Biopolymers

Research article

Insight into the redox partner interaction mechanism in cytochrome P450BM-3 using molecular dynamics simulations

Rajni Verma[a, b], Ulrich Schwaneberg[b], Danilo Roccatano[a]*

[a]School of Engineering and Science, Jacobs University Bremen, Campus Ring 1, 28759 Bremen, Germany; [b]Department of Biotechnology, RWTH Aachen University, Worringer Weg 1, 52074 Aachen, Germany.

***Corresponding Author:**

E-mail: d.roccatano@jacobs-university.de

Telephone: +49 421 200-3144

Fax: +49 421 200-3249

Address: School of Engineering and Science, Jacobs University Bremen, Campus Ring 1, Bremen D-28759, Germany

Abstract

Flavocytochrome P450BM-3, is a soluble bacterial reductase composed of two flavin (FAD/FMN) and one HEME domains. In this paper, we have performed molecular dynamics simulations on both the isolated FMN and HEME domains and their crystallographic complex, with the aim to study their binding modes and to garner insight into the inter-domain electron transfer (ET) mechanism. The results evidenced an inter-domain conformational rearrangement that reduces the average distance between the FMN and HEME cofactors from 1.81 nm, in the crystal structure, to an average value of 1.41 ± 0.09 nm along the simulation. This modification is in agreement with previously proposed hypotheses suggesting that the crystallographic FMN/HEME complex is not in the optimal arrangement for favorable ET rate under physiological conditions. The calculation of the transfer rate along the simulation, using the Pathway method, demonstrated the occurrence of seven ET pathways between the two redox centers, with three of them providing ET rates (K_{ET}) comparable with the experimental one. The sampled ET pathways comprise the amino acids N319, L322, F390, K391, P392, F393, A399, C400 and Q403 of the HEME domain and M490 of the FMN domain. The values of K_{ET} were closer to the experiment were calculated along the pathways FMN(C7) → F390 → K391 → P392 → HEME(Fe) and FMN(C8) → M490 → F393 → HEME(Fe). Finally, the analysis of the collective modes of the protein complex evidences a clear correlation of the first two essential modes with the activation of the most effective ET pathways along the trajectory.

Introduction

The cytochrome P450BM-3 is an important representative of the large family of Cytochrome P450 monooxygenases.[1-3] It is a NADPH-dependent fatty acid hydroxylase enzyme isolated from soil bacterium *Bacillus megaterium*.[4,5] P450BM-3 is an attractive target and a model system for biochemical and biomedical applications for different reasons.[6,7] First, it is a stable, catalytically self-sufficient protein with a convenient multidomain structure that allows easier production and handling than other monooxygenases of this family.[8] Second, it is a water-soluble enzyme with a high catalytic efficiency and oxygenase rate and it can be readily expressed recombinantly.[9,10] Third, it resembles to eukaryotic diflavin reductase such as the human microsomal P450s. As a pivotal member of P450 superfamily, it has been widely studied as an important model system for the comprehension of structure-function-dynamics relationships. The wealth of structural and kinetic data makes it one of the most studied enzyme.[11,12]

P450BM-3, being a multidomain protein has two reductase, flavin adenine dinucleotide (FAD)- and flavin mononucleotide (FMN)- binding domains and a HEME-binding domain arranged as HEME-FMN-FAD on a single polypeptide chain.[13,14] The main catalytic function of P450s is to transfer oxygen atom from molecular oxygen to their substrates. During the reaction, the enzyme is reduced by NADPH, with the electrons first transferred to the FAD cofactor of the FAD-binding domain and then to the HEME iron of the substrate bound HEME-binding domain mediated by the FMN cofactor of the FMN-binding domain. The crystallization of the whole P450BM-3 protein has been proven difficult due to the presence of flexible linker regions between domains. However, the crystallographic structures of the isolated HEME domain[15], FAD domain[16] and a non-stoichiometric complex with one FMN and two HEME domains[15] are

available in the PDB database. In the crystallographic FMN/HEME complex (PDB ID: 1BVY), the edge-to-edge shortest distance between redox centers is 1.81 nm.[15] However, it has been shown from a survey of electron transfer (ET) in the oxidoreductase proteins that a distance less than 1.40 nm is required between the redox centers for an efficient ET tunneling in the protein environment.[17] In addition, Munro *et al*. [11], have proposed that the rearrangement of the FMN domain in the structure of the crystallographic HEME/FMN complex is essential to decrease the distance between the FMN and HEME cofactors within the physiological range (less than 1.40 nm) for ET.[11] This hypothesis was corroborated by previous experimental evidences suggesting that the catalytic efficiency of P450BM-3 is determined by a optimal arrangement of the HEME and FMN domains.[18]

In this study, we aim to extend our knowledge regarding structure-function-dynamics relationships in P450BM-3 at molecular level using molecular dynamics (MD) simulations of the isolated HEME and FMN domains and of the FMN/HEME complex in solution. Previous MD simulations studies from our group have focused on the structure and dynamics of the isolated HEME and FMN domains.[19-21] Herein, the analysis will be focused on the relative rearrangement of the FMN and HEME domains in their complex and how it affects the ET tunneling from the FMN isoalloxazine ring to the HEME iron. To the best of our knowledge, this is the first MD study of this type on this system.

The paper is organized as follows. The details of the force field and MD simulations are reported in the Method section. In the first part of the subsequent Results and Discussion section, the general structural properties of the simulated systems are reported and discussed. In particular, cluster analysis is used to identify representative conformations and to evidence the structural differences of the domains in the solution and complex. The crystallographic structure

and selected conformations from the trajectory are then used to analyze the possible FMN/HEME ET pathways using the Pathways method.[22] The structural variations of other relevant part of the enzyme as the substrate access channel and the water coordination with the HEME iron are also reported. Hence, the collective dynamics of the system is analyzed using the principal component analysis of the trajectories and the essential modes are correlated with the K_{ET} calculated using the more efficient pathways. Finally, the conclusion section summarizes the outcome of the study.

Methods

Starting coordinates

The non-stoichiometric FMN/$(HEME)_2$ complex of one FMN domain and two HEME domains without substrate (PDB ID: 1BVY with resolution 0.203 nm)[15] were used to obtain the starting coordinate for the MD simulation. One of the two HEME domains (chain A: 20 - 450) was in the close proximity of the FMN domain (chain F: 479 - 630) in the crystal structure. Hence, the latter A and F chains were extracted from the crystal structure (including crystallographic water within 0.60 nm from the domains) and used as the starting coordinates for the MD simulation. 1,2-ethanediol molecules were removed from the crystallographic structure and replaced by water molecules.

Molecular dynamic simulations

The GROMOS96 43a1 force field[23] was used for all the simulations. The MD simulations performed in this study are summarized in Table 1. Figure 1 shows the FMN and HEME cofactors in stick representation. The parameters for the ferric iron of the HEME cofactor were

adopted from Helms *et al*.[24] and have been used already for the MD simulation of the P450BM-3 HEME domain by Roccatano *et al*..[19,20] Some of the partial charges were redistributed on the porphyrin ring of the HEME cofactor to adjust the parameters to the topology of GROMOS96 43a1 force field[23] with explicit hydrogen atoms bound to the bridging carbon in the porphyrin ring (see Table S1 in supporting information (SI)).

The FMN cofactor is considered being in oxidized state in the FMN domain. In a previous publication, we improved the GROMOS model of FMN cofactor by adding additional improper dihedrals to adopt the conformation of the FMN isoalloxazine ring as observed in the crystallographic structure and in quantum mechanical calculations of flavin in both redox states.[21,25] Here, the MD simulation of the isolated FMN domain is the continuation of the one reported in the previous publication. [21]

The isolated domains and their complex were centered in a cubic periodic box and set to have at least a minimal distance between the domain and any side of the box larger than 0.80 nm. They were solvated by stacking the equilibrated boxes of the solvent molecules to completely fill the simulation box. All the solvent molecules within the distance of 0.15 nm from the atoms of the domain were removed. The simulations were performed by using SPC model[26] of water. Sodium counter ions were added by replacing water molecules at the most negative electrostatic potential to obtain a neutral simulation box. The protonation state of the residues in the protein was assumed to be the same as of the isolated amino acids in the solution at pH 7. All bond lengths were constrained by the LINCS[27] algorithm. The SETTLE[28] algorithm was used for the solvent molecules. The electrostatic interactions were calculated by using the Particle Mesh Ewalds (PME) method.[29] For the long-range interactions, a grid spacing of 0.12 nm combined with a fourth-order B-spline interpolation were used to compute the potential and forces between

grid points. A pair-list for non-bonded interactions within the cutoff of 1.3 nm was used and updated at every 5 time-steps. The simulated systems were first energy minimized, using the steepest descent algorithm, for at least 2000 steps in order to remove clashes between atoms that were too close. After energy minimization, initial velocities obtained from a Maxwell-Boltzmann velocity distribution at 300 K were assigned to all atoms. All systems were initially equilibrated for 100 ps with position restraints on the heavy atoms of the solute for the relaxation of the solvent molecules. Berendsen's thermostat[30] was used to keep the temperature at 300 K by weak coupling the systems to an external thermal bath with a time constant of 0.1 ps. The pressure of the system was kept at 1 bar by using the Berendsen's barostat[30] with a time constant of 1 ps. A time step of 2 fs was used to integrate the equations of motions. After the equilibration procedure, position restraints were removed and the system was gradually heated from 50 K to 300 K in 200 ps. Finally, a production run of 100 ns was performed at 300 K for all the systems. All the simulations and analysis of the trajectories were performed by using the GROMACS (version 4.07) software package.[31] The crystal structure of the P450BM-3 domains was used as reference structure for the analysis of all the simulations. During the simulations, the conformational changes occurred in the P450BM-3 domains were examined by analyzing the root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration (Rg) and secondary structure elements (using DSSP criteria[32]) with respect to the crystal structure.

Cluster analysis

Cluster analysis was performed to characterize the conformational diversity of the structures generated during the MD simulations. It was performed using the Gromos clustering

algorithm that is based on the RMSD of the selected atoms of the conformations obtained from the simulations.[33] A structure is assigned to a cluster if its RMSD from the cluster median structure is within a given cutoff. In this work, the method was applied to the backbone atoms and a RMSD cutoff of 0.10 nm was used. For the analysis, 1400 structures were sampled at every 50 ps in the last 70 ns of the trajectories.

Electron transfer tunneling

Electron transfer (ET) tunneling from the FMN to HEME cofactor was calculated using the Pathways program.[22,34] For a given protein conformation, the program identifies an effective ET coupling by evaluating the highest electronic tunneling coupling (T_{DA}) through different pathway connecting the donor and the acceptor through bonds and space.[22] The Pathways program uses the graph theory to identify the series of steps to maximize T_{DA} values by assigning different step contributions weather it is mediated by covalent bonds (ε^{cb} in equation 2), hydrogen bonds (ε^{hb} in equation 3) and through space jump (ε^{sj} in equation 4). Hence, T_{DA} (equation 1) is proportional to the product of the latter contributions:

$$T_{DA} \alpha \prod_i \varepsilon_i^{cb} \prod_j \varepsilon_j^{hb} \prod_k \varepsilon_k^{sj} \quad (1)$$

$$\varepsilon_i^{cb} = 0.60 \quad (2)$$

$$\varepsilon_j^{hb} = 0.36 e^{-1.70(R-2.80)} \quad (3)$$

$$\varepsilon_k^{sj} = 0.60 e^{[-1.70(R-1.40)]} \quad (4)$$

where R is the distance of the step i, j and k in Angstrom. The ET pathways between FMN/HEME were calculated in the crystal structure and by taking different conformations sampled along the MD trajectories. For the calculations, the C8 or C7 atom of the FMN isoalloxazine ring as donor and the HEME iron as acceptor were used. The ET pathways were calculated and visualized using the plugin Pathways[34] and the visualization program VMD.[35] The possible non-adiabatic ET reaction rate (K_{ET}) was estimated using the equation 5.[36]

$$K_{ET} = \frac{2\pi}{\hbar} \frac{e^{\frac{\left[-(\Delta G+\lambda)^2\right]}{4\lambda k_B T}}}{\sqrt{4\pi\lambda k_B T}} \left|T_{DA}\right|^2 \quad (5)$$

where ΔG is the driving force and λ is the Marcus reorganization energy for the ET reaction, $\hbar = h/2\pi$ with h the Plank constant, and k_B is the Boltzmann constant. The difference in the reduction potential of the FMN and HEME cofactor (-0.224 eV) was used as ΔG.[11,14,17] λ was considered equal to 0.7 eV as a good approximation of the inter protein ET.[11,17,37]

The average values of T_{DA} and K_{ET} were calculated by evaluating along the trajectories the minimum lengths of the intermediate paths of the selected pathways. These distances were used to estimate, at each trajectory conformation the value of T_{DA} and, hence, of K_{ET} using equation 1-5. Finally, the time series were used to calculate the mean and its standard deviation.

Principal component analysis

Principal component analysis (PCA) was performed to access the conformational space in the biomolecules during the MD simulation. The details of the PCA, also called as essential dynamics, can be found elsewhere.[38,39] The backbone atoms (C_α, C and N) of the domains

(considering 1773, 1317 and 456 atoms for AF, A and F chain, respectively) were used for the calculation of the covariance matrix. The PCA analysis was performed on the last 70 ns of the trajectories. For the 3D displacements along different eigenvectors were calculated by projecting the atomic coordinates along the trajectory on eigenvectors and by extracting sampled conformations along the two extremes of the projection. UCSF Chimera visualization tool is used for representation.[40] The comparison of eigenvectors obtained from the different simulations was performed using the root-mean-square inner product (RMSIP in equation 6).[41]

$$RMSIP = \sqrt{\frac{1}{N}\sum_{i=1}^{m}\sum_{j=1}^{m}(\mathbf{v}_i \cdot \mathbf{u}_j)^2} \qquad (6)$$

where v_i and u_j are the i^{th} and j^{th} eigenvectors of the two different m dimensions essential subspaces of the two systems. RMSIP gives a simple measure to assess the dynamical similarity of eigenvectors.[41] The convergence of the essential modes was performed by comparing the RMSIPs calculated from the MD trajectories.

Results and Discussion

Structural properties

Figure 2a shows the backbone RMSD of both the P450BM-3 domains as a function of time. The RMSD curves of both the AF chains and the single A chain converge to average RMSD values of 0.41 ± 0.03 nm and 0.36 ± 0.03 nm, respectively. The isolated A chain shows a slightly lower average RMSD value of 0.33 ± 0.02 nm. In the FMN/HEME complex, the RMSD of the F chain increases to an average value of 0.25 ± 0.02 nm in the last 10 ns. For the isolated F

chain, the average RMSD value of 0.26 ± 0.02 nm was observed with a slight little variation at the end of simulation.[21,25]

In Figure 2b, the time series of the Rg values for the P450BM-3 domains are reported. The Rg values of the complex show a decrease of ~3.7% from the crystallographic value (2.42 nm) in the first 10 ns with an average value of 2.33 ± 0.01 nm. The Rg variations of the A chain, in the complex and in solution with respect to the crystal structure (2.16 nm) is less than 1.8% (2.12 ± 0.01 nm and 2.14 ± 0.01 nm, respectively). The Rg value of the F chain does not vary significantly from the crystal structure (1.45 nm) with an average of 1.45 ± 0.01 nm and 1.46 ± 0.01 nm for the single domain and the complex simulation, respectively.

The cytochrome P450BM-3 has structurally conserved P450- and flavodoxin- like protein folds in the A and F chains, respectively. Figure 3c shows the structure of FMN/HEME complex with the labeled helices of A (A to L) and F (α1 to α4) chain, and FMN binding loops (Lβ1, Lβ3 and Lβ4). The loop regions are named according to the secondary structure element (α helix or β sheet) preceding them. The analysis of the secondary structure along the trajectories is reported in Figure S1 in SI. The results of the analysis clearly indicate that the crystallographic secondary structure of the P450BM-3 domains remains fairly preserved in all the simulations.

Figure 3a and 3b show the per-residue RMSD and RMSF with respect to the crystal structure, respectively. The cofactor binding sites show smaller deviations and fluctuations from the crystal structure in the isolated (in red color) and complex (black color) simulations. For both the domains, the loop regions, and N- and C-termini show larger deviations. The isolated domains in solution deviate more than the one in the complex except the region between the helices, A - B, B' - C, H - I, and K - L and in G helix of the A chain and Lβ3 of the F chain. The isolated F chain shows largest deviation in the Lβ2 and Lβ4 regions. In both the systems, the F

chain shows higher fluctuation in the Lβ2 and Lα2 loops. In the simulation of the complex, the loop regions A/B, and F/G have larger fluctuations. Finally, in the simulation of the isolated F chain, the inner FMN cofactor binding loop (Lβ3) shows slightly higher fluctuations.

Cluster analysis

For the AF chain, the first two clusters account for 46.16 % and 27.12 %, respectively. The analysis of trajectories from the isolated A and F chains in solution produced, 6 clusters in both the cases, whereas the one of the chain in the complex has returned 7 and 8 clusters, respectively. The first two clusters of the isolated A chain account for 46.53 % and 30.12 % of the population of conformers, respectively. However, the percentage for A chain changes to 76.87 % and 10.99 %, respectively in the complex. For the F chain, the first cluster comprises the ~64 % and the second 21.05 % and 23.05 % of the total conformations in the isolated and the complex simulation, respectively. The HEME domain is more liable for conformational changes when it is isolated than in the complex, while the F chain shows negligible difference in the conformational space in both the simulations.

In the complex simulation, the median structure of the first cluster for the AF chain shows a conformational rearrangement in both the domains that resulted into an increase in the compactness of the FMN/HEME complex (see Figure 2b and 4). Major differences were observed in the loop regions of the domains in all simulations. In Figure 4, the arrow shows the deviation from the crystal structure after inter-domain rearrangement that mainly involve G helix and H/I and K/L loops (residue 380 - 390) in the HEME domain and α2 helix of the FMN domain. The rearrangement decreases the edge-to-edge distance between both the FMN and HEME cofactors. The G helix and, H/I and K/L loop regions constitute the important part of the

P450BM-3 HEME domain. The H/I and K/L loops are involved in the binding of the FMN domain and the residues 380 – 390 of K/L loop region head the HEME cofactor binding site therefore they might influence ET tunneling from FMN to HEME.

In the crystal structure, the minimum distance between the HEME and FMN domains is 0.46 nm with a total number of 20 contacts. In the simulation of the complex, the latter distance decreased within the first 10 ns, then it stabilized to an average value of 0.37 ± 0.01 nm for the rest of the simulation (see Figure 5a). The number of contacts shows a sharp increase in the first 10-20 ns and then it stabilizes to an average of 167 ± 14 contacts in the last 50 ns of the simulation (see Figure 5b). The change of the minimum distance and the number of contacts between the two domain strongly affect the ET tunneling (*vide infra* in the ET tunneling pathways section).

Substrate access channel

Pro45 and Ala191 are located at the mouth of the substrate access channel. In the crystal structure of the P450BM-3 complex, the P45C_α and A191C_α are 1.61 nm apart (0.87 nm in the A chain of 1BU7). Chang *et. al*. observed using MD simulations and docking approaches that the substrate binding was not dramatically affected by the closeness of the substrate access channel in P450BM-3.[42] The behavior of the substrate access channel has been previously evaluated by monitoring the distance during the simulation between these two residues of the HEME domain (PDB code 1BU7[15]) in solution by Roccatano *et al*..[19] In Figure 6, the P45C_α - A191C_α distances were calculated in the isolated domain and complex simulations are reported. Both the simulations show higher variations in the P45C_α - A191C_α distance in the first 20 ns simulation. Conversely, in the isolated A chain, it stabilizes to the average distance of 1.11 ± 0.10 nm. In the

simulation of the complex, the P45C_α - A191C_α distance continuously decreased until 32 ns and then it stabilizes to an average value of 0.59 ± 0.10 nm. The decrease in the P45C_α - A191C_α distances during the simulation indicating the closing of the mouth of the substrate access channel was observed and discussed previously by Roccatano *et al*..[19] During the simulations, the more extended closure of the substrate access channel when HEME domain was in the complex then in solution that might be caused by the higher deviation of F/G loop in the complex simulation than in the solution.

Water coordination with HEME iron

Previous simulation by Roccatano *et al*.[19] starting from the crystal structures of the individual P450 BM3 HEME domain (PDB ID: 1BU7),[15] have shown that although the water was not covalently bound to the HEME in the simulation model, the force field non covalent interactions between the iron and the water molecules could still provide a good approximation of the water coordination with an average Fe-O distance of ~0.3 nm compared to the 0.26 in the crystal.[15]

Contrary to the 1BU7, the crystal structure of the HEME, in the FMN/HEME complex do not contain crystallographic water bound to the iron. However, the water molecules coordinating to the HEME iron were observed (consistently with the previous simulations of the 1BU7 strcture) at an average distance of 0.28 ± 0.13 nm and of 0.34 ± 0.14 nm (see Figure S2 in SI) in the HEME domain during the isolated and the complex simulations, respectively. The coordination of iron by the water molecule is consistent with the configuration of the enzyme without the presence of the substrate.

ET tunneling pathways

The minimum distance between the FMN isoalloxazine ring (using the heavy atoms only) and the HEME cofactor is reported in Figure S3 in SI (the simulation of the complex was extended till 150 ns to check the FMN/HEME distance convergence). During the complex simulation, the FMN/HEME distance decreases from 1.81 nm (in the crystal structure) to an average distance of 1.41 ± 0.09 nm, with the minimum distance of 1.02 nm. The decreased distances are within the range of expected ET between the redox centers[17] (less than 1.50 nm) as proposed by Munro *et al*.[11] and consistent with experimental and theoretical observations.[11,17]

Table 2 summarizes the ET pathways in AF chain calculated using the Pathways model (see Methods) in the crystal structure, the median conformation of the first cluster, the conformation with minimum FMN/HEME contact distance and those sampled every 10th ns from the 100 ns trajectory. The distributions of the total distances along the considered ET pathways are represented in the Figure S4 of SI. All the sampled pathways are characterized by through space (see equation 3). The most effective pathways are the first, second and seventh with high K_{ET} values (see Table 2) of 39.95 ± 0.7 s^{-1}, 65 ± 1.0 s^{-1} and 20 ± 1.0 s^{-1}, respectively. Figure 7a, 7b and 7c show the FMN/HEME ET pathways identified by the Pathways model[34] in the crystal structure (FMN/HEME distance of 1.81 nm), the median conformation of the first cluster (FMN/HEME distance of 1.41 nm) and the last conformation of the simulation (FMN/HEME distance of 1.27 nm), respectively. In the crystal structure, the ET tunneling from FMN to HEME cofactor is mediated by solvent molecules. After the conformational rearrangement in the AF chain, the FMN cofactor comes close to the HEME cofactor and eliminates the involvement of water molecules in the FMN/HEME ET tunneling for most of the time during the simulation. In fact, only in the pathway from the configuration at 100 ns (see Figure 7c), the ET tunneling is

mediated by Met490 and Gln403 residues with the involvement of a solvent molecule. When the FMN/HEME distance decrease from 1.8 nm (in crystal structure see Figure 7a) to 1.4 nm (just after rearrangement, see Figure 7b), a K_{ET} value of 65 ± 1.00 s^{-1} (Pathway 2 in Table 2) close to the experimental FMN/HEME K_{ET} value of 80 s^{-1} was obtained.[43]

Principal component analysis

In the simulations of the isolated domain and of the FMN/HEME complex, the cumulative sum of the relative positional fluctuation (RPF) is greater than 69% for the first 50 eigenvectors of the A and F chains (reported in Figure S5 of SI). For the first twenty eigenvectors of A and F chains, the RMSIP was less than 0.53 in both the simulations. The inner product values were less than 0.25 and 0.43, respectively for the first three eigenvectors of the A and F chains. These values indicate a low similarity in the first three larger collective modes especially for the A chain.

Figure 8a, 8b and 8c represent components associated with the first three eigenvectors of A and F chains for the isolated domain (in red color) and complex (in black color) simulations. Figure 9 shows the RMSF associated with the first three eigenvectors (a, b and c) of the A (in sky blue) and F (in tan color) chains in solution (a1, b1 and c1) and complex (a2, b2 and c2) simulations, respectively.

In the complex simulation, the first collective motion (Figure 9a1) of the A chain involves the turn region between the residues 44 – 48 (the highest eigenvector component is for the residue R47 that is involved in the substrate binding), the loop regions D/E (residues 130 – 138), F/G (residues 190 – 196), K/L loop (residue 385 – 390) and the C-terminus loops (residues 425 – 432 and 452 – 458). The collective mode involves the turn region 44 – 48 and F/G loop is

related to the closing and opening of the substrate access channel.[44] Different residues (F390, K391, P392, F393, A399, and C400) of the latter K/L loop region are present in most of the ET pathways found in the complex simulation. The residue F393, a part of the K/L loop later region, is involved in the FMN domain binding and found to be involved in the ET tunneling in the first cluster of the AF chain (Pathway 1). F393, a conserved HEME-binding residue, is considered a key residue in the thermodynamics control of the P450BM-3 catalytic activity.[45,46] In the complex simulation, the first collective mode of the F chain involves the major contribution of the Lα2 loop with slightly higher components of the inner FMN binding loop Lβ2 and Lβ3. The collective modes involving Lα2 and Lβ3 might facilitate the ET tunneling from the FMN to HEME cofactors. In the FMN/HEME complex, the collective modes of both the domains were synchronized to relate the ET tunneling and the change in the substrate binding region. The effect was clearly seen when the first eigenvectors of the AF chain was compared with that of the isolated A and F chains in the complex simulation (reported in Figure S6a and S7a in SI). In both AF and, individual A and F chain, the first eigenvector show fluctuations in the same regions with higher values for the AF chain. The second collective mode of the A chain involves mainly the motion in D/E and G/H loops, β- sheets in K/L regions and A/B region and the third collective motion was restricted to D/E and G/H loops and C-terminus loop (residues 425 – 432). The F chain shows the involvement of Lα2 and Lβ2 loops and the C-terminus region in the second collective mode, while involvement of Lα2, Lβ3 and Lβ5 loops in the third eigenvector. In the AF chain, the collective modes associated with the first two eigenvectors belongs to the movement of the F chain towards the A chain to decrease the distance between the FMN and HEME cofactors and show slightly higher fluctuation than in the individual chains. In the third eigenvector, the major difference was observed mainly in Lβ3 and Lα2 loops of the F chain with

higher fluctuations. The collective motions associated with the first three eigenvectors of the AF chain are reported in Figure S7a, S7b and S7c of SI, respectively.

In the isolated A chain, the first collective mode (Figure 9a2) has higher values of components at the end of C helix (residue 103 – 107) and C-terminus (residues 452 – 458). Other regions involved in the first collective mode were D/E, E/F, F/G and K/L loops (residue 385 – 390). Altogether, the motion is related to the change in the binding regions for the substrate and the FMN domain.

The first collective mode of the isolated F chain shows higher components in Lα2 and Lβ2 loops, and slightly high values in Lβ3 and Lβ4 loops. In the simulations of individual domains, the collective motions is more related to the binding of the FMN cofactor in the F chain and is restricted to the substrate binding region in the A chain. The second collective mode of the A chain involves mainly the motion in D/E, E/F and F/G loops and it is restricted to the F/G region only in the third collective motion. In the F chain, the Lα2 and Lβ2 loops are involved in the second collective mode and Lα2 and Lβ3 loops in the third mode.

Figure 10 shows the projection of the AF chain trajectory on the first and second eigenvectors. The projection is characterized by a V shaped distribution. The left and right strokes will be named Region I and II, respectively. The Region I in the interval [-6:0] of the x-axis correspondes, approximately, to the first 50 ns of the simulation and the other the last 50 ns. In Figure 10b, 10c and 10d, the values of K_{ET} , calculated along the pathways 1, 2 and 7 for each conformation along the trajectories, are averaged on the eigenvector plane using a square grid of bin 0.05 nm. For the pathway 1, the highest values of K_{ET} are localized in Region I (see Figure 10b) On the contrary, the pathway 7 (sampled at 100 ns, see Table 2) has the largest value K_{ET} in the Region II of the projection (see Figure 10d). Finally, Pathway 2 shows a uniform distribution

of the high value of K_{ET} in the Region I and partial distribution in the Region II (see Figure 10c). The results indicate that the essential modes do have an effect on the activation of different pathways responsible for ET tunneling along the trajectory that is consistent with the previous study on the effect of the protein dynamics on the ET process.[47-51]

Conclusions

We performed the MD simulations of the P450BM-3 HEME and FMN domains as isolated domains or in the complex. The secondary structure and tridimensional structure of the two domains do not significant change from crystallographic structure during the 100 ns simulations. The isolated FMN domain shows major conformational change in Lα2 loop as observed in a previous study.[21] In the isolated HEME domain, the major conformational changes were observed in the FMN binding region especially in C helix and H/I and K/L (residue 385 – 395) loops. Conversely, the FMN/HEME complex undergoes an inter-domain conformational rearrangement in the first 10 ns of the simulation that increased its compactness (with the change in the Rg values from 2.42 nm (in the crystal structure) to 2.33 nm) and reduced of 22% in the FMN/HEME minimum distance from 1.81 nm (in the crystal structure) to an average 1.41± 0.09 nm. The change of the average distance between the FMN and HEME domains confirm the previous theoretical and experimental observation[11,52,53] about the non-competent arrangement for an efficient ET of the crystallographic FMN/HEME complex. The conformations obtained from our complex simulation have minimum distances between the FMN to HEME cofactors that provide ET rate consistent with experimental data.[43]

Both the FMN and HEME domains show difference in the collective modes in solution compared to the ones in the FMN/HEME complex. In the latter, the collective motions were clearly associated with the change of the ET pathways observed during the simulations.

In summary, the results of this theoretical study are consistent with the available experimental data and provide further insight at the atomistic level to understand the structure and dynamics of this complex enzyme. In particular, the structural determinants of the inter-domain ET mechanism can put forward important information to extend the knowledge of the P450BM-3 enzyme for a better exploitation of the enzyme in biotechnological applications. In fact, the results of this study have identified the residues, N319, L322, F390, K391, P392, F393, A399, C400 and Q403 of HEME domain, and M490 of FMN domain, to be involved in the ET tunneling from FMN to HEME cofactor. These amino acids can be the target of the site directed evolution experiments to prove their significance in the ET mechanism.

Associated content

Supporting Information

The partial charge on the HEME cofactor with the ferric iron for GROMOS96 43a1 force field[23], the secondary structure elements of the HEME and FMN domains, the distance between water molecule and the HEME iron, the distance between the heavy atoms of FMN isoalloxazine ring and the HEME cofactor, the distribution of distance along the different FMN/HEME ET pathways, the cumulative sum of the relative positional fluctuation (RPF) of the first 50 eigenvectors of the A and F chains, components for the first, second and third eigenvectors of the AF, A and F chains in solution and complex simulation and the RMSF of the protein backbone

atoms along the first, second and third eigenvectors after projecting the trajectory on the corresponding eigenvector of AF chain in the complex simulation.

Author Information

Corresponding Author

*Email: d.roccatano@jacobs-university.de

Acknowledgement

We thank European Union 7th framework program (project "OXYGREEN", Project Reference: 212281) for financial support. This study was performed using the computational resources of Computer Laboratories for Animation, Modeling and Visualization (CLAMV) at Jacobs University Bremen.

References

1. Chefson, A.; Auclair, K. Mol Biosyst 2006, 2, 462-469.
2. Wong, L. L. Curr Opin Chem Biol 1998, 2, 263-268.
3. Guengerich, F. P. Chem Res Toxicol 2001, 14, 611-650.
4. Narhi, L. O.; Fulco, A. J. J Biol Chem 1986, 261, 7160-7169.
5. Narhi, L. O.; Fulco, A. J. J Biol Chem 1987, 262, 6683-6690.
6. Whitehouse, C. J.; Bell, S. G.; Wong, L. L. Chem Soc Rev 2012, 41, 1218-1260.
7. Di Nardo, G.; Gilardi, G. Int J Mol Sci 2012, 13, 15901-15924.
8. Munro, A. W.; Girvan, H. M.; McLean, K. J. Biochim Biophys Acta 2007, 1770, 345-359.

9. Munro, A. W.; Lindsay, J. G.; Coggins, J. R.; Kelly, S. M.; Price, N. C. Febs Letters 1994, 343, 70-74.

10. Warman, A. J.; Roitel, O.; Neeli, R.; Girvan, H. M.; Seward, H. E.; Murray, S. A.; McLean, K. J.; Joyce, M. G.; Toogood, H.; Holt, R. A.; Leys, D.; Scrutton, N. S.; Munro, A. W. Biochem Soc Trans 2005, 33, 747-753.

11. Munro, A. W.; Leys, D. G.; McLean, K. J.; Marshall, K. R.; Ost, T. W.; Daff, S.; Miles, C. S.; Chapman, S. K.; Lysek, D. A.; Moser, C. C.; Page, C. C.; Dutton, P. L. Trends Biochem Sci 2002, 27, 250-257.

12. Girvan, H. M.; Waltham, T. N.; Neeli, R.; Collins, H. F.; McLean, K. J.; Scrutton, N. S.; Leys, D.; Munro, A. W. Biochem Soc Trans 2006, 34, 1173-1177.

13. Peterson, J. A.; Sevrioukova, I.; Truan, G.; GrahamLorence, S. E. Steroids 1997, 62, 117-123.

14. Munro, A. W.; Daff, S.; Coggins, J. R.; Lindsay, J. G.; Chapman, S. K. Eur J Biochem 1996, 239, 403-409.

15. Sevrioukova, I. F.; Li, H. Y.; Zhang, H.; Peterson, J. A.; Poulos, T. L. P Natl Acad Sci USA 1999, 96, 1863-1868.

16. Joyce, M. G.; Ekanem, I. S.; Roitel, O.; Dunford, A. J.; Neeli, R.; Girvan, H. M.; Baker, G. J.; Curtis, R. A.; Munro, A. W.; Leys, D. FEBS Journal 2012, 279, 1694-1706.

17. Page, C. C.; Moser, C. C.; Chen, X.; Dutton, P. L. Nature 1999, 402, 47-52.

18. Hazzard, J. T.; Govindaraj, S.; Poulos, T. L.; Tollin, G. J Biol Chem 1997, 272, 7922-7926.

19. Roccatano, D.; Wong, T. S.; Schwaneberg, U.; Zacharias, M. Biopolymers 2006, 83, 467-476.

20. Roccatano, D.; Wong, T. S.; Schwaneberg, U.; Zacharias, M. Biopolymers 2005, 78, 259-267.

21. Verma, R.; Schwaneberg, U.; Roccatano, D. J Chem Theory Comput 2013, 9, 96-105.

22. Beratan, D. N.; Betts, J. N.; Onuchic, J. N. Science 1991, 252, 1285-1288.

23. van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hunenberger, P. H.; Kruger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. Biomolecular Simulation: The GROMOS96 Manual and User Guide., 1996.

24. Helms, V.; Deprez, E.; Gill, E.; Barret, C.; Hui Bon Hoa, G.; Wade, R. C. Biochemistry 1996, 35, 1485-1499.

25. Zheng, Y.-J.; Ornstein, R. L. J Am Chem Soc 1996, 118, 9402-9408.

26. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. Intermolecular Forces 1981, 331-342.

27. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. J Comput Chem 1997, 18, 1463-1472.

28. Miyamoto, S.; Kollman, P. A. J Comput Chem 1992, 13, 952-962.

29. Darden, T.; York, D.; Pedersen, L. J Chem Phys 1993, 98, 10089-10092.

30. Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. J Chem Phys 1984, 81, 3684-3690.

31. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. J Chem Theory Comput 2008, 4, 435-447.

32. Kabsch, W.; Sander, C. Biopolymers 1983, 22, 2577-2637.

33. Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. Angew Chem Int Edit 1999, 38, 236-240.

34. Balabin, I. A.; Hu, X.; Beratan, D. N. J Comput Chem 2012, 33, 906-910.

35. Humphrey, W.; Dalke, A.; Schulten, K. J Mol Graphics 1996, 14, 33.

36. Marcus, R. A.; Sutin, N. Biochim Biophys Acta 1985, 811, 265-322.

37. Moser, C. C.; Keske, J. M.; Warncke, K.; Farid, R. S.; Dutton, P. L. Nature 1992, 355, 796-802.

38. Amadei, A.; Linssen, A. B.; Berendsen, H. J. Proteins 1993, 17, 412-425.

39. Amadei, A.; Linssen, A. B.; de Groot, B. L.; van Aalten, D. M.; Berendsen, H. J. J Biomol Struct Dyn 1996, 13, 615-625.

40. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. J Comput Chem 2004, 25, 1605-1612.

41. Amadei, A.; Ceruso, M. A.; Di Nola, A. Proteins 1999, 36, 419-424.

42. Chang, Y. T.; Loew, G. H. J Biomol Struct Dyn 1999, 16, 1189-1203.

43. Sevrioukova, I.; Shaffer, C.; Ballou, D. P.; Peterson, J. A. Biochemistry 1996, 35, 7058-7068.

44. Li, H.; Poulos, T. L. Acta Crystallogr D 1995, 51, 21-32.

45. Ost, T. W.; Munro, A. W.; Mowat, C. G.; Taylor, P. R.; Pesseguiero, A.; Fulco, A. J.; Cho, A. K.; Cheesman, M. A.; Walkinshaw, M. D.; Chapman, S. K. Biochemistry 2001, 40, 13430-13438.

46. Chen, Z.; Ost, T. W.; Schelvis, J. P. Biochemistry 2004, 43, 1798-1808.

47. Balabin, I. A.; Onuchic, J. N. Science 2000, 290, 114-117.

48. Jones, M. L.; Kurnikov, I. V.; Beratan, D. N. J Phys Chem A 2002, 106, 2002-2006.

49. Balabin, I. A.; Beratan, D. N.; Skourtis, S. S. Phys Rev Lett 2008, 101, 158102.

50. Beratan, D. N.; Skourtis, S. S.; Balabin, I. A.; Balaeff, A.; Keinan, S.; Venkatramani, R.; Xiao, D. Acc Chem Res 2009, 42, 1669-1678.

51. Skourtis, S. S.; Waldeck, D. H.; Beratan, D. N. Annu Rev Phys Chem 2010, 61, 461-485.

52. Sevrioukova, I.; Peterson, J. A. Biochimie 1996, 78, 744-751.

53. Aigrain, L.; Fatemi, F.; Frances, O.; Lescop, E.; Truan, G. Int J Mol Sci 2012, 13, 15012-15041.

Table titles

Table 1: Simulation summary of P450BM-3 domains in solution

Table 2: ET tunneling pathways in the FMN/HEME complex of P450BM-3

Figure legends

Figure 1: a) HEME cofactor and b) FMN cofactor are in stick representation, colored by elements such as, oxygen in red, nitrogen in blue, hydrogen in green, iron or phosphorus in orange and carbon in gray, with the atomic labeling according to their GROMOS96[23] topology.

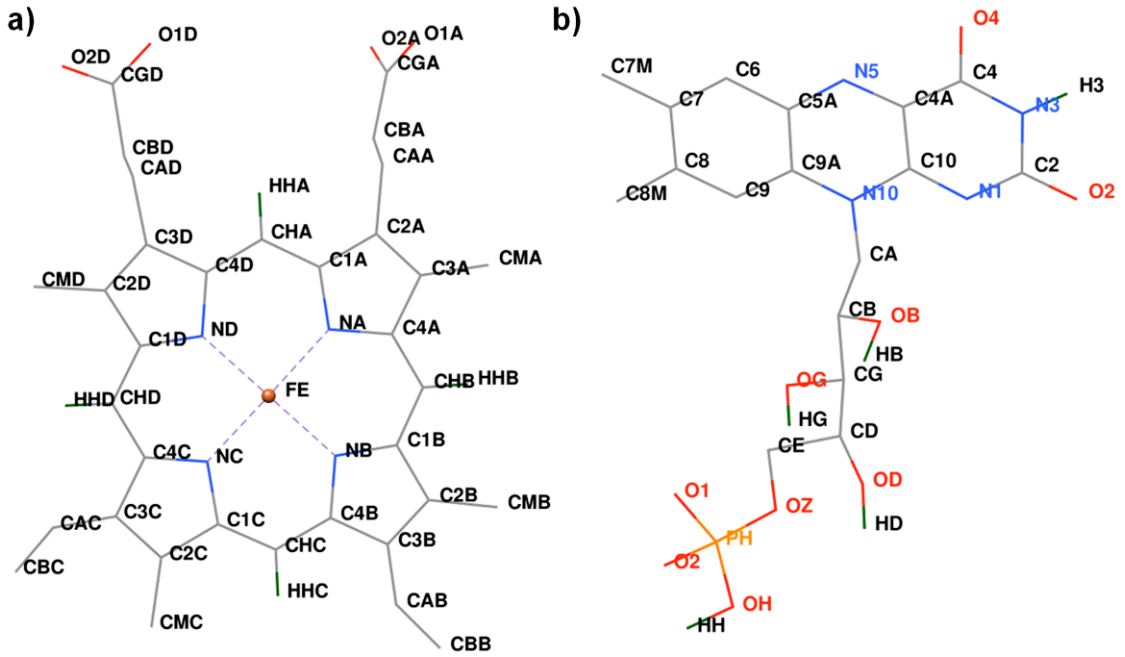


Figure 2: a) Backbone RMSD and b) Rg with respect to the crystal structure as a function of time for the AF chain (black), A of AF chain (red), F of AF chain (green), A chain (blue) and F chain (orange).

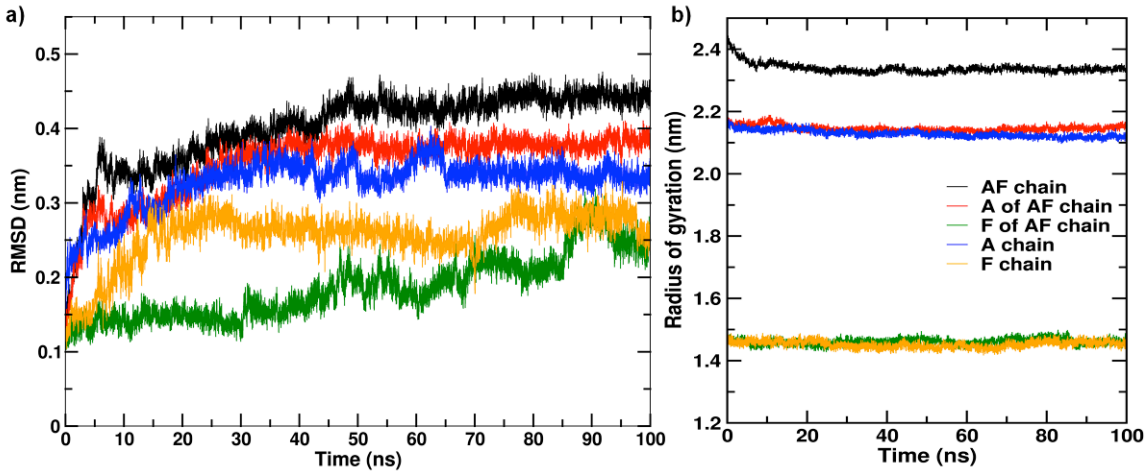


Figure 3: Backbone RMSD (a) and RMSF (b) per residue with respect to the crystal structure for the isolated domains (in red) and in the complex (in black) MD simulations. The green vertical line separates the HEME and FMN domains. Horizontal bars, in blue and orange color represent helices (labeled) and beta sheets, respectively. The regions involved in the cofactor binding are

represented by horizontal bars in purple color. Vertical bars in grey color show the interaction between the HEME and FMN domains. (c) The HEME and FMN domains are in the cartoon representation in sky blue and tan color, respectively. The HEME and FMN cofactors are in red and green color, respectively. Helices, cofactors, FMN binding regions and, N- and C- termini are labeled.

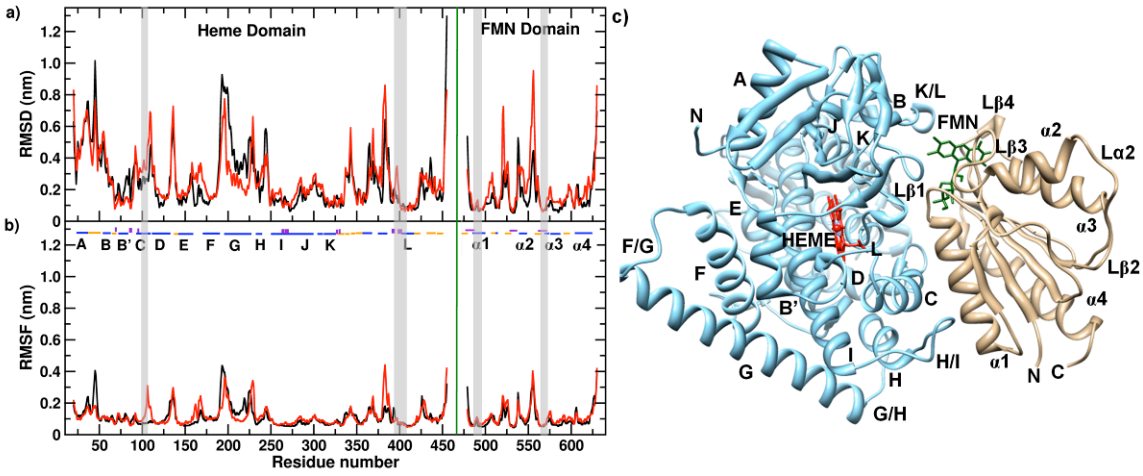


Figure 4: The median AF chain conformation of the first cluster superimposed with the crystal structure. In the crystal structure, the A and F chains are in the cartoon representation in dark and light blue color, respectively. For the complex simulation, the A and F chain are in surface (in tan color) and cartoon representation (in orange color), respectively. The HEME and FMN cofactors and the FMN/HEME distance (by dotted line) are represented in green and red color in the crystal structure and in the equilibrated conformation, respectively. The purple colored arrow shows the displacement of the FMN domain towards the HEME domain after conformational rearrangement.

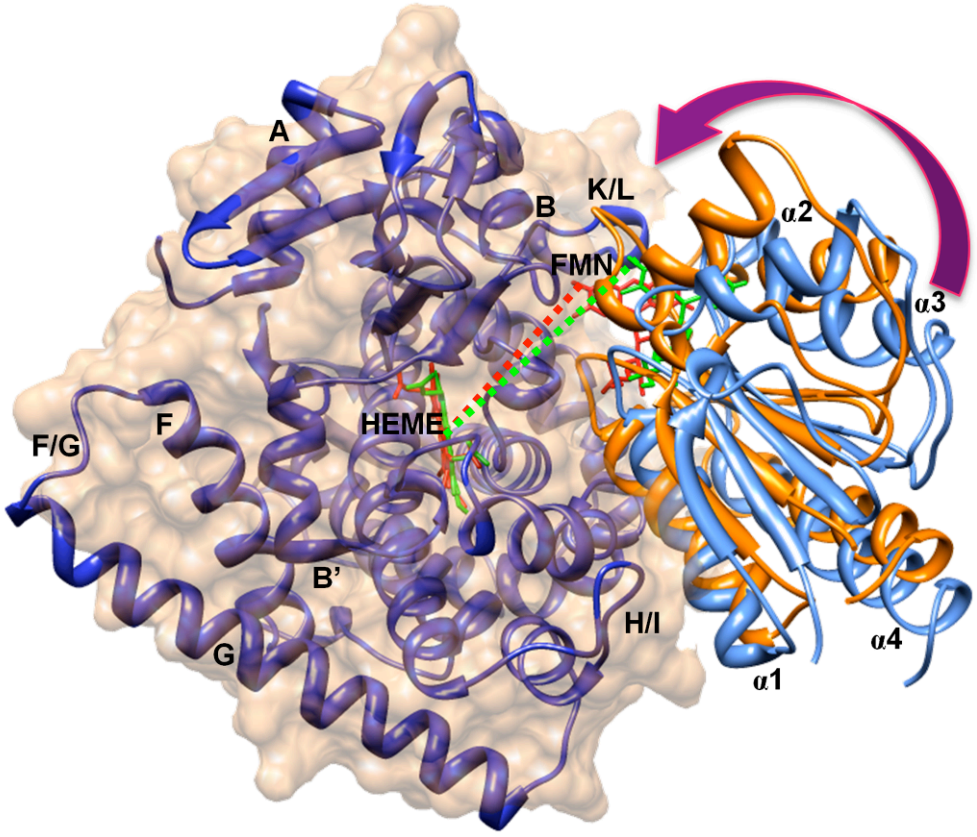


Figure 5: (a) Minimum distance and (b) the number of contacts (less than 0.6 nm) between the backbone atoms of the A and F chains. Horizontal red color line shows the minimum distance and the number of contacts in the crystal structure.

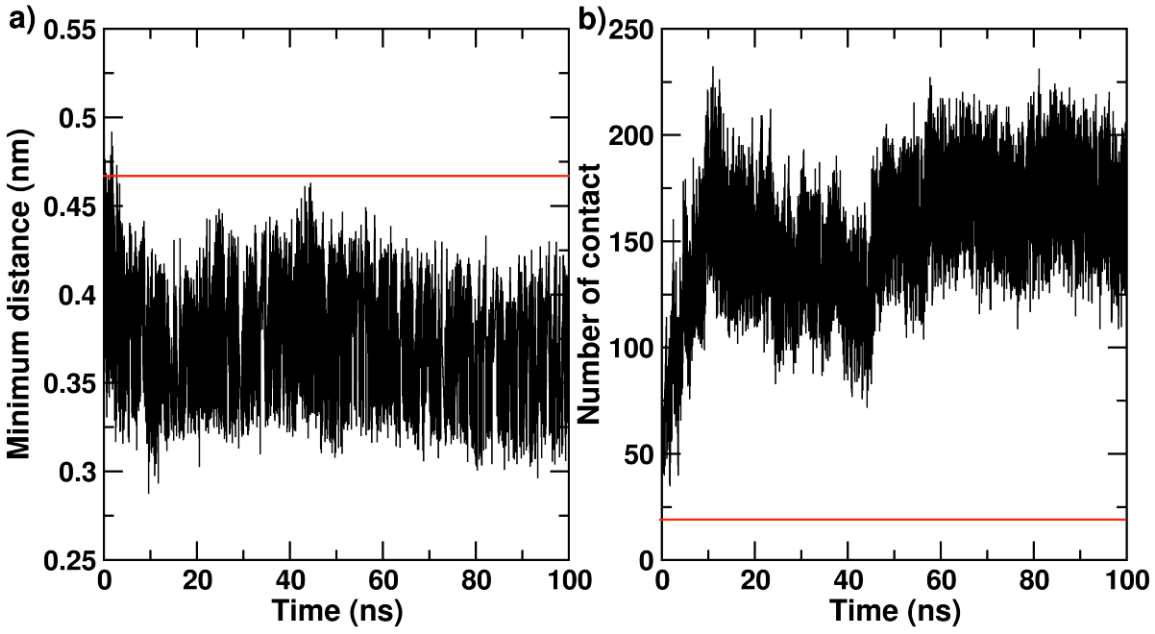


Figure 6: Minimum distance between P45C_α and A191C_α as a function of time for the A chain in the solution (in red color) and complex (in black color) simulation.

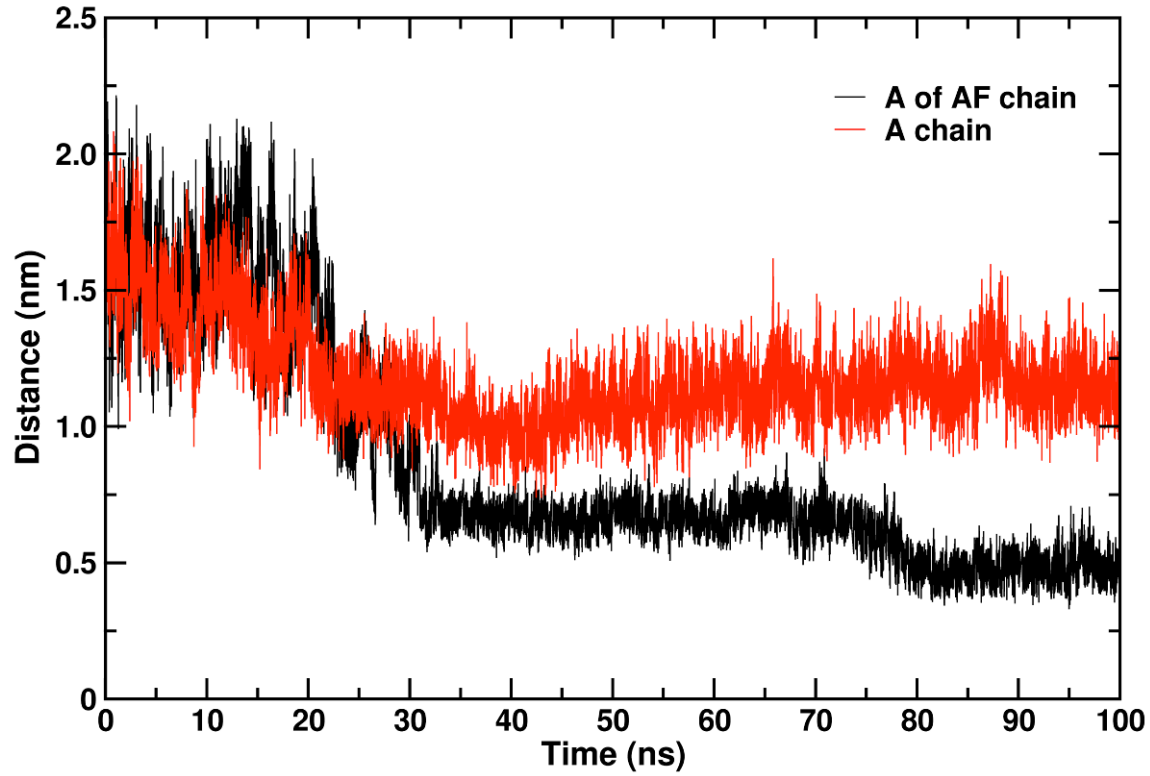


Figure 7: ET tunneling from the FMN isoalloxazine ring (C8 atom) (in gray color) to the HEME iron (in black color) is represented by red color tubes in a) the crystal structure, b) the conformation of the first cluster (50.6 ns) and c) the conformation at the end of simulation 100 ns. The amino acids within the distance of 0.50 nm from both the cofactors are shown in licorice representation and colored by element type (oxygen in red, carbon in cyan and nitrogen in blue color) and their associated secondary structure in the cartoon representation in sky blue color for the HEME and FMN domains in orange color. The residues involved in the FMN/HEME ET tunneling are represented and labeled in green color.

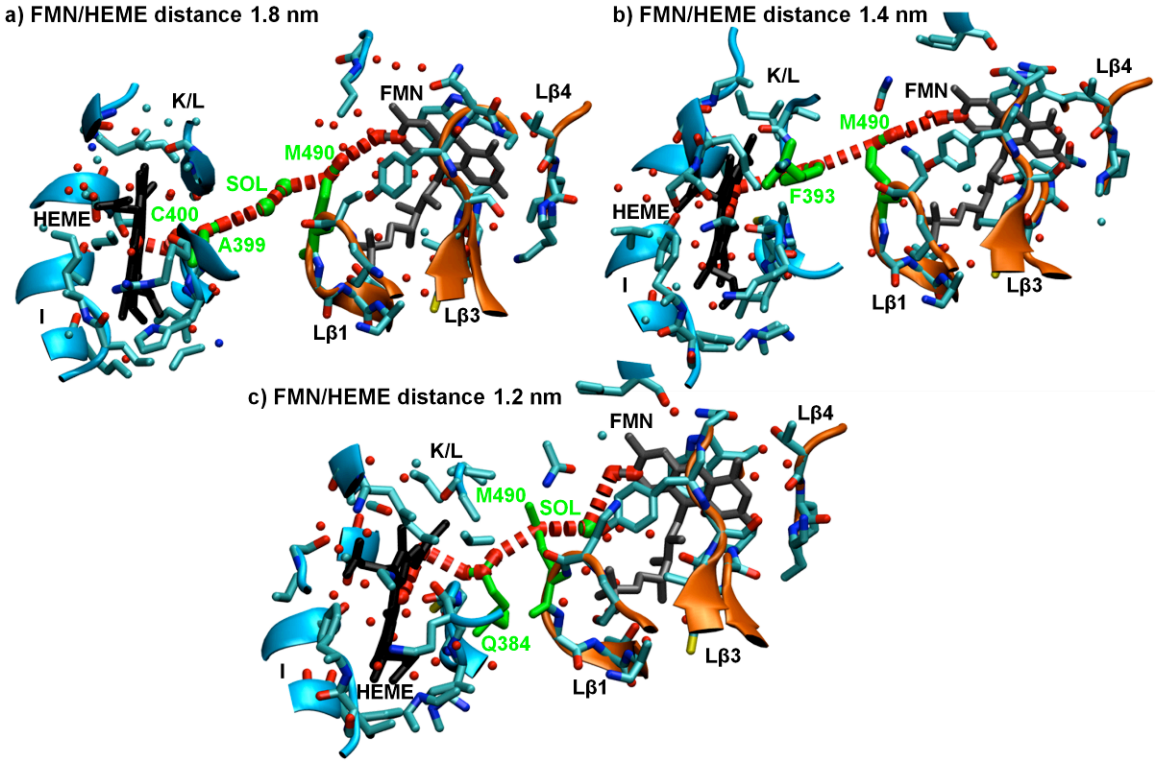


Figure 8: The components for the (a) first, (b) second and (c) third eigenvectors of the A and F chain in the solution (red color) and complex (black color) simulations. The green vertical line separates the HEME and FMN domains. Horizontal bars represent helixes (labeled) and beta sheets in blue and orange color, respectively. The regions involved in the cofactor binding are represented by horizontal bars in purple color.

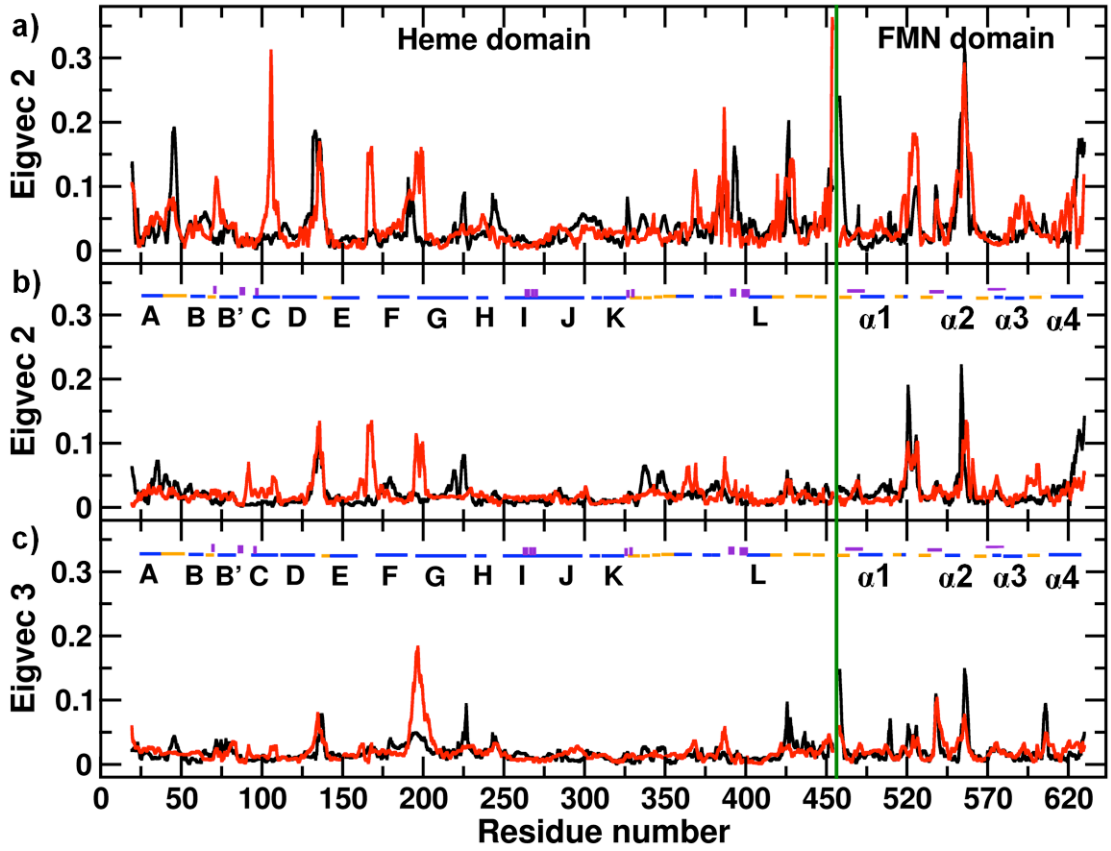


Figure 9: The RMSF of the protein backbone atoms along the first (a), second (b) and third (c) eigenvectors after projection of the trajectory on the corresponding eigenvector of the A and F chain in the complex (a1, b1 and c1) and the isolated domain (a2, b2 and c2) simulation. The 10 sequential frames represent the extension of the fluctuations in the trajectories along the eigenvectors. The first extreme conformation is shown in green color and last extreme in violet color. Other conformations of the A and F chain are in sky blue and tan color, respectively. Helices and loops in the FMN domain are labeled. N and C indicate the N- and C-termini of the P450BM-3 domains (labeled in red color).

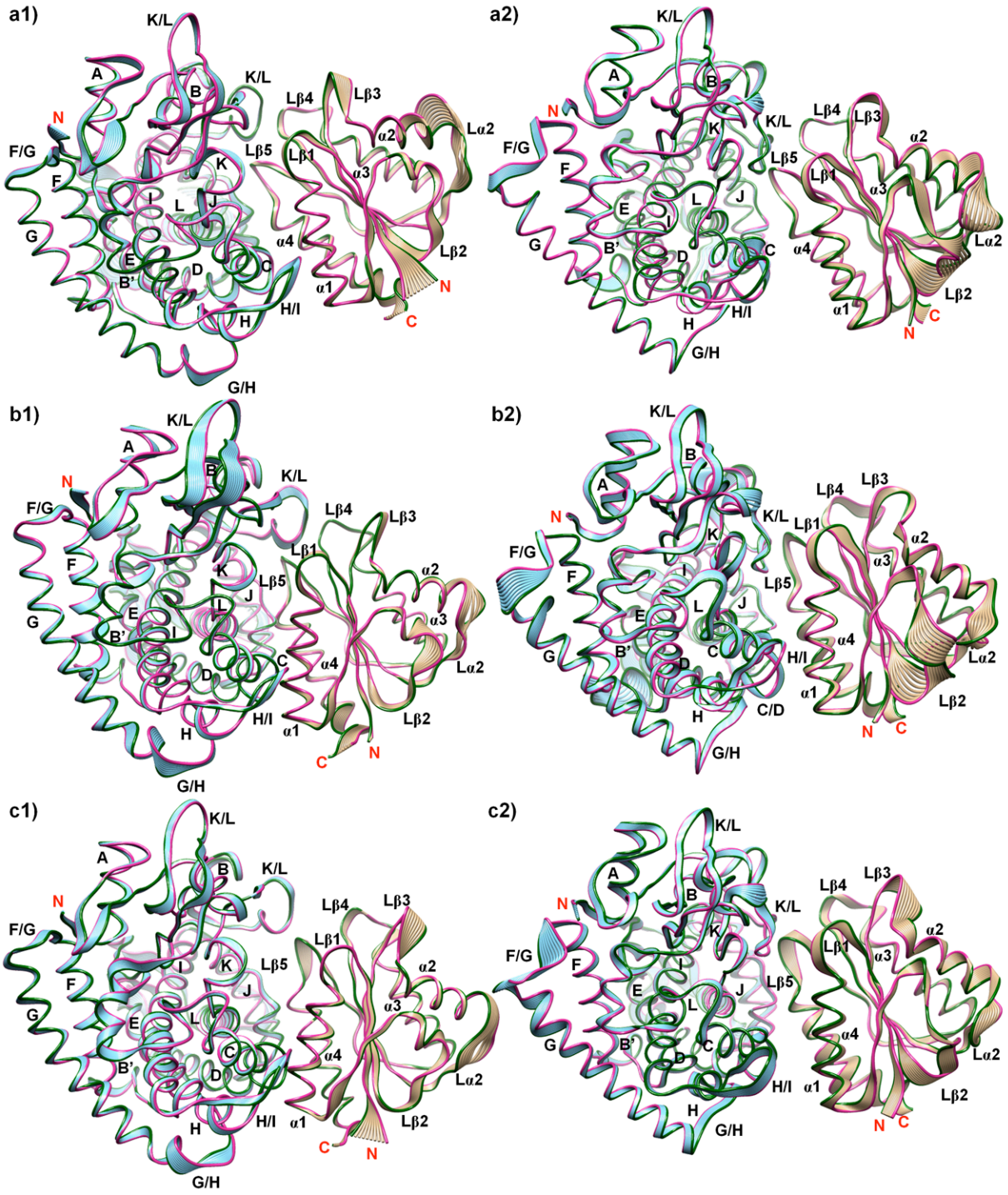
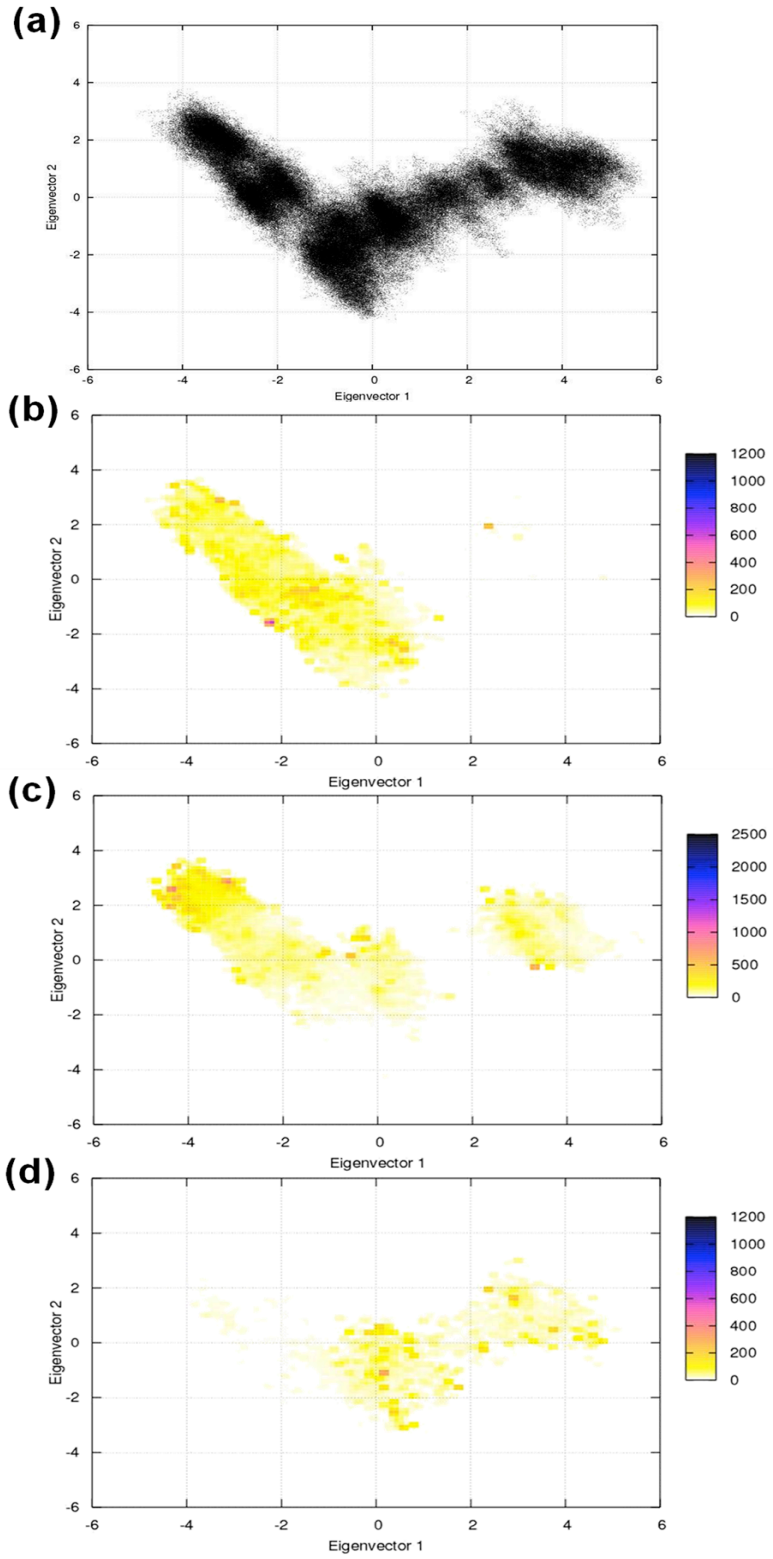


Figure 10: (a) The 2D projection of AF chain trajectory onto the corresponding first two eigenvectors obtained from the backbone atoms covariance matrixes. (b), (c) and (d) show the projection of ET pathway 1, 2 and 7 on the first and second eigenvectors of AF chain.


(a)
Eigenvector 2
Eigenvector 1
(b)
Eigenvector 2
Eigenvector 1
1200
1000
800
600
400
200
0
(c)
Eigenvector 2
Eigenvector 1
2500
2000
1500
1000
500
0
(d)
Eigenvector 2
Eigenvector 1
1200
1000
800
600
400
200
0