4

Molecular dynamics of the HSA-heme complex

T.R. Cuya Guizado, S.R.W Louro and C. Anteneodo
accepted European Biophysics Journal (2011)
4.1 Summary

In order to analyze the interaction heme-HSA, molecular dynamics simulations of HSA with bound heme has been performed. Based on results of X-ray diffraction, the binding site of the heme, localized in subdomain IB, has been considered. We analyze the fluctuations and their correlations along trajectories to detect collective motions. Complementarily, the role of H-bonds and salt bridges in the stabilization of the heme in its pocket is also investigated. Also a comparative analysis between the bonded/unbounded HSA is shown in the Appendix D.

4.2 Introduction

The heme binds HSA, in subdomain IB (71, 70), allowing the protein to store the heme in excess in blood serum. This capacity of HSA makes it a target of diverse studies aiming to create the conditions to turn it a transporter of oxygen (artificial blood) (45, 69). Therefore, it is crucial to get insights on the interaction between HSA and heme, which motivates the present work.

Correlated movements have been reported for several proteins (40, 13, 47). These movements are important as they may affect the access and binding of ligands (2). Fujiwara et al. (32) studied HSA and the effect of fatty acids in 10 ns simulations finding possible collective movements in the HSA molecule. Here we pursue the study of collective modes when the heme is bound in its pocket.

We performed molecular dynamics simulations of the HSA-heme complex in explicit water. The results of trajectories and their comparison with experiments furnish a reference for future MD simulations containing that complex. Furthermore, by means of the dynamical cross-correlation map (DCCM) and principal component analysis (PCA), we determine the intradomain and interdomain correlated motions of HSA, in particular those regions that are correlated with the residues composing the heme cavity. We also show the role of H-bonds and salt bridges in the interaction of the heme with its pocket in HSA.
4.3
Methods

4.3.1 Molecular dynamics

Molecular dynamics (MD) is a powerful tool to analyze the interaction ligand-protein at the atomic level. Classical trajectories were obtained by means of the GROMACS 3.2 package (67) with the Gromos96 53A6 force-field.

The initial coordinates of HSA were obtained from the structure 1O9X deposited in the Protein Data Bank (PDB), obtained from X-ray diffraction, with a resolution of 0.32 nm (71, 55). In the PDB file, the protein is found complexed with hemin and fatty acids. Then, in a first step, the fatty acids were removed. Atoms were included in incomplete residues by means of Swiss PDB Viewer (34). The topology of HSA was the one given in Gromos96 parameter base, while that of the heme was obtained in previous work (21).

The system was solvated using the SPC water model (10) in a dodecahedron box. The distance from the protein to the dodecahedron walls was 1.2 nm. The HSA has a negative charge of -13e and the heme a negative charge of -2e at physiological pH, thus, the system HSA-heme was balanced with 15 Na\(^+\) counter ions.

A NPT (fixed number of atoms, pressure and temperature) thermodynamic ensemble was considered. The systems were thermodynamically coupled to a 310 K bath with the Berendsen thermostat, at a coupling pressure of 1 bar using a Berendsen barostat (9). Electrostatic and van der Waals interactions were evaluated up to a cut-off of 1 nm. The long range electrostatic interactions were treated through the PME (Particle Mesh Ewald) method (28).

Previous to MD simulations, energy minimization (EM) was performed following steps:

1. EM with solute position restriction using steepest-descent algorithm;
2. EM without solute position restriction using steepest-descent algorithm;
3. EM without solute position restriction using LBFGS algorithm (limited memory variation of the Broyden-Fletcher-Goldfarb-Shanno algorithm) (52).

A short MD of 500 ps was run with solute position restriction in order to avoid the overlapping of the van der Waals radii of water molecules and for orientation of hydration shells. The outcome of this stage is considered the starting structure \((t = 0)\). After the initial preparation of the system, a MD
of 100 ns was run with a simulation time-step of 2 fs to obtain useful data. The trajectory was saved every 4 ps.

4.3.2 Data analysis

To have a rough measure of the compactness of HSA, we calculated the radius of gyration

\[ R_g(t) = \left[ \frac{1}{M} \sum_{i=1}^{N} m_i || \mathbf{r}_i(t) - \mathbf{r}_{CM}(t) ||^2 \right]^{1/2}, \quad (4-1) \]

where \( \mathbf{r}_i(t) \) is the position of atom \( i \) at MD frame \( t \), \( m_i \) is the mass of atom \( i \), \( M = \sum_{i=1}^{N} m_i \) and \( \mathbf{r}_{CM} \) is the position of the center of mass of HSA. \( R_g \) was computed for alpha carbons (\( C_\alpha \)) with the program \texttt{g.gyrate} of GROMACS. As a function of time it gives information about breath motions and conformational changes of the protein.

To quantify the structural distance between two conformations, we consider the weighted root mean square deviation (RMSD)

\[ RMSD(t) = \left[ \frac{1}{M} \sum_{i=1}^{N} m_i || \mathbf{r}_i(t) - \mathbf{r}^*_i ||^2 \right]^{1/2}, \quad (4-2) \]

where \( \mathbf{r}^*_i \) is the position of atom \( i \) in the reference structure (fitted to minimize the RMSD). It was computed, for \( C_\alpha \) atoms only, with the program \texttt{g.rms} of GROMACS (67).

To determine which aminoacids of HSA are in closer contact with the heme, we computed the intermolecular surface contact (ISC) between the protein and the ligand. The ISC is determined by the intersection of the surfaces accessible to the solvent of both the protein and the ligand. The ISC furnishes information on the interaction between the heme and given residues of HSA. The ISC was obtained by means of the software SURF (64). This software is based on the Connolly algorithm (15), using a 1.4 Å test radius and 1 point/Å².

Additionally, the formation of H-bonds and salt bridges was also analyzed by means of \texttt{g.hbond} and \texttt{g.salbr} of GROMACS, respectively.

To detect the flexibility of different regions of the protein, we considered the RMSF (root mean square fluctuation of atomic positions)

\[ RMSF_i = \left( || \mathbf{r}_i(t) - \langle \mathbf{r}_i \rangle ||^2 \right)^{1/2}, \quad (4-3) \]

where the brackets denote time average. The RMSF was computed with the program \texttt{g.rmsf} of GROMACS, with the option of previously eliminating
global rotations and translations.

Correlated motion is crucial in the performance of proteins (as examples, see (59, 48)). Hence we aimed to detect the regions of HSA with correlated fluctuations and their potential modifications by the heme. Then, we computed the dynamic cross-correlation matrix (DCCM)

$$DCCM_{ij} = \frac{\langle (r_i - \langle r_i \rangle)(r_j - \langle r_j \rangle) \rangle}{\sigma_i \sigma_j},$$  \hspace{1cm} (4-4)

where \( \sigma_i = (\langle (r_i - \langle r_i \rangle)^2 \rangle^{1/2} \) is the standard deviation. Moreover, since only internal motions are relevant to find collective movements, global rotations and translations were previously eliminated from the trajectories by means of program \texttt{trjconv}. These coefficients \( DCCM_{ij} \) measure the linear correlation between residues. A vanishing coefficient means absence of linear correlation. A unitary value indicates a positive correlation (movements in the same direction), while -1 indicates a negative correlation (correlated motions in opposite directions). Calculations were performed by means of \texttt{g_covar} of GROMACS.

Another technique to identify and characterize correlated motions is the principal component analysis (PCA) applied to the covariance matrix. In order to do so, we obtained the covariance matrix

$$C_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle,$$  \hspace{1cm} (4-5)

where \( \{x_i, 1 \leq i \leq 3N\} \) are the atomic coordinates. The covariance matrix was calculated by considering the alpha-carbon (C\(_\alpha\)) of each HSA residue (hence \( N = 579 \) is the number of residues in our case, starting from the fourth residue, namely serine), by means of \texttt{g_analyze} and \texttt{g_anaeig} of GROMACS. Global rotations and translations were previously eliminated from the trajectories. PCA relies on the diagonalization of the covariance matrix, which generates \( 3N \) modes, thus yielding 1737 modes. The eigenvalues \( \lambda_i \) (1 \( \leq i \leq N \)) represent the amplitude of the fluctuations along their corresponding eigenvectors. Then, the main collective movements occur along the directions of the eigenvectors associated to the largest eigenvalues. Usually the first 20 eigenvectors (ordered by decreasing eigenvalues) represent more than 90\% of the total motion (23). Such reduction of the original dimensionality arises due to the large number of internal constraints and restrictions due to covalent bonds, weak non-bonded interactions, etc.
4.4 Results and Discussion

4.4.1 Radius of gyration ($R_g$) and root mean square deviation (RMSD)

The time evolution of $R_g$ is depicted in Fig. 4.1 (top panel). It fluctuates around $R_g = 2.78$ nm, a bit below the initial value (approx. 2.87 nm, which corresponds to the structure 1O9X). It presents a tendency to decay with time approaching a mean value close to those reported for the free HSA, namely, 2.74 ± 0.035 nm, obtained from small-angle neutron scattering (43) and 2.72 ± 0.023 nm from MD simulations (32). This implies that the presence of the heme does not provoke a significant expansion of the protein.

The RMSD between the overall complexed protein and the experimental structure 1O9X was computed. The $C_\alpha$ RMSD plot is displayed in Fig. 4.1 (bottom panel). Stabilization is fairly attained after about 20 ns, around 0.35 nm.

4.4.2 Intermolecular surface contact (ISC)

The time averaged ISC was computed for the trajectories within the interval 20-100 ns. The ISC of all the aminoacids possessing non-null ISC is displayed in Fig. 4.2, although low values (below 10 Å²) may not be significant.
All of them belong to the IB subdomain, as expected, as long as the heme binds to this subdomain. Our analysis indicates that the aminoacids that more strongly interact with the heme, in order of decreasing ISC, are: LEU-115, ARG-186, TYR-161, TYR-138, ILE-142, HIS-146, ARG-114, ARG-145, LYS-190 and PHE-149 (all of them with ISC > 10 Å²).

Figure 4.2: Intermolecular surface contact between HSA and heme. Solid and error bars indicate time average and standard deviation, respectively.

These results are consistent with experimental ones. For the crystallographic methemalbumin obtained by X-diffraction at 1.9 Å, it was reported that there are five residues displaying close interaction with the heme and appearing to be key contributors to the high binding affinity, namely, TYR-161, ILE-142, TYR-138, HIS-146, and LYS-190 (70). In fact, from our results, these five residues are among those with ISC > 15 Å², however, our analysis reveals that mainly LEU-115 and ARG-186 also interact strongly. These residues are included in a larger list of domain IB residues having close interaction with the heme or contributing to the hydrophobic surface of the binding pocket, cited in Ref. (70): TYR-161, PHE-157, ARG-186, LEU-182, ARG-117, PHE-134, LEU-135, LEU-154, PHE-149, ILE-142, HIS-146, ARG-114, LYS-190, SER-193, ALA-158, TYR-138, LEU-115, MET-123, PHE-165, and PRO-118. The results in Fig. 4.2 recover the above list, but extra residues with non-null ISC additionally arise: LEU-112, PRO-113, VAL-116, LEU-139, GLU-141, ARG-145, LEU-185, GLY-189. Notice however that they have ISC values below 10 Å² (hence, may be not significant), except ARG-145 whose ISC is close to 20 Å². In an experimental study at 3.2 Å resolution, it was shown that two tyrosines are essential for heme binding: TYR-138 and TYR-161, while other
residues, LEU-139, HIS146, ILE-142 and LEU-154, exhibit slight adjustments upon heme binding (71). Therefore, our MD simulations may reveal the importance of a wider set of aminoacids in the sub-domain IB, than the crystal structure analysis does.

Figure 4.3: Energy of interaction with the heme of the residues presenting ISC > 10Å².

Figure 4.4: ISC vs. interaction energy (between each residue and the heme).
We also computed the energy of interaction with the heme (Coulomb and Lennard-Jones) of each residue presenting a value of ISC > 10Å². The results are presented in Fig. 4.3, displaying values of the order of 6 kcal/mol in average. As expected a correlation exists between the values of ISC and interaction energy, the larger the ISC the larger the absolute value of the energy, as depicted in Fig. 4.4.

The study of H-bonds and salt bridges that will be presented in the next section allows to confirm the importance of some aminoacids in the interaction with the heme.

**H-bonds and salt bridges**

H-bonds and salt bridges are essential in the folding process as well as in the binding of ligands. H-bonds are specific and directional. In this study a cutoff of 0.35 nm for the donor - acceptor distance and a cutoff angle of 30° for acceptor - donor - hydrogen were chosen. The H-bond has an energetic contribution of 5-30 kJ/mol, it is stronger than a van der Waals interaction but weaker than a covalent bond. Salt bridges are electrostatic interactions between residues with opposite charges. They are weak interactions if compared to H-bonds, but in buried sites the energies can reach about 30 kJ/mol. Differently from H-bonds, salt bridges are not directional. A cutoff of 0.5 nm was used in its evaluation.

Table 4.1: Permanence of H-bonds. Only the residues with more than 5% of permanence are shown. Lifetimes are also shown.

<table>
<thead>
<tr>
<th>donor</th>
<th>acceptor</th>
<th>permanence (%)</th>
<th>lifetime (ps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEU-115-N</td>
<td>heme-O1D</td>
<td>32</td>
<td>24.8</td>
</tr>
<tr>
<td>LEU-115-N</td>
<td>heme-O2D</td>
<td>37</td>
<td>28.5</td>
</tr>
<tr>
<td>HIS-146-NE2</td>
<td>heme-O1A</td>
<td>17</td>
<td>15.0</td>
</tr>
<tr>
<td>TYR-161-OH</td>
<td>heme-NB</td>
<td>12</td>
<td>2.8</td>
</tr>
</tbody>
</table>

The permanence of H-bonds is defined as the percentual fraction of time that the H-bond remains formed. The permanence of the H-bonds established between the IB sub-domain and the heme were calculated and are shown in Table 4.1. Lifetimes are also exhibited. According to the above criteria, LEU-115 forms two H-bonds whose permanence is larger than 30% and it is the main aminoacid to fix the heme via this interaction. HIS-146 and TYR-161 also make H-bonds but with lower permanence. Notice that HIS-146 has been
A computational approach to the structure and dynamics of human serum albumin: effects of the heme

proposed before as contributing to heme stability through this interaction (70). In fact, it also fulfils the salt-bridge criteria.

Moreover, we observe that the heme suffers a process of desolvation when it enters the pocket in the IB subdomain. In fact, the average number of H-bonds that the heme inside its main site forms with water is about $1.6 \pm 0.9$, much smaller than in aqueous environment (20). But, despite site hydrophobicity, there are water molecules forming H-bonds with the attached heme.

Table 4.2: Permanence of salt bridges. OA and OD refer to the carboxyl groups of the heme. Only salt bridges with more than 10% permanence are shown.

<table>
<thead>
<tr>
<th>salt bridge</th>
<th>permanence (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARG-114 heme-OA</td>
<td>13</td>
</tr>
<tr>
<td>ARG-114 heme-OD</td>
<td>66</td>
</tr>
<tr>
<td>ARG-145 heme-OA</td>
<td>45</td>
</tr>
<tr>
<td>ARG-145 heme-OD</td>
<td>67</td>
</tr>
<tr>
<td>ARG-186 heme-OA</td>
<td>12</td>
</tr>
<tr>
<td>ARG-186 heme-OD</td>
<td>13</td>
</tr>
<tr>
<td>LYS-190 heme-OA</td>
<td>69</td>
</tr>
</tbody>
</table>

Salt bridges are shown in Table 4.2. LYS-190, ARG-145, ARG-114 are the main residues that might fix the heme in the cavity via salt bridges. All of them can form two salt bridges with heme, one with permanence greater than 50%. Instead, ARG-186 salt bridges are weak. Experimental results based on X-ray diffraction (71, 70) also show that LYS-190 and ARG-114 (together with HIS-146) play important roles in heme stabilization by salt bridges. Our MD study shows that besides them, ARG-145, and in a lower degree also ARG-186, contribute to the heme stabilization too.

4.4.3 Root mean square fluctuations

The RMSF was computed in the window from 20 to 100 ns, to exclude the transient relaxation before 20 ns. The RMSF of the $C_{\alpha}$ of each residue is displayed in Fig. 4.5. For comparison, the RMSF of experimental data computed from the b-factors ($RMSF = \sqrt{3B/(8\pi^2)}$, where $B$ is the experimental b-factor) of the structure 1O9X is also displayed. The RMSF profiles are in good agreement. The N-terminus region has less mobility than the C-terminus. In comparison with free HSA (5 ns) MD simulations (3) the presence of the heme stabilizes subdomain IA, reducing its fluctuations. Modifications of the
Figure 4.5: Cα RMSF for each residue of the complexed protein from MD simulations. The RMSF of structure 1O9X (computed from experimental b-factors) is also included for comparison.

RMSF profile of subdomain IIB are also observed. The heme does not affect the RMSF of residues in subdomain IB.

4.4.4 Analysis of collective motions

In the present section we aim to detect how the heme affects the correlated motions in the protein, through the analysis of the DMCC and of the PCA of the covariance matrix. We considered the trajectories between 20-100 ns. As reference structure to eliminate global translations and rotations, we considered the respective structures at \( t = 20 \text{ ns} \).

The dynamic cross-correlation matrix (DCCM) is shown in Fig. 4.6. Outstanding off-diagonal regions correspond to (positive) correlations only between subdomains A and B of each one of the three domains (intradomain correlations). The main inter-domain correlation is observed between IB and IIIB, which are negatively correlated. Since, in MD simulations, subdomains IB and IIIB are at an average distance of 3.7 nm, there must be communication pathway between them. In fact, by comparing X-ray structures 1O9X and 1AO6 (hemin free) (63), through the analysis performed by means of DynDom (41), one notices that the heme enhances that distance (from 3.46 to 3.98 nm), through conformational changes of domains I and III around domain II, without intradomain deformations. Subdomain I suffers a twist motion of 22.8° and subdomain III a hinge motion with a rotation angle of 19.1° (depicted in Fig. 4.7). The key residues (hinge residues) contributing to the twist inter-domain motion are GLY-189 and LYS-190. In the relative movement of subdomain III with respect to subdomain II, the hinge residues (located in their interface)
are ASN-386 to LYS-391, MET-446 to ASP-451, CYS-476 to THR-478, and LEU-481 to CYS-486. The aperture conformational change could facilitate the

Figure 4.6: DCCM map for HSA-heme, shown below the diagonal. Above the diagonal only elements with absolute value larger than 0.5 are displayed.

Figure 4.7: Comparison of structures 1AO6 (light gray) and 1O9X (colored). The two arrows represent the rotation axes of domains I (red) and III (yellow) with respect to domain II (blue). Hinge residues are displayed in green.
binding of another substrate. Let us recall that the region between domains I and III has been proposed as a possible weak secondary binding site for heme binding (25).

Figure 4.8: Spectrum of eigenvalues as a function of eigenvector index. Main frame: cumulative normalized eigenvalues $\sum_{j=1}^{i} \lambda_j / \sum_{j=1}^{N} \lambda_j$ vs index $i$. Inset: $\lambda_i$ vs $i$.

Results of PCA are presented in Fig. 4.8 and Table 4.3. In Fig. 4.8, the spectrum of eigenvalues is represented in terms of the cumulative normalized eigenvalues, which quantify the relative contribution to protein fluctuations. Table 4.3 shows the numerical values for the 10 first modes. One observes that these modes contribute with 76.2% of the total movement. Moreover, the first 4 eigenvectors are already responsible for more than 50% of the motion.

We analyzed stationarity by performing the calculations in separate time intervals. The overlap (53) between the subspaces spanned by the first 10 eigenvectors in the intervals 20-60 ns and 60-100 ns is 0.32 indicating that the system is still subject to conformational changes in the interval 20-100 ns, then the results must be interpreted with caution. Analyzing separate 5 ns intervals and evaluating the overlap of all pairs, we generated a $16 \times 16$ matrix. The off-diagonal elements were found to belong to the interval $0.34 \pm 0.03$. These relative low value confirms the above mentioned non-stationarity. In fact, a 585-aminoacid protein, could require a very long time for stabilization. But there is also the possibility that the overlap is not a good measure of convergence, then, we also considered the cosine content (c.c.) (39), a measure of convergence through the evaluation of the degree of diffusive components. The cosine content (see Table 4.3) of the first mode is high, indicating that the
Table 4.3: Contribution of the first 10 modes to the total motion of HSA-heme. The % motion is given by the percentual cumulative normalized eigenvalues. The ordered eigenvalues $\lambda_i$ and cosine content (c.c.) corresponding to each mode $i$ are also displayed.

<table>
<thead>
<tr>
<th>Mode $i$</th>
<th>$\lambda_i$</th>
<th>% motion</th>
<th>c.c.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9.8</td>
<td>30.3</td>
<td>0.604</td>
</tr>
<tr>
<td>2</td>
<td>3.5</td>
<td>41.2</td>
<td>0.261</td>
</tr>
<tr>
<td>3</td>
<td>3.0</td>
<td>50.3</td>
<td>0.036</td>
</tr>
<tr>
<td>4</td>
<td>2.9</td>
<td>59.1</td>
<td>0.172</td>
</tr>
<tr>
<td>5</td>
<td>2.1</td>
<td>65.5</td>
<td>0.002</td>
</tr>
<tr>
<td>6</td>
<td>1.0</td>
<td>68.6</td>
<td>0.026</td>
</tr>
<tr>
<td>7</td>
<td>0.8</td>
<td>71.0</td>
<td>0.001</td>
</tr>
<tr>
<td>8</td>
<td>0.7</td>
<td>73.0</td>
<td>0.031</td>
</tr>
<tr>
<td>9</td>
<td>0.5</td>
<td>74.7</td>
<td>0.039</td>
</tr>
<tr>
<td>10</td>
<td>0.5</td>
<td>76.2</td>
<td>0.017</td>
</tr>
<tr>
<td>1-10</td>
<td>24.8</td>
<td>76.2</td>
<td></td>
</tr>
</tbody>
</table>

associated motion is mainly diffusive. For the other modes, the cosine content is low.

![Figure 4.9: Superposition of the frames corresponding to the first 4 modes of HSA-heme. Colors correspond to different time steps, from the first (red) to the last (blue).](image)

The changes in the dynamics of the first 4 modes, that concentrate about 50% of the total variability, are shown in Fig. 4.9. The movements projected along the first 4 eigenvectors are also represented in Fig. 4.10 through porcupine graphs (57). Both sets of figures display the correlated fluctuations. In comparison with the results shown in Ref. (32), the dominant motion are clearly associated to different modes. In particular, one also observes in Figs. 4.9 and 4.10 that, in the complex, a cooperative motion between
4.5 Conclusions and remarks

Molecular dynamics simulations were carried out for the HSA-heme complex, starting from the crystal structure 1O9X. By means of the analysis of ISC, H-bonds and salt bridges, we determined the residues that strongly interact with the heme. Residues LEU-115, HIS-146, LYS-190, ARG-145, ARG-114, ARG-186 help to fix the ligand in the hydrophobic pocket via either H-bonds or salt bridges, representing molecular latches. The dynamics reveals that the heme induces anti-correlated fluctuations between domains IB and IIIB, a dominant mode which is not present in the free HSA. We expect that the present results give insights on the mechanisms that rule the interaction HSA-heme and can be useful on future research to turn HSA into an oxygen carrier. Moreover, the good agreement between simulations and crystallographic experimental results sets a basis for MD simulations of systems containing the HSA-heme complex.