ORIGINAL PAPER

Analysis of the structure and dynamics of human serum albumin

T. R. Cuya Guizado

Abstract Human serum albumin (HSA) is a biologically relevant protein that binds a variety of drugs and other small molecules. No less than 50 structures are deposited in the RCSB Protein Data Bank (PDB). Based on these structures, we first performed a clustering analysis. Despite the diversity of ligands, only two well defined conformations are detected, with a deviation of 0.46 nm between the average structures of the two clusters, while deviations within each cluster are smaller than 0.08 nm. Those two conformations are representative of the apoprotein and the HSA-myristate complex already identified in previous literature. Considering the structures within each cluster as a representative sample of the dynamical states of the corresponding conformation, we scrutinize the structural and dynamical differences between both conformations. Analysis of the fluctuations within each cluster set reveals that domain II is the most rigid one and better matches both structures. Then, taking this domain as reference, we show that the structural difference between both conformations can be expressed in terms of twist and hinge motions of domains I and III, respectively. We also characterize the dynamical difference between conformations by computing correlations and principal components for each set of dynamical states. The two conformations display different collective motions. The results are compared with those obtained from the trajectories of short molecular dynamics simulations, giving consistent outcomes. Let us remark that,

beyond the relevance of the results for the structural and dynamical characterization of HAS conformations, the present methodology could be extended to other proteins in the PDB archive.

Introduction

Human serum albumin (HSA) is a 66.5 kDa (585 aminoacids) helicoidal protein, consisting of three structurally similar domains I, II, and III), each of them formed by two subdomains (A and B), (see Fig. 1). It represents about 60 % of the total protein content in the blood serum.

The main function of HSA is to regulate the colloid osmotic pressure. Other physiologic functions include binding and transport of molecules, as well as anti-oxidant and anti-inflammatory actions [1]. HSA has the ability to bind insoluble substances such as fatty acids, porphyrins, and a wide variety of drugs. The pharmacokinetics and pharmacodynamics of any drug depends strongly of its interaction with the HSA, these interactions can be monitored using fluorescence techniques, circular dichroism, and spectroscopic techniques [2]. HSA binds protoporphyrin IX, heme, and synthetics porphyrins in the IB domain, these porphyrins have both regions hydrophobic and hydrophilic [3–5]. This makes HSA of great interest to investigate its carrier properties for drug delivery. Also HSA binds the heme and other porphyrins, this feature converts this protein into a target for photo dynamic therapy and a potential oxygen carrier [6–8].

There are two main binding sites with high affinity for diverse substances, known as Sudlows's sites. The first one is located in the core of the IIA subdomain and the second in

T. R. C. Guizado
Physics Department, Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro, Brazil

T. R. C. Guizado (✉)
Chemistry Department, Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro, Brazil
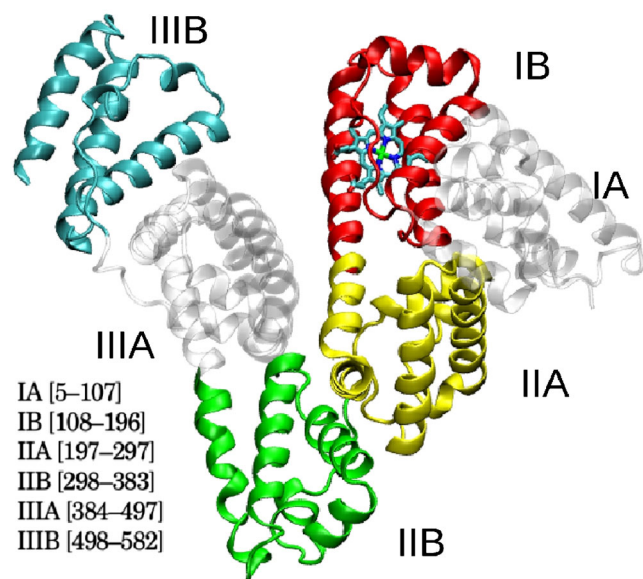e-mail: trcuyag@gmail.com.br

Fig. 1 HSA structure

IA [5–107]
IB [108–196]
IIA [197–297]
IIB [298–383]
IIIA [384–497]
IIIB [498–582]

the IIIA subdomain, more exposed to the solvent than the former one [9]. Besides these two sites, there are another five sites, with high affinity for fatty acids [10, 11].

Based on experimental research, the existence of mainly two conformations of HSA is known, the first one corresponding to the defatted HSA and the second to the fatted protein (HSA-myristate conformation) [10–15]. Zunszain et al. [14] proposed that fatty-acids in subdomain IIA might contribute to achieve an intermediate conformation. However, this state could have arisen as a consequence of the crystallization procedure.

In this work we aim to gain insights about the differences among conformations, the allosteric effects of ligands, the hinge residues involved in the changes between conformational states, the correlated regions that might be involved in allosteric modulation, among others.

In order to do so, we analyze the tertiary structure of HSA from X-ray crystallographic data deposited in the Protein Data Bank [16], where over 50 structures are available. Through this analysis two dissimilar conformations are clearly detected.

Complementarily, molecular dynamics (MD) simulations of HSA in explicit water were carried out starting from each typical conformation. Therefore, a structural and dynamic characterization of the main conformations of HSA is presented.

We found correlations between domain IB and II, coincident with the allosteric modulation reported in the experimental literature [17–19] but not in previous simulations [20, 21].

Also a strong anti-correlation between domains I and IIIB can be observed, given by a motion of approximation and separation of domains I and III.

Material and methods

HSA structures

The coordinates of different structures of HSA deposited in PDB were obtained from X-ray crystallography. More than 50 structures can be found. Each structure corresponds to the HSA under different experimental conditions, either unbound or bound to diverse ligands (mainly anesthetics and fatty acids). Sequences from Ser-5 up to Ala-578 were considered in calculations. Structures with missing residues (C_α) within that sequence were discarded.

Then 42 structures were kept. Single lacking atoms were completed with the SPDBV program [22]. Table 1 shows all the selected structures and the corresponding ligands. As will be shown below, they can be grouped in two clusters, C1 and C2.

Clustering analysis

Over the set of 42 X-ray structures of HSA, a clustering analysis was carried out by taking into account the C_α positions only. For this purpose, we used the g_cluster tool of GROMACS, with the linkage method [23, 24], which includes a structure into a cluster when its distance to any element of the cluster is below a given threshold (0.1 nm was used). In this way two clusters were found, called cluster 1 (C1) and cluster 2 (C2). The first one is composed of 12 elements and the second of 28 (shown in Table 1). Two structures were observed to fall out of those two clusters: 3JQZ and 1O9X, although they could be assigned to clusters C1 and C2, respectively, they were excluded from computations, as will be explained below.

The average structure of each cluster was obtained with the g_rmsf tool. We chose as a representative member (named Xi) of each cluster Ci the structure with the minimal root mean square deviation (RMSD) with respect to the average one. The tool g_confirms was used to evaluate the RMSD between the respective average and each element of the cluster.

Molecular dynamics

The GROMACS 4.5.4 package with the Gromos96/53a6 force field was used for molecular dynamics simulations [24, 25].

Gromos96/53A6 force field was parameterized using the free enthalpy of hydration, was validate for relevant proteins like hen egg-white lysozyme, and shown that it is better at describing the folding—unfolding balance of the peptide [25]. Also the HSA-heme complex was simulated with this force field, reproducing experimental values like the gyration radius, hydrogens bonds, and molecular environment [8].

Table 1 Structures of HSA and ligands. Ligands are indicated by a 3-letter code: aza= azapropazone, azt=3'-azido-3'-deoxythymidine, cap=capric acid, cei=cis-5,8,11,14-eicosatetraenoic acid, dia=diazepam, dii=3,5-diiodosalicylic acid, dla=dansyl-l-arginine, dls=dansyl-l-asparagine, dlg=dansyl-l-glutamate, hal=halothane, hem=hemin, ibu=ibuprofen, imx=[(1R,2R)-2-{[(5-fluoro-1H-indol-2-yl)carbonyl]amino}-2,3-dihydro-1H-inden-1-yl]acetic acid, ind=indomethacin, iod=iodipamide, lau=lauric acid, lys=lysophosphatidyletanolamine, myr=myristatic acid, ole=oleic acid, oxy=oxyphenbutazone, pal=palmitic acid, phe=phenyl-butazone, pro=propofol, rew=r-(+) enantiomer of warfarin, sal=salicylic acid, sew=s-(−)enantio-mer of warfarin, ste=stearic acid, thy=thyroxine, tia=tri-iodobenzoic acid, war=warfarin. Only the subdomains and inter-faces with presence of ligands are shown

PDB	Res (nm)	IA	IA/IIA	IB	IIA	IIA/IIB	IIIA	IIIA/IB	IIIB
C1									
2BXC	0.310				phe				
1E7B	0.240					hal	hal		
1E78	0.260								
1E7A	0.220						pro		pro
2BXD	0.305				war				
2BXF	0.295						dia		
2BXB	0.320				oxy				
2BX8	0.270			aza	aza				
2BXG	0.270					ibu	ibu		
1A06	0.250								
1BM0	0.250								
3 LU6	0.270				imx	imx			
C2									
2BXK	0.240		**myr**	myr	aza ind	myr	myr(2)		myr
2BXM	0.250		**myr**	ind myr	myr ind	myr	myr(2)		myr
1H9Z	0.250		**myr**	myr	rew	myr	myr(2)		myr
2BXN	0.265		**myr**	myr	iod iod	myr	myr(2)	iod	myr
1HA2	0.250		**myr**	myr	sew	myr	myr(2)		myr
2XVW	0.265		**myr**	dla myr	dla	myr	myr(2)		myr
1BKE	0.315		**myr**	tia myr	tia		myr(2)		myr
2BXQ	0.260		**myr**	ind myr	ind phe	myr	myr(2)		myr
2BXO	0.260		**myr**	myr	oxy	myr	myr(2)		myr oxy
1HK4	0.240		**myr**	myr	myr	myr	myr(2)	thy	myr
2BXP	0.230		**myr**	myr	phe	myr	myr(2)		myr
1Bj5	0.250		**myr**	myr			myr(2)		myr
2XS1	0.270		**myr**	myr dlg	dlg	myr	myr(2)		myr
1GN1	0.240		**ole**	ole	ole	ole	ole(2)		ole
1E7E	0.250		**cap**	cap	cap	cap(3)	cap(2)	cap	cap
1E7C	0.240	hal	**myr**	myr hal	hal(2)	hal(3)	myr(2)		myr
1GNJ	0.260		**cei**	cei	cei	cei	cei(2)		cei
2BXI	0.250		**myr**	aza	aza	myr	myr(2)		myr
1E7I	0.270		**ste**	ste	ste	ste	ste(2)		ste
1E7G	0.250		**myr**	myr	myr	myr	myr(2)		myr
1E7H	0.245		**pal**	pal	pal	pal	pal(2)		pal
2BXL	0.260		**myr**	myr	dii	dii	myr(2)		myr
2XVV	0.240		**myr**	myr dls	dls	myr	myr(2)		myr
1E7F	0.245	lau	**lau**	lau	lau	lau	lau(2)		lau
3CX9	0.280		**myr**		lys	myr	myr(2)		myr
1N5U	0.190		**myr**	hem		myr	myr(2)		myr
3B9M	0.270		**myr**	myr sal	sal azt	myr	myr(2)		myr
3B9L	0.260		**myr**	myr azt	myr azt	myr	myr(2)		myr

Each representative structure was solvated in explicit solvent using the SPC water model in a dodecahedron box [26]. The minimal distance from the protein to a dodecahedron wall was 1.2 nm.

Van der Waals interactions were considered up to a cut-off of 1 nm. For the long range electrostatic interactions, the particle mesh Ewald (PME) treatment was used [27]. The systems were balanced with 13 Na^+ counter ions.

A NPT (fixed number of atoms, constant pressure, and temperature) thermodynamic system was considered coupled to a 310 K bath with the v-rescale method [28], at a coupling pressure of 1 bar by means of a Berendsen barostat [29].

Previous to the MD simulation, three cycles of energy minimization (EM) were performed, where the output of each cycle was the input of the next one. Each cycle consisted of the following steps: (1) EM with protein position restriction using steepest-descent algorithm; (2) EM without position restriction using steepest-descent algorithm; (3) EM without position restriction using the LBFGS method.

In each cycle a convergence energy criterion of 10^{-6} kJ mol^{-1} was used between successive steps.

The MD simulation consisted of the followings steps: (1) MD of 500 ps with position restriction in order to avoid the overlapping of the van der Waals radii of the water molecules and for orientation of hydration shells. (2) MD of 20 ns to obtain the useful data was performed. We use this short simulation time in order to obtain conformations approximate to the typical structures (X1, $X2$). The simulation time-step was set to 2 fs and data were recorded every 1 ps.

Covariance analysis

Correlated motions are crucial in the performance of proteins [30, 31]. To detect the regions of HSA with correlated fluctuations and their differences either between clusters or between MD trajectories, we first computed the dynamic cross-correlation matrix (DCCM):

$$DCCM_{ij} = \left\{ < (r_i - < r_i >).(r_j - < r_j >) > \right\}/(\sigma_i \sigma_j) \quad (1)$$

where r_i is the position of the C_α of residue i and $\sigma_i = \cdot (r_i - \cdot r_i)^2 \cdot 1/2$ is the standard deviation. Moreover, since only internal motions are relevant to find collective movements, global rotations and translations were previously removed from the trajectories by means of tool *trjconv*. The coefficients $DCCM_{ij}$ measure the linear correlation between residues i and j. A vanishing coefficient means absence of linear correlation. Positive values indicate movements in the same direction, while negative ones mean anti-correlation (correlated motions in opposite directions). Calculations were performed by means of tool *xypder* plugin for Pymol [32, 33].

Another technique to identify and characterize correlated motions is the principal component analysis (PCA) applied to the covariance matrix:

$$C_{ij} = (x_i - x_i)(x_j - x_j) \quad (2)$$

where $\{x_i, 1 \leq i \leq 3N\}$ are the atomic coordinates. The covariance matrix was calculated by considering only the C_α of each HSA residue (hence $N=578$, the number of residues in our case) by means of *g_covar*. Each cluster was converted into a pseudo-trajectory (multi pdb), for all the cases rotations and translations were previously eliminated.

PCA relies on the diagonalization of the covariance matrix, which generates 3 N modes, thus yielding 1734 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 a few eigenvectors represent most of the total motion [34]. This reduction of the original dimensionality arises from the large number of internal constraints and restrictions due to covalent bonds, weak non-bonded interactions, etc. An interesting references about the application of PCA to the HSA in simulations time of 200 ns can be found in the work of Paris et al. [35].

Dynamic domains and relative inter-domain motions

To identify dynamic domains, we employed the program DynDom 1.5 [36]. The procedure described by the authors of the program was thoroughly followed [37]. The treatment by DynDom is based on the description of the motion of a rigid-body as a screw motion that can be decomposed into a rotation about an axis and a translation along that same axis [36].

Rotation vectors (in the direction of the screw axis) and translation lengths were calculated for all residues, considering N, C, C_α, and C_β atoms (a pseudo-C_β position was assumed for glycines). Residues within a cluster of rotation vectors are considered to belong to the same dynamic domain provided that this domain consists of at least 20 residues and that it is rigid enough to be considered as a rigid body. The residues connecting different clusters of rotation vectors define bending regions.

For each pair of accepted domains, DynDom characterizes the inter-domain motion by calculating the inter-domain screw axis and one or several bending regions [36]. If an interdomain screw axis passes within 0.55 nm of the C_α atom of any of the residues in a bending region, this axis is considered an effective hinge axis. This indicates that the residues in the bending region act as a mechanical hinge. An axis parallel to the connection of the centers of mass of the two domains describes a twist motion (twist axis), while a perpendicular axis describes a closure motion (closure axis) [36]. Any axis can be decomposed into components parallel and perpendicular to the line joining the centers of mass of the domains. Then, following the definition of closure by Hayward et al. [38], the degree of closure (% closure) of an axis is measured as the square of its projection on the closure axis×100.

In order to evaluate the relative inter-domain motions we used the Domain Select program of the DynDom server. This tool allows one to select the domains and analyze the rigid-body movement of one domain (the moving domain) relative to another one (the fixed domain), in the same way as the DynDom program does [36]. Coils were not included. Figure 2
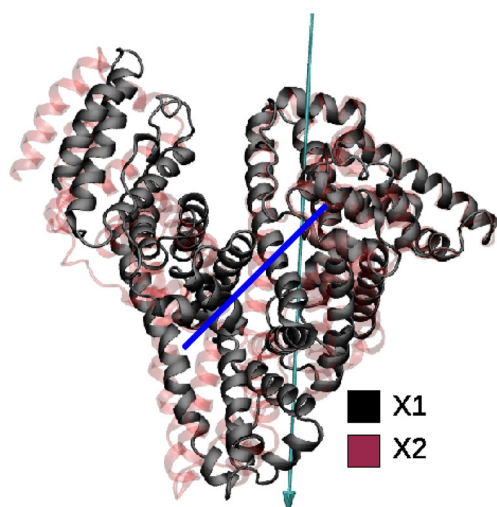
Fig. 2 Screw axis (*green vector*) for the pair of dynamic domains detected by DynDom from the analysis of the representative structures X1 and *X*2 [36]. The (*blue*) segment connects the CM of both dynamic domains

shows the screw axis that represents the relative movement of structures X1 and *X*2.

Results and discussion

We considered the coordinates of 42 different structures of HSA deposited in PDB, obtained from X-ray crystallography, as explained in Materials and methods. Table 1 shows all the selected structures and the corresponding ligands. As will be shown below, they can be grouped in two clusters, C1 and C2.

Cluster analysis

Applying the linkage procedure (see Materials and methods), two clusters were found, based on C_α: C1 with 12 elements and C2 with 28 (see Table 1). The average structure for each cluster was obtained. The RMSDs of each element in a cluster with respect to the average structure were computed and listed in decreasing order, as shown in Table 2. The top ones (named X1 and *X*2) were considered representative of each cluster.

These two clusters typify two well defined conformations, observed in X-ray diffraction of HSA subject to various experimental conditions and bound to diverse substances. For both clusters, the maximal (intracluster) deviation from the average does not exceed 0.08 nm, while the RMSD between X1 and *X*2 is 0.46 nm.

The conformation X1 is typical of the apoprotein (1AO6, 1BM0, and 1E78), hence it can be identified with the native structure. Most part of the structures in its cluster are complexed with anesthetics, which have preference for subdomains IIA, IIIA and the interface IIA/IIB, and no fatty acids are present. Considering that the deviation with respect to the average structure of C1 is smaller than 0.080 nm, this indicates that the binding of anesthetics does not affect the native structure.

The main interactions between the anesthetics and the apo protein are hydrogen bonds, for instance in the experimental work of Bhattacharya et al., [13] refer to the bind of proposol, one molecule binds in subdomain IIIA, and the other binds in subdomain IIIB. The propofol molecule in IIIA binds in an a polar pocket with the phenolic hydroxyl group, making a H-bond with the Leu-430 and with the aromatic ring of the anesthetic sandwiched between the side chains of Leu-453 and Asn-391 [13]. It also makes close contacts with several
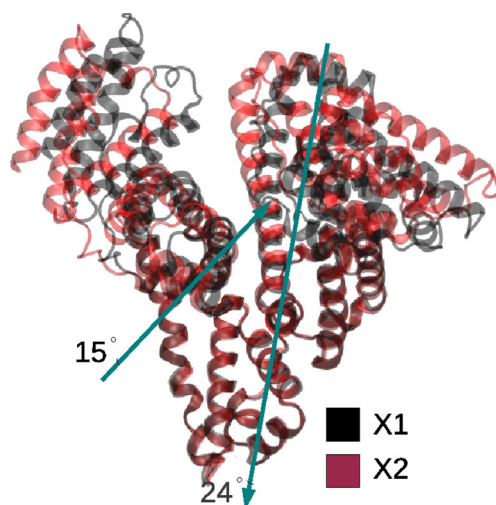
Table 2 RMSD of each structure. RMSD was computed with respect to the average structure of each cluster. Structures marked with * (named X1 and *X*2, respectively), presenting the lowest RMSD, will be considered as representative of each cluster

C1		C2			
PDB	RMSD (nm)	PDB	RMSD (nm)	PDB	RMSD (nm)
*2BXC	0.0328	*2BXK	0.0209	1E7E	0.0317
1E7B	0.0340	2BXM	0.0220	1E7C	0.0331
1E78	0.0347	1H9Z	0.0221	1GNJ	0.0338
1E7A	0.0353	2BXN	0.0232	2BXI	0.0341
2BXD	0.0364	1HA2	0.0242	1E7I	0.0344
2BXF	0.0370	2XVW	0.0243	1E7G	0.0393
2BXB	0.0399	1BKE	0.0245	1E7H	0.0407
2BX8	0.0408	2BXQ	0.0257	2BXL	0.0423
2BXG	0.0572	2BXO	0.0267	2XVV	0.0435
1AO6	0.0690	1HK4	0.0278	1E7F	0.0442
1BM0	0.0701	2BXP	0.0282	3CX9	0.0527
3 LU6	0.0787	1BJ5	0.0298	1N5U	0.0630
		2XSI	0.0308	3B9M	0.0650
		2CNI	0.0316	3B9L	0.0680

Fig. 3 Superposition of the X1 (*black*) and X2 (*red*) structures, based on the matching of the Cα of domain II

side chains (Asn-391, Leu-407, Arg-410, and Tyr-411) [13]. The second propofol molecule binds in a pocket located in sub-domain IIIB which is mainly occupied by aromatic residues (Phe-502, Phe-507, Phe-509, and Phe-551). The entrance to the binding pocket is quite polar with a number of polar residues in close proximity [13].

In another study performed by Ghuman et al.[15] the azapropazone binds the subdomain IA and IB, but it displaces the fatty acid in the HSA-myristate complex. This molecule has a H-bond with the Lys-199 and is surroundings by aromatics residues (His-242, Try-214, Tyr-150) in the IB subdomain [15].

In the structure 2BXC, phenylbutazone binds the IIA subdomain and has well defined H-bonds with the Arg-257 and Tyr-150 and is surroundings by polar residues (Arg-218, Arg-222, Lys-195, Lys-199, Ile-264) [13].

In general several interactions like H-bonds, polar contacts, and pi-stacking, could be responsible for the binds of the anesthetic in the HSA and eventually it would compete with fatty acid.

C2 contains 28 elements. In most of them, molecules of fatty acids together with other ligands are present. The cluster presents a deviation smaller than 0.068 nm with respect to the average.

This confirms that the presence of fatty acids induces a well-defined conformation, different from the native one. The conformation X2 is typical of the HSA-(fatty-acids) conformation, known in the related literature as HSA-myristate [13]. The superposition of X1 and X2 is shown in Fig. 3.

Moreover, we found two structures that do not present a good fit with either of the two clusters: 1O9X and 3JQZ, obtained with resolutions of 0.32 nm and 0.33 nm. The RMSD of 1O9X with respect to the average of C1 and C2 is 0.60 nm and 0.25 nm, respectively, while the RMSD of 3JQZ is 0.17 nm and 0.49 nm, respectively. Also, taking into account that the distance between clusters is 0.46 nm, 1O9X and

3JQZ can be considered as belonging to C2 and C1, respectively. However, they will be excluded in the following calculations.

Let us also note that Zunszain et al. [13], obtained 1O9X in 2003 and proposed the found HSA-myristate-hemin structure to be an intermediate state between the apo and the HSA-myristate conformations. However, the authors emphasize that the difference between the HSA-myristate and HSA-myristate-hemin conformations may be due to differences in crystal packing. They also argue that a weak density of the fatty-acid in subdomain IIA might contribute to the failure in achieving a typical HSA-myristate conformation. In the structures for the complex HSA-myristate-others deposited in PDB, the subdomains IIA and IIB are not always occupied. Moreover, notice that the 1N5U structure [39], HSA-myristate-hemin (released later, in 06/24/2003), without fatty-acid molecule in the subdomain IIA, belongs to C2. The structure 3JQZ corresponds to a dimer of HSA. In each monomer a lidocaine molecule is located in the interface between the domains I and III. From our previous analysis of the components of C1, lidocaine is not expected to generate
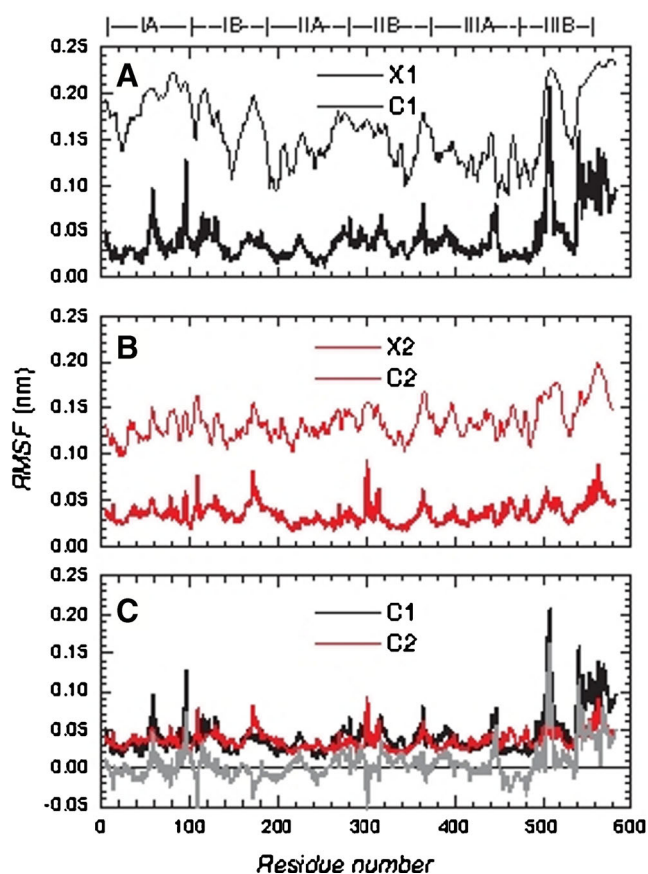

Fig. 4 RMSF for each cluster. In panels A and B, full lines correspond to the RMSF computed for all the members of the cluster with respect to the average structure, dashed lines to the RMSF of the representative structure computed from experimental b-factors. In panel C, C1, and C2 RMSF together with their difference profile (light, *green line*) are shown. In all cases, only Cα where considered

a significant conformational change. Then, the discrepancy of 0.17 nm between 3JQZ and C1 may be due to the poor resolution of the structure or crystallization packing.

The observation of only two structures is interesting because due to the size (585 amino acids) and number of subdomains, one could expect more than two conformations. Also we would expect that metalporphyrins (as in 1N5U) were bound strongly to certain residues to change the conformation and dynamics of the protein [40], but, apparently these effects are very local in HSA and do not induce global changes with respect to the HSA-myristate.

In contrast, fatty acids (myristic, oleic, capric acid, etc.) do induce the change X1→X2. From Table 1, other conclusions can be drawn: the regions poorly populated by ligands in cluster C2 (mainly subdomain IA and also interface IIIA/IB) cannot be responsible for the conformational change X1→X2. Those regions which are occupied in all members of C2 are subdomain IIIA (with two fatty acid molecules), subdomain IIIB and interface IA/IIA (where fatty acid tail penetrates IIA from IA). Both regions IIIB and IA/IIA

interface are key to induce the conformational change. Subdomains IB, IIA, and interface IIA/IIB are not occupied in all members of C2, therefore they alone cannot in principle be responsible for the conformational change. Hence, mutations in the key regions might strongly affect the structure and dynamics of the protein.

Root mean square fluctuations

For each cluster, the RMSF by residue was obtained from the pseudo trajectory (considering all the structures in the cluster) and the average structure was taken as reference.

In panels A and B of Fig. 4, one observes that the fluctuations among members of clusters C1 and C2 follow the experimental ones (obtained from b-factors) for the representative structures X1 and X2, respectively. The fluctuations are smaller for the pseudo-trajectories since frames (deposited X-ray structures) correspond to already averaged structures. Also notice in panel C that the presence of ligands does not change significantly the RMSF profile of the apoprotein, up to about

Fig. 5 RMSF by residue with respect to the average for separate domains of C1 and C2. In C1 the more flexible regions are 1a: 55–60, 1b: 94–97, 1c: 114–116, 3a: 439–447, 3b: 498–512, 3c: 517–529, and 3d: 538–582. In C2, 1a: 93–96, 1b: 108–110, 2a: 300–302, 2b: 311–314, and 3a: 556–566. In the lower panel we exhibit the position of those portions on the two representative structures. Notice that the flexible regions are typically associated to random coils
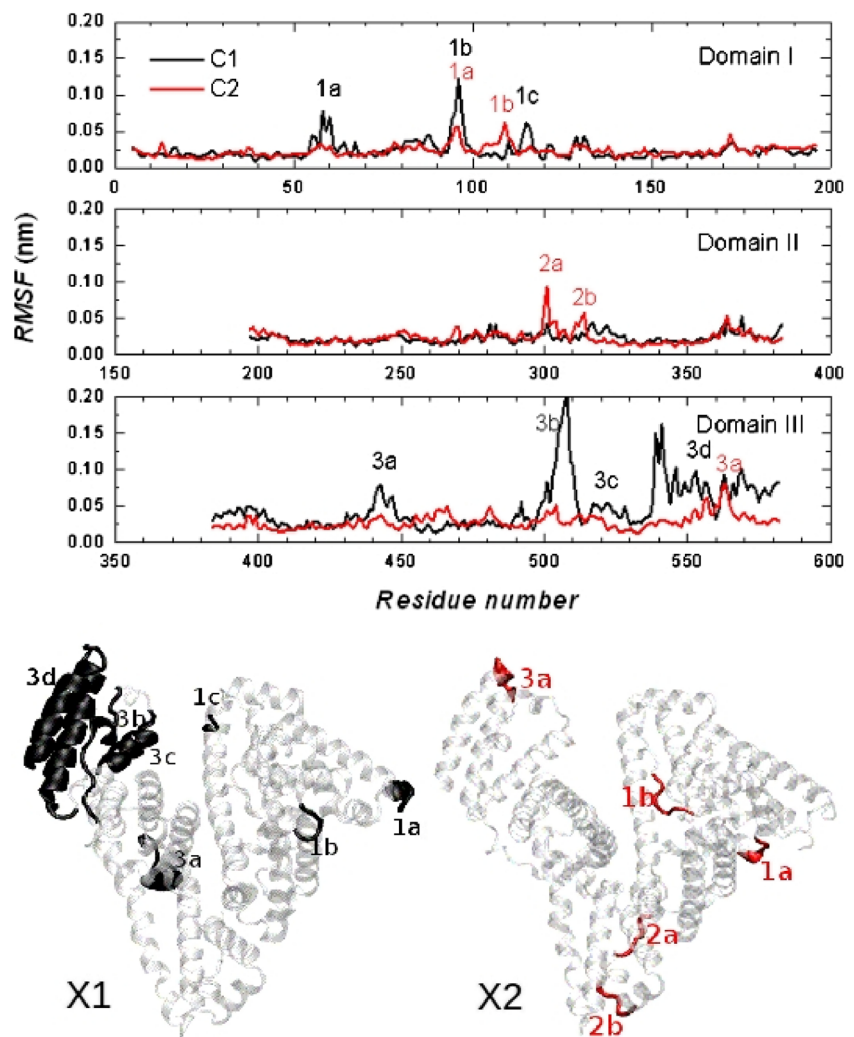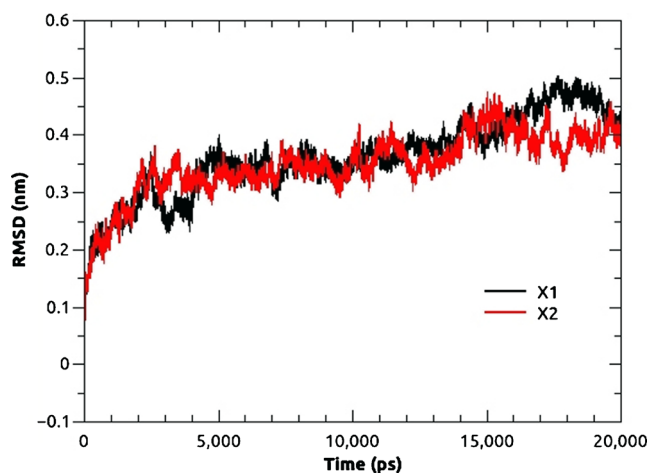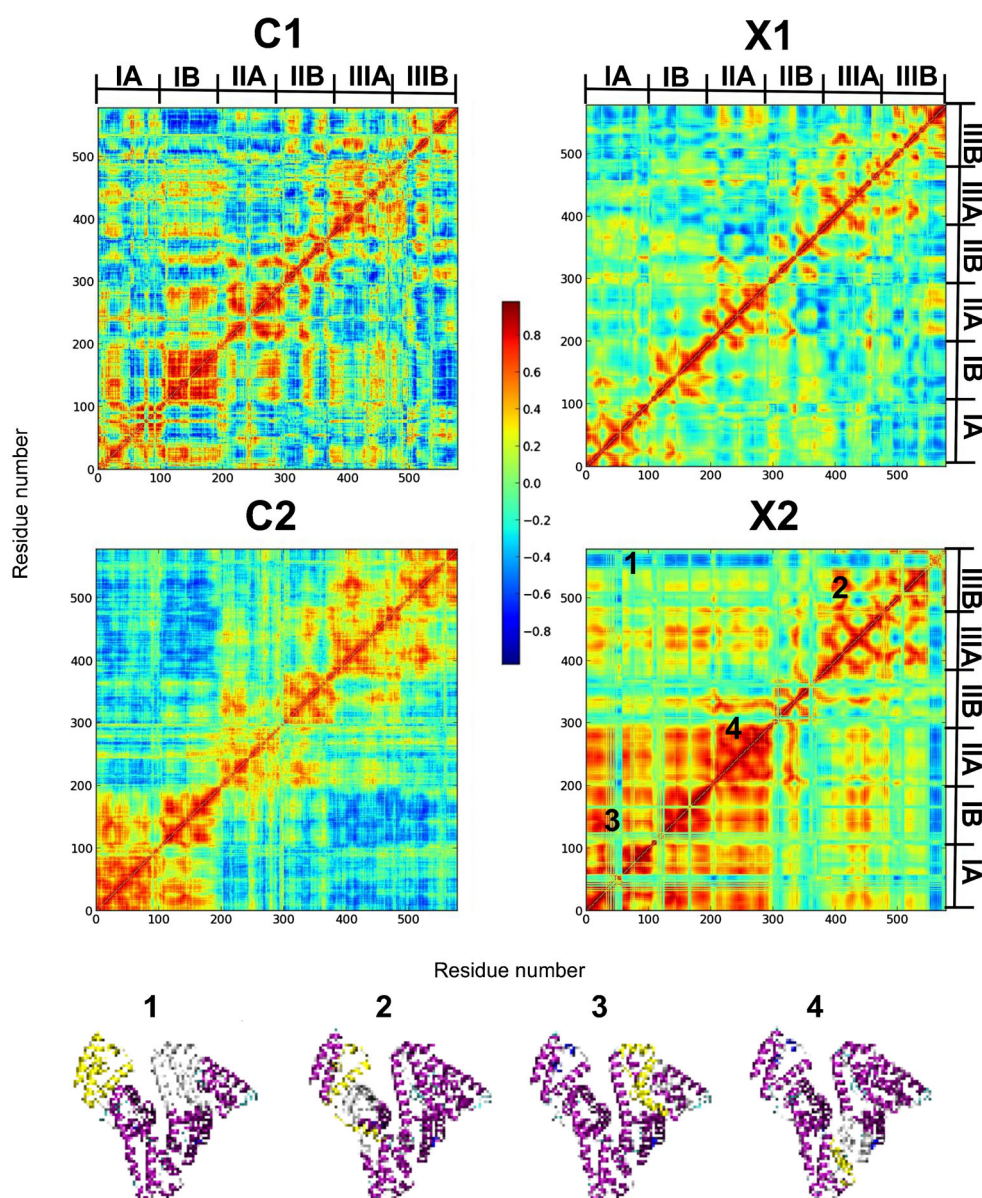
Fig. 6 RMSD of X1 and *X*2 vs time. For both systems a plateau is achieved in about 5 ns

residue 500. In this region the maximal deviation is restricted to isolated residues and does not exceed 0.03 nm. This is interesting considering that the subdomains IIA and IIIA are the main binding sites for drugs. Fatty acids mainly stabilize subdomains IA (N-terminal) and IIIB (C-terminal).

To estimate the conservation of each domain within its cluster, the RMSD with respect to the average structure was evaluated for each domain of the protein separately, as depicted in Fig. 5.

Domain II presents a low level of fluctuations which is very similar for both clusters. It is the most rigid region of HSA. Similar profiles for both clusters are also observed in domain I. Differences in the fluctuation profiles appear mainly in domain III: while C1 shows large fluctuations in the region 500–582, they are quenched in C2. The effect of ligands in C2

Fig. 7 Dynamic cross correlation for C1 and C2 (considering all members in each cluster) and for X1 and *X*2 (from MD). The four more noticeably correlated regions are shown in the HSA structures in *white and yellow colours*

is to reduce the fluctuations of coil portions when compared to C1 (regions 1a, 1b, 3a, 3b). Also the presence of ligands in subdomain IIIB strongly stabilizes that region. This suggests that changes in this subdomain contribute to the transition $X1 \rightarrow X2$.

Considering that domain II is the most rigid one, we matched the C_α of domain II of structures X1 and $X2$, in order to visualize the relative displacements of the other two domains, as depicted in Fig. 3. The best fit of domain II was 0.068 nm.

Root mean square deviation

For molecular dynamics (MD) trajectories, the root mean square deviation (RMSD) of X1 and $X2$ with respect to the corresponding X-ray structures was obtained using the g-_rmsd_ tool, where only the C_α were considered. The time evolution of the RMSD gives a measure of the drift from the crystalline structure. For both structures, in Fig. 6, a plateau is achieved in about 5 ns. However, $X2$ is only locally stable, since for larger times it is expected to evolve toward X1, which is the stable conformation in the absence of ligands. We performed calculations with data in the interval 5–20 ns, in

order to sample the dynamics in a time interval where the structures stabilize close to the respective initial states.

Covariance analysis

The dynamic cross correlation matrix (DCCM) was obtained for each cluster. Since the clusters (mainly cluster 1), contain a small number of members, calculations were complemented with those from MD of X1 and $X2$, as exhibited in Fig. 7.

For structures X1 and $X2$ there are common positive correlations between nearby helices belonging to subdomains IA and IB (region marked 3), IIA and IIB (region 4), and IIIA and IIIB (region 2). These regions are also observed in C1 and C2. A strong negative correlation between subdomains IB and IIIB, for system $X2$ (region 1) is more pronounced that for X1.

This anti-correlated region is also present in C2, while it is not clear in the noisy map of C1 caused by the poor statistics. Given the large distance between both subdomains, changes on the H-bond network and electrostatic interactions could account for this correlation. In X1, two very localized anti-correlated regions (with absolute values greater than 0.5) exist between IB and IIA and between IB and the interface IIA/IIB. For those regions, only correlations lower than 0.5 were found

Fig. 8 First principal component obtained from X-ray and MD data, in porcupine representation, where the cones point in the direction of movement
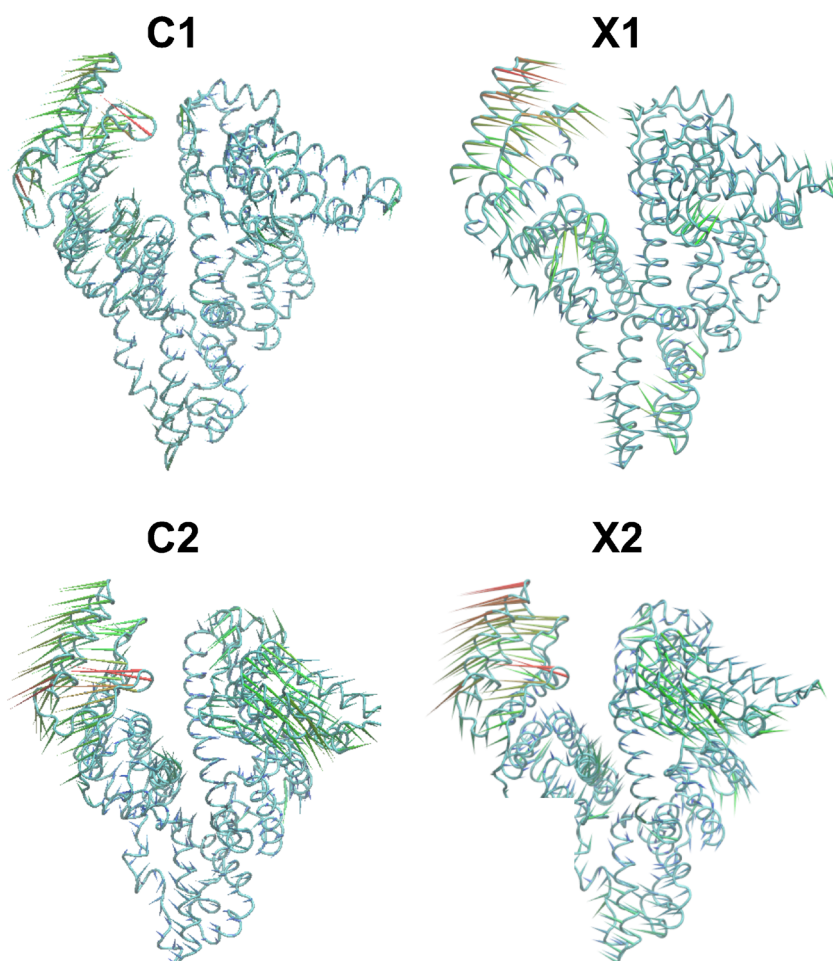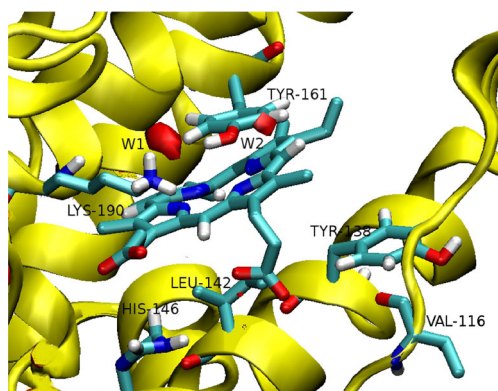
Fig. 9 The regions in red (W1 and W2) show the isosurfaces of probability to find water molecules, these water molecules are important for the heme stability in the HSA. Figure obtained from reference [8]

in X2. According to the experimental literature [41–43], saturation of Sudlow's site I (localized in IIA) with warfarin produced allosteric effects in the heme clef (localized in IB), decreasing in one order of magnitude the affinity constants.

Moreover, studies based on both optical absorbance and NMR spectroscopy show that the observed conformational change of Mn(III)heme-HSA (where Mn(III)heme is bound in IB) is allosterically induced by myristic acid bound to Sudlow's site I [44]. Moreover, experimental works show that IIA and IB are functionally coupled [10, 17]. There are also other studies showing that the fatty acid molecule in the interface IA/IIA could induce the change from X1 to X2 [14]. In all the cases, significant anti-correlations were found between IA/IIA and IB, but this effect is more noticeable in the HSA-myristate system (corresponding to X2 and C2).

The analysis of principal components provides insights about the collective motions of the protein. In Fig. 8, the first principal component was plotted in porcupine representation.

The analysis was performed using the *g-covar* and *g-_anaeig* tools, where a representative trajectory of 5000 frames of each MD simulation was used as input. In the case of the clusters, all the frames were used as a pseudo-trajectory.

A collective (anti-correlated) motion between IB and IIIB is observed for the C2 cluster (HSA-myristate conformation), confirmed by MD simulations (X2). This is in accord with the strong negative correlation observed in the DCCM matrix. No similar collective motion is observed for X1.

The contribution of the ten first PCs is shown in Table 3. The sample of C2 seems sufficient to obtain information (comparable with MD) about the first modes of the protein.

In contrast, the small sample size of C1 leads to overestimate the contribution of the modes. Comparing the spectra of X1 and X2, the first mode is dominant in the protein with about 30 %, the second mode contributes in about 13 %. The first ten modes account for about 70 % of the total motion.

Dynamic domains and relative inter-domain motions

For the analysis of dynamic domains the DynDom server was used (see Material and Methods) [31]. Twelve different pairs of elements of each cluster were selected and their elements were compared by means of the server. We excluded coil regions from the input sequences. For domain I, the rigid regions were 5–76, 87–92, 98–105, 121–193; for domain II: 197–291, 305–383, and for domain III 384–490, 515–582. Results are shown in Table 4.

Ala-194, Lys-195 and Gln-196 (bending residues in Table 5 appear to be key residues in the conformational difference between both clusters. These residues are named hinge residues. Modifications (not shown) occur in the psi dihedral angle of Ala-194 (about 14°) and the phi angle of Lys-195 (about −20°).

The differences observed in three cases, the structures 2BXD-1HA2, 2BX8-2BXQ, and 2BXG-2BX0, are related to a fixed domain of five residues which is not detected in the other cases. Such residues are: Glu-280, Lys-281, Leu-284, Glu-285, and Lys-286.

Figure 2 shows the dynamic domains that result from the comparison of X2 with respect to X1. In all cases analyzed in

Table 3 Ten first eigenvalues. Their percent contribution to the total motion is shown

Index	C1		C2		X1		X2	
	Eingenv.	%	Eingenv.	%	Eingenv.	%	Eingenv.	%
1	0.702	48.9	0.284	34.8	2.673	29.6	3.038	33.0
2	0.286	68.9	0.095	46.5	1.058	41.3	1.273	46.8
3	0.129	77.9	0.074	55.6	0.581	47.7	0.539	52.7
4	0.085	83.9	0.060	62.9	0.462	52.8	0.424	57.3
5	0.073	89.0	0.045	68.4	0.340	56.6	0.270	60.2
6	0.052	92.7	0.041	73.5	0.291	59.8	0.260	63.1
7	0.032	94.9	0.030	77.2	0.267	62.8	0.220	65.5
8	0.027	96.8	0.026	80.4	0.247	65.5	0.191	67.5
9	0.024	98.5	0.024	83.3	0.219	67.9	0.169	69.4
10	0.013	99.4	0.018	85.5	0.166	69.8	0.152	71.0

Table 4 Dynamic domain analysis. Statistics over pairs of members of each cluster

C1	C2	Fixed residues	Moving residues	Rotation (degrees)	Bending (residues)	Closure (degrees)
*2BXC	*2BXK	5–194	195–582	25.87	194–195	55.52
1E7B	2BXM	7–195	196–580	26.18	195–196	53.07
1E78	1H9Z	7–195	196–580	25.47	195–196	55.68
1E7A	2BXN	7–195	196–580	26.38	195–196	55.30
2BXD	1HA2	7–195	196–580	25.74	195–196	55.25
		281–284	285–580		280–281	
					284–285	
2BXF	2XVW	7–195	196–580	25.77	195–196	53.15
2BXB	1BKE	7–195	196–580	25.84	195–196	52.33
2BX8	2BXQ	7–195	196–580	25.87	195–196	57.31
		281–284	285–578		280–281	
					284–285	
2BXG	2BXO	7–201	202–280	26.22	197–204	55.77
		281–285	286–576		280–281	
					285–286	
1AO6	1HK4	7–195	196–580	25.13	195–196	62.05
1BM0	2BXP	7–194	195–580	25.36	194–195	62.08
Averages				25.8		56.13

Table 5, the rotation angle of the mobile domain with respect to the fixed one is about 26°, the % closure is between 52 and 62°, indicating a twist motion, however, since the distance of the line between the CM of the domains to the axis is greater than 0.55 nm, it cannot be considered an effective hinge axis.

Our analysis of dynamic domains indicates that Gln-196 could be a key residue in the dynamics of HSA. MD also shows that this residue changes its H-bond profile with the Tyr-150, when comparing X1 and X2 trajectories. Based in X-ray diffraction experiments, Ascenzi et al. [17], concluded that, due to the fatty acid presence in IB, the Tyr-150 changes its H-bonds starting the network reordering, increasing the volume of Sudlow's site and altering its polarity distribution. Precisely, Gln-196 is one of the residues involved in such change. This could indicate that the Gln-196 acts as an effective hinge residue.

From Figs. 4 and 5, one notes that domain II is more rigid than the other ones. Then, we chose it to match the two representative structures X1 and X2, in order to visualize the relative displacements of domains I and III with respect to domain II, as shown in Fig. 3. The inter-domain motion was evaluated using the first 12 structures of C2 in Table 1 and compared them with C1. Mean (and standard deviation) values were computed (see Table 5). In all cases, domain II was taken as reference. Only the rigid regions were considered, that is, coils were excluded. Therefore, each domain was considered as a rigid body. The best matching of the domains II of X1 and X2 was 0.076 nm (calculated with *g_confirms*). When domain II is selected as reference, two predominant motions of X2 with respect to X1 are observed, the first one shows a separation hinge motion of domain III, the second a twist motion of domain I.

The role of the water in the HSA conformations

Two different conformations were observed in our experimental analysis. The MD simulation of these conformations gives an energy barrier associated with this change of ~−500 kJ mol^{-1} from X1 in relation to X2. In the X2 conformation (HSA-myristate) the domains I and III are more separate in relation to X1 (apo conformation) and have more exposed hydrophobic solvent accessible surface (SAS). The average difference in the total SAS between X2 and X1 is 6.96 nm^2 and 78 % is hydrophobic contribution. It means the solvation energy is predominant and important in the stabilization of both conformations.

Table 5 Inter-domain relative motions

Domains	Rotation (degrees)	Translation (nm)	% Closure (degrees)	Angle axis/CM (nm)	Distance axis/CM (nm)	Motion type
3 to 2	14.6±0.5	−0.041±0.011	99.1±1.4	85.7±3.6	0.877±0.091	Hinge
1 to 2	23.7±0.8	0.111±0.017	32.7±2.5	34.8±1.5	0.644±0.060	Twist

Also it is known that the water has different roles in the binds process to form the complex HSA-ligand. In the HSA-myristate conformation the waters penetrate the main binding site of porphyrins in the subdomain IB [8]. These waters play an important role to orient the heme in relation to both proximal and distal tyrosine, Fig. 7 shows isosurfaces of probability to find these water molecules in such positions [8]. There are no other theoretical studies about the function of the water in the different subdomains of the apo conformation.

Conclusions

Clustering analysis of the X-ray structures available in the RCSB PDB yields only two well defined clusters (named C1 and C2). Structures denoted X1 and $X2$ were taken as representative of each cluster. They represent the apo and the HSA-myristate conformations with a deviation of 0.46 nm between them.

From the analysis of cluster C1, one concludes that the binding of anesthetics does not induce significant changes in the native structure as long as the RMSD is lower than 0.08 nm. Fatty acids in IIIA (two sites), IIIB, and the interface IA/IIA do lead to a different conformation (HSA-myristate) characteristic of cluster C2. The deviation between members of C2 is lower than 0.07 nm. Ligands like metal-porphyrins do not induce a conformation different from the HSA-myristate.

Electrostatic potential maps (not shown) display some slightly electronegative regions exposed to the solvent, but the more striking feature is an electropositive pocket in the breach between domains I and III, which is more exposed in X1 than in $X2$. The presence of (negatively charged) fatty acids may favor the closure of the electropositive pocket and contribute to the conformational change X1→$X2$.

Structurally, this change is characterized by a twist motion with an angle of rotation of 24° between domains I and II and a hinge motion with an angle of 15° between domains III and II. These motions might have their origin in a reorganization of H-bonds around Tyr-152, due to the presence of fatty acids in its surroundings, making residues 195–196 act as a molecular hinge.

Gln-196 is a key residue involved in the conformational change from X1 to $X2$, an in silicon mutation and subsequent MD could confirm our hypothesis. A future work could include this issue.

The members of each cluster were regarded as samples of the dynamic states of each conformation, for the purpose of dynamical characterization. PCA analysis reveals that the collective motions are different in both clusters. In both cases the first mode represents about 30 % of the total motion and the first four about 60 %. In the HSA-myristate the first mode represents a collective anticorrelated motion between the C-terminal and N-terminal regions, which does not occur in the apo conformation. We additionally performed molecular dynamics simulations of each typical conformation (X1 and $X2$) in the absence of ligands.

A molecular dynamics of 20 ns was considered in order to remain close to each initial conformation. Fluctuation and correlation analyses yielded outcomes in accord with those obtained from the PDB clusters. Finally, let us note that the present approach could be extended to other proteins in PDB.

References

1. Wong F (2007) Drug insight: the role of albumin in the management of chronic liver disease. Nat Clin Pract Gastroenterol Hepatol 4:43–51
2. Divsalar A, Saboury AA, Ahadi L, Zemanatiyar E, Mansouri-Torshizi H, Ajloo A, Sarma RH (2011) Biological evaluation and interaction of a newly designed anti-cancer Pd(II) complex and human serum albumin. J Biomol Struct Dyn 29:283–296
3. Guizado TRC, Louro SRW, Pascutti PG, Anteneodo C (2010) Solvation of anionic water-soluble porphyrins: a computational study. Int J Quantum Chem 110:2094–2100
4. Guizado TC, Pita SR, Louro SRW, Pascutti PG (2008) Int J Quantum Chem 108:2603–2607
5. Guizado TRC, Louro SRW, Anteneodo C (2011) Hydration of hydrophobic biological porphyrins. J Chem Phys 134:055103,1–9
6. Wang RM, Komatsu T, Nakagawa A, Tsuchida E (2005) Human serum albumin bearing covalently attached iron(II) porphyrins as O2-coordination sites. Bioconjug Chem 16:23–26
7. Tominaga TT, Yushmanov VE, Borissevitch IE, Imasato H, Tabak M (1997) Aggregation phenomena in the complexes of iron tetraphenylporphine sulfonate with bovine serum albumin. J Inorg Biochem 65:235–244
8. Guizado TRC, Louro SRW, Anteneodo C (2012) Dynamics of heme complexed with human serum albumin: a theoretical approach. Eur Biophys J 41:1033–1042
9. Sudlow G, Birkett DJ, Wade DN (1975) The characterization of two specific drug binding sites on human serum albumin. Mol Pharmacol 11:824–832
10. Curry S, Mandelkow H, Brick P, Franks N (1998) Crystal structure of human serum albumin complexed with fatty acid reveals an asymmetric distribution of binding sites. Nat Struct Biol 5:827–835
11. Petitpas I, Grune T, Bhattacharya AA, Curry S (2001) Crystal structures of human serum albumin complexed with monounsaturated and polyunsaturated fatty acids. J Mol Biol 314:955–960
12. He XM, Carter DC (1992) Atomic structure and chemistry of human serum albumin. Nature 358:209–215
13. Bhattacharya AA, Curry S, Franks NP (2000) Binding of the general anesthetics propofol and halothane to human serum albumin: high-resolution crystal structures. J Biol Chem 275:38731–38738
14. Zunszain PA, Ghuman J, Komatsu T, Tsuchida E, Curry S (2003) Crystal structural analysis of human serum albumin complexed with hemin and fatty acid. BMC Struct Biol 3:6 available at http://www.biomedcentral.com/1472-6807/3/6

15. Ghuman J, Zunszain PA, Petitpas I, Bhattacharya AA, Otagiri M, Curry S (2005) Structural basis of the drug-binding specificity of human serum albumin. J Mol Biol 353:38–52

16. Available at http://www.pdb.org/

17. Ascenzi P, Fasano M (2010) Allostery in a monomeric protein: the case of human serum albumin. Biophys Chem 148:16–22

18. Ascenzi P, di Masi A, De Sanctis G, Coleta M, Fasano M (2009) Ibuprofen modulates allosterically NO dissociation from ferrous nitrosylated human serum albumin by binding to three sites. Biochem Biophys Res Commun 387:83–86

19. Ascenzi P, di Masi A, Coletta M, Ciaccio C, Fanali G, Nicoletti FP, Smulevich G, Fasano M (2009) Ibuprofen impairs allosterically peroxynitrite isomerization by ferric HAS. J Biol Chem 284: 31006–31017

20. Artali R, Bombieri G, Calabi L, Del Pra A (2005) A molecular dynamics study of human serum albumin binding sites. Il Farmacol 60:485–495

21. Fujiwara S, Amisaki T (2006) Molecular dynamics study of conformational changes in human serum albumin by binding of fatty acids. Proteins 64:730–739

22. Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18:2714–2723

23. Szekely GJ, Rizzo ML (2005) Hierarchical clustering via join between-within distances: extending ward's minimum variance method. J Classif 22:151–183

24. van der Spoel D, Lindahl E, Hess B, van Buuren AR, Apol E, Meulenhoff PJ et al. (2010) Gromacs user manual version 4.5 www.gromacs.org

25. Oostenbrink C, Villa A, Mark AE, Van Gunsteren EF (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. J Comput Chem 25:1656–1676

26. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J (1981) In: Pullman B (ed) Intermolecular forces. Reidel, Dordrech

27. Essmann U, Pereira L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) A smooth particle mesh Ewald method. J Chem Phys 103: 8577–8593

28. Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity-rescaling. J Chem Phys 126:014101, FALTA

29. Berendsen HJC, Postma JPM, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. J Chem Phys 81: 3684–3690

30. Scheer A, Cotecchia S (1997) J Recept Signal Transduct Res 17:57–73

31. Lee AL, Wand AJ (2001) Microscopic origins of entropy, heat capacity and the glass transition in proteins. Nature 411:501–504

32. Pasi M, Tiberti M, Arrigoni A, Papaleo E (2012) xPyder: a PyMOL plugin to analyze coupled residues and their networks in protein structures. J Chem Inf Model 52(7):1865–1874

33. DeLano WL (2002) The PyMOL molecular graphics system. http://www.pymol.org

34. de Groot BL, van Aalten DMF, Amadei A, Berendsen HJC (1996) The consistency of large concerted motions in proteins in molecular dynamics simulations. Biophys J 71:1707–1713

35. Paris G, Ramseyer C, Enescu M (2014) A principal component analysis of the dynamics of subdomains and binding sites in human serum albumin. Biopolymers 101:561–572

36. Available at http://fizz.cmp.uea.ac.uk/dyndom/

37. Hayward S, Berendsen HJC (1998) Systematic analysis of domain motions in proteins from conformational change: new results on citrate synthase and T4 Lysozyme. Proteins 30:144–154

38. Hayward S, Kitao A, Berendsen HJC (1997) Model-Free methods of analyzin domains motions in proteins from simulation: a comparison of normal model analysis and molecular dynamics simulation of Lyysozome. Proteins 22:425–437

39. Wardell M, Wang Z, Ho JX, Justin R, Ruker F, Ruble J, Carter C (2002) The atomic structure of human serum albumin at 1.9 C. Biochem Biophys Res Commun 291:813–819

40. Lehninger AL, Nelson DL, Cox MM (1982) Principles of biochemistry. Worth, New York, p 145

41. Bocedi A, Notari S, Menegatti E, Fanali G, Fasano M, Ascenzi P (2002) Allosteric modulation of anti-HIV drug and ferric heme binding to human serum albumin. FEBS J 272:6287–6296

42. Fanali G, Bocedi A, Ascenzi P, Fasano M (2007) Modulation of heme and myristate binding to human serum albumin by anti-HIV drugs. An optical and NMR spectroscopic study. FEBS J 274:4491–4502

43. Ascenzi P, Bocedi A, Notari S, Fanali G, Fesce R, Fasano M (2006) Allosteric modulation of drug binding to human serum albumin. Mini Rev Med Chem 6:483–489

44. Fanali G, Fesce R, Agrati C, Ascenzi P, Fasano M (2005) Allosteric modulation of myristate and Mn(III)heme binding to human serum albumin. Opt NMR Spectrosc Charact FEBS J 272:4672–4683