows that as the Xe position along the reaction coordinate moves further from Xe2 more Xe populate Xe1, however the location of the TS varies for each system. A TS was identified by where there was a close to 50% split in the distribution of Xe. For US system 0100, 0110, and 0101 it resides 4.6 Å, while for system 0111 it resides at 4.8 Å. This was in agreement with the energy barrier peaks along the reaction coordinate as seen in Figure 1. For system 0110 the distribution of Xe favours Xe2 more at 5.0 Å than at 4.8 Å, this was a consequence of a energy plateau between these two coordinates as the difference was 2.1%. Table S9 shows that for unbiased systems 0100 and 0110 the TS was at 4.6 Å and for systems 0101 and 0111 at 4.8 Å along the reaction coordinate. When compared to the TS taken from US reaction coordinates the TS are closer to Xe2 but a similar trend can be seen, that when Xe4 it occupied the TS shifts 0.2 Å

towards Xe1. Together this indicates that the TS scarcely shifts between the unbiased and US systems for the Xe2 to Xe1 transition.

The downhill simulation comparison was also repeated for the Xe1 to Xe2 transition for unbiased Systems. Table S10 shows that for systems 1000, 1010 the TS was about 5.0 Å from Xe1, and for 1001 and 1011 it resides at 4.8 Å. While all the TSs are in near agreement with the energy barrier peaks in Figure 5 the position of coordinate of the TS was modulated by Xe4 occupancy, which shifts the TS 0.2 Å closer to Xe1. Interestingly, for system 1010 Xe favours Xe1 at 5.2 Å more than at 5.0 Å by 7.8%, this was indicative of a potential saddle point in the energy barrier as can be seen in the 1010 unbiased FEP curve of Figure 5. There are slight differences between the distribution of Xe for the Xe1 to Xe2 migration initiated from unbiased simulations compared to the US distribution. Table S11 shows that the TSs are within 0.2 Å of each other, for system 1000 and 1011 it was at 4.8 Å and system 1010 and 1001 it was at 5.0 Å. This suggests that there is a characteristic TS when Xe3 and Xe4 are both either empty or occupied by Xe however when only one or the other is occupied the TS shifts towards Xe2.

the relative differences between ΔG , as seen in Figures 1 and 5, for different systems during the Xe1 and Xe2 transition at times were as small as 1 kcal/mol. Given the CHARMM22 force-field errors it would be difficult to confer whether the differences between these barriers are thermodynamically different. However, it has been shown that the CHARMM22 force-field is robust for solvated Mb systems and that FEPs estimated from US or implicit ligand sampling have operated within sub 1 kcal/mol ranges and yielded similar FEP ranges for Xe and other physiological Mb molecules.^{13,22,25} It should also be noted that much of the analysis is orientated around the US systems and as such a reaction coordinate had to be predetermined which contributed to differences between the unbiased and US FEP. However While a number of analytical tools have been utilised during this research only some of which

have been suited for the question. This has been made apparent by the PCA shown in S7, in the case that the analytical techniques described above were sufficient to describe the first few eigenvectors, to capture 60% of the variance in the data at least seven eigenvalues need to be accounted for. This indicates that there is still much work to be done in precisely and comprehensively mapping the allosteric mechanism of Mb.

This work highlights how a specific analytical toolkit is required for identifying allosteric mechanisms within Mb. Analysis which considers secondary structure and α helical domain reorganisation did not yield quantifiable changes as Xe migrated between Xe1 and Xe2. This indicates that unlike, for example TetR family,²⁰ where its allosteric function is dependent on these protein features, Mb requires a different analytical method. This proposes that the allosteric function of proteins whose physiological molecule is oligomeric exhibit allosteric activity different to those which are functional as monomers, and therefore require a specific analytical lens.

Together these findings make can be reduced to three scientific contributions.

1. Proposition of a method for resolving allostery that synergistically depends on system, protein, domain, secondary structure, residue, and side chain resolution analysis to make allosteric inference.
2. Identified a computational toolkit which is necessary for probing the structural and thermodynamic properties of Mb allosteric mechanisms.
3. Pinpointed elements of Mb which are likely to play a role in the allosteric mechanism of Mb and sorted them into different functional groups.

These findings were supported by unbiased simulations and comparison to previous research. Similar FEPs for each system were calculated from unbiased simulations, indicating that the US FEP is reflective of the Xe migration from Xe2 to Xe1 but the energetic barrier has been

overestimated as can be the case with US calculated FEPs.^{13,22} Furthermore the distribution of TS is in agreement with the average transition time as seen in Table S1 and Figure S2 for specific ΔG as informed by transition state theory.

Conclusion

This work presents a systematic and analytical approach to identifying which residues in Mb are involved in managing atom diffusion and storage between Xe1 and Xe2. FEP curves are used to identify the energetic barriers between Xe1 and Xe2, how they're modulated by Xe presence in other cavities, and if the barriers are the same in both directions. It has been shown that the ΔG is heightened upon Xe presence in other pockets, and these findings are consistent with those found from the unbiased FEP albeit the magnitude differs. It has also been demonstrated that the ΔG itself is dependent on the direction which Xe is migrating.

This led to inquiries of what structural components are responsible for i) the ΔG profile and ii) the difference in profile depending on migration direction. There has been analysis of the mechanical relationship between the volume of cavities Xe1, Xe2, Xe3, and Xe4. For the Xe2 to Xe1 only when Xe3 and Xe4 are both either empty or occupied does Xe1, Xe3, and Xe4 respond to Xe migration by increasing in volume with ΔG . While from Xe1 to Xe2 only when either Xe3 or Xe4 are occupied is there a migration related increase in the cavity volumes in relation to the ΔG . Both these examples suggest that it is Xe pockets are required to be occupied in order to trigger a breathing motion in response to Xe migration between Xe1 and Xe2.⁷⁷

The first point and second were also addressed through the χ_1 angle analysis. It has been shown that residues Leu89 and Phe138 along the reaction pathway are involved in constitu-

tion the ΔG . They both take on characteristic shifts in their χ_1 in response to increases as well as Xe position along Xe1 and Xe2 migration channel. However for the Xe2 to Xe1 migration Leu89 and Phe138 take on a χ_1 to -170° for longer periods and more consistently, indicating a potential source for the difference in ΔG .

The characterisation of the subpockets which manifest along the Xe1 and Xe2 channel were shown in both the FEP and χ_1 data. Further inspection of the structural region allowed for a more precise description of the gating mechanisms regulate Xe diffusion along the Xe1 and Xe2 channel.

To more precisely identify which residues play a role in the difference in Xe dynamics as it passes through the Xe1 to Xe2 channel through different directions the sRMSD, U_{tors} , U_{LJ} , and U_{elec} for all residues was calculated. This provided unique insight, independent of the previous results, which help nominate other residues which respond to Xe migration, allowing for the grouping of residues depending on their location and response to Xe migration as seen in Figure 4 and 8.

Together this research impinges on important questions regarding Mb and allosteric. Through computational investigations it has resolved which regions of Mb respond to Xe migration through systematic analysis of the structural and energetic components of the transition. Furthermore, that the allosteric networks within Mb are vast and require careful analysis to tease apart which is aided by a multi-dimensional analytical approach. This has been useful when inquiring upon the allosteric of a protein and trying to identify sites which function in the allosteric network of Mb.

Supplementary information

Table S1: Number of unbiased simulations (η) analysed for the FEP curve, average transition time \bar{X} (ps) and standard deviation, as well as ΔG (kcal/mol) for unbiased and US simulations for systems involved in Xe Xe2 to Xe1 and Xe1 to Xe2 transition.

Xe2 to Xe1 FEP					Xe1 to Xe2 FEP				
System	η	\bar{X}	ΔG	ΔG (US)	System	η	\bar{X}	ΔG	ΔG (US)
0100	71	46.1 ± 22.9	1.4	2.0	1000	73	52.9 ± 70.5	1.8	3.1
0110	74	473.2 ± 170.9	3.8	5.3	1010	71	150.8 ± 113.6	2.9	4.5
0101	70	164.9 ± 81.5	3.3	4.4	1001	72	74.2 ± 59.8	2.0	3.0
0111	72	489.7 ± 234.5	4.2	5.4	1011	72	177.2 ± 103.7	3.4	4.2

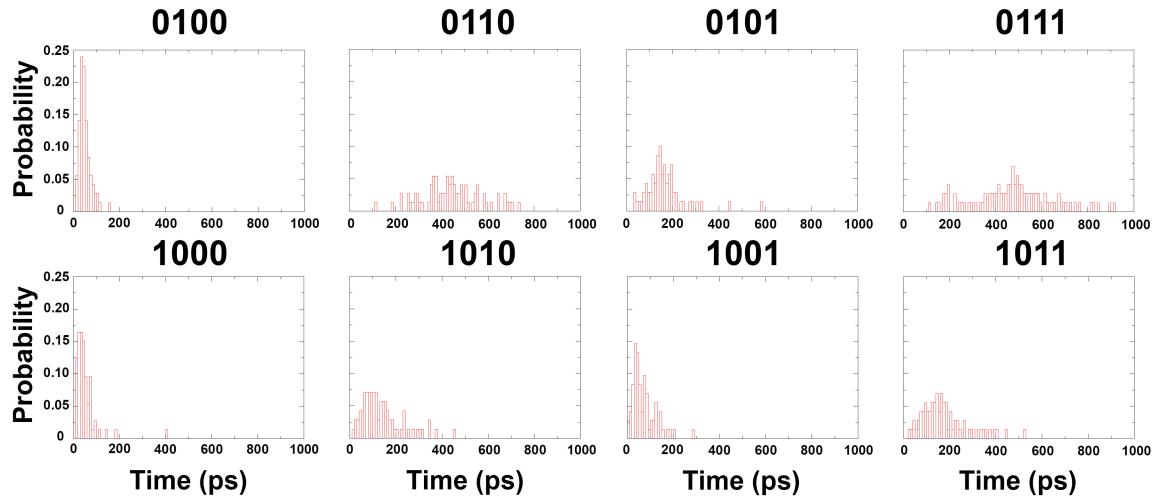


Figure S2: Histograms of distribution of transition times rounded to the nearest 10 ps for Xe migrations for Xe2 to Xe1 and Xe1 to Xe2 systems.

Table S3: Average volume of cavities Xe1, Xe2, Xe3, and Xe4 along with standard deviation as Xe transitions from Xe2 to Xe1 and Xe1 to Xe2 for various Xe systems.

	Xe2 to Xe1 transition \bar{X} cavity size (\AA^3)				Xe1 to Xe2 transition \bar{X} cavity size (\AA^3)			
	0100	0110	0101	0111	1000	1010	1001	1011
Xe1	72 \pm 23	65 \pm 17	72 \pm 19	70 \pm 20	101 \pm 32	69 \pm 21	96 \pm 30	87 \pm 25
Xe2	49 \pm 18	48 \pm 13	60 \pm 18	47 \pm 13	51 \pm 21	61 \pm 23	98 \pm 30	50 \pm 15
Xe3	126 \pm 37	145 \pm 32	118 \pm 37	141 \pm 47	181 \pm 55	163 \pm 51	158 \pm 31	181 \pm 34
Xe4	66 \pm 26	60 \pm 23	82 \pm 23	66 \pm 18	70 \pm 26	55 \pm 19	61 \pm 27	80 \pm 20

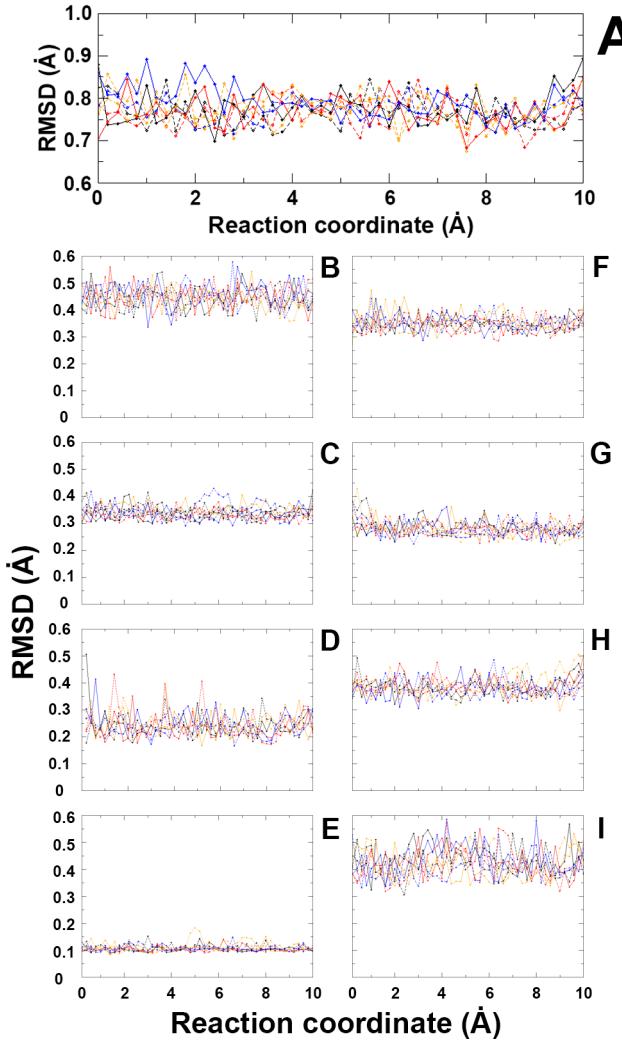


Figure S4: Panel A shows all atom RMSD between US systems 0100, while panels B through to I exhibits the all atom dRMSD for domains $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7$, and α_8 as defined in Table respectively with the crystal structure as reference (PDB: 2W6W)⁴² for each window across Xe1 to Xe2 migration route. Each system is represented by a different colour and line appearance as follows: (solid black), 0110 (solid red), 0101 (solid blue), 0111 (solid orange), 1000 (dashed black), 1010 (dashed red), 1001 (dashed blue), and (dashed orange).

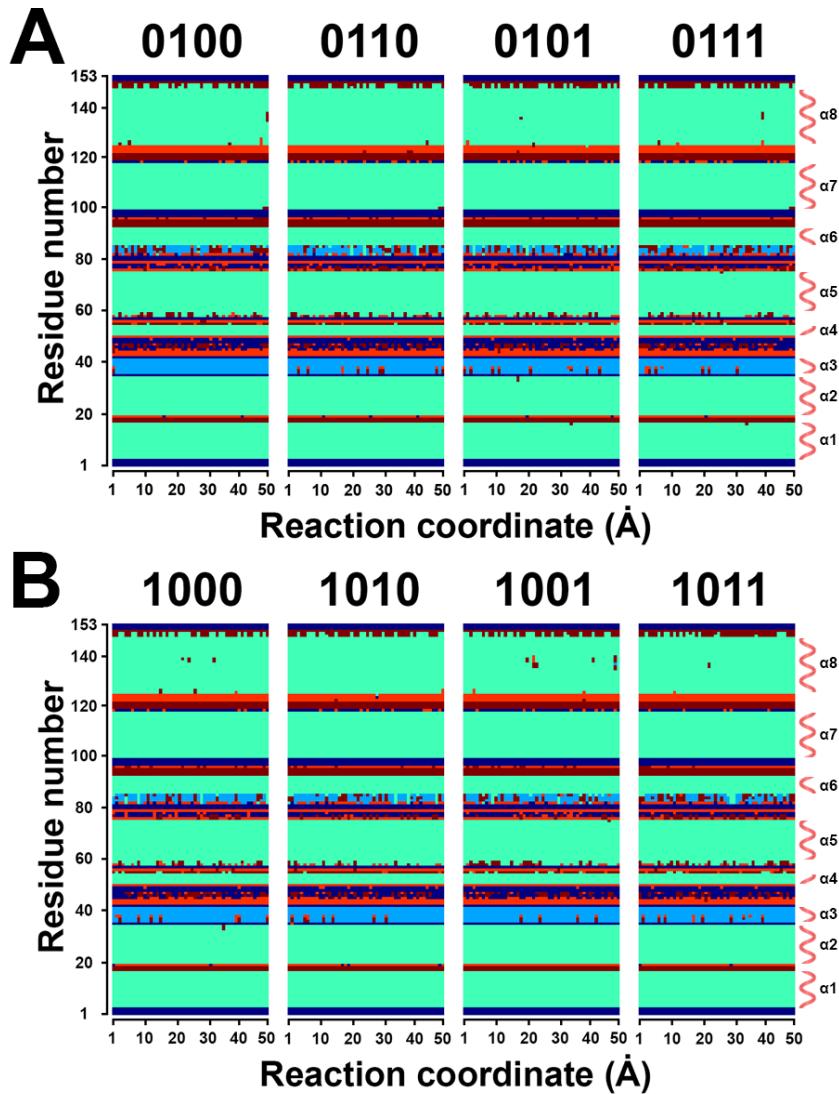


Figure S5: Panel A shows secondary structure assignment for each residue at each window along the reaction coordinate for as Xe migrates between Xe2 to Xe1 for US systems with the experimentally observed alpha helical domain definitions illustrated alongside while Panel B shows the same but for the Xe1 to Xe2 migration. Each colour corresponds to a different secondary structure, for coil (*blue*), α -helix (*turquoise*), turn (*maroon*), bend (*orange*), and 3-helix (*light blue*).

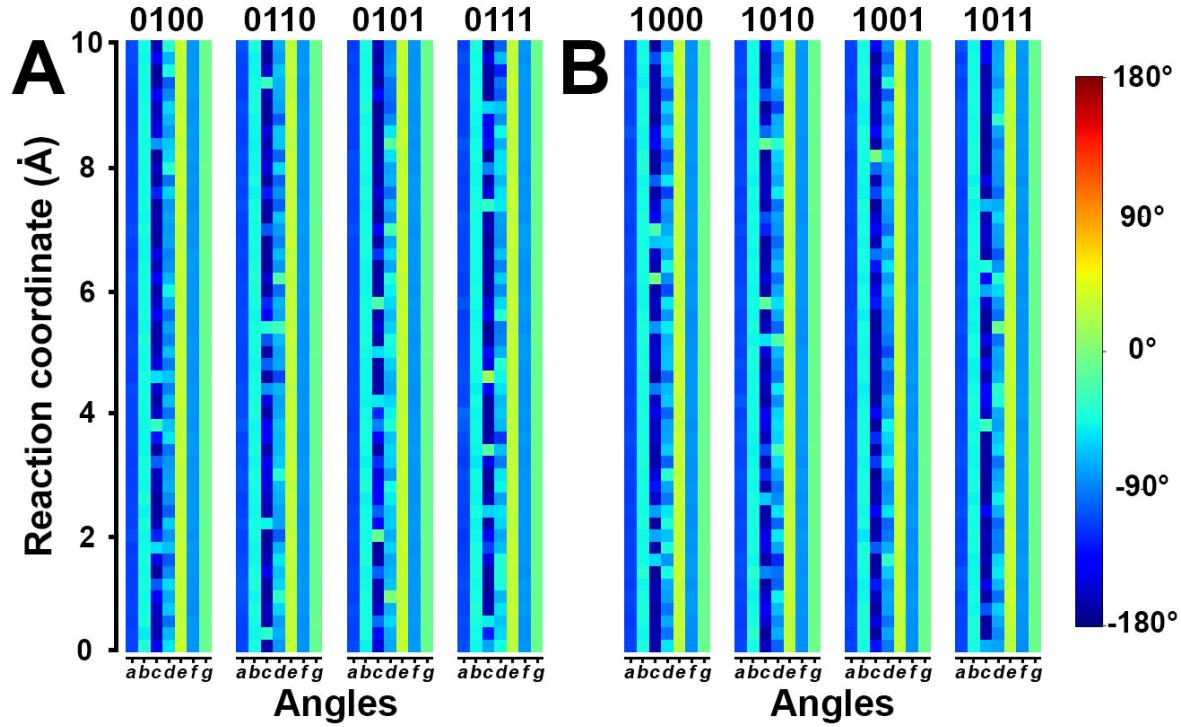


Figure S6: Panel A displays the dihedral angle between consecutive alpha helical domains for each system for Xe at each point along the Xe1 to Xe2 reaction coordinate for the Xe2 to Xe1, and panel B the same but for the Xe1 to Xe2 migration. *a* as the dihedral between α_1 and α_2 , *b* as the dihedral between α_2 and α_3 , *c* as the dihedral between α_3 and α_4 , *d* as the dihedral between α_4 and α_5 , *e* as the dihedral between α_5 and α_6 , *f* as the dihedral between α_6 and α_7 , and *g* as the dihedral between α_7 and α_8 .

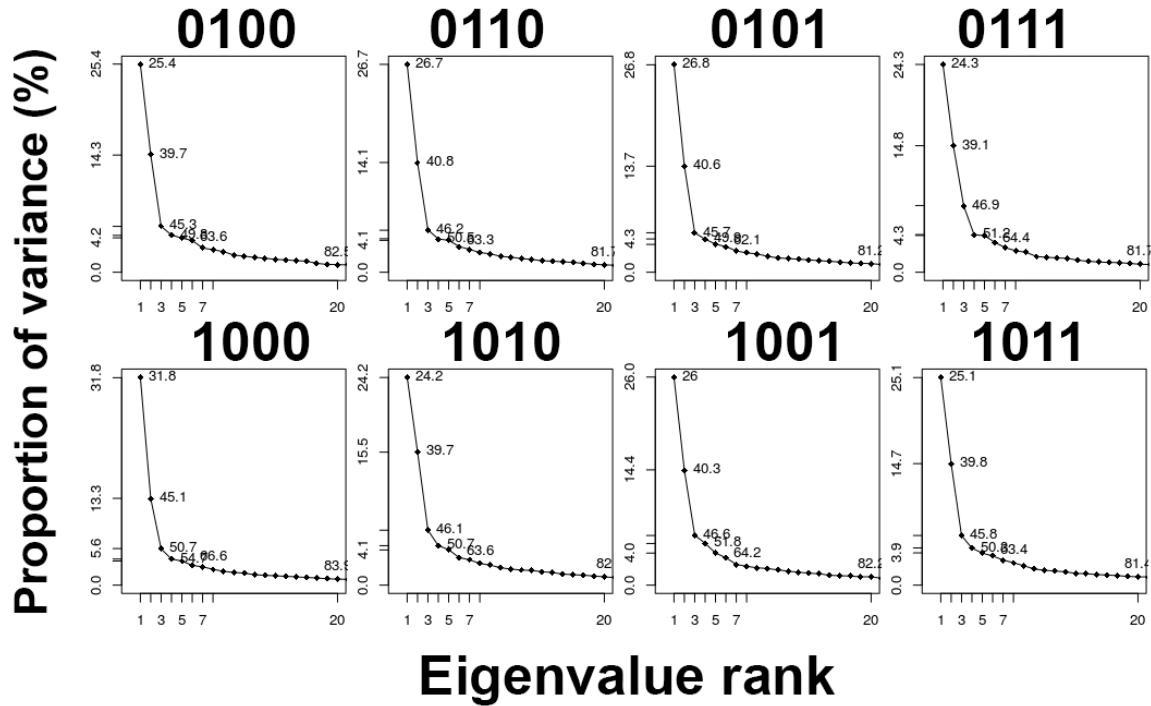


Figure S7: PCA calculated using Mb C α for each Mb system.

Acknowledgments

This work was supported by the Swiss National Science Foundation, NCCR MUST, and the University of Basel.

References

- (1) Giordano, D.; Boron, I.; Abbruzzetti, S.; van Leuven, W.; Nicoletti, F. P.; Forti, F.; Bruno, S.; Cheng, C. H.; Moens, L.; di Prisco, G. et al. Biophysical Characterisation of Neuroglobin of the Icefish, a Natural Knockout for Hemoglobin and Myoglobin. Comparison with Human Neuroglobin. *PLoS ONE* **2012**, *7*.
- (2) Anselmi, M.; Di Nola, A.; Amadei, A. The effects of the L29F mutation on the ligand and migration kinetics in crystallized myoglobin as revealed by molecular dynamics simulations. *Proteins* **2011**, *79*, 867–879.
- (3) Šrajer, V.; Ren, Z.; Teng, T. Y.; Schmidt, M.; Ursby, T.; Bourgeois, D.; Pradervand, C.; Schildkamp, W.; Wulff, M.; Moffat, K. Protein conformational relaxation and ligand migration in myoglobin: A nanosecond to millisecond molecular movie from time-resolved Laue X-ray diffraction. *Biochem.* **2001**, *40*, 13802–13815.
- (4) Brunori, M.; Vallone, B. Neuroglobin, seven years after. *Cell. Mol. Life Sci.* **2007**, *64*, 1259–1268.
- (5) Bourgeois, D.; Vallone, B.; Schotte, F.; Arcovito, A.; Miele, A. E.; Sciara, G.; Wulff, M.; Anfinrud, P.; Brunori, M. Complex landscape of protein structural dynamics unveiled by nanosecond Laue crystallography. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 8704–8709.
- (6) Parak, F. G.; Nienhaus, G. U. Myoglobin, a paradigm in the study of protein dynamics. *ChemPhysChem* **2002**, *3*, 249–254.

- (7) Tilton, R. F.; Singh, U. C.; Weiner, S. J.; Connolly, M. L.; Kuntz, I. D.; Kollman, P. A.; Max, N.; Case, D. A. Computational studies of the interaction of myoglobin and xenon. *J. Mol. Biol.* **1986**, *192*, 443–456.
- (8) Hummer, G.; Schotte, F.; Anfinrud, P. A. Unveiling functional protein motions with picosecond x-ray crystallography and molecular dynamics simulations. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 15330–15334.
- (9) Hamdane, D.; Kiger, L.; Dewilde, S.; Uzan, J.; Burmester, T.; Hankeln, T.; Moens, L.; Marden, M. C. Hyperthermal stability of neuroglobin and cytoglobin. *FEBS J.* **2005**, *272*, 2076–2084.
- (10) ElGamacy, M.; Riss, M.; Zhu, H.; Truffault, V.; Coles, M. Mapping Local Conformational Landscapes of Proteins in Solution. *Structure* **2019**, *27*, 853–865.e5.
- (11) Traut, T. A vertebrate globin expressed in the brain. *Chemtracts* **2001**, *14*, 328–331.
- (12) Ostermann, A.; Waschipky, R.; Parak, F. G.; Nienhaus, G. U. Ligand binding and conformational motions in myoglobin. *Nature* **2000**, *404*, 205–208.
- (13) Cohen, J.; Arkhipov, A.; Braun, R.; Schulten, K. Imaging the migration pathways for O₂, CO, NO, and Xe inside myoglobin. *Biophys. J.* **2006**, *91*, 1844–1857.
- (14) Nienhaus, K.; Deng, P.; Olson, J. S.; Warren, J. J.; Nienhaus, G. U. Structural Dynamics of Myoglobin: Ligand migration and binding in valine 68 mutants. *J. Biol. Chem.* **2003**, *278*, 42532–42544.
- (15) Olson, J. S.; Soman, J.; Phillips, G. N. Ligand pathways in myoglobin: A review of trp cavity mutations. *IUBMB Life* **2007**, *59*, 552–562.
- (16) Scott, E. E.; Gibson, Q. H.; Olson, J. S. Mapping the Pathways for O₂ Entry Into and Exit from Myoglobin. *J. Biol. Chem.* **2001**, *276*, 5177–5188.

- (17) Frauenfelder, H.; McMahon, B. H.; Austin, R. H.; Chu, K.; Groves, J. T. The role of structure, energy landscape, dynamics, and allostery in the enzymatic function of myoglobin. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 2370–2374.
- (18) Cui, Q.; Karplus, M. Allostery and cooperativity revisited. *Protein Sci.* **2008**, *17*, 1295–1307.
- (19) Lee, M. W.; Meuwly, M. Molecular dynamics simulation of nitric oxide in myoglobin. *J. Phys. Chem. B* **2012**, *116*, 4154–4162.
- (20) Motlagh, H. N.; Wrabl, J. O.; Li, J.; Hilser, V. J. The ensemble nature of allostery. *Nature* **2014**, *508*, 331–339.
- (21) Nutt, D. R.; Meuwly, M. Studying reactive processes with classical dynamics: Rebinding dynamics in MbNO. *Biophys. J.* **2006**, *90*, 1191–1201.
- (22) Plattner, N.; Meuwly, M. Quantifying the importance of protein conformation on ligand migration in myoglobin. *Biophys. J.* **2012**, *102*, 333–341.
- (23) Nutt, D. R.; Meuwly, M. CO migration in native and mutant myoglobin: Atomistic simulations for the understanding of protein function. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 5998–6002.
- (24) Schmidt, M.; Nienhaus, K.; Pahl, R.; Krasselt, A.; Anderson, S.; Parak, F.; Nienhaus, G. U.; Srager, V. Ligand migration pathway and protein dynamics in myoglobin: A time-resolved crystallographic study on L29W MbCO. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 11704–11709.
- (25) Diamantis, P.; Unke, O. T.; Meuwly, M. Migration of small ligands in globins: Xe diffusion in truncated hemoglobin N. *PLoS Comput. Biol.* **2017**, *13*, 1–22.
- (26) Wang, Y.; Baskin, J. S.; Xia, T.; Zewail, A. H. Human myoglobin recognition of oxygen: Dynamics of the energy landscape. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 18000–18005.

- (27) Scott, E. E.; Gibson, Q. H. Ligand migration in sperm whale myoglobin. *Biochem.* **1997**, *36*, 11909–11917.
- (28) Kim, S.; Lim, M. Protein conformation-controlled rebinding barrier of NO and its binding trajectories in myoglobin and hemoglobin at room temperature. *J. Phys. Chem. B* **2012**, *116*, 5819–5830.
- (29) Vallone, B.; Nienhaus, K.; Matthes, A.; Brunori, M.; Nienhaus, G. U. The structure of carbonmonoxy neuroglobin reveals a heme-sliding mechanism for control of ligand affinity. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 17351–17356.
- (30) Brittain, T. The Anti-Apoptotic Role of Neuroglobin. *Cells* **2012**, *1*, 1133–1155.
- (31) Greenberg, D. A.; Jin, K.; Khan, A. A. Neuroglobin: an endogenous neuroprotectant. *Curr. Opin. Pharmacol.* **2008**, *8*, 20–24.
- (32) Maragliano, L.; Cottone, G.; Ciccotti, G.; Vanden-eijnden, E. Mapping the Network of Pathways of CO Diffusion in Myoglobin. *J. Am. Chem. Soc.* **2010**, *132*, 1010–1017.
- (33) Brunori, M.; Vallone, B.; Cutruzzola, F.; Travaglini, A.-C.; Berendzen, J.; Chu, K.; Sweet, R. M.; Schlichting, I. The role of cavities in protein dynamics: Crystal structure of a photolytic intermediate of a mutant myoglobin. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 2058–2063.
- (34) Frauenfelder, H.; McMahon, B. H.; Fenimore, P. W. Myoglobin: The hydrogen atom of biology and a paradigm of complexity. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 8615–8617.
- (35) Nishihara, Y.; Sakakura, M.; Kimura, Y.; Terazima, M. The escape process of carbon monoxide from myoglobin to solution at physiological temperature. *J. Am. Chem. Soc.* **2004**, *126*, 11877–11888.

- (36) Chu, K.; Vojtchovský, J.; McMahon, B. H.; Sweet, R. M.; Berendzen, J.; Schlichting, I. Structure of a ligand-binding intermediate in wild-type carbonmonoxy myoglobin. *Nature* **2000**, *403*, 921–923.
- (37) Bossa, C.; Amadei, A.; Daidone, I.; Anselmi, M.; Vallone, B.; Brunori, M.; Nola, A. D. Molecular dynamics simulation of sperm whale myoglobin: Effects of mutations and trapped CO on the structure and dynamics of cavities. *Biophys. J.* **2005**, *89*, 465–474.
- (38) Desdouits, N.; Nilges, M.; Blondel, A. Principal Component Analysis reveals correlation of cavities evolution and functional motions in proteins. *J. Mol. Graph. Model.* **2015**, *55*, 13–24.
- (39) Huang, X.; Boxer, S. G. Discovery of new ligand binding pathways in myoglobin by random mutagenesis. *Nat. Struct. Biol.* **1994**, *1*, 226–229.
- (40) Bourgeois, D.; Vallone, B.; Arcovito, A.; Sciara, G.; Schotte, F.; Anfinrud, P. A.; Brunori, M. Extended subnanosecond structural dynamics of myoglobin revealed by Laue crystallography. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 4924–4929.
- (41) Ruscio, J. Z.; Kumar, D.; Shukla, M.; Prisant, M. G.; Murali, T. M.; Onufriev, A. V. Atomic level computational identification of ligand migration pathways between solvent and binding site in myoglobin. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 9204–9209.
- (42) Savino, C.; Miele, A. E.; Draghi, F.; Johnson, K. A.; Sciara, G.; Brunori, M.; Vallone, B. Pattern of cavities in globins: The case of human hemoglobin. *Biopolymers* **2009**, *91*, 1097–1107.
- (43) Hou, Y.; Niemi, A. J.; Peng, X.; Ilieva, N. Myoglobin ligand gate mechanism analysis by a novel 3D visualization technique. *J. Math. Chem.* **2019**, *57*, 1586–1597.
- (44) Danielsson, J.; Banushkina, P.; Nutt, D. R.; Meuwly, M. Computer simulations of struc-

- tures, energetics and dynamics of myoglobinligand complexes. *Int. Rev. Phys. Chem.* **2006**, *25*, 407–425.
- (45) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187–217.
- (46) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S. et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Journal of Physical Chemistry B* **1998**, *102*, 3586–3616.
- (47) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (48) Huang, J.; MacKerell, Jr., A. D. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* **2014**, *34*, 2135–2145.
- (49) Mark, P.; Nilsson, L. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *J. Phys. Chem. B* **2001**, *105*, 9954–9960.
- (50) Tilton, R. F.; Kuntz, I. D.; Petsko, G. A. Cavities in Proteins: Structure of a Metmyoglobin-Xenon Complex Solved to 1.9 Å. *Biochem.* **1984**, *23*, 2849–2857.
- (51) Park, S. Y.; Yokoyama, T.; Shibayama, N.; Shiro, Y.; Tame, J. R. 1.25 Å Resolution Crystal Structures of Human Haemoglobin in the Oxy, Deoxy and Carbonmonoxy Forms. *J. Mol. Biol.* **2006**, *360*, 690–701.
- (52) Anselmi, M.; Nola, A. D.; Amadei, A. The kinetics of ligand migration in crystallized

- myoglobin as revealed by molecular dynamics simulations. *Biophys. J.* **2008**, *94*, 4277–4281.
- (53) Cazade, P. A.; Meuwly, M. Oxygen migration pathways in NO-bound truncated hemoglobin. *ChemPhysChem* **2012**, *13*, 4276–4286.
- (54) Ewald, P. P. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Annalen der Physik* **1921**, *369*, 253–287.
- (55) Petersen, H. G. Accuracy and efficiency of the particle mesh Ewald method. *Phys. Fluids* **1967**, *10*, 3668–3679.
- (56) Hess, B.; Bekker, H.; Berendsen, H. J.; Fraaije, J. G. LINCS: A Linear Constraint Solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.
- (57) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Chem.* **1977**, *23*, 327–341.
- (58) Zhang, F.; Li, J. W. Y.; Xu, K.; Lin, J. Intra-channel fiber nonlinearity mitigation based on DBP in single channel and wdm coherent optical transmission systems using different pulse shapes. *Frequenz* **2011**, *65*, 307–311.
- (59) Soloviov, M.; Meuwly, M. Reproducing kernel potential energy surfaces in biomolecular simulations: Nitric oxide binding to myoglobin. *J. Chem. Phys.* **2015**, *143*.
- (60) Hartigan, J. A.; Wong, M. A. Algorithm AS 136 : A K-Means Clustering Algorithm Author (s): J . A . Hartigan and M . A . Wong Published by : Blackwell Publishing for the Royal Statistical Society Stable URL : <http://www.jstor.org/stable/2346830>. *J. Royal Stat. Soc. Series C (Applied Statistics)* **2009**, *28*, 100–108.
- (61) 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*.
- (62) Laskowski, R. A. SURFNET: A program for visualizing molecular surfaces, cavities, and intermolecular interactions. *J. Mol. Graph.* **1995**, *13*, 323–330.
- (63) Cazade, P. A.; Berezovska, G.; Meuwly, M. Coupled protein-ligand dynamics in truncated hemoglobin N from atomistic simulations and transition networks. *Biochim. Biophys. Acta* **2015**, *1850*, 996–1005.
- (64) Grant, B. J.; Rodrigues, A. P.; ElSawy, K. M.; McCammon, J. A.; Caves, L. S. Bio3d: An R package for the comparative analysis of protein structures. *J. Bioinform.* **2006**, *22*, 2695–2696.
- (65) Schotte,; Friedrich,; Lim,; Manho,; Jackson,; A., T.; Smirnov,; V., A.; Soman,; Jayashree, et al. Watching a Protein as it Functions with 150-ps Time-Resolved X-ray Crystallography. *Science* **2003**, *300*, 1944–1947.
- (66) Tomita, A.; Sato, T.; Ichiyanagi, K.; Nozawa, S.; Ichikawa, H.; Chollet, M.; Kawai, F.; Park, S. Y.; Tsuduki, T.; Yamato, T. et al. Visualizing breathing motion of internal cavities in concert with ligand migration in myoglobin. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 2612–2616.
- (67) El Hage, K.; Brickel, S.; Hermelin, S.; Gaulier, G.; Schmidt, C.; Bonacina, L.; Van Keulen, S. C.; Bhattacharyya, S.; Chergui, M.; Hamm, P. et al. Implications of short time scale dynamics on long time processes. *Structural Dynamics* **2017**, *4*.
- (68) Zoete, V.; Meuwly, M.; Karplus, M. Investigation of Glucose Binding Sites on Insulin. *Proteins: Structure, Function and Genetics* **2004**, *55*, 568–581.
- (69) Kriegl, J. M.; Bhattacharyya, A. J.; Nienhaus, K.; Deng, P.; Minkow, O.; Nien-

- haus, G. U. Ligand binding and protein dynamics in neuroglobin. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 7992–7997.
- (70) El Hage, K.; Mondal, P.; Meuwly, M. Free energy simulations for protein ligand binding and stability. *Mol. Simulat.* **2018**, *44*, 1044–1061.
- (71) Tosqui, P.; Colombo, M. F. Neuroglobin and cytoglobin: Two new members of globin family. *Rev. Bras. Hematol. Hemoter.* **2011**, *33*, 307–311.
- (72) Rapp, C. S.; Pollack, R. M. Crystal packing effects on protein loops. *Proteins: Struct., Funct., Genet.* **2005**, *60*, 103–109.
- (73) Anselmi, M.; Brunori, M.; Vallone, B.; Di Nola, A. Molecular dynamics simulation of the neuroglobin crystal: Comparison with the simulation in solution. *Biophys. J.* **2008**, *95*, 4157–4162.
- (74) 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*, 2105–2118.
- (75) Laidler, K. J.; King, M. C. The development of transition-state theory. *J. Phys. Chem.* **1983**, *87*, 2657–2664.
- (76) Begun, A.; Molochkov, A.; Niemi, A. J. Protein tertiary structure and the myoglobin phase diagram. *Sci. Rep.* **2019**, *9*, 1–8.
- (77) Ansari, A.; Jones, C. M.; Henry, E. R.; Hofrichter, J.; Eaton, W. A. Conformational Relaxation and Ligand Binding in Myoglobin. *Biochem.* **1994**, *33*, 5128–5145.