Human Serum Albumin: *A Paradox of Structural Simplicity and Functional Complexity*

What is Human Serum Albumin?

Human serum albumin (HSA) is globular protein of 585 amino acids and 66,438Da. It is encoded by a single copy gene ALB gene on the long arm of homo sapiens chromosome 4. Due to the addition of an N-terminus pro-peptide (6 residues) and signal peptide (18 residues) the translated protein is 609 amino acids, until processing in the hepatocyte cytoplasm [1]. HSA is a member of a protein family of homologous proteins with an extraordinary ligand binding capacity and distinctive structural features [2]. This facilitates the protein as a depot and carrier for many endogenous and exogenous compounds [3]. HSA's primary function is as the main determinant of plasma oncotic pressure and modular of fluid distribution between body compartments. Its secondary function is to bind ligands like water, zinc, calcium, sodium, potassium, fatty acids, hormones, bilirubin and drugs; facilitated by its unique and intricate structure. It is the most abundant protein in human blood plasma, contributing to around half of serum protein [4].

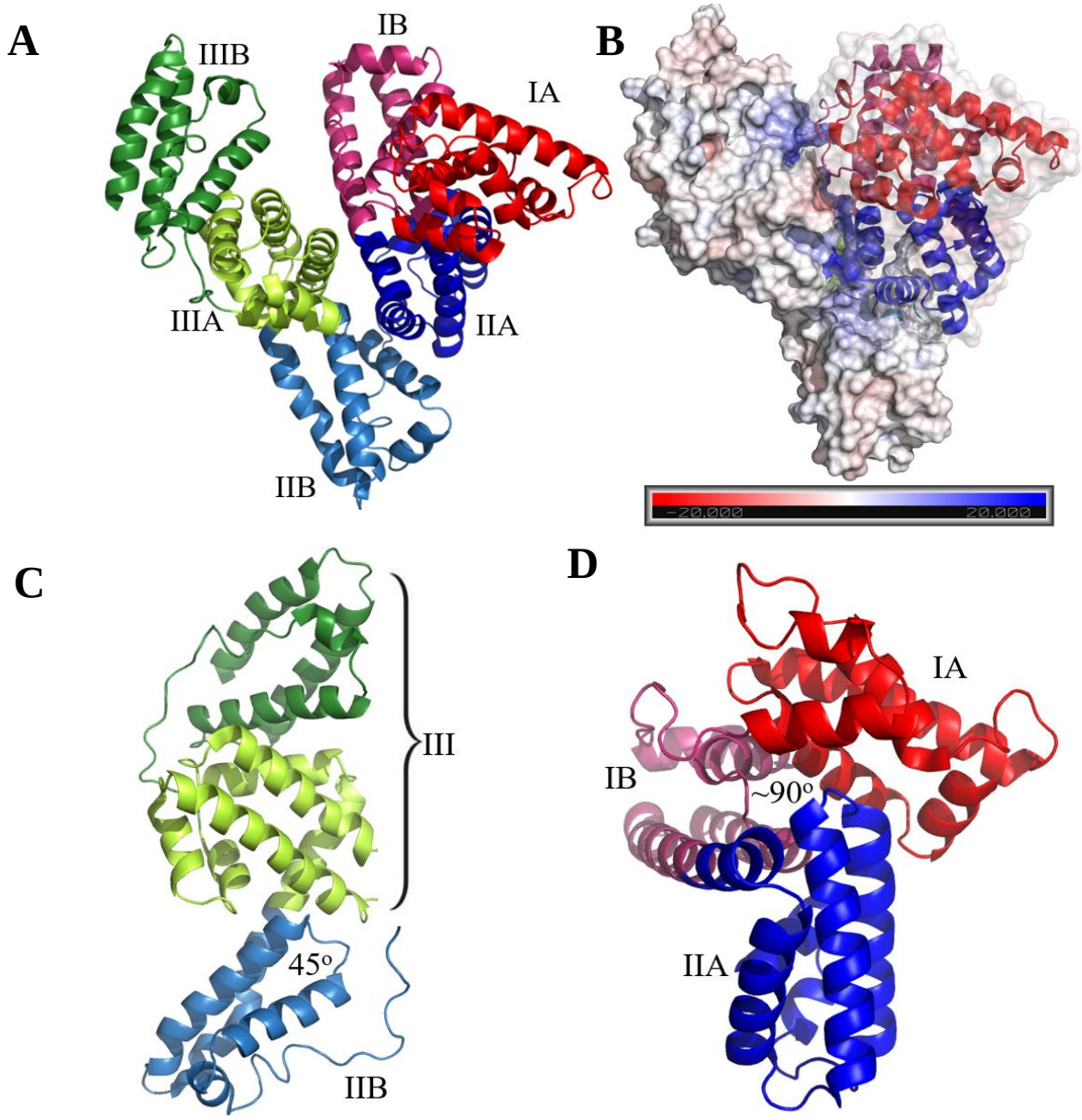


Figure 1: Cartoon structure of HSA made in Pymol [5] A) This image depicts the 6 subdomains within the 3 domains of HSA. The first domain is coloured in shades of red, second in shades of blue and third in shades of green. **B)** This image was created using an **APBS plugin** for pymol which depicts the electrostatics of HSA's surface, colour coded according to the legend below. **C)** This is the Y junction between III and IIB discussed on the next page. The angle is shown as ~45 degrees **D)** This is the 'T' configuration discussed on the next page between I and IIB. The angle is 90 degrees (perpendicular).

Human Serum Albumin General Structure

The secondary structure of HSA is 68% alpha helices and has no B sheet element. It contains three homologous domains indicated as I (1-195), II (196-383) and III (384-585) (**fig.1a**). However, the structure is not so simple as each domain contains two separate subdomains (a and b) of ten helices: six (h1-h6) and four (h7-h10), respectively, connected by a long-extended loop (fig1b). The total number of helices is, however 28 (not 30) because the last helices of domains I and II are fused to the first helices of the next domain. [5]. The structure is held together by 17 pairs of intramolecular disulphide bridges but there is one free cysteine residue, Cys34, which is the only thiol group in the protein that does not take part [2]. This cysteine bonds are essential to curve the subdomains and hold the structure together to cluster aliphatic and aromatic side chains in the inner surface to form each subdomain's hydrophobic core. This topological feature is seen in every subdomain except Ia-helix 3: the lone Cys34 resides between Ia-helix 2 and Ia-helix 3 [5]. This cysteine can form intermolecular disulphide linkages and it is oxidised by cysteine/glutathione in 30-40% of blood HSA molecules [1].

The primary factor which facilitates different active site affinities is the positioning of the subdomains relative to each other: despite structural similarity, the domains interact with neighbour domains in different ways which creates a highly asymmetric environment with a variety of binding sites. A perpendicular T-Shaped arrangement is formed between subdomain IIA and the interface region between subdomains IA and IB, facilitated by hydrophobic interactions and hydrogen bonds (**fig.1d**). A unique Y-Shaped assembly between domains II and III is formed via a 45^{o} protrusion of domain III from subdomain IIb (**fig.1c**). The relative motions of these domain structures imply flexibility of HSA in different conditions [5].

Human serum albumin has a long lifetime of 28-36 days that can introduce structural modifications that affect ligand binding and anti-oxidant properties [6]. In addition to a higher proportion of acidic to basic residues, HSA's primary structure is predominantly composed of ionized residues which give it a net charge of around -15 and an acidic isoelectric point (pI) of 4.7 at physiological PH7 (**fig.1b**). This is essential in its function and regulation in the blood [7]. This high number of acidic and basic residues allows HSA structure to undergo reversible transitions at different pH values: an extended (E) conformation below 2.7, a fast-migrating (F) form between 2.7 and 4.3, the neutral (N) form between pH 4.3 and 8.0 and finally a basic (B) form at a greater pH, characterised by the loss of alpha helix and increased in affinity for some ligands [8].

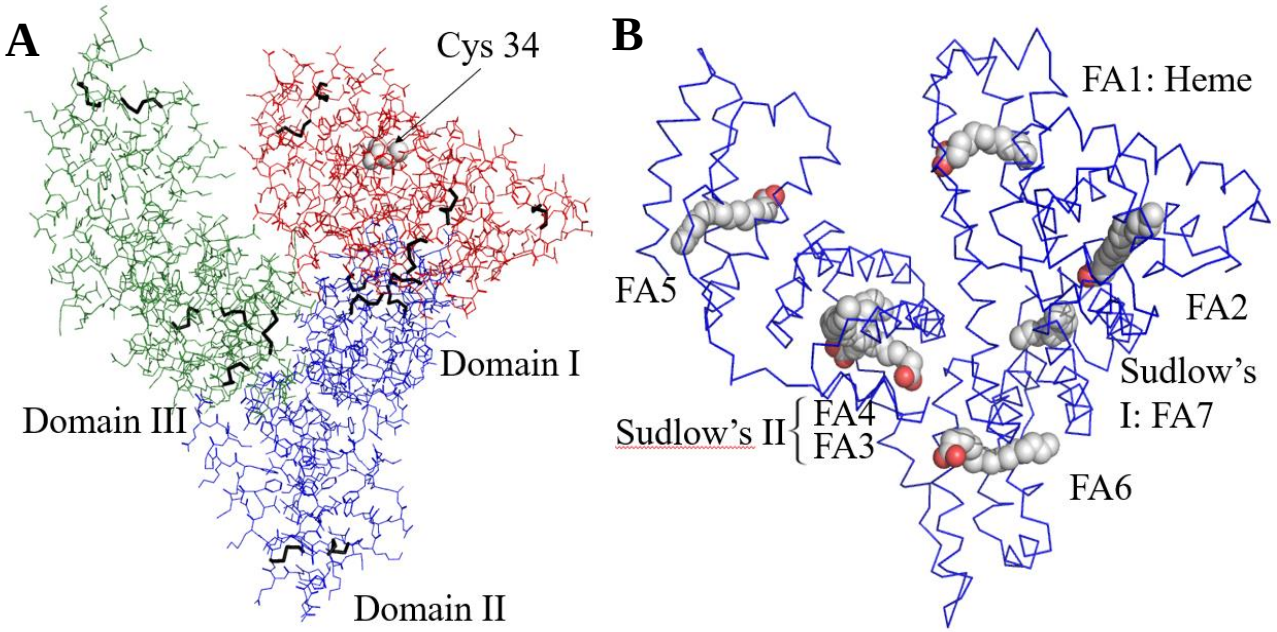


Figure 2: Ribbon structure of HSA made in Pymol [5][10]. A) This image depicts all 17 disulphide bond pairs as black lines within the HSA structure. There are, however, 35 cysteine residues in HSA and the free Cys34 position is depicted as 'ball and stick; and labelled. [A] **B)** This image depicts all 7 fatty acid binding sites in HSA. Sites 3 and 4 are grouped as 'Sudlow's Site I' and site 7 is 'Sudlows Site II'. The palmitic fatty acid moieties are depicted as ball and stick, coloured by element [B].

Function of Human Serum Albumin

Aliphatic Fatty Acids – Eggs, Nuts, Fish, Milk

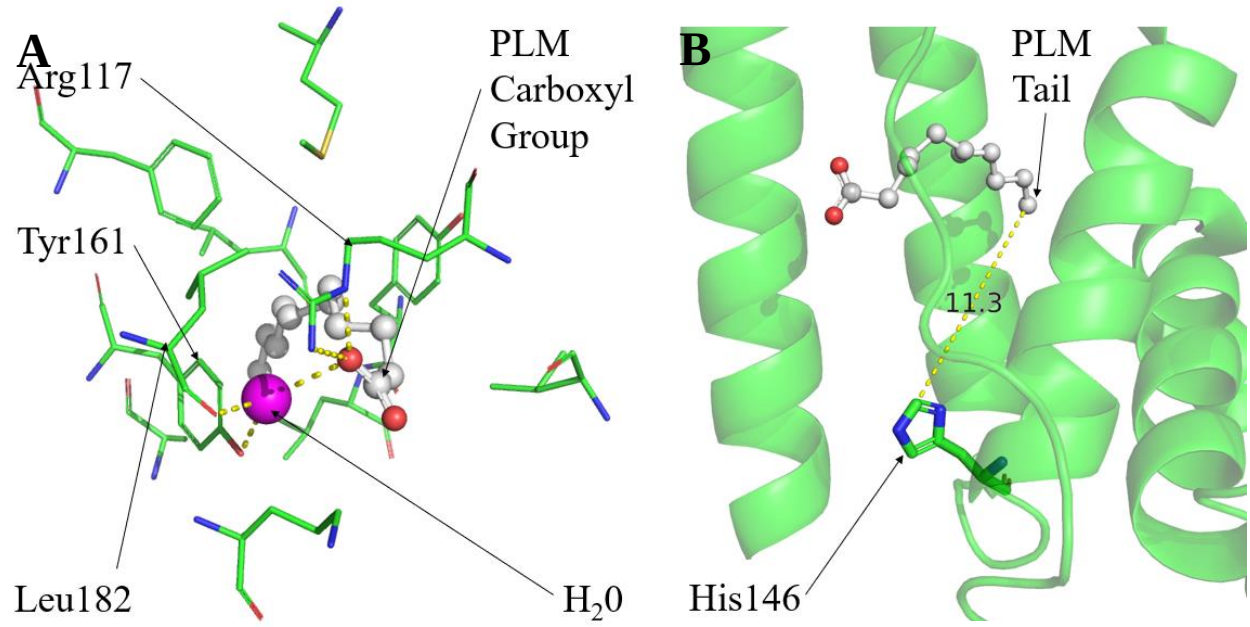
An essential function of Human Serum Albumin is to carry substrates in the blood and its structure is specifically designed for essential fatty acids ligands. Figure 3 pictures palmitic acid within the active sites of serum albumin [A]. Palmitic acid is a hexadecenoic acid with a 16-carbon chain; it is the most common saturated fatty acid found in animals, plants and microorganisms and makes up 44% of total fats. The palmitate anion is the observed from of palmitic acid at physiological pH (7.4) [9]. There are 7 binding sites on the molecule (11 for medium-chain fatty acids C10:14) occupied in different ways by fatty acids, facilitated by folding and arrangement of the subdomains to create environments of high affinity (**fig.2b**). Sites 1-5 have well placed clusters of amino acids which co-ordinate the binding of the fatty acid and anchor its carboxylate moiety by electrostatic interactions to basic and polar side-chains. A notable difference in these 5 sites is how the fatty acid is folded in the site: the longer the methylene tail of the fatty acid, the stronger electrostatic and hydrophobic interactions with the pocket. Palmitic acid is a 'long fatty acid' (>16C) and binds strongly. Sites 6 and 7 are 'lower affinity' [10].

The interaction homology of the first five sites can be seen by the binding of palmitic acid in the most open and accessible site, site I within subdomain IB (**fig.3a**). Fatty acids in these sites bind in the same orientation with the carboxylate group hydrogen-bonded to a water molecule. In site 1, this is coordinated by the side chain hydroxyl of Tyr161 and carboxyl oxygen of Leu182 which a water molecule in place that interacts and anchors the fatty acid carboxyl group [10]. Arg117 also contributes to hydrogen bonding interactions with the palmitic acid carboxylate moiety [5]. For longer chain fatty acids, like palmitic acid, the hydrophobic tail curls around the inside surface of the cavity toward His146 at the lower end of the cavity opening [10]. The distance of 11.3Å can be reduced quickly to the ~5Å maximum distance of Van der Waals interactions by the folding of the fatty acid chain which is 43.7Å in length (**fig.3b**) [11]. Site 2 is between IIa and IIb and similar but it is one of the most enclosed fatty acid binding sites on HSA and long fatty acid chains extend further into the pocket. Site 3 is also similar but broader running from subdomain IIIa to IIb: long fatty acid chains fold into a U-bend configuration. Site 4 is narrow and long and longer chains just extend further into the hydrophobic tunnel of subdomain IIIa; site 5 is the exact same, running through sub-domain IIIb [10].

The unique structure of site 6 is located in a shallow trench on the surface of the protein, between IIa and IIb (**fig.3c**). It lacks a basic/polar amino acid clusters to anchor the carboxylate moiety as in sites 1-5. The ligand density is stronger on the IIb side, including side chains Arg209, Lys351 and Ser480, which transiently interact with the carboxyl group in a variety of positions, with no conserved electrostatic interaction. However, the middle portion of the methylene tail is well anchored via a said chain 'strap' holding it in position, formed between Arg209 to both Asp324 and Glu354 (**fig.3c**)[10]. Site 7 is significantly smaller and has been hypothesised as a binding site specific for shorter chain fatty acids: this is supported by evidence which tried to covalently modify lys220 in site 7 subdomain IIa as this was blocked by small chain fatty acids but not long chain fatty acids (>16C), confirming binding [12].

Phenolic Acids – Fruit, Vegetables, Cereals

Many plant-based foods introduce phenolic acids into the diet; these are not aliphatic and have a phenolic ring in their structure. Gallic acid (trihydroxybenzoic acid) has antioxidant effects and low toxicity; it is a hydrolysed tannin which primarily enters the body from tea and fruits like strawberries, grapes and bananas [13]. Gallic acid binds to serum albumin in the blood after intestinal absorption and which introduces the importance of the unusual single tryptophan residue, Trp214, within the structure of HSA (most proteins have more) (**fig.4a**) [1]. This tryptophan residue helps position the gallic acid moeity in subdomain IIA of HSA to a position ~10.4Å away where it interacts with hydrophobic residues concentrated in the pocket at a distance of ~3Å (**fig4**). This is because the nitrogen of the tryptophan is attracted to the hydroxyl group of gallic acid. Within the pocket, gallic acid forms many hydrogen bonds between its protruding hydroxyl groups and nitrogen of basic amino acids which shield the gallic acid such as Arg257 and Arg222. The tyrosine 150 in the pocket is particularly important as it has an aromatic component which can form pi-pi stacking hydrophobic interactions with the aromatic ring in the gallic acid moiety. These hydrophobic interactions can also extend from gallic acid to Ala261 and Leu219/238 via induced Van der Waals forces with their methyl groups. The shielding of gallic acid by the pocket decreases its absorbance [13].



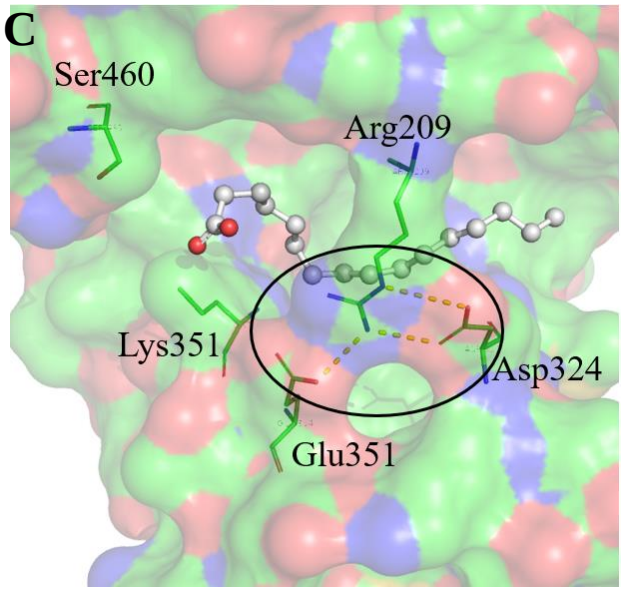


Figure 3: Active Site representation of HSA made in *PyMol* [10]. A) Active site of fatty acid binding site 1, within four helices at the centre of subdomain IB. An additional Arg117 is located in a crevice formed by the helix cluster of subdomain Ib and the extended loop between subdomains Ia and Ib. The water molecule is pink and polar interactions are shown by dashed yellow lines. **B)** The 11.3A distance depicted between the bottom His146 residue and the PLM tail. Histidine is often important in catalytic triads and His146 is the gate to the cavity. **C)** The 'strap' formed by carboxyl groups on Asp324 and Glu341 with the nitrogen from Arg209 is circled. This holds the palmitic acid in place.

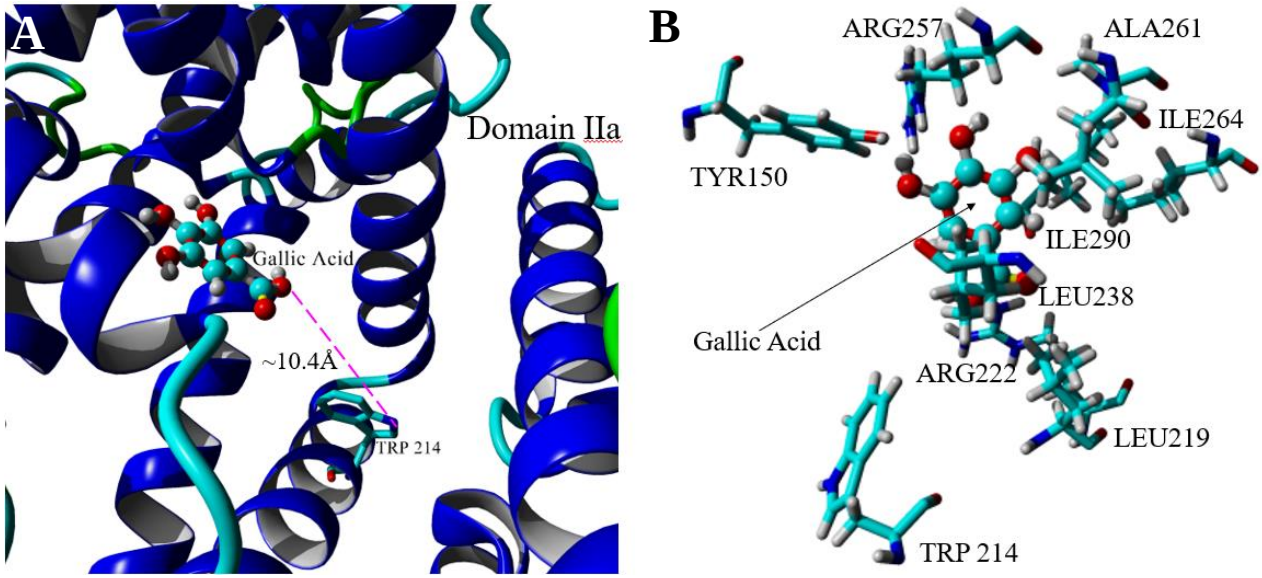


Figure 4: Active site representation of HSA made in *Yasara* [5][13]. A) A combined structure using a ligand free crystallised pdb of HSA and a gallic acid molecule built from a smile string calculated using its structure as C1=C(C=C(C(=C1O)O)O)C(=O)O [13]. The unique single TRP 214 residue is labelled and its distance between its nitrogen and the hydroxyl of gallic acid is marked in pink. **B)** The active site of HSA (FA6) within subdomain IIA. Relevant amino acids are labelled.

Anion Binding of Human Serum Albumin (400 words)

HSA is an essential transition metal ion transporter in plasma: it binds approximately 45% of circulating calcium, 45% of circulating magnesium and 80% of circulating zinc. It has the highest affinity to zinc, followed by calcium and then magnesium. The binding is facilitated by potentially more than 2 calcium binding sites; there is a shared binding site between zinc and calcium at Asp273 which also suggests a crosstalk between zinc and calcium transport in the blood [14]. HSA was long described as a 'sponge' which binds all and everything non-specifically. However, in the last decades, along with regulatory roles of the protein, specific binding sites with specific metal preferences were inferred from spectroscopic techniques and limited analysis of the structural amino acid environment. There are four metal binding sites in most mammalian albumins including HSA and BSA: the N-terminal site (NTS), Cys34 residue (and surrounding environment), Site A (MBS – multi metal binding site) and Site B (unknown location) (**fig.4a**) [15].

The **NTS** site is the hallmark of albumin metal binding. It is composed of the first three Asp-Ala-His residues in the albumin sequence (hence N-terminal). These make up four donor atoms involved in coordination which prefer metal ions capable of a square-planar formation like Cu(II) and Ni(II): NTS bound to HSA is the second largest pool of copper in human blood serum (after ceruloplasmin). The donor atoms include the His-3 and free N-terminal amine group which facilitate the formation of peptide bonds in the ion coordination [15]. The **Cys34 site** involves the lone cysteine mentioned earlier, that is not involved in HSA's 17 disulphide bridges [2]. In reduced HSA, it contains a free thiol group with a very low pKa of 5. This Cys34 is located in a cleft between helices 2 and 3 and subdomain IA. This site is specific for metal ions that bind effectively to HSA via a single, high enthalpy, metal-sulphur bond. **Site A** is an extensively studied site. It is the 'multi binding site (MBS)' because it is the preferential site for Zn(II), one of the two thermodynamically equivalent sites for Cd(II) and a secondary weaker binding site for both Cu(II) and Ni(II). This site consists of imidazole nitrogens (from His247 and His67) located in the axial positions of a distorted trigonal bipyramidal arrangement around the metal ion. Asn99 and Asp249 provide carboxyl ligands and the final, 5th, ligand is water (**fig.4b**). The His67 and Asn99 are from domain I whereas His247 and Asp249 are from domain II. There is also a 6th, more distant ligand which is provided by the His247 amid oxygen [15]. **Site B** is non-localised and is specific for cadmium (Cd(II)) with the same affinity as site A. Chemical shift shows that there is not more than one nitrogen ligand involved in binding. It is a primary site for Mn(II) ions and can also bind Zn(II), albeit with much lower affinity.

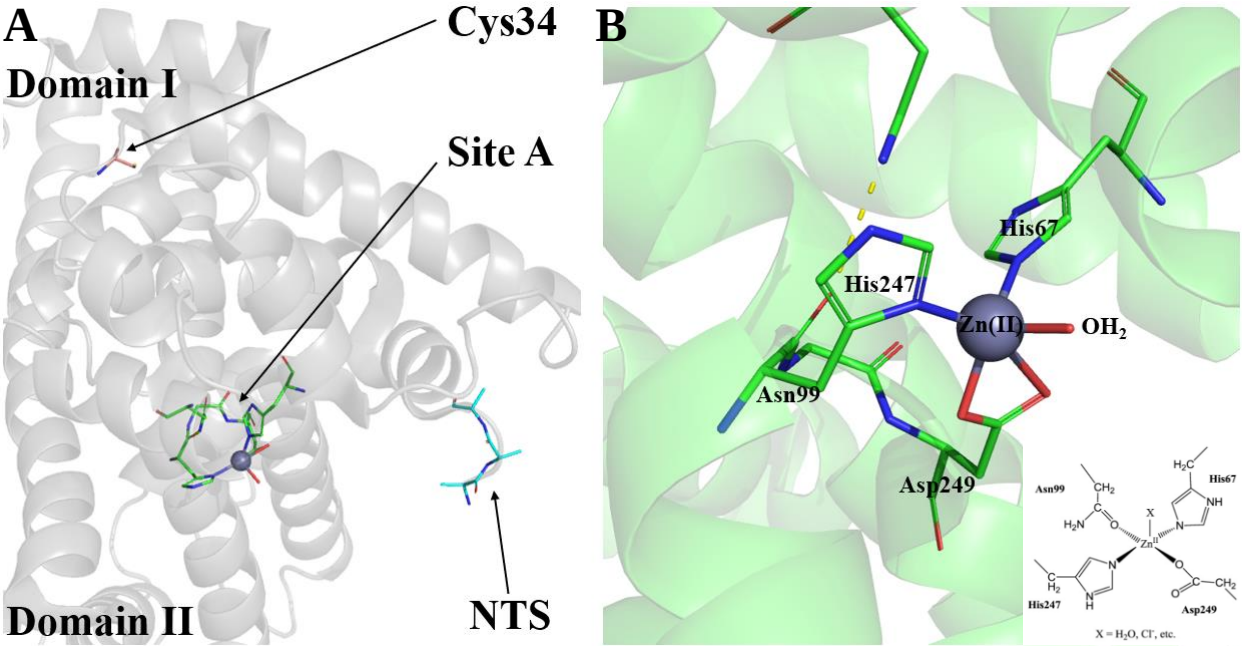


Figure 4: Ion Binding Sites made in *PyMol* [15][16][17]. A) Discovered locations of 3 of the 4 sites (Cys34, Site A and NTS) all within Domain II. The location of Site B has not yet been determined. **B)** The Site A active site with a Zn(II) ion bound. The key residues and water are labelled, forming a bipyramidal arrangement around the metal ion.

Conclusion

The future and utilisation of our knowledge lies in albumin-based drug delivery. The two major drug binding sites are named Sudlow's site 1 (binds drugs like warfarin) and Sudlow site II (binds drugs like ibuprofen) (**fig.2b**). Furthermore, drug metabolites (furosemide and salicylic acid) and Nonsteroidal Anti-Inflammatory Drugs (ibuprofen) can react covalently with HSA: the free thiol group of cysteine 34 is an attractive feature for forming such bonds in drug delivery [17]. Albumin-associated drugs are being developed to enhance drug biodistribution and bioavailability due to the reversibility of their binding. These drugs incorporate components, like fatty acids above, to potentiate albumin association and increase drug efficacy. Ligand binding capacity of HSA (as defined by crystallographic studies to date) are depicted in table 1: the sites correspond to sites visualised in figure 2b.

FA1	FA3/4: Sudlow Site II	FA5	FA6	FA7	CLEFT
Hemin 2º: Azapopazone 2º: Indomethacin 2º: TIB	Thyroxine 4 Diflunisal Diazepam Halothane Ibuprofen Indoxyl Sulphate Propofol 2º: CMPF	2º: Oxyphe-nbutzaone 2º: Propofol	2º: Diflunisal 2º: Halothane 2º: Ibuprofen	Azapropazone CMPF DIS Indomethacin Oxyphenobutazone Phenylbutazone TIB Warfarin 2º: Indoxyl-Sulphate 3º: Diffunisal	Thyroxine 5 2º: Indipamide

Table 1: Ligan Binding Capacity of HSA: Sites correspond to those visualised in figure 2b [18]**.** The 'cleft' is an additional site, mainly for thyroxines that is located in the interface between subdomains IIA and IIB.

Overall, serum albumin is the most abundant soluble protein in the body of all vertebrates and one of the longest known and most studied of all proteins. However, the ability of this protein to have such a myriad of affinities despite only two major binding sites is a paradox that still puzzles scientists [1]. There is great research into fatty acid binding and ion binding discussed in this report. However, the details of thermodynamics and kinetics of binding between metal ions and HSA is limited since there is a lack of X-ray structures of HSA in metal bound states; Site B's localisation is also yet to be discovered [15]. Furthermore, there is even possibility of an interplay between ions like Zn(II) and fatty acid binding that can 'translate certain aspects of the organismal energy state into global zinc signals'. This is mediated by interdependent binding between Zn(II) and fatty acid binding at the FA2 site: fatty acids may serve as an allosteric switch for Zn(II) binding to plasma [19]. Thus, ligand binding is not so simple as one compound. These factors all makes understanding serum albumin ever more complex and a true paradox of functional complexity despite structural simplicity .

References:

[1]https://books.google.co.uk/books?hl=en&lr=&id=i1DC3KlTAB8C&oi=fnd&pg=PP2&dq=All+about+Albumin:+Biochemistry,+Genetics+and+Medical+Applications&ots=WYHebQ9obf&sig=TqlEVQAZlJZB04L5Vsw8egBSqLo&redir_esc=y#v=onepage&q&f=false: Chapter 4.1
[2] https://www.sciencedirect.com/science/article/pii/S0098299711000896#b1255
[3] https://iubmb.onlinelibrary.wiley.com/doi/abs/10.1080/15216540701474523
[4] https://www.uniprot.org/citations/19021548
[5] https://academic.oup.com/peds/article/12/6/439/1487202 [1AO6]
[6] https://www.sciencedirect.com/science/article/pii/S0014579308003967
[7] Book, Chapter 2.1

[8] https://www.sciencedirect.com/science/article/pii/S0301462210000608
[9] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5682332/
[10] https://www.sciencedirect.com/science/article/pii/S0022283600941585?via%3Dihub#bBIB10 [1E7H]
[11] https://pubmed.ncbi.nlm.nih.gov/18210100/
[12] https://cir.nii.ac.jp/crid/1571417124224806144
[13] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7700232/
[14] https://www.uniprot.org/uniprotkb/P02768/entry
[15] https://www.sciencedirect.com/science/article/pii/S030441651300278X?via%3Dihub
[16] https://pubmed.ncbi.nlm.nih.gov/28567254/ [5IJF]
[17] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4769556/
[18] https://pubmed.ncbi.nlm.nih.gov/23726993/
[19] https://www.sciencedirect.com/science/article/pii/S0022283605008855?via%3Dihub

Pdb Files used:

[5] 1AO6 : https://www.rcsb.org/structure/1ao6
[10] 1E7H: https://www.rcsb.org/structure/1e7h
[16] 5IJF https://www.rcsb.org/structure/5ijf