Journal of Molecular Graphics and Modelling 54 (2014) 164–173

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Probing the binding site characteristics of HSA: A combined molecular dynamics and cheminformatics investigation

Prapasiri Pongprayoon, M. Paul Gleeson*

Department of Chemistry, Faculty of Science, Kasetsart University, 50 Ngam Wong Wan Rd, Chatuchak, Bangkok 10900, Thailand

ARTICLE INFO

Article history:
Accepted 11 October 2014
Available online 23 October 2014

Keywords:
Human serum albumin ADMET
Molecular dynamics (MD)
Principal components analysis (PCA)

ABSTRACT

Human serum albumin is a remarkable protein found in high concentrations in the body. It contains at least seven distinct fatty acid binding sites and two principle sites for drugs. Its primary function is to act as a fatty acid transport system, but it also shows the capacity to bind a diverse range of acidic, neutral and zwitterionic drug molecules. In this paper we investigate the ligand binding selectivity of HSA using cheminformatics analyses and molecular dynamics simulations. We compare and contrast the known ligand binding specificities as obtained from X-ray structural data using PCA, with additional direct analyses of the seven key binding pockets using analyses derived from molecular simulations. We assess both the fatted and defatted states of HSA using 100 ns simulations of the APO and HOLO forms, as well as structures containing one, three and seven myristic acid molecules. We find that differences in fatty acid binding can have a dramatic effect on the flexibility of the protein and also the pocket characteristics. We discuss how the remarkable selectivity of the HSA pockets towards both endogenous fatty acids and exogenous drug molecules is not simply controlled by bulk property effects such as ionization state and lipophilicity.

1. Introduction

The binding of molecules to plasma proteins is important from a pharmaceutical perspective since this property can be used to estimate the total body clearance in man [1]. There are two main proteins implicated in the binding of drug molecules in plasma, namely human serum albumin (HSA) and α1-acid glycoprotein (AGP). The former is the most abundant protein found in plasma, making up approximately 60% of the total protein present. HSA is an important transport protein that binds numerous molecules of endogenous and exogenous molecules. Fatty acid (FA) binding to HSA has been long studied due to the critical role these molecules play in energy metabolism and membrane synthesis [2]. HSA is well known for its capacity to bind drug molecules and better understanding of the structure and function of the protein is important for future drug development [3,4]. As a result, the binding characteristics and specificity of HSA have received much attention both experimentally and computationally. To date approximately 100 HSA structures have been published in the Protein Data Bank (www.rcsb.org). In addition, NMR [5,6], EPR [7] and biochemical methods [8–10] have been widely used to explore aspects of ligand binding to the protein.

HSA has a molecular mass of 66 kDa (585 amino acids). It exists as a monomer with three similar α-helical domains (I–III) (Fig. 1) [3]. Each domain can be divided further into subdomains A (four α-helices) and B (six α-helices). Subdomains IIA and IIIA serve as a drug binding sites [11–13]. The two primary sites for drugs are known as Sudlow (or drug sites) I and II [4,5,14]. Recently, circular dichroism (CD) studies have revealed subdomain IB as a third drug binding site for steroidal compounds [15,16]. Additionally, HSA contains numerous sites for long chain fatty acids distributed across the protein. Kragh-Hansen et al. noted that Lauric acid (12 carbons) binds to 8 unique sites and decanoic acid (10 carbons) can bind up to ten [17]. Simard et al. observed that seven FA sites exist for FAs with chain lengths >14 carbons [4,5]. Further studies revealed that three of the sites displayed higher affinities for FA (sites FA2, FA4 and FA5, where FA5 is the highest) [4,5]. Nevertheless, even the low affinity site FA1 can bind very large acidic molecules such as HEME with high affinity [18]. The presence of FA bound to HSA is also known to induce conformational change in the protein and this affects drug binding [13]. Recent spectroscopic studies have highlighted the flexibility of HSA to accommodate both anionic and neutral ligands in subdomain IIA [19]. Thus, understanding the origin of the FA-binding

* Corresponding author. Tel.: +66 2 562 5555x2210; fax: +66 2 5793955.
E-mail address: paul.gleeson@ku.ac.th (M.P. Gleeson).

http://dx.doi.org/10.1016/j.jmgm.2014.10.007
1093-3263/

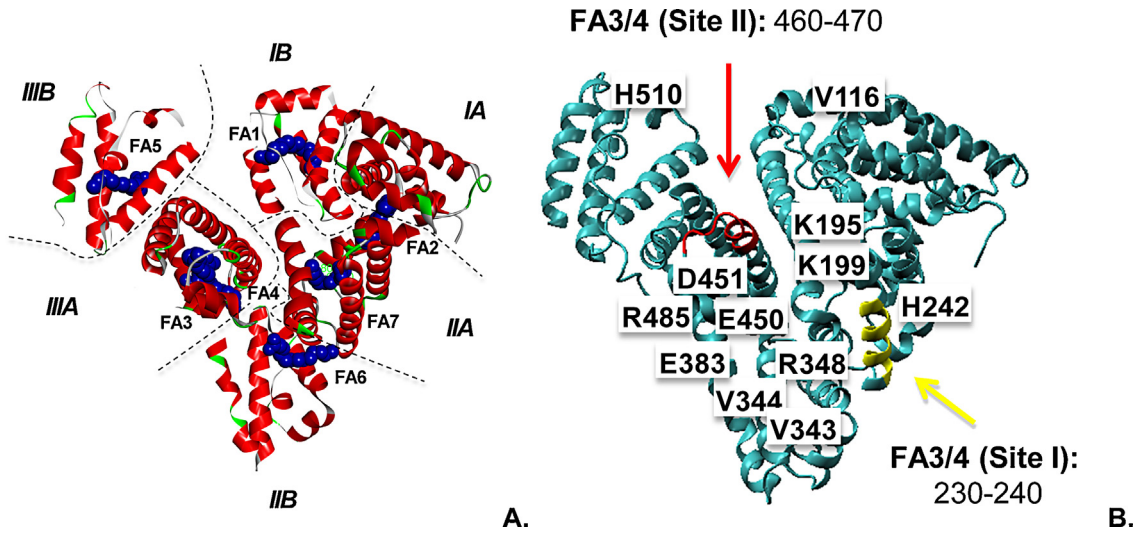

Fig. 1. Cartoon views of the key regions and binding sites (A) and important residues in HSA (B).

differences among these seven sites remains a source of major interest [5,13,18,20–22].

QSAR approaches have been extensively used to rationalize both HSA and whole plasma protein binding [9,23,24]. However, the specific interactions and pocket shapes give rise to ligand selectivity [13] which bulk physico-chemical property QSAR models cannot easily account for. However, incorporating specific details of the binding interactions at a molecular level may help improve this situation in the future [25–27]. Furthermore, there are an increasing number of computational studies on HSA in which molecular docking or molecular dynamics (MD) simulations have been used to gain insight into HSA-ligand binding. For example, Fujikara et al. employed MD simulations and principle component analysis (PCA) to evaluate the possible conformational changes of HSA in the presence of bound myristic acids [28,29]. MD simulations have also been used to postulate drug-binding sites or understand the drug-binding mechanisms at specific HSA pockets (i.e. warfarin [30], hydroxyquinoline derivatives [31], betulinic acid [32]). More recently Castellanos et al. have investigated the role of disulphide bonds on the structural flexibility of the protein [33]. For an extensive review of the most recent molecular modelling studies conducted on HSA please see the following reference [34]. These studies have focused their attention on the macroscopic conformational changes of HSA in the presence or absence of natural or synthetic ligands. However, the characteristics of the individual binding pocket remain largely unstudied computationally.

In this study we aim to shed further light on the structural characteristics and binding pocket selectivities of HSA. We have extracted and analyzed known drug molecules bound to each of the seven principle binding pockets of HSA. PCA was then used to understand what physical properties are most important in defining the selectivity for each site. We then contrast this result with those derived from 100 ns MD simulations of HSA to assess how the seven different pockets evolve over time and under different conditions. Simulations have been performed on the closed form of HSA (APO model), the open form with seven bound myristic acid (MYR) molecules (FA(1–7) model). We have also assessed intermediate states by partially or fully removing the FA molecules from the open protein structure; (a) the three highest affinity sites (FA(2,4,5) model), (b) the single highest affinity site (FA(5) model) (c) zero FA molecules bound (HOLO model).

2. Computational procedures

2.1. MD simulations

HSA models for the closed and open form or HSA were created from the coordinates 1e78 and 1e7 g, respectively [35], which were sourced from the RCSB databank [36]. The closed, non-ligand bound structure was used to generate the APO HSA model. The open structure, containing bound MYR molecules, was used to create a HSA model containing ligands at the seven principle FA sites (FA(1–7) model). The latter X-ray structure was also used to create an additional model of HSA in the open form but containing no fatty acids (HOLO model). Two other HSA models were created which contained either one or three fatty acids at the most high affinity sites. These are termed the FA(2) and FA(2,4,5) models.

The N- and C-terminal ends of each model were capped with acetyl (ACE) and methyl amino (NME) groups, respectively, and the overall quality confirmed using PROCHECK v3.5.4 [37]. The protonation state of ionizable residues were determined using PROPKa [38] and a visual analysis of the environment surrounding each ionizable residue. Residues were treated as HID unless otherwise stated: HIE (39, 288, 340) and HIP (3, 67, 105, 128, 146, 247, 338, 367, 510). Each model was placed in a cubic simulation box containing water and neutralized with counter ions by randomly replacing water molecules.

Simulations were carried out using GROMACS v4 [39] with the AMBER99SB forcefield [40] Ligand topology files were generated using ACPYPE script [41] and the GAFF force field [42]. Electrostatic charges for FA molecules were generated using a HF/6-31G(d) optimized structures (Gaussian 09 Revision D01) [43] and the RESP protocol of Ambertools1.5 [44]. Long range electrostatic interactions were calculated using the Particle Mesh Ewald (PME) method with a 0.12 nm cut-off [45] while van der Waals interactions utilized a 1 nm cut-off. Simulations were performed at constant temperature, pressure and number of particles (NPT). The temperature of the protein and solvent were each coupled separately. The Berendsen thermostat [46] was applied at 300 K with a coupling constant of $\tau = 0.1$ ps [47]. Coordinates and velocities were saved every 2 ps. The LINCS algorithm was used to restrain bond lengths and a 2 fs timestep was used for integration [48]. Five ns of protein backbone-restrained dynamics were employed initially, followed by 100 ns production runs. Simulations for the APO, HOLO and FA(1–7) models were performed in duplicate to assess the sensitivity of the results to different starting conditions. Analyses on the resulting

MD coordinates were performed using GROMACS routines, Discovery Studio [49], and locally written scripts. Molecular graphics images were produced using VMD [50]. The statistical significance of any reported differences have been confirmed using paired Student's *t*-test or *F*-test, respectively at >95% confidence level.

2.2. Cheminformatics analyses

The physico-chemical determinants of selectivity for the seven HSA binding pockets were assessed by curating a list of ligands known to bind in each pocket. This was done based on an analysis of structures deposited in the Protein databank structures. A range of commonly used physico-chemical descriptors were calculated for the list of fifty seven inhibitors identified as binding to at least one of the known binding sites. All descriptors were calculated using the Chemaxon JChem suite [51] and analyzed in Vortex [52] (Supplementary Information, Table S1).

PCA models were generated using data reported in Tables 1–3 to (a) investigate the structural differences between the difference proteins simulations and (b) investigate the differences in pockets of HSA and the effect of different simulation conditions. Models were generated in SIMCA-P10 [53] using the default settings (i.e. mean centred and scaled descriptors, with the auto-fitting of components) and report in the form of either the traditional loading and scores plot or a combined loadings bi-plot (i.e. both the scores and loadings projected onto the same plot).

3. Results and discussion

We begin this section with a discussion of the global structural features of the HSA. In the subsequent section we discuss the use of FA molecules to assess the dynamic characteristics of the seven pockets of HSA over the course of the MD simulations. This of course assumes that FA molecules are suitable probes for the range of substrates that FAs bind. As such, in the final section we employ a complementary, ligand-based approach to further probe the binding site specificity of the protein. We compare and contrast the calculated physico-chemical descriptors of the known substrates of each pocket, to the binding site parameters obtained from the MD simulations. In this way we can gain insight into the unique characteristics of the pockets and assess their dynamic character to a level of detail yet reported.

3.1. HSA structural characteristics

The overall size, shape, and dynamics of the HSA protein are thought to influence the behaviour of its binding pockets, which in turn will affect the binding affinity of fatty acids and other drug molecules [54]. Based on a series of simulations performed in this study, the global dynamic properties of HSA under each condition were investigated and compared in order to elucidate the key dynamic features that are essential for structure and function. With this study, we investigate the microscopic details of the binding pockets that contribute to the ligand binding using MD simulations.

The constant total energy of each simulation indicates the convergence of each run after 5-ns pre-equilibration (see Supplementary Information, Fig. S1). Both C_α RMSDs and RMSFs of the proteins show the flexibility of HSA (Figs. 2 and 3). Large RMSDs and RMSFs are observed at sub-domain I and II in HOLO form (Fig. 3). These findings are indicative of significant protein flexibility in both these sub-domains which agrees well with recent experimental work [19]. The RMSDs of the duplicate simulations performed on the APO and HOLO structures show slight differences in the backbone flexibility, this being related to when the inter-domain movement begins. Unlike the APO and HOLO models, the presence of seven bound FA molecules is found to reduce the magnitudes of both the RMSDs and RMSFs significantly, which indicates a more rigid protein. We observe that the key structural features are preserved well, although certain individual residues or loop movements can differ between duplicates as noted elsewhere [29].

Simulations with either one or three FA molecules at the highest binding sites (FA(5) and FA(2,4,5) models) are more akin to the physiologically FA binding conditions experienced in the body [55,56]. Experimental X-ray crystallographic studies have shown that HSA in this state is found to adopt a structure intermediate between the open and closed forms of the protein [13,35]. Recent spectroscopic studies showed that such conformational adjustment in subdomain IIA was required for ligand binding [19]. In the presence of FA, our simulations show a reduction in the RMSD and an increase in the number of hydrogen bonds (Table 1) indicating an increase in protein rigidity as a result of conformational changes. This finding also implies that significant conformational adjustment follows ligand binding. The higher rigidity observed over the closed protein conformation is due between ~3 and 6 additional hydrogen bonds in the FA bound open conformation. Taken overall, the similarities between the number of hydrogen bonds among amino acids, the radius of gyration (indicating the compactness of protein structure), and the residue surface area, suggest the partially and fully FA-bound HSA show similar global properties (Table 1). Furthermore, it appears that the HOLO structure does not transition to the closed conformation, as might be expected. The 100 ns timescale appears to be insufficiently long to observe the complete conformational change from FA-bound to the native HSA structure.

A number of interactions have been identified as key indicators of the conformational state of the protein in the fatted and de-fatted states [4,5,13]. The H510-V116 distance is a notable parameter, being indicative of the distance between interface of domain IIIB and IB (H510 on domain IIIB and V116 on domain IB, Fig. 1B). A change in the distance between H510-V116 can indicate the movement of both domains, which is an indicator for the open and closed states. When FA binds, the H510-V116 distance increases from 2.12 nm (APO) to 2.63 nm for FA(1–7), 2.4 nm for FA(2,4,5), and 2.28 nm for FA(5) (Table 2), respectively. Binding of FA molecules clearly induces the movement of domain IIIB and IB, which agree well with previous studies [57,58] and shows that the partially occupied structures are in a more intermediate state as expected experimentally. The direction of movement can be captured by PCA analysis (Supplementary Information, Fig. S2). The dominant motion obtained (i.e. component one) demonstrates a scissor-like motion of domain IIB and IB in both the absence and presence of bound FAs. This motion can induce conformational adjustment in subdomain IIB as found by previous spectroscopic studies [19] and is in a good agreement with previous theoretical studies [28,30].

Compared to domain IB and IIIB, the other domains seem to be rather rigid. As noted by Ghuman et al. [13] FA binding at site FA3 in domain IIIA leads to the disruption of a salt-bridge between R348-E450, resulting in a stronger H-bond interaction between E450 with the backbone of V343/V344 (Fig. 1B). The relocation of E450 sidechain leads to the displacement of D451, resulting in an interaction with K195. In our simulations we observed an increase in R348-E450 distance in the HOLO and FA(1–7) (~1.2 nm) models, while the two FA3-free simulations (FA(5) and FA(2,4,5)) displayed shorter distances (~0.9 nm) (Table 2). The disruption of R348-E450 when the FA3 site is occupied leads to E450 to interact with V343, resulting in lower distances on average (~0.6 nm), and an improved interaction between K195-D451 (reducing from ~0.8 nm in FA3-free models to ~0.69 nm in FA(1–7) (Table 2).

Table 1
Comparison of key parameters obtained for each of the 5 simulations. All values reported are average values obtained over the course of the 100 nS simulations. Standard deviation given in parenthesis.

Model	RMSD	RMSF	H-bonds	Ryd. Gyr.	Area
APO	0.38 (0.07)	0.18 (0.12)	456 (11.4)	2.72 (0.03)	0.59 (0.39)
HOLO	0.31 (0.07)	0.17 (0.07)	459 (11.4)	2.76 (0.03)	0.60 (0.38)
FA (1–7)	0.32 (0.08)	0.15 (0.06)	461 (11.3)	2.77 (0.02)	0.61 (0.37)
FA (2,4,5)[a]	0.24 (0.04)	0.16 (0.06)	461 (11.9)	2.77 (0.02)	0.60 (0.38)
FA (5)[a]	0.27 (0.04)	0.19 (0.08)	462 (10.4)	2.78 (0.02)	0.60 (0.38)

[a] Results derived from a single simulations.

Table 2
Centre of mass (COM) distances between key residues within HSA. Distances are reported in nm.

Descriptor	APO	HOLO	FA(1–7)	FA (2,4,5)[a]	FA (5)[a]
H510-V116	2.12 (0.67)	2.09 (0.39)	2.63 (0.27)	2.40 (0.29)	2.28 (0.24)
E383-R485	0.74 (0.06)	0.85 (0.15)	0.78 (0.08)	0.89 (0.11)	0.83 (0.09)
K195-D451	0.81 (0.08)	0.69 (0.05)	0.69 (0.05)	0.82 (0.14)	0.84 (0.12)
R348-E450	0.91 (0.17)	1.24 (0.12)	1.18 (0.04)	0.89 (0.05)	0.87 (0.03)
V343-E450	0.89 (0.07)	0.76 (01.2)	0.66 (0.04)	0.79 (0.04)	0.83 (0.06)
V344-E450	0.62 (0.06)	0.69 (0.10)	0.67 (0.05)	0.58 (0.02)	0.59 (0.04)
Y138-Y161	0.37(0.02)	0.51(0.03)	0.45(0.10)	0.47(0.07)	0.90(0.06)

[a] Results derived from a single simulations.

Table 3
The mean number of H_2O contacts, H-bonds, Total contacts observed for FA bound in the FA(1–7) simulations. Also reported are the mean pocket sizes for each simulation.

Simulation	Descriptor	FA Site. 1	FA Site. 2	FA Site. 3	FA Site. 4	FA Site. 5	FA Site. 6	FA Site. 7
FA(1–7)	H2O-contacts	5.58 (0.75)	4.10 (0.93)	2.58 (0.75)	6.14 (0.99)	5.84 (0.66)	7.75 (0.97)	8.43 (0.83)
FA(1–7)	H-bonds	3.45 (0.38)	3.35 (0.29)	3.24 (0.11)	2.29 (0.53)	1.75 (0.20)	1.83 (0.31)	1.83 (0.32)
FA(1–7)	Total-contacts	182.1 (14.2)	203.3 (16.5)	212.9 (14.2)	185.3 (17.1)	214.4 (14.6)	184.4 (14.7)	169.7 (15.2)
FA(1–7)	FA length	0.88 (0.18)	1.4 (0.09)	1.24 (0.05)	1.3 (0.07)	1.34 (0.03)	1.16 (0.10)	0.78 (0.12)
FA(1–7)	Volume	5.74 (0.75)	5.89 (0.74)	5.27 (0.53)	5.28 (0.56)	4.76 (0.61)	4.44 (0.45)	5.01 (0.58)
HOLO	Volume	5.35 (0.71)	5.94 (0.77)	5.42 (0.53)	5.29 (0.52)	4.73 (0.58)	4.53 (0.46)	4.91 (0.60)
APO	Volume	5.58 (0.70)	5.99 (0.76)	5.09 (0.52)	5.04 (0.48)	4.78 (0.58)	4.36 (0.45)	4.93 (0.60)
FA(5)	Volume	5.02 (0.19)	5.97 (0.23)	5.14 (0.21)	5.10 (0.17)	4.88 (0.16)	4.60 (0.13)	4.96 (0.19)
FA(2,4,5)	Volume	4.76 (0.62)	5.52 (0.69)	5.00 (0.47)	5.32 (0.51)	4.88 (0.58)	4.59 (0.44)	5.01 (0.50)

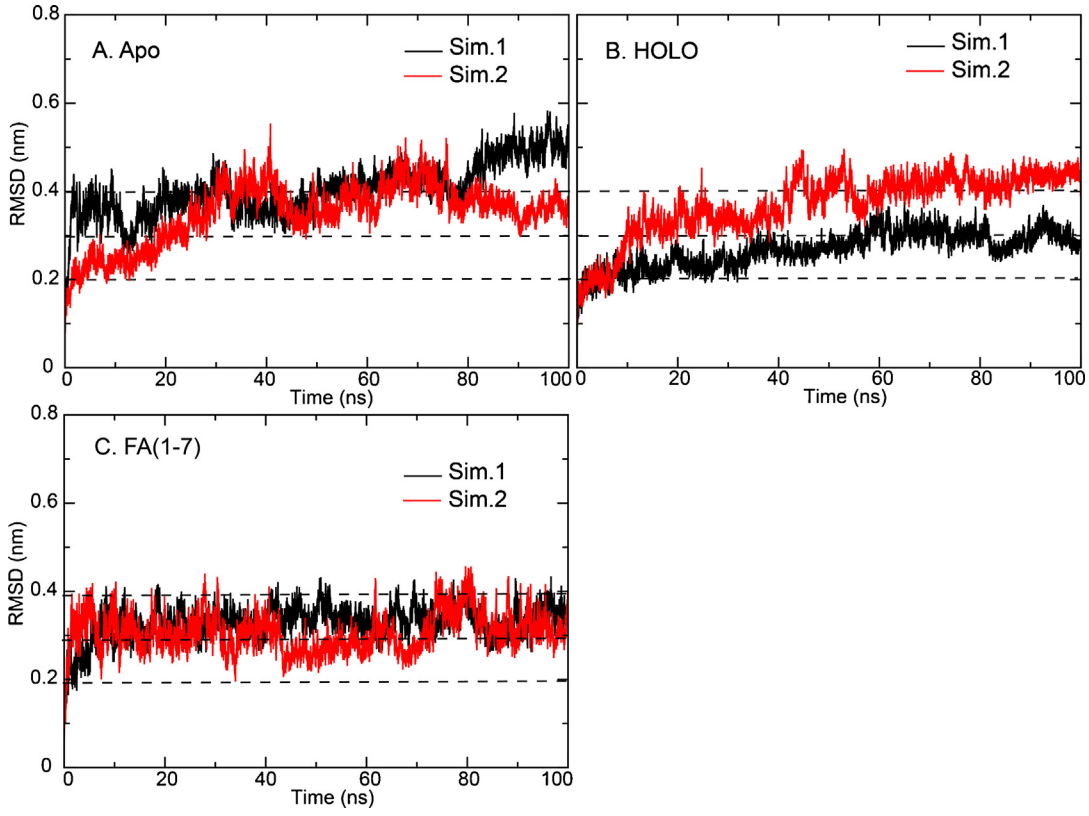

Fig. 2. (A–C) Clockwise from top left, Cα RMSDs observed in the APO (open conformation with no FA), HOLO (closed conformation with FAs removed) and FA(1–7) (closed conformation containing seven FAs) models.

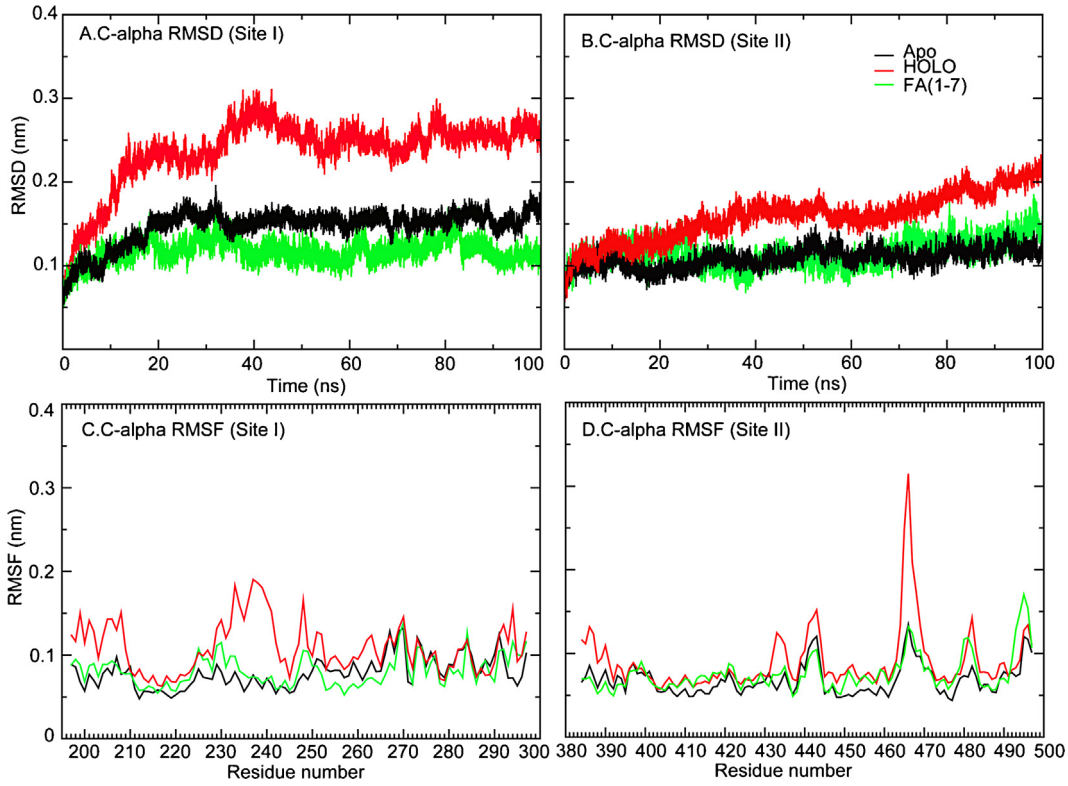

Fig. 3. The average RMSD and RMSF values associated with Sudlow sites 1 and 2 sites for the APO, HOLO and FA(1–7) simulations of HSA.

3.2. HSA binding site characteristics from MD

The binding of ligands to Sudlow sites I (FA 3&4) and II (FA7) have been extensively reported in the literature from a structural, biochemical and computational perspective [13,14,20,57]. Other binding sites play a role in ligand binding, although often secondary, and are therefore less well studied. For some sites the intricacies of the interactions have not been elaborated on, or for others the primary amino acids that bind substrates were not been conclusively confirmed due to the dynamic nature of the protein [4]. In this section we aim to describe the different binding pockets of HSA in terms of the size, shape and interactions, and how they evolve over the course of the simulations. We also discuss the effect of FA binding on the characteristics of these sites and the implications this will have on more rigorous calculations to estimate binding affinities.

The C_α RMSDs of residues surrounding both binding pockets for the APO, HOLO and FA(1–7) have been analyzed in Fig. 3. RMSDs indicate that the Sudlow site II is much less sensitive to the conformation state of the protein. The RMSD of surrounding site I are broadly comparable for the APO and FA(1–7) structures. However, removing the FA from the latter (i.e. the HOLO structure) leads to a dramatically larger RMSD and RMSF. In contrast, site II shows a smaller effect, with a smaller deviation between the different HSA configurations. This suggests site II is less sensitive to dynamic effects (i.e. the open, closed or intermediate conformation).

All of the seven known fatty acids sites can bind drug-like molecules, albeit most with much lower frequency that the Sudlow sites. It is of interest to us to explore the differences between these sites in greater detail and how they are affected by random dynamic fluctuation. To better characterize each pocket, the distances between each FA and its neighbouring key residues, hydrogen bonds, water and total contacts were extracted and reported in Table 3 as well as graphically in Figs. 4 and 5 and Supplementary Information, Figs. S3 and S4.

The FA1 pocket located in subdomain IB is a large (relative to the other seven pockets), curved cavity (5.89 nm^2) with moderate solvent accessibility with 5.58 contacts on average. The FA molecules are on average found to maintain approximately three non-covalent bonds. Distances of ~0.4 nm between FA1 and R117 (before 70 ns), R186, and Y161 indicate FA1 maintains two salt bridges and one hydrogen bond consistently (Fig. 4 and Table 3). The fluctuating distance observed after 70 ns of R117 sidechain indicates the high flexibility which enables interactions with residues in subdomain IB as mentioned above. FA2 is also a large but linear pocket, with a volume of ~5.9 nm^2 and is located between subdomains IA and IIA (Table 3). It is the second most enclosed of the HSA sites and maintains a single salt bridge (R257) and two H-bond interations with Y150 and S287 over the course of the simulations (Fig. 4).

FA3 is a medium sized pocket with a volume of ~5 nm^2 located in subdomain IIIA. The lowest water contacts of 2.58 is as expected observed for the most buried site for FA3. However, FA3 can make two salt bridges with R348 and R485. FA4 sits in subdomain IIIA and has a pocket volume of ~5 nm^2. This site makes up Sudlow site II in combination with FA3. FA4 site is rather solvent accessible with the water contacts of 6.14 (Table 3). Based on closed distances and computed H-bonds, FA4 can form approximately two H-bonds with N391 and Y411 and a salt bridge to R410 even though such interactions are transient. FA5 is a small linear pocket (volume of ~4.7 nm^2) located in subdomain IIIB. It is however moderately accessible with the 5.84 water contacts. A H-bond with Y401 and a salt bridge with K525 are observed.

FA6 is a small pocket with the volume of ~4.4 nm^2 in HSA, located in domain IIB. However, it is also the second most water exposure (water contacts is 7.75). From earlier X-ray and NMR results it was not clear which residues helped to anchor the bound carboxyl group of the bound FA [4]. Here, FA6 is found to make a salt bridge with K351 consistently and a H-bond to S480 to a lesser extent. FA7, located in subdomain IIA, is a medium pocket with a

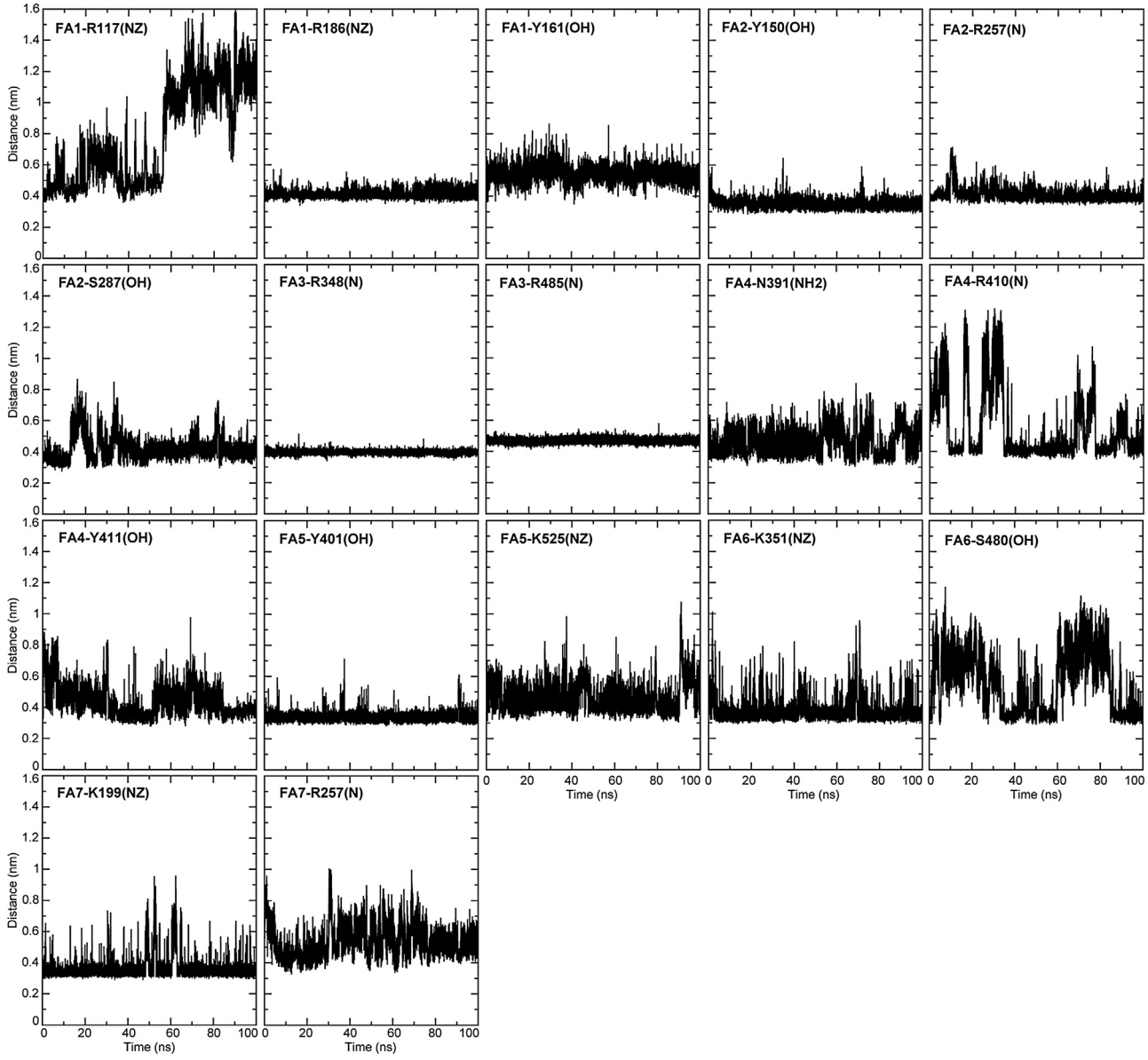

Fig. 4. Distances between the carboxylic carbon of each FA in FA(1–7) and the centre atom in functional group on a sidechain of key neighbours in an individual pocket (NZ is an amine nitrogen, N is a quanidinium nitrogen, and OH is a hydroxyl oxygen).

volume of ~5 nm^2 and is also known as the Sudlow site II. FA7 is the most solvent accessible FA pocket with two salt bridges with K199 and R257. Despite the small pocket size and low number of protein-FA interactions, the high degree of water exposure permit FA molecules to easily diffuse out of the FA6 and FA7 pockets which would help to explain their lower binding affinity.

The results here suggest that the more enclosed sites are important for high affinity are broadly in line with previous NMR [4]. The findings also support more recent NMR displacement studies which showed that FA bound to the low affinity FA6 and FA7 sites were easily displaceable [6]. Our results suggest that the more facile substrate displacement observed from sites FA6 and FA7 is a result of their greater solvent exposure (water contacts = 7.75 at FA6 and 8.43 at FA7) and fewer protein-FA hydrogen bonds and supports the findings from recent spectroscopic experiments [19].

While all sites are found to have at least two electrostatic bonds (at least one has to be salt bridge and another can be either a H-bond or salt bridge), the number of electrostatic interactions or H-bonds does not correlate with the known site affinities for FA. A more detailed calculations of the electrostatic interaction strength with the programme APBS [59] appears to confirm this (Supplementary Information, Table S2). We find the strength of the electrostatic interactions follow the order FA7 > FA6 = FA3 > FA5 > FA2 = FA4 > FA1. FA sites 2,4, and 5 are known to be the most high affinity FA sites, yet site 1, 2, and 3 have the greatest number of H-bonds on average, sites 4, 6 and 7 are the most solvent accessible, and sites 1, 2 are the largest in size. Furthermore, sites FA1, FA2 and FA3 show the greatest degree of volume fluctuation between the different protein configurations (APO, HOLO, FA(1–7)). However, we can see that FAs bound to FA1, FA6, and FA7 bind in a more distorted, non-linear fashion (Table 3 & Fig. 5B), indicating greater internal strain. In addition, these ligands show greater fluctuation, as given by the standard deviation, suggesting there are less effectively bound. FA3 lies in an intermediate state, presumably adversely affected by the bonding of FA to the adjacent site 4. It would therefore appear that the selectivity displayed by HSA pockets towards

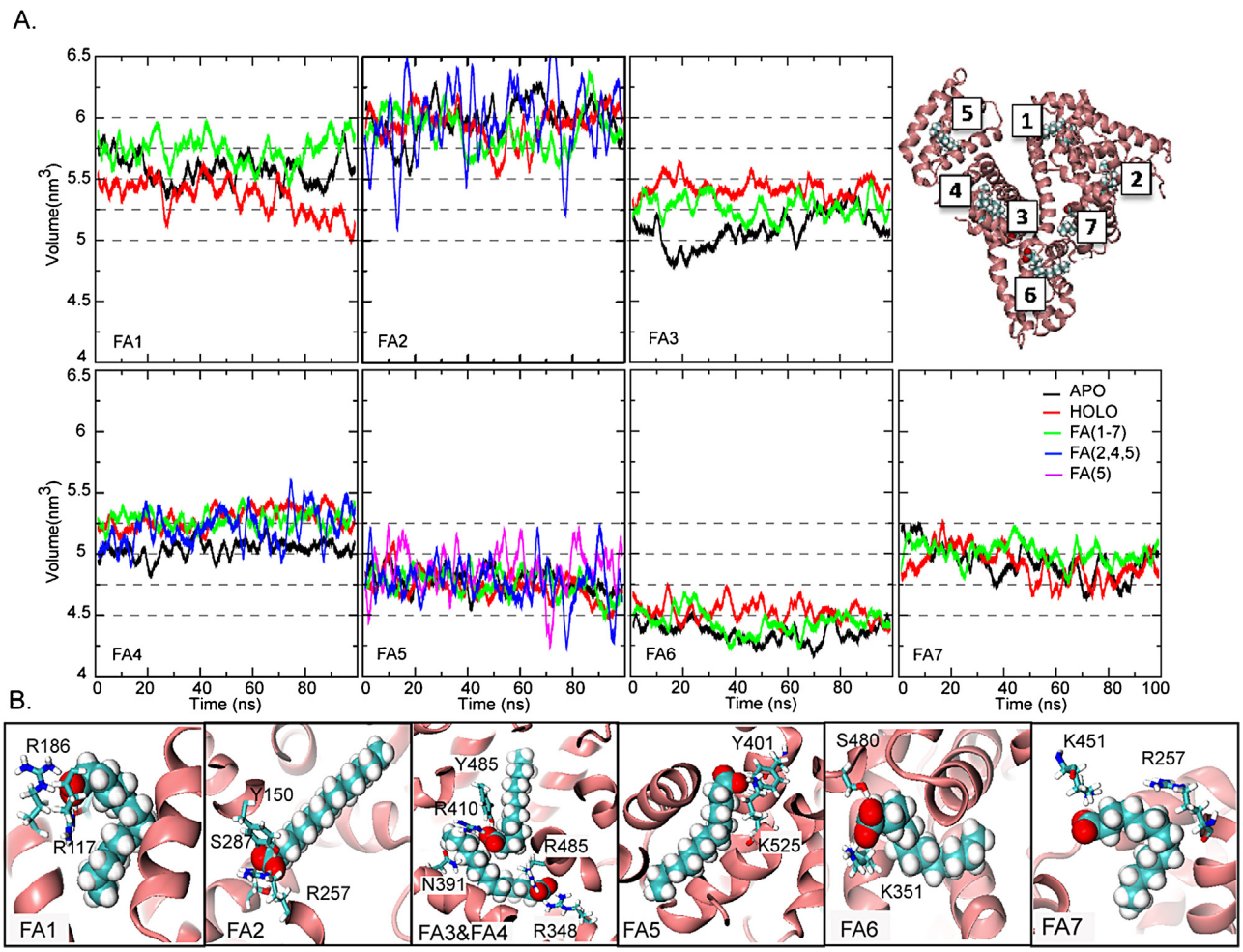

Fig. 5. (A) Computed volume of fatty acid binding sites 1–7 for APO, HOLO, FA(1–7), FA(2,4,5), and FA(5) models. (B) Final snapshot of FA alignment in each pocket with key residues for FA(1–7).

much more diverse exogenous drug-like molecules will be much more complex, and not as easily described by bulk molecular features.

Another consideration is the effect of concomitant ligand binding (Table 3 & Supplementary Information, Fig. S5). It is found that the FA2 pocket derived from the FA (1–7) model shows much greater fluctuation than in the partially occupied FA(2,4,5) model. In contrast, for FA5, ligand binding to this pocket increases the distances between two the ring centres of the π-staked residues Y138 and Y161 significantly (0.37 nm in Apo to ~0.9 nm in FA(5) (Table 2), increasing the volume of the pocket, and presumably also having an impact of the structural flexibility of the protein due to its location at the hinge point for the inter-domain movement of III and I.

3.3. Cheminformatics analysis of HSA ligand binding sites

The focus of this study has thus far been on gaining insight into the unique characteristics of the seven different HSA binding sites using atomic information derived from MD simulation data. A limitation of the univariate discussion above is that it neglects the inter-correlation between the multiple different parameters that were extracted. To rectify this limitation we have assessed (a) the overall structural features of the protein and (b) the individual pocket characteristics using PCA. We also used an orthogonal, ligand based approach to assess the features of the different sites.

3.3.1. PCA analysis of the MD structural information

A PCA model was generated using the eight distinct simulations performed and the eighteen common simulation derived descriptors (Tables 1–3). The model fitted three components which describe 87% of the total variation in the dataset. The information content of each component corresponds to 42%, 36% and 13% respectively (Supplementary Information, Tables S3–S5). In addition to the correlation between different models, this analysis also offers us a way to assess the impact the different starting configurations have on an individual simulation. The results for the two key components are presented in as a loadings bi-plot Fig. 6A.

The clustering of the simulations in different areas of the plots is indicative of their relative similarity or dissimilarity. On the first component we see that the APO and FA(1–7) simulations are separated. Both replicates are closely clustered together indicating the two simulations differ only to a limited degree compared to the difference between the different simulations. The key descriptors that give rise to the separation are found at the extreme of the *X*-axis. For example, the radius of gyration, area, total hydrogen bonds etc. are all greater in the FA(1–7) model, which are indicative of the open conformation. In contrast, the HOLO, FA(5) and FA(2,4,5) all occupy a position intermediate between the fatted and de-fatted states as is expected. Component two, as described by the *Y*-axis, describes the physical differences between the FA(1–7) and APO from the FA(5), FA(2,4,5) and HOLO structures. It should be noted that the HOLO structure shows the greatest deviation in terms of the distance between replicates. This difference reflects the

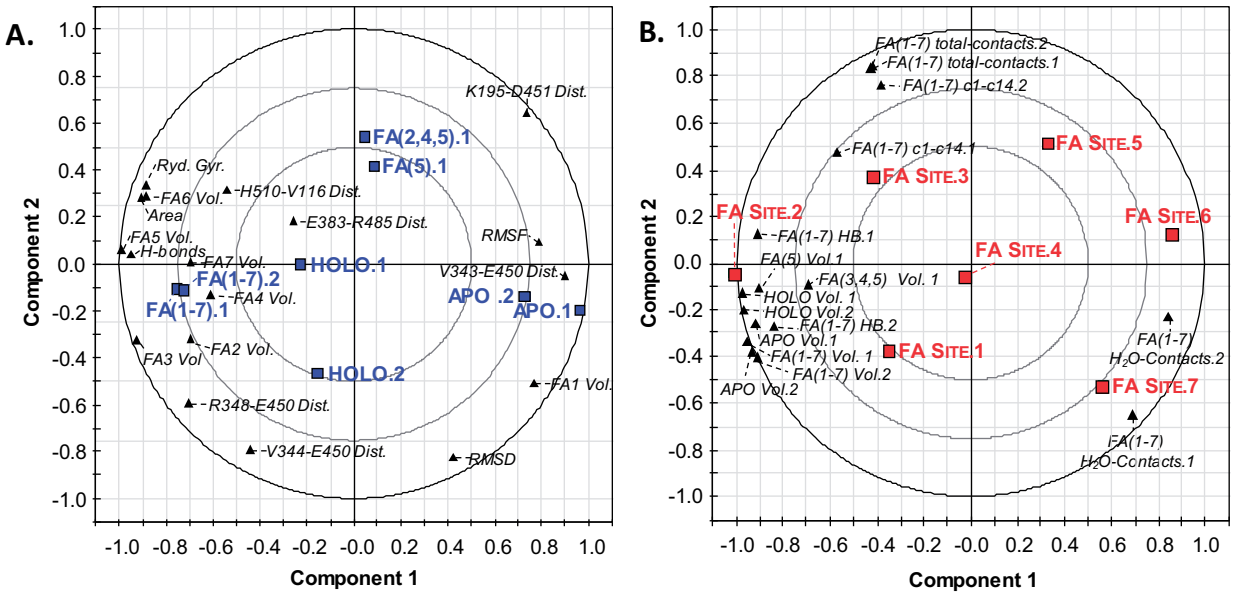

Fig. 6. Loadings bi-plot for the PCA models describing the differences between the different protein configurations (right) and the differences between the binding pockets (left).

observations in Fig. 2, where there is a dramatic change in RMSD due to domain movements between IIB and IB caused by the change of a salt bridge between R117 with either E119 or E520 located nearby. Duplicate simulations were not undertaken for FA(5) and FA(2,4,5), however, but we appear to observe the opposite behaviour with significantly lower RMSDs than either of the FA(1–7) simulations. It is also apparent from Fig. 6A that the change in fatty acid binding has a dramatic change on the flexibility of the structure, key distances and the pocket volumes. This has dramatic implications for simulations of HSA since it is known that the effect of fatty acid can change the affinity of exogenous ligands for HSA [55,56].

3.3.2. PCA analysis of the MD based pocket data

We next assessed the similarities and differences between the different pockets using an analysis of the FA site data extracted from the MD simulations. The FA site PCA model was generated using information from the seven distinct simulations performed for which twelve descriptors were extracted. The model fitted three components which describe 94% of the total variation in the dataset. The information content of each component corresponds to 67%, 19% and 8% respectively (Supplementary Information, Table S6).

The multivariate analysis of the FA binding pockets presented in Fig. 6B shows the similarities between the pockets in terms of the volume, H-bond interactions, total interactions and water contacts over the course of the eight different simulations performed. While this analysis does not take into account the specific 3D volume, or the precise nature of the H-bonds (i.e. salt bridge), they provide at least a crude estimate or their overall properties. Component one is dominated by the volume and H-bond, and appears to be a descriptor for the overall size of the pockets (i.e. the total no. of H-bonds correlates with the volume). In contrast, the total VDW contacts and the H_2O contacts dominate the second component. As discussed above FA2 displays the largest volume and FA5 and FA6 are the smallest. Sites FA1 and FA7 make the most VDW contacts with sites FA3 and FA5 being the least. The solvent exposure descriptor is encompassed in both Components one and two, lying in the bottom right quadrant. Thus sites FA6 and FA7 have the greatest exposure and FA2 and FA3 the least. Considering Fig. 6B holistically, sites FA5 and FA6 cluster together roughly, as do sites FA1, FA4 and FA7, while sites FA3 and FA2 seem more unique.

3.3.3. PCA analysis of the X-ray based ligand-pocket data

An orthogonal approach to exploring the HSA binding pockets is to consider the ligand properties of those molecules that are confirmed binders. Thus, a second PCA model was generated using fifty seven different crystallographic determined ligands of HSA (at different sites) for which fourteen physico-chemical descriptors were calculated, for contrasting with the orthogonal MD based approach described above. The model fitted four components which describe 81% of the total variation in the dataset. The information content of each component corresponds to 37%, 20%, 13% and 11% respectively (Supplementary Information, Table S7). Again, we limit our discussion to the two most significant components.

Of the fifty seven ligands identified as binding to HSA [60], 43 are acidic, ten are neutral and three are zwitterionic. HEME, a forth known zwitterionic molecule was not included in the study due to the inability to calculate a number of descriptors due to the presence of the Fe^{3+} metal cofactor. The zwitterions Lysophosphatidylethanolamine and dansyl-L-arginine, are found to bind to site 1 while the latter is also found to bind to site 7. The six unique neutral molecules are found to occupy sites FA1, FA2, FA3, FA6 and FA7. No basic molecules were identified from the crystallographic studies. The multivariate analysis of the HSA ligands broken down by binding site is presented in Fig. 7. Firstly, Component one primarily encode for H-bonding characteristics, with larger molecules capable of more H-bonds or salt bridge interactions (i.e. acid class indicator) being found to the left of the *X*-axis. Component two of the PCA plot is dominated by descriptors that encode hydrophobicity, with polar molecule found at the bottom of the Y-axis and hydrophobicity (see Table S1b for descriptor meanings). Molecular size descriptors contribute to both component one and two, with large, hydrophobic molecule capable of multiple H-bonds being found in the top left quadrant. Three non-polar FA molecules (C14, C15 and C16) which contain one H-bonding substituent were added as references, and these are located in the right hand upper quadrant.

In contrast to the model presented in Fig. 6B, we do not see a dramatic separation of the ligands that bind to HSA based on the

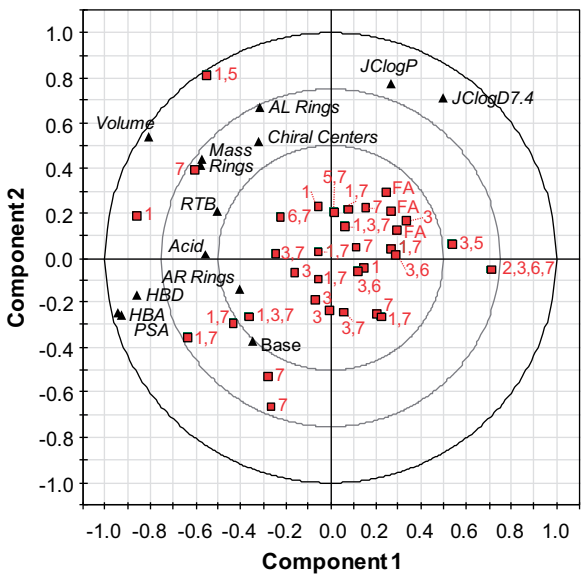

Fig. 7. PCA scores (top) and loadings plots (bottom) generated from the key physicochemical properties of substrates known to bind at each binding site. JclogP = calculated logP, $JclogD_{7.4}$ = calculated distribution coefficient $pH_{7.4}$, RTB = rotatable bonds, PSA = polar surface area, HBA = H-bond acceptor, HBD = H-bond donor, volume = molecular volume, AR Rings = aromatic rings, AL rings = aliphatic rings, while Acid and Base or ionization state indicator variables.

pockets that they are found to occupy (Fig. 7). Indeed, analysis of the ligands in greater detail showed that it was not possible to separate them based on bulk properties, suggesting that binding to HSA must involve a large degree of pocket specificity. Interestingly, from Fig. 7 it can be seen that drug-like molecules that bind to FA1 often do so in combination with FA7 or FA3. This is consistent with the PCA model results based on the known ligands as these two sites have very similar bulk properties, which leads them to cluster together. While FA3 does not cluster with FA1 on both components, the pockets are very similar on Component one which indicates they have similar H-bonding and volume characteristics.

4. Conclusions

Ligand binding to HSA is a remarkably exciting, but very difficult task due to rationalize [5,13], let alone predict due to (a) the diverse conformational states the protein can occupy (open, intermediate and closed), (b) the seven distinct binding pockets known for endogenous and exogenous ligands and (c) the interaction effect other ligands binding at other pockets can have on the binding affinity (d) good configurational sampling of protein-ligand space via replicate runs and longer sampling times [61–63]. While bulk ligand molecule properties can describe binding to a moderate degree of accuracy from QSAR analyses [9,23,24], it is clear that greater understanding of the specificity of binding is needed before we can have confidence in predicting either the likely binding pocket or overall affinity from molecular dynamics.

In this work we have shown that different starting configurations of HSA will have implications for simulation results obtained. Indeed, this is consistent with experimental data since it is known that the effect of fatty acid can change the overall structure or affect the affinity of ligands for HSA. Additionally, the overall size, shape and dynamics of the protein is expected to have a dramatic effect on the binding pockets within HSA, and this will affect the affinity fatty acids and other drug molecules have for them. Our results confirm that differences in fatty acid binding can have a dramatic effect on the flexibility of the structure and the pocket volumes. We see that over the course of a 100 ns simulation, varying the amount of FA bound to the protein allows the open form of the protein to adopt the experimentally expected intermediate conformational state, between the open and closed. Interestingly, FA sites 2, 4 and 5 are known to be the most high affinity FA sites, yet sites 1, 2 and 3 have the greatest number of H-bonds on average, sites 4, 6 and 7 are the most solvent accessible and sites 1 and 2 are the largest in size. However, sites FA1, FA2 and FA3 show the greatest degree of volume fluctuation between the different simulations. Analysis of the linearity of the pockets reveals that FA2, FA4 and FA5 can bind long chained fatty acids with the lowest strain and this would help to explain their higher affinity status.

Our comparison of the properties of the HSA pockets using MD parameters or physic-chemical properties of known pocket ligands provides interesting results. In terms of their gross properties, sites FA5 and FA6 cluster together, as do sites FA1, FA4 and FA7, while sites FA3 and FA2 appear more unique. In contrast, a ligand based analysis of bound drug-like molecules shows that we do not see a dramatic separation of the ligands according to physiochemical properties and the pockets that they occupy. This confirms that binding to HSA involves a degree of specificity above and beyond the bulk properties of its ligands.

Acknowledgment

This work was supported by the Thailand Research Fund grant RSA5480016 (MPG).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jmgm.2014.10.007.

References

[1] J.R. Gillette, Factors affecting drug metabolism, Ann. N. Y. Acad. Sci. 179 (1971) 43–66.
[2] J.A. Hamilton, NMR reveals molecular interactions and dynamics of fatty acid binding to albumin, Biochim. Biophys. Acta 1830 (2013) 5418–5426.
[3] D. Carter, X. He, Structure of human serum albumin, Science 249 (1990) 302–303.
[4] S. Curry, P. Brick, N.P. Franks, Fatty acid binding to human serum albumin: new insights from crystallographic studies, Biochim. Biophys. Acta 1441 (1999) 131–140.
[5] J.R. Simard, P.A. Zunszain, C.E. Ha, J.S. Yang, N.V. Bhagavan, I. Petitpas, et al., Locating high-affinity fatty acid-binding sites on albumin by X-ray crystallography and NMR spectroscopy, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 17958–17963.
[6] E.S. Krenzel, Z. Chen, J.A. Hamilton, Correspondence of fatty acid and drug binding sites on human serum albumin: a two-dimensional nuclear magnetic resonance study, Biochemistry 52 (2013) 1559–1567.
[7] M.J.N. Junk, H.W. Spiess, D. Hinderberger, The distribution of fatty acids reveals the functional structure of human serum albumin, Angew. Chem. Int. Ed. 49 (2010) 8755–8759.
[8] K. Valko, S. Nunhuck, C. Bevan, M.H. Abraham, D.P. Reynolds, Fast gradient HPLC method to determine compounds binding to human serum albumin. Relationships with octanol/water and immobilized artificial membrane lipophilicity, J. Pharm. Sci. 92 (2003) 2236–2248.
[9] G. Colmenarejo, A. Alvarez-Pedraglio, J.L. Lavandera, Cheminformatic models to predict binding affinities to human serum albumin, J. Med. Chem. 44 (2001) 4370–4378.
[10] N.A. Kratochwil, W. Huber, F. Muller, M. Kansy, P.R. Gerber, Predicting plasma protein binding of drugs: a new approach, Biochem. Pharmacol. 64 (2002) 1355–1374.
[11] K.J. Fehske, U. Schläfer, U. Wollert, W.E. Müller, Characterization of an important drug binding area on human serum albumin including the high-affinity binding sites of warfarin and azapropazone, Mol. Pharmacol. 21 (1982) 387–393.
[12] G. Sudlow, D.J. Birkett, D.N. Wade, Further characterization of specific drug binding sites on human serum albumin, Mol. Pharmacol. 12 (1976) 1052–1061.
[13] J. Ghuman, P.A. Zunszain, I. Petitpas, A.A. Bhattacharya, M. Otagiri, S. Curry, Structural basis of the drug-binding specificity of human serum albumin, J. Mol. Biol. 353 (2005) 38–52.
[14] G. Sudlow, D.J. Birkett, D.N. Wade, The characterization of two specific drug binding sites on human serum albumin, Mol. Pharmacol. 11 (1975) 824–832.

[15] F. Zsila, Circular dichroism spectroscopic detection of ligand binding induced subdomain IB specific structural adjustment of human serum albumin, J. Phys. Chem. B 117 (2013) 10798–10806.
[16] F. Zsila, Subdomain IB is the third major drug binding region of human serum albumin: toward the three-sites model, Mol. Pharmaceut. 10 (2013) 1668–1682.
[17] U. Kragh-Hansen, H. Watanabe, K. Nakajou, Y. Iwao, M. Otagiri, Chain length-dependent binding of fatty acid anions to human serum albumin studied by site-directed mutagenesis, J. Mol. Biol. 363 (2006) 702–712.
[18] R. Subramanyam, A. Gollapudi, P. Bonigala, M. Chinnaboina, D.G. Amooru, Betulinic acid binding to human serum albumin: a study of protein conformation and binding affinity, J. Photochem. Photobiol. B 94 (2009) 8–12.
[19] S. Datta, M. Halder, Detailed scrutiny of the anion receptor pocket in subdomain IIa of serum proteins toward individual response to specific ligands: HSA-pocket resembles flexible biological slide-wrench unlike BSA, J. Phys. Chem. B 118 (2014) 6071–6085.
[20] I. Petitpas, A.A. Bhattacharya, S. Twine, M. East, S. Curry, Crystal structure analysis of warfarin binding to human serum albumin: anatomy of drug site I, J. Biol. Chem. 276 (2001) 22804–22809.
[21] K. Honma, M. Nakamura, Y. Ishikawa, Acetylsalicylate-human serum albumin interaction as studied by NMR spectroscopy – antigenicity-producing mechanism of acetylsalicylic acid, Mol. Immunol. 28 (1991) 107–113.
[22] Y. Kurono, H. Yamada, K. Ikeda, Effects of drug binding on the esterase-like activity of human serum albumin. V. Reactive site towards substituted aspirins, Chem. Pharm. Bull. (Tokyo) 30 (1982) 296–301.
[23] S.B. Gunturi, R. Narayanan, A. Khandelwal, In silico ADME modelling 2: computational models to predict human serum albumin binding affinity using ant colony systems, Bioorg. Med. Chem. 14 (2006) 4118–4129.
[24] E. Estrada, E. Uriarte, E. Molina, Y. Simon-Manso, G.W. Milne, An integrated in silico analysis of drug-binding to human serum albumin, J. Chem. Inf. Model. 46 (2006) 2709–2724.
[25] G.J.P. van Westen, J.K. Wegner, A.P. Ijzerman, H.W.T. van Vlijmen, A. Bender, Proteochemometric modeling as a tool to design selective compounds and for extrapolating to novel targets, MedChemComm 2 (2011) 16–30.
[26] M.P. Gleeson, S. Modi, A. Bender, R.L. Marchese Robinson, J. Kirchmair, M.<ET-AL.> Promkatkaew, The challenges involved in modeling toxicity data in silico: a review, Curr. Pharm. Des. 18 (2012) 1266–1291.
[27] O. Deeb, M.C. Rosales-Hernández, C. Gómez-Castro, R. Garduño-Juárez, J. Correa-Basurto, Exploration of human serum albumin binding sites by docking and molecular dynamics flexible ligand–protein interactions, Biopolymers 93 (2010) 161–170.
[28] S. Fujiwara, T. Amisaki, Molecular dynamics study of conformational changes in human serum albumin by binding of fatty acids, Proteins 64 (2006) 730–739.
[29] S. Fujiwara, T. Amisaki, Identification of high affinity fatty acid binding sites on human serum albumin by MM-PBSA method, Biophys. J. 94 (2008) 95–103.
[30] S. Fujiwara, T. Amisaki, Steric and allosteric effects of fatty acids on the binding of warfarin to human serum albumin revealed by molecular dynamics and free energy calculations, Chem. Pharm. Bull. (Tokyo) 59 (2011) 860–867.
[31] O.K. Abou-Zied, N. Al-Lawatia, M. Elstner, T.B. Steinbrecher, Binding of hydroxyquinoline probes to human serum albumin: combining molecular modeling and forster's resonance energy transfer spectroscopy to understand flexible ligand binding, J. Phys. Chem. B 117 (2013) 1062–1074.
[32] C. Malleda, N. Ahalawat, M. Gokara, R. Subramanyam, Molecular dynamics simulation studies of betulinic acid with human serum albumin, J. Mol. Model. 18 (2012) 2589–2597.
[33] M.M. Castellanos, C.M. Colina, Molecular dynamics simulations of human serum albumin and role of disulfide bonds, J. Phys. Chem. B 117 (2013) 11895–11905.
[34] S. Fujiwara, T. Amisaki, Fatty acid binding to serum albumin: molecular simulation approaches, Biochim. Biophys. Acta 1830 (2013) 5427–5434.
[35] A.A. Bhattacharya, T. Grüne, S. Curry, Crystallographic analysis reveals common modes of binding of medium and long-chain fatty acids to human serum albumin, J. Mol. Biol. 303 (2000) 721–732.
[36] RCSB Protein Data Bank: www.rcsb.org
[37] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, J. Appl. Crystallogr. 26 (1993) 283–291.
[38] H. Li, A.D. Robertson, J.H. Jensen, Very fast empirical prediction and rationalization of protein pKa values, Proteins: Struct. Funct. Bioinform. 61 (2005) 704–721.
[39] B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation, J. Chem. Theory Comput. 4 (2008) 435–447.
[40] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, Comparison of multiple Amber force fields and development of improved protein backbone parameters, Proteins: Struct. Funct. Bioinform. 65 (2006) 712–725.
[41] A.W. Sousa da Silva, W.F. Vranken, ACPYPE – AnteChamber PYthon Parser interfacE, BMC Res. Notes 367 (2012).
[42] J. Wang, W. Wang, P.A. Kollman, D.A. Case, Automatic atom type and bond type perception in molecular mechanical calculations, J. Mol. Graph. Model. 25 (2006) 247–260.
[43] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, et al., Gaussian 03, Revision C.02, Gaussian, Inc., Wallingford CT, 2004.
[44] Case, D.A., Darden, T.A., Cheatham, T.E., Simmerling, C., Wang, J., Duke, R.E., et al. AMBER 12, Francisco, CA.
[45] T. Darden, D. York, L. Pedersen, Particle Mesh Ewald: an N·log(N) method for Ewald sums in large systems, J. Chem. Phys. 98 (1993) 10089–10092.
[46] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola, J.R. Haak, Molecular dynamics with coupling to an external bath, J. Chem. Phys. 81 (1984) 3684–3690.
[47] S. Nosé, A molecular-dynamics method for simulations in the canonical ensemble, Mol. Phys. 52 (1984) 255–268.
[48] B. Hess, H. Bekker, H. Berendsen, J.L.I.N.C.S. Fraaije, A linear constraint solver for molecular simulations, J. Comput. Chem. 18 (1997) 1463–1472.
[49] Discovery Studio 2.5: www.accelrys.com
[50] W. Humphrey, A. Dalke, K. Schulten, VMD – Visual molecular dynamics, J. Mol. Graph. 14 (1996) 33–38.
[51] ChemAxon JChem: www.chemaxon.com
[52] Vortex 2014: www.dotmatics.com
[53] Umetrics. SIMCA-P 11.5. Umetrics, Uvistevägen 48, Box 7960, SE-907 19 Umeå, Sweden.
[54] A. Nicholls, G.B. McGaughey, R.P. Sheridan, A.C. Good, G. Warren, M. Mathieu, et al., Molecular shape and medicinal chemistry: a perspective, J. Med. Chem. 53 (2010) 3862–3886.
[55] D.J. Birkett, S.P. Myers, G. Sudlow, Effects of fatty acids on two specific drug binding sites on human serum albumin, Mol. Pharmacol. 13 (1977) 987–992.
[56] H. Vorum, H.R.I. Jørgensen, R. Brodersen, Variation in the binding affinity of warfarin and phenprocoumon to human serum albumin in relation to surgery, Eur. J. Clin. Pharmacol. 44 (1993) 157–162.
[57] S. Curry, Lessons from the crystallographic analysis of small molecule binding to human serum albumin, Drug Metab. Pharmacokinet. 24 (2009) 342–357.
[58] S.I. Fujiwara, T. Amisaki, Molecular dynamics study of conformational changes in human serum albumin by binding of fatty acids, Proteins – Struct. Funct. Bioinform. 64 (2006) 730–739.
[59] N.A. Baker, D. Sept, S. Joseph, M.J. Holst, J.A. McCammon, Electrostatics of nanosystems: application to microtubules and the ribosome, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 10037–10041.
[60] G. Amidon, H. Lennernäs, V. Shah, J. Crison, A theoretical basis for a biopharmaceutic drug classification: the correlation of in vitro drug product dissolution and in vivo bioavailability, Pharm. Res. 12 (1995) 413–420.
[61] C.D. Christ, A.E. Mark, W.F. van Gunsteren, Basic ingredients of free energy calculations: a review, J. Comput. Chem. 31 (2010) 1569–1582.
[62] A. Pohorille, C. Jarzynski, C. Chipot, Good practices in free-energy calculations, J. Phys. Chem. B 114 (2010) 10235–10253.
[63] A. de Ruiter, C. Oostenbrink, Free energy calculations of protein–ligand interactions, Curr. Opin. Chem. Biol. 15 (2011) 547–552.