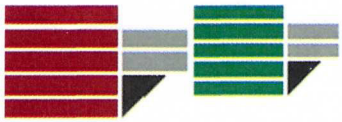
UNIVERSITÀ DELLA CALABRIA



Dipartimento di Fisica

Doctorate School of Science and Technique

"Bernardino Telesio"

A thesis submitted for the degree of Doctor in Physics of Complex Systems

XXVII Cycle

FIS/07

"Molecular Simulation of Transport Proteins in Interaction with Physiological and Pharmacological Ligands"

School Director

Prof. Roberto Bartolino

Supervisors

Dr. Rita Guzzi

Dr. Bruno Rizzuti

Curriculum Coordinator

Prof. Vincenzo Carbone

Candidate

Stefania Evoli

December 2014

TABLE OF CONTENTS

Abbreviations

βLG	Beta lactoglobulin
HSA	Human serum albumin
FA	Fatty acid
IBP	ibuprofen
MD	molecular dynamics
SD	stochastic dynamics
LINCS	LINear Constraints Solver
PME	particle mesh Ewald
TI	thermodynamic interaction
EXP	exponential averaging
BAR	Bennett acceptance ratio
MBAR	multistate Bennett acceptance ratio
PDB	protein data bank
BFGS	Broyden-Fletcher-Goldfarb-Shanno
PCCA	Perron-cluster-cluster analysis
HB	hydrogen bond
R_g	radium of gyration
RMSD	root mean square deviation
RMSF	root mean square fluctuation
NMR	nuclear magnetic resonance

Abbreviations for amino acids

Ala	Alanine	Leu	Leucine
Arg	Arginine	Lys	Lysine
Asn	Asparagine	Met	Methionine
Asp	Aspartate	Phe	Phenylalanine
Cys	Cysteine	Pro	Proline
Glu	Glutamate	Ser	Serine
Gln	Glutamine	Thr	Threonine
Gly	Glycine	Trp	Tryptophan
His	Histidine	Tyr	Tyrosine
Ile	Isoleucine	Val	Valine

Summary

Molecular complexes of transport proteins with small compounds have been studied by using docking techniques and molecular dynamics simulations. The macromolecules considered are β-lactoglobulin and albumin, i.e. the most abundant proteins in bovine milk and human blood serum, respectively. The ligands are long-chain fatty acids of different length and ibuprofen, a molecule of pharmaceutical interest.

Simulations of β-lactoglobulin with fatty acids, ranging from caprylic to stearic acid, revealed the key protein residues that contribute to the binding process. In particular, a rationale was found for the high binding affinity of both stearic and palmitic acid compared to shorter lipids. Moreover, the location of two low-affinity external binding sites was predicted for palmitic acid, by comparing docking results with those obtained for vitamin D_3, for which an external site has already been identified in crystallography.

For human serum albumin, docking results suggest different candidate binding locations for both charged and neutral ibuprofen. An alchemical free energy approach has been used to estimate the binding affinity for each pose. The results show that charged ibuprofen has a greater affinity for albumin compared to the ligand in the neutral form, suggesting that the former corresponds to the physiological binding state. The simulation findings were compared to experimental results and show an overall good agreement, predicting details of the protein-ligand interaction that include binding geometries and contacts with specific amino acid residues.

The overall findings reveal significant features of the binding of well-known ligands to two extensively investigated transport proteins, and show how computational tools can be used to support experimental techniques in a variety of cases.

Sommario

Sono stati studiati complessi molecolari di proteine di trasporto con piccoli composti per mezzo di tecniche di docking e simulazioni di dinamica molecolare. Le macromolecole considerate sono la β-lattoglobulina e l'albumina, ossia la proteina più abbondante nel latte bovino e nel siero di sangue umano, rispettivamente. I ligandi sono acidi grassi a catena lunga di differente estensione e l'ibuprofene, una molecola di interesse farmaceutico.

Simulazioni della β-lattoglobulina con acidi grassi, che vanno dal caprilico allo stearico, hanno rivelato i residui proteici chiave che contribuiscono al processo di associazione. In particolare, è stata trovata una spiegazione della alta affinità di legame dell'acido palmitico e stearico rispetto a lipidi con catena più corta. Inoltre, è stata predetta la posizione di due siti esterni a bassa affinità per l'acido palmitico, comparando risultati di docking con quelli ottenuti per la vitamina D_3, per la quale un sito esterno è già stato identificato in cristallografia.

Per l'albumina del siero umano, i risultati di docking suggeriscono differenti posizioni candidate ad essere di legame, sia per l'ibuprofene carico e sia per quello neutro. Un approccio al calcolo di energia libera di tipo alchemico è stato utilizzato per stimare l'affinità di legame per ogni posa. I risultati hanno mostrato che l'ibuprofene carico ha un'affinità maggiore per l'albumina rispetto al ligando in forma neutra, suggerendo che il primo dei due corrisponde allo stato associato in condizioni fisiologiche. I risultati di simulazione sono stati comparati con quelli sperimentali e mostrano un buon accordo complessivo, consentendo di predire dettagli dell'interazione proteina-ligando che includono le geometrie di legame e i contatti con residui aminoacidici specifici.

I risultati complessivi rivelano caratteristiche significative dell'associazione di ligandi ben noti con due proteine di trasporto estesamente studiate, e mostrano come tecniche computazionali possono essere utilizzate per supportare quelle sperimentali in un'ampia varietà di casi.

Introduction

Proteins are complex systems that perform essential functions in living organisms. They may have structural functions, catalyze chemical transformations as enzymes, are responsible of muscular movements, coordinate transport of other molecules and defend the organism from external agents.

Transport proteins are responsible for the transfer of ligands and are also known as 'carriers'. Numerous carriers exist according to the different kind of ligands that they have to transport. In some cases, the interaction is also important because some ligands acquire their functionality in consequence of the binding. The interaction between proteins and ligands involve a complex molecular process including recognition, binding, transport and release of the compound to the target site. In this interaction both structural and dynamical aspects need to be explored. Studies on the nature of the interaction between proteins and small molecules have a great importance in the biomedical and pharmaceutical research fields.

Within this framework, the proteins we are interested in are β-lactoglobulin (βLG) and Human Serum Albumin (HSA), model proteins that act as carriers of fatty acids and other small organic molecules. They differ in size and degree of structural complexity. βLG (162 amino acids) belongs to the lipocalin family and it is the most abundant protein in the milk of ruminants. It is a very important system for food industry and for biotechnological applications. The biological function of βLG is not completely clarified, although it certainly has a role as carrier of fatty acids [Perez and Calvo, 1995; Rocha et al., 1996; Brownlow et al., 1997] and other small hydrophobic compounds. It also shows affinity for a number of bioactive molecules, such as vitamins (e.g., folic acid, cholecalciferol) and polyphenols (e.g., resveratrol). The main binding site of βLG is an internal central cavity, known as calyx. In

addition to the calyx, the existence of other binding sites on the protein surface have been suggested by several authors [Narayan et al., 1998; Wang et al., 1998; Yang et al., 2008; Zhang et al., 2014].

The first part of this thesis deals with the investigation of the interaction of fatty acids of different length (ranging from 8 to 18 carbon atoms) in the βLG calyx and with the theoretical prediction of secondary binding sites for palmitic acid. This ligand is a natural lipid component and it is the most abundant in βLG isolated from bovine milk [Barbiroli et al., 2011]. It also has a high binding affinity for this protein compared to other fatty acids [Frapin et al., 1993; Collini et al., 2003].

The second part of this work is concerned with a more complex molecular system, HSA in interaction with a pharmacological molecule, ibuprofen (IBP), which is the active ingredient in numerous drugs (e.g., Antalgil, Brufen, Moment) with anti-inflammatory effect [McKee et al., 2008; Palma et al., 2009; Nanau and Neuman, 2010]. HSA (585 amino acids) is the most abundant protein in blood plasma and interstitial fluids [Petitpas et al., 2001]. It binds several physiological compounds such as fatty acids [Petitpas et al., 2001] and, importantly, a high number of drug molecules. HSA has seven binding sites that can be occupied by medium and long-chain saturated fatty acids [Battacharya et al., 2000] and two sites recognized as drug binding sites [Sudlow et al., 1976].

Although important, the knowledge of the location of protein binding sites is only part of the issue of understanding the interaction between protein and ligands. In fact, despite the X-ray structures of both βLG and HSA are known, and even for ligands whose interaction sites are established, the dynamical characteristics of the protein-ligand complex are mostly unknown and can play a relevant role in determining its functionality. Questions associated

with molecular recognition, binding affinity, transport and ligand release require the use of combined experimental and computational approaches.

In this work, we use two computational techniques, docking and molecular dynamics (MD) [Karplus and McCammon, 2002] to clarify these issues. Docking allows to identify the most probable ligand binding sites in βLG and HSA, whereas MD is used to study the dynamics of the βLG-fatty acids and HSA-IBP complexes. Moreover, absolute binding free energy methods are used to evaluate protein-ligand binding affinity of the HSA-IBP complex in simulation.

Docking is a method that allows to predict the structure of a protein-ligand complex and it is based on a geometric and energetic analysis of the intermolecular interactions [Kitchen et al., 2004]. The structures suggested by molecular docking can be the starting configurations for subsequent MD simulations, which allow to study the dynamical features of the molecular complex. The combination of docking and MD provides information not easily accessible to the various experimental techniques, and therefore complete them providing useful additional predictions.

The results on the βLG-fatty acid interaction show that the ligand binding in the βLG main site increases the protein flexibility in the loops surrounding the calyx, compared to the unliganded form. All of the fatty acids are anchored with the head-group at the entrance of the protein calyx, but only palmitic and stearic acid are found in a fully extended conformation in simulation, whereas shorter fatty acids fluctuate more in correspondence with the tail. Two additional binding sites for palmitic acid have been identified on the external surface of βLG.

Free energy calculations on the HSA-IBP complex for both charged and neutral IBP show that this ligand can bind HSA in several poses that can be ranked according to their

binding affinity. These poses correspond to binding sites previously identified by crystallography for fatty acids and drugs [Bhattacharya et al., 2000; Ghuman et al., 2005]. The simulation results agree very well with experimental data and predict the location of additional IBP binding sites available for this ligand in solution.

This thesis is organized as follows. In the first chapter, the basic principles of molecular docking, MD and alchemical free energy calculations are introduced. The second chapter is concerned with the results obtained on the interaction of βLG with fatty acids, and also describes the structure of the protein, conditions of simulation, and methods of data analysis. Finally, the molecular basis of the interaction of IBP with HSA, as obtained by using accurate free energy calculations, are described in the third chapter.

1. Docking, Molecular Dynamics and binding free energy

1.1. Docking

1.1.1. Docking algorithms

Molecular docking is a computational technique that allows to predict the interaction sites between two biological macromolecules, for instance between a protein and a small ligand. This technique needs knowledge of the structure of a binding site, which is normally obtained by crystallography and NMR techniques or for homology with known structures. The ligand is placed into the binding site through an optimization of steric, hydrophobic, electrostatic and hydrogen bond (HB) interactions and the resulting binding free energy is evaluated. The binding affinity is the energetic difference between the complex and the sum of two uncomplexed molecules, and the entropic and enthalpic variation guides the complex formation:

$$\Delta G = \Delta H - T\Delta S \qquad (\mathbf{1.1})$$

where ΔG is the binding energy, ΔH is the change in the enthalpy of the system, T is the temperature and ΔS is the change in the entropy of the system.

Food and pharmaceutical research requires high performances and plausible results in a reasonable time. Therefore, all docking techniques aim to use all known information of the complex receptor-ligand, by using specific algorithms to predict the binding geometry and parametric functions for the evaluation of the binding affinity. The function that estimates the affinity between the target and the ligand is known as scoring function. Most docking programs use a series of search approaches applied to the ligand and the receptor. These

methods include systematic or stochastic torsional searches about rotatable bonds, MD simulations and genetic algorithms.

Among the docking software currently employed, AutoDock Vina is one of the fastest and most accurate [Trott and Olson, 2010]. AutoDock Vina combines a Lamarckian genetic algorithm with an empirical free energy force field, obtaining fast prediction of bound conformations and the corresponding free energies of association [Morris et al., 1998].

1.1.2. AutoDock Vina

AutoDock Vina combines an empirical free energy force field with the Lamarckian genetic algorithm to predict bound conformations with free energies of association. A scoring function is used to approximate the standard chemical potential of the system [Trott and Olson, 2010]. The binding free energy can be written as:

$$\Delta G = \Delta G_{gauss} + \Delta G_{rep} + \Delta G_{hbond} + \Delta G_{hydroph} + \Delta G_{tors} \qquad (\mathbf{1.2})$$

where ΔG_{gauss} is the attractive term for dispersion (described by two gaussian functions), ΔG_{rep} is the term for steric repulsion, ΔG_{hbond} refers to HB and is described by a ramp function, as well as the hydrophobic term, $\Delta G_{hydroph}$, and ΔG_{tors} performs the restriction of the internal rotors and global translations [Morris et al., 1998].

The conformation-dependent component of the scoring function *c* can be expressed as the summation over all pairs of atoms moving relative to each other, excluding 1–4 interactions:

$$c = \sum_{i<j} f_{t_i t_j}(r_{ij}) \qquad (\mathbf{1.3})$$

where for each atom i a type t_i is assigned, and $f_{t_i t_j}$ is a symmetric set of interaction functions of the interatomic distance r_{ij}. The general function c depends on the intermolecular and intramolecular interactions so it can be also defined as:

$$c = c_{intramol} + c_{intermol} \qquad (\mathbf{1.4})$$

The predicted binding free energy is represented by the intermolecular component of the lowest-scoring function conformation:

$$s_1 = g(c_1 - c_{1,intramol}) = g(c_{1,intermol}) \qquad (\mathbf{1.5})$$

where g is an increasing smooth non-linear function, and the subscript 1 indicates the lowest scoring conformation. By using the $c_{1,intermol}$ value it is possible to assign a ranking for the other lower-score conformations s_i:

$$s_i = g(c_i - c_{1,intramol}) \qquad (\mathbf{1.6})$$

The atom typing scheme is the same of X-score [Wang et al., 2002]: H-bond donors (O and N atoms bonded to H atoms); H-bond acceptors (O and sp2 or sp hybridized N atoms with lone pairs); H-bond donors/acceptors (O and N atoms which can be H-bond donor or H-bond acceptor); polar atoms (O and N atoms that are neither H-bond donor nor H-bond acceptor, and C atoms bonded to hetero-atoms); and 'hydrophobic' atoms (C atoms that cannot be considered belonging to the 'polar atom' group) [Wang et al., 2002].

The functions $f_{t_i t_j}$ (see eq. 1.3) depends on the surface distance $d_{ij} = r_{ij} - R_{t_i} - R_{t_j}$[Jain, 1996] where R_t is the van der Waals radius of atom type t:

$$f_{t_i t_j}(r_{ij}) \equiv h_{t_i t_j}(d_{ij}) \qquad (\mathbf{1.7})$$

where $h_{t_i t_j}$ is the weighted sum of the steric interactions.

The function $g(c_{1,intermol})$ depends on N_{rot}, the number of active rotatable bonds between the ligand and the heavy atoms, weighted by a coefficient w:

$$g(c_{intermol}) = \frac{c_{intermol}}{1+wN_{rot}} \qquad (\mathbf{1.8})$$

To find the global minimum of c and other low-scoring conformations an optimization algorithm is needed. The Broyden-Fletcher-Goldfarb-Shanno (BFGS) [Nocedal and Wright, 1999] method, an iterative algorithm for solving unconstrained nonlinear optimization problems, is used in AutoDock Vina for the local optimization.

The BFGS method uses the derivatives of the scoring function with respect to the position and orientation of the ligand and the torsions for the active rotatable bonds in the ligand. The derivatives would represent the negative total forces acting on the ligand, the negative total torque and the negative torque projections (torque applied to the branch moved by the torsion, projected on its rotation axis). The number of the steps in a run depends on the complexity of the search. Runs start from random conformations and can be performed through multithreading. The different minima found with the optimization algorithm are combined and used during a clustering stage.

AutoDock Vina calculations are performed in several steps: (1) the preparation of coordinate files of both the protein and the ligand, (2) the definition of the search space within a volume including the protein and of the rotatable bonds for the ligand by using an auxiliary

software such as AutoDock Tools [Morris et al., 1998]; (3) the actual docking of the ligand using AutoDock Vina; and (4) the visualization of the resulting docking poses by using the molecular graphic system PyMOL [DeLano, 2012]. For the ligand and receptor coordinates, AutoDock Vina uses the file format PDBQT, which is an extension of the PDB file format [Berman et al., 2000] additionally containing atomic type definition, atomic charges and, for the ligand, topological information.

Since the ligand needs a large conformational searching space around the protein, AutoDock Vina uses a grid-based method to evaluate the binding energy of trial conformations. The receptor is placed in a grid, a probe atom is sequentially confined in each grid point and the overall interaction energy between the probe and the receptor is computed.

1.2. Molecular Dynamics

1.2.1. Simulation of a system of particles

Molecular dynamics (MD) is one of the most used computational techniques that describe the structural and dynamical properties of solvated proteins and biomolecules on a time scale longer than the nanosecond [Karplus and McCammon, 2002]. Computer simulations describe the molecular system in the atomic detail. The MD technique is based on the calculation of the Newtonian equations of motion for a system composed of N atoms:

$$m_i \frac{\partial^2 \vec{r}_i}{\partial t^2} = m_i \vec{a}_i = \vec{F}_i \qquad i = 1,2,\dots,N \qquad (\mathbf{1.9})$$

where $m_i, \vec{r}_i, \vec{a}_i$ and $\vec{F}_i$ are, respectively, the atomic mass, position, acceleration and force acting on the *i*-th particle.

The force depends on the overall atomic positions $\vec{R}$ through the potential function V:

$$\vec{F}_i = -\nabla_i V(\vec{R}) = -\nabla_i V(\overrightarrow{r_1}, \overrightarrow{r_2}, \dots, \overrightarrow{r_N}) = -\left(\frac{\partial V}{\partial x_i}, \frac{\partial V}{\partial y_i}, \frac{\partial V}{\partial z_i}\right) \qquad (\mathbf{1.10})$$

MD produces the trajectory of the system, i.e. a set of states of the simulated ensemble. This approach allows to calculate physical quantities through the time average of the values obtained during the simulation. According to the ergodic hypothesis, the average of the values of an ensemble is the same as the average of the system values during the time evolution.

The equations of motion for the N atoms must be solved numerically, not only because of the high number of particles constituting the molecular systems, but also because there is no analytical solution for the motion of a system of three or more bodies. Numerous numerical algorithms have been developed for integrating the equations of motion. Because of its time-reversibility and symplectic nature, the leap-frog algorithm is commonly used in MD simulations [Hockney and Goel, 1974]. The leap-frog Stochastic Dynamics integrator (SD) [Van Gunsteren and Berendsen, 1988], which adds a friction and a noise term to the Newtonian equations of motions, can alternatively be used [Goga et al., 2012]. The integration time step is limited by the fast motions in the system, hence it should be on a femtosecond time scale to ensure stability of the integration.

Periodic boundary conditions are used to avoid boundary effects due to the finite size of the system and simulate bulk conditions. The molecular system is placed in a unitary cell spatially surrounded by other identical and translated copies. The set of replicas forms a tridimensional and periodic lattice that is virtually infinite. To avoid the effects caused by periodicity, the minimum-image convention [Rahman and Stillinger, 1971] is used. According to this convention, each atom interacts only with the nearest copy. An atom can

virtually leave the central cell during the simulation, but it will be substituted by its image coming from the opposite side of the cell. Therefore, the number of atoms into the cell is constant. Different cell shapes tessellating the 3-D space exist. One of the most convenient shape to simulate a globular macromolecule in hydrated condition is the rhombic dodecahedron, because it minimizes the amount of solvent required [Bekker, 1997].

1.2.2. Potential functions

Empirical potential functions are used for molecular dynamics calculations. These functions reproduce the essential physical properties of the system, achieving a good compromise between accuracy and computational efficiency. The functions representing the potential energy and the parameters used constitute the so-called force field.

The potential energy of the system is described by the sum of bonded and non-bonded interactions:

$$V(\vec{R}) = V_{bonded} + V_{non-bonded} \qquad (\mathbf{1.12})$$

Bonded and non-bonded interactions correspond to distinct physical terms; in additions, in MD simulations special interactions can be introduced. These interactions allow to restrict or fix the atomic positions and mutual geometry through the use of restraints or constraints, respectively [van Gunsteren and Karplus, 1982].

1.2.3. Bonded interactions

Bonded potential is the sum of the following three terms:

$$V_{bond} = V_{bond-stretch} + V_{angle-bend} + V_{torsion} \quad (\mathbf{1.13})$$

The bond-stretch represents the interaction with nearest neighbors, i.e. between pairs of atoms connected by a covalent bond. In the GROMOS 53a6 force field [Oostenbrink et al., 2004] some H atoms are not explicitly considered, but they are included within the C atom which they are attached to. This model is called united-atom or extended-atom [Ponder et Case, 2003], and it can be applied only to non-polar H atoms, since polar ones are important to keep the electrostatic properties of the system. In contrast, in other force field such as the AMBER99SB-ILDN force field [Lindorff-Larsen et al., 2010] non-polar H atoms are explicitly considered.

The bond-stretch can be described by using an harmonic potential:

$$V^{bond} = \sum_{bond} \frac{1}{2} K_b (b - b_0)^2 \quad (\mathbf{1.14})$$

where b is the bond distance, b_0 its equilibrium value and $K_b$is the force constant.

An alternative way to treat these interactions is to apply constraints to the bond distance [Ryckaert et al., 1977]. The use of constraints is convenient since, at room temperature (~ 300 K), the vibration frequency corresponding to the classical limit is given by:

$$h\nu = k_B T \Longrightarrow \nu \sim 10^2\ cm^{-1} \quad (\mathbf{1.15})$$

where h is Planck constant, ν is the frequency limit and k_B is the Boltzmann constant. For most of biologically interesting molecules, vibration frequencies of the bonds are over the classical limit. Thus, the constraints in this case are to be preferred to a potential of harmonic type. The use of constraints allows to increase the time-step in the integration of the equations motion up to $\Delta t \approx 2fs$ [van Gunsteren and Berendsen, 1977]. One of the most widely used algorithm to keep the constraints is P-LINCS (*Parallel LINCS*) [Hess, 2008a], an optimized version of LINCS (LINear Constraints Solver) [Hess et al., 1997] for parallel computation.

The bond-angle vibration between a triplet of atoms can be described by a harmonic potential on the angle:

$$V^{angle} = \sum_{angle} \frac{1}{2} K_\theta (\theta - \theta_0)^2 \qquad (\mathbf{1.16})$$

where θ, θ_0 and K_θ are the bond angle, its equilibrium value and the force constant, respectively.

Biological molecules can have different conformations differing in rotational orientation around the covalent bonds, which can be described by a *proper* torsion term. In addition, *improper* torsions can be defined to maintain a group of four atom (one central atom connected with other three atoms) in a defined geometry, either planar or tetrahedral. A set of planar geometries is required to maintain the conformation of rings or for other atomic groups, such as in peptide bonds, whereas a tetrahedral geometry is essential in the case of a force field that uses united atoms, to prevent transition to a configuration of opposite chirality.

$$V_{torsion} = V_{proper} + V_{improper} \qquad (\mathbf{1.17})$$

The function describing the proper torsion is:

$$V_{proper} = \sum_{\substack{proper \\ torsions}} \frac{1}{2} V_n \left(1 \pm \cos(n\phi - \phi_0)\right) \qquad (\mathbf{1.18})$$

where $\phi = \phi_{ijkl}$ is dihedral angle, ϕ_0 is its reference value, n is the multiplicity (an integer that determines the periodicity of the rotation), and V_n is the barrier height.

The potential for the improper torsion for the AMBER force field [Lindorff-Larsen et al., 2010] is described by the same term used for proper dihedrals (eq. 1.18), whereas GROMOS [Oostenbrink et al., 2004] uses the form:

$$V_{improper} = V_{id}(\varphi_{ijkl}) = \frac{1}{2} k_{\varphi} (\varphi_{ijkl} - \varphi_0)^2 \qquad (\mathbf{1.19})$$

where the angle of equilibrium φ_0 is 0^o for planar configurations and 35.26^o for the tetrahedral ones.

1.2.4. Non-bonded interactions

The non-bonded interactions play an important role in biomolecules despite being weaker than the bonded interactions. The two terms considered are:

$$V_{non-bonded} = V_{van\ der\ Waals} + V_{electrostatic} \qquad (\mathbf{1.20})$$

In simulations, these interactions are not taken into account for first and second neighbors, which are already involved in bond and angle interactions, respectively. For third neighbors, the so-called 1-4 interactions are either evaluated or not for torsions according to the model and force field type used. In particular, GROMOS [Oostenbrink et al., 2004] uses

special terms for calculating the non-electrostatic and electrostatic 1-4 parameters, whereas AMBER [Lindorff-Larsen et al., 2010] scales these two terms to 1/2 and 5/6, respectively.

Van der Waals interactions can be modeled with a Lennard- Jones potential:

$$V_{van\ der\ Waals} = \sum_{couple\ i,j} 4\varepsilon_{ij}\left(\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right) \qquad (\mathbf{1.21})$$

where r_{ij} is the distance between the i–th e the j–th atom, ε_{ij} is the depth of the potential well and σ_{ij} is the finite distance at which the inter-particle potential is zero. The first term of the summation depends on r^{-12} and takes account the repulsion between two atoms due to the impenetrability of the electron clouds. The second term, which depends on r^{-6}, describes the mutual attraction caused by London forces. It should be noted that GROMOS [Oostenbrink et al., 2004] uses geometric combination rule (with several *ad hoc* exceptions) for heteroatomic pairs, whereas AMBER used the so-called Lorentz-Berthelot rules, consisting in arithmetic and geometric averages for and σ_{ij} and ε_{ij}, respectively.

Coulomb's law describes the electrostatic interactions:

$$V_{Coulomb} = \sum_{couple\ i,j} \frac{1}{4\pi\varepsilon_0\varepsilon_r}\frac{q_i q_j}{r_{ij}} \qquad (\mathbf{1.22})$$

where q_i and q_j are the partial charges of the two interacting particles, ε_0 and $\varepsilon_r = 1$ are the permittivity in vacuum and in the medium, respectively.

The non-bonded interaction term is the most expensive one from a computational point of view. These interactions, especially the electrostatic ones, decrease slowly and, in principle, they should be evaluated for each couple of atoms of the system. The computational cost in this case would be proportional to N^2, where N is the number of atoms of the system.

Techniques reducing the computational cost exist and affect only minimally the accuracy of biomolecular simulations in the presence of solvent [Schlick, 2010].

To reduce the calculation time of the van der Waals interactions, the contribute of the interactions among distant atoms can be neglected by introducing a *cut-off radius*, R_{cutoff}. With the introduction of R_{cutoff}, the computational complexity is reduced and scales as N. Van der Waals potential decreases as r^{-3}, so it is possible to calculate all the interactions within a cut-off radius $R_{cutoff,1}$~1 nm and to neglect them entirely outside a cut-off radius $R_{cutoff,2}$ ~ 1.4 nm. In the intermediate zone between the two cut-off radii $R_{cutoff,1}$ and $R_{cutoff,2}$ the interactions are evaluated every N_{list} time-steps. To obey to the minimum-image convention the cell dimensions will have a minimum size that depends on the cut-off radius $R_{cutoff,2}$ [Leach, 2001]:

$$R_{cutoff,2} < \frac{L_{box}}{2} \qquad (\mathbf{1.23})$$

Ewald summation [Ewald, 1921] is a technique to calculate the electrostatic interactions in MD simulations. This method describes the charge-charge interaction into the central cell, and between the central cell and each image cell. The term accounting for these interactions can be written as:

$$V_{Coulomb}(r_{ij}) = \frac{1}{2} \sum_{|n|^*=0} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{4\pi\varepsilon_0} \frac{q_i q_j}{|r_{ij} + n|} \qquad (\mathbf{1.24})$$

where n=(n_xL,n_yL,n_zL), with n_x, n_y and n_z integer numbers and L the distance between the central simulation box and the adjacent replicas. The star indicates that the series is not including the term i=j for n=0. The summation converges slowly and conditionally (i.e., it contains a mixture of positive and negative terms, each of the two forming a divergent series,

and overall the results depends on the order of summation), but it can be divided in two summations that converge both fast and absolutely. To this aim, a Gaussian neutralizing distribution is added to each charge, and an analogous and opposite charge distribution is further added. Therefore, the summation can be expressed through three terms: the *real-space* term, the *reciprocal-space* one, and the *self-term*. If the medium around the periodic cells is considered non-conductive then another corrective term is added. When the real term converges quickly, the reciprocal one converges slowly and *vice versa*. Thus, a balance between reciprocal and real terms is necessary in the calculation. To improve the Ewald summation method for the irregular charge distribution typical of MD, Darden et al. introduced the *Particle Mesh Ewald* (PME) method [Darden et al., 1993]. The PME calculates direct-space interactions within a finite distance by using a modification of Coulomb's law and in reciprocal space by using a Fourier transform to build a mesh of charges, interpolated into a grid. In this way, the PME methods scales as $N\ lnN$ instead of N^2as in the original Ewald scheme.

1.3. Binding free energy

1.3.1. Introduction

Free energy calculation comprises a set of computational procedures to estimate free energy differences between different thermodynamic states. The importance of free energy is tightly connected to the stability of a system, because when the system reaches an equilibrium with its environment the free energy is minimized. Free energy calculations are useful to determine transfer free energies and partitioning coefficients for small molecules. The binding

free energy of a small molecule to a receptor can easily be converted to obtain the dissociation constant for that molecule.

MD simulation methods are used to produce independent samples from equilibrium, and these are used to estimate free energy differences. Due to the statistical nature of the simulations, the free energy results are not exact and, hence, error analysis must be carefully performed.

3.3.1. Calculating free energy differences from simulations

The statistical probability that a molecule (or a system) is found in some state i depends on the energy of the system. Given two different thermodynamic states in a constant volume ensemble, the free energy difference can be expressed as:

$$\Delta A_{ij} = -k_B T \ln \frac{Q_i}{Q_j} = -k_B T \ln \frac{\int V_j \, e^{-\frac{U_j(\vec{q})}{k_B T}} d\vec{q}}{\int V_i \, e^{-\frac{U_i(\vec{q})}{k_B T}} d\vec{q}} \qquad (\mathbf{1.25})$$

where ΔA_{ij} is the Helmholtz free energy difference between state i and j, k_B the Boltzmann constant, T the temperature expressed in Kelvin, Q the canonical partition function, U_i and U_j the potential energies depending on the coordinates and momenta and V_i and V_j the phase-space volumes. Replacing U_i and U_j with $U_i + P\mathcal{V}_i$ and $U_j + P\mathcal{V}_j$, respectively, and integrating over all container volumes, the Gibbs free energy and the isobaric-isothermal partition function can be calculated.

Alchemical free energy calculations originated with Kirkwood [Kirkwood, 1935] and Zwanzig [Zwanzig, 1954] relationships. According to the Kirkwood approach, a coupling

parameter (λ) controls the interaction strength between the ligand and the rest of the system. In this case, free energy differences are computed using *thermodynamic interaction* (TI). Zwanzig's method, known as *exponential averaging* (EXP), calculates the free energy difference between two states through an exponential average of energy differences over an ensemble of configurations [Zwanzig, 1954]. The free energy between two states with potentials $U_0(\vec{q})$ and $U_1(\vec{q})$ over a momentum space $\vec{q}$ can be expressed as:

$$\Delta A = -k_B T \ln\langle e^{-(k_B T)^{-1}(U_1(\vec{q})-U_o(\vec{q}))}\rangle_0 \qquad (\mathbf{1.26})$$

In this case the phase volume for both states is the same and the EXP converges very poorly if the two systems have substantial energy differences, or if the configuration space for the two systems is substantially different. This relationship can be used when the difference between potential energies is small. EXP method is more accurate when V_j is a subset of V_i [Lu and Kofke, 1999; Lu et al., 2003; Wu and Kofke, 2005; Jarzynski, 2006]. When the states have very little phase-space overlap, it is necessary to introduce a series of intermediate states that overlap with each other. The introduction of intermediate states implies the individual calculations of ΔA for each of them. Since only the difference between the starting and final state is important, the form of the intermediates is not relevant. For this reason, it is possible to choose entirely unphysical states that have good overlap with one another. The most convenient way to consider the intermediate states is belonging to a continuous pathway that links the initial and final states. The distance along this pathway is λ, i.e. the earlier introduced coupling parameter in the TI approach, with the initial state corresponding to $\lambda = 0$ and the final one to $\lambda = 1$. Therefore, by simulating the potential function depending on λ and $\vec{q}$, $U(\lambda, \vec{q})$, the estimation of each ΔA is possible.

The Bennett Acceptance Ratio (BAR) method can be thought of as providing the minimum variance/maximum likelihood estimate of the free energy difference between two thermodynamic states, given simulations conducted in both states. It also can be thought of as providing the optimal way of combining two different EXP free energy estimates of the free energy difference between state A and B, obtained from simulations in both states A and B, into a single estimate of the free energy difference between those states. This method has considerably improved results over EXP. Given two different states along the pathway, the potential energy difference of the same configuration $\vec{q}$ is $\Delta U_{ij}(\vec{q})$. The relationship between the distribution of potential energy differences $\Delta U_{ij}(\vec{q})$ sampled from the state j and $\Delta U_{ji}(\vec{q})$ from the state i represents the BAR:

$$\Delta A_{ij} = -\ln k_B T \frac{Q_j}{Q_i} = k_B T \ln \frac{\langle \alpha(\vec{q}) e^{-k_B T \Delta U_{ij}(\vec{q})} \rangle_j}{\langle \alpha(\vec{q}) e^{-k_B T \Delta U_{ji}(\vec{q})} \rangle_i} \qquad (\mathbf{1.27})$$

where $\alpha(\vec{q})$ is a positive function for all $\vec{q}$. By minimizing the variance of the free energy, Bennett found a suitable choice for $\alpha(\vec{q})$ so that ΔA is an implicit function that can be solvable numerically:

$$\sum_{i=1}^{n_i} \frac{1}{1 + e^{\left(\ln\left(\frac{n_i}{n_j}\right) + k_B T \Delta U_{ij} - k_B T \Delta A\right)}} - \sum_{j=1}^{n_j} \frac{1}{1 + e^{\left(\ln\left(\frac{n_j}{n_i}\right) + k_B T \Delta U_{ji} - k_B T \Delta A\right)}} = 0 \qquad (\mathbf{1.28})$$

where n_i and n_j are the number of samples for each state. Shirts and Lu demonstrated the theoretical and pratical superiority of BAR with respect to EXP in MD [Lu et al., 2003; Shirts and Pande, 2005] and BAR converges to EXP when all samples are from a single state [Bennett, 1976; Shirts et al., 2003].

An extension of BAR is the *multistate Bennett acceptance ratio* (MBAR) [Shirts and Chodera, 2008]. This method uses weighting functions $\alpha_{ij}(\vec{q})$ that minimize the variance during the reweighting, and reduces to BAR in the case of a single pair of states. Essentially, it generalizes BAR to the case where we are interested in free energy differences between a number of different states, and a simulation in any state can provide some information (even if small) about the other states.

2. Molecular simulations of β-lactoglobulin complexed with fatty acids

2.1. Description of βLG

βLG is one of the most important members of the lipocalins, a large family of proteins involved in some processes such as insect camouflage and transport of small hydrophobic compounds [Flower et al., 1996]. It is the principal whey protein in the milk of mammals such as cows (2−4 g/L) [Farrell et al., 2004], pigs, camels and dogs [Pervaiz and Brew, 1985]. The biological function of βLG is the transport of fatty acids [Perez and Calvo, 1995; Rocha et al., 1996; Wang et al., 1997; Kontopidis et al., 2002], retinoids and other small hydrophobic molecules. Under physiological conditions (neutral pH and protein concentration > 50 μM), βLG is dimeric. At pH lower than 3 and under salt-free conditions, βLG is monomeric [Timasheff and Townend, 1961; Baldini et al., 1999; Sakurai et al., 2001]. Below room temperature in the pH range 3.7 to 5.2, βLG forms oligomers [Timasheff and Townend, 1961; Kumosinski and Timasheff, 1966; McKenzie et al., 1967; Piazza and Iacopini, 2002]. βLG exists in three variants. Variant A and B differ for two residues, Asp vs Gly at position 64 and Val vs Ala at position 118, respectively. Variant C differs from B for a His/Gln substitution at position 59.

From a structural point of view, βLG is a small globular protein with 162 amino acid residues (molecular mass 18,400 Da) folded into a predominantly β-sheet structure (Fig. 2.1) with an additional 3-turn α-helix on the outer surface [Kontopidis et al., 2004].

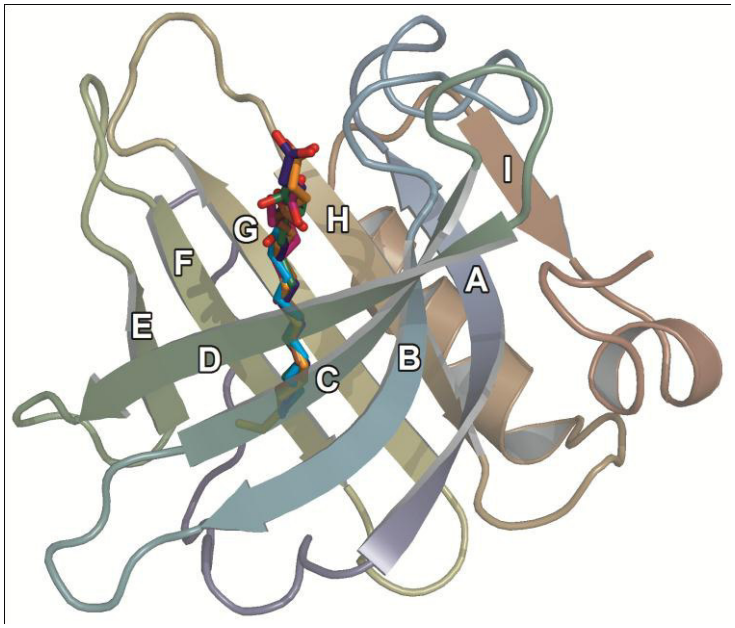


Fig.2.1 *– Crystallographic structure [Loch et al., 2011; Loch et al., 2012] of βLG complexed with fatty acids. Overlaid molecules are stearic (orange), palmitic (blue), myristic (green), lauric (brown), capric (cyan) and caprylic acid (purple).*

The eight β-strands A-H form a barrel with a conical cavity called calyx, which constitutes the primary binding site for hydrophobic ligands, whereas β-strand I is involved in dimer formation. The β-strands are connected by loops, each named according to the two strands it separates. The loops BC, DE and FG are at the closed end of the protein, whereas the loops AB, CD, EF and GH are at the calyx entrance and regulate the access to the binding site. In particular, at pH higher than 7.0 the EF loop (residues 85-90) adopts an open conformation that allows the ligands to enter into the calyx [Qin et al., 1998]. Gln89 is the residue involved in this "gate" function of the EF loop [Qin et al., 1998]. The βLG structure is stabilized by two disulfide bridges, between Cys106-Cys119, connecting strands G and H, and Cys66-Cys160, linking CD loop and C-terminus [Loch et al., 2011].

Since the middle of the last century the binding of hydrophobic ligands to the βLG was identified [McMeekin et al., 1949]. Studies conducted with a variety of experimental methods, including electron spin resonance [Guzzi et al., 2012], nuclear magnetic resonance spectroscopic [Ragona et al., 2000], spectrophotometry [Hu et al., 2011], affinity chromatography [Pelletier et al., 1998], equilibrium dialysis and fluorescence [Muresan et al., 2001], demonstrated the possibility of binding to βLG for numerous ligands [Sawyer et al., 1998; Sawyer, 2003]. Crystallographic and spectroscopic studies show that βLG binds, inside the central cavity, hydrophobic linear molecules [Wu et al., 1999; Kontopidis et al., 2002; Kontopidis et al.,2004; Loch et al., 2012], such as retinol and fatty acids with different aliphatic chain length, ranging from 8 (short fatty acids) to 20 C atoms (long fatty acids). The protein binding pocket is formed by hydrophobic residues. Lys60, Glu62 and Lys69 are the only charged residues located on (or close to) the CD loop at the entrance of the calyx. Fatty acids bound to βLG are found in an extended conformation into the calyx, with the carboxylate group anchored at the cavity entrance [Wu et al., 1999; Loch et al., 2012]. The fatty acid binding affinity depends on the chain length [Loch et al., 2012], the highest value being found for palmitic (C16:0) and stearic acid (C18:0) [Frapin et al., 1993; Loch et al., 2012]. In addition to the calyx, the existence of lower affinity external binding sites has been suggested by several authors [Narayan and Berliner, 1998; Qin et al., 1998; Wang et al., 1998; Wu et al., 1999; Yang et al., 2008a; Yang et al., 2008b]. However, several alternative positions on the surface of the protein have been proposed for fatty acids, as well as for folic acid [Liang and Subirade, 2010; Liang et al., 2011] cisparinaric acid [Dufour et al., 1992] and phosphatidylcholine [Mandalari et al., 2009]. The position of the potential binding sites proposed are: (1) close to the C-terminal loop, β-strand C and D [Yang et al., 2008], (2) close to the C-terminal α-helix pocket, β-strand F, G, H and A [Wu et al., 1999], (3) in a surface

hydrophobic cavity between the C-terminal region of the α-helix and β-strand I [Yang et al., 2008], (4) near the β-strand H and (5) in between CD and DE loops [Yang et al., 2008].

2.2. Computational methods

2.2.1. Protein modeling and ligand docking

The unliganded form of βLG and complexes with either caprylic, capric, lauric, myristic, palmitic or stearic acid (C8:0 to C18:0) were obtained starting from the X-ray structure crystallized in the presence of stearic acid (3UEX entry [Loch et al., 2012] in the Protein Data Bank (PDB)). The position of three missing residues in the loop GH was reconstructed on the basis of the corresponding region in the unliganded protein (3NPO entry [Loch et al., 2011] in the PDB).

Docking of fatty acids to βLG was investigated by using AutoDock Vina [Trott and Olson, 2010]. The graphical interface AutoDock Tools 1.5.6 [Morris at al., 1998] was used to convert the structures from the PDB format, add polar hydrogens to the protein and determine the center of the grid box, for which a grid spacing of 37.5 pm was applied. A search space including the entire βLG molecule was considered, and full flexibility was allowed for the ligand.

The docking procedure consisted of two independent runs, each determining the best ten docking conformation ranked according to binding affinity, and uncertainty ranges were determined by the differences among energy values for the same pose. Docking of fatty acids on the outer surface of βLG, in locations other than the protein calyx, was investigated for palmitic acid. For comparison, the same procedure was applied for vitamin D_3, to assess the

docking to an external βLG binding site in the case of a ligand for which a crystallographic complex is available [Yang et al., 2008].

2.2.2. Molecular dynamics

The simulation package GROMACS 4.0.7 [Hess et al., 2008] was used for trajectory production and analysis, in combination with the GROMOS 53a6 force field [Oostenbrink et al., 2004]. The softwares VMD [Humphrey et al., 1996] and PyMOL [DeLano, 2002] were used for molecular visualization. The details of the topologies of fatty acids were previously described [Rizzuti et al., 2010; Guzzi et al., 2012]. In brief, the fatty acid head-group is modeled like the carboxylate moiety of Asp/Glu residues, the aliphatic chain replicates methylene groups as in the side chain of Glu (their number depending on the length of each fatty acid), and the tail resembles the methyl moiety of Ala.

In all simulations, βLG was placed at the center of a rhombic dodecahedron box with a minimum distance of 1 nm with respect to cell walls. The protein-ligand complex (or unliganded protein) was surrounded with about 6400 water molecules, for which the SPC model was used [Berendsen et al., 1981]. The addition of eight Na^+ counterions (seven for the simulation of the unliganded protein) allowed to neutralize the overall charge of the system. Periodic boundary conditions were applied along the three spatial directions to prevent edge effects. The system was energy minimized with a steepest descent method for 200 steps.

Initial atomic velocities were extracted from a Maxwell-Boltzmann distribution corresponding to 250 K and, subsequently, temperature was increased up to 300 K in 50 ps. The temperature was controlled by using a velocity rescaling thermostat [Bussi et al., 2007], with coupling constant 0.1 ps. The Berendsen barostat was used to control the pressure [Berendsen et al., 1984], with reference pressure 10^5 Pa and coupling constant 1 ps. The

particle-mesh Ewald (PME) method was used for computing the electrostatic interactions [Darden et al., 1993; Essmann et al., 1995]. Bond distances were constrained by using the P-LINCS algorithm [Hess, 2008] and a time step of 2 fs was used to integrate the equations of motion. The production runs were carried out for 30 ns.

A single simulation run was performed for βLG complexed with each of the six fatty acids considered, whereas six runs starting from different atomic velocities were carried out for the unliganded protein. Results for both sets of simulations, either on the liganded or unliganded protein, were averaged and directly compared to estimate the range of variability obtained in the two cases.

2.2.3. Data analysis

The equilibration of the protein structures was evaluated by analyzing the radius of gyration (R_g) and the root mean square deviations (RMSD) and fluctuations (RMSF) of atomic positions. The radius of gyration is calculated as:

$$R_g = \left(\frac{\sum_{i=1}^{N} m_i r_i^2}{\sum_{i=1}^{N} m_i} \right)^{1/2} \qquad (\mathbf{2.1})$$

where m_i and r_i are the mass for each atom and distance between it and the center of mass of the molecule, respectively. The atomic RMSD and RMSF are calculated for C^{α} atoms as:

$$\Delta R(t) = \left(\frac{1}{N} \sum_{i=1}^{N} \langle \Delta r_i^2(t) \rangle \right)^{1/2} \qquad (\mathbf{2.2})$$

where N is the number of C^{α} atoms of the protein and Δr_i is the difference between the instantaneous and starting atomic positions of the i-th atom in the case of the RMSD, and of the instantaneous and average position in the case of the RMSF. The angular brackets $\langle ... \rangle$ indicate the time average on the simulation.

In the calculation of both atomic RMSD and RMSF, protein rototranslation was eliminated by a mass-weighted least squares fit with respect to the reference starting structure.

To analyze the collective motion of protein residues, correlated fluctuations were calculated by using the Essential Dynamics technique [Garcìa, 1992; Amadei et al., 1993]. This technique is based on the principal component analysis of protein inner motions during the simulation. The correlation between atomic motions can be expressed by using the covariance matrix C_{ij} of the positional deviations:

$$C_{ij} = < (X_i - <X_i>) (X_j - <X_j>) > \qquad (\mathbf{2.3})$$

where X are the three-dimensional coordinates of the C^{α} atoms i and j, and the angular brackets indicate a time average on the simulation trajectory. Orthogonal eigenvectors, representing collective modes of fluctuation, and eigenvalues, corresponding to their mean square values, are obtained from diagonalization of the covariance matrix. The first eigenvectors represent the directions with the largest positional deviations, and most of the atomic fluctuations take place in an 'essential space' spanned by the first few eigenvectors.

2.3. Results

2.3.1. Protein dynamics in the presence of a fatty acid

The dynamics of unliganded βLG was compared with the behavior of the protein in the presence of a single fatty acid molecule in the protein calyx, either caprylic, capric, lauric, myristic, palmitic or stearic acid (C8:0 to C18:0). The overall stability of the protein structure was monitored by inspecting R_g and the deviations of the C^{α} atoms with respect to the starting structure, both as a function of the simulation time. The R_g (Fig. 2.2) and RMSD (Fig. 2.3) values were obtained by using an adjacent-averaging on 500 points of the data.

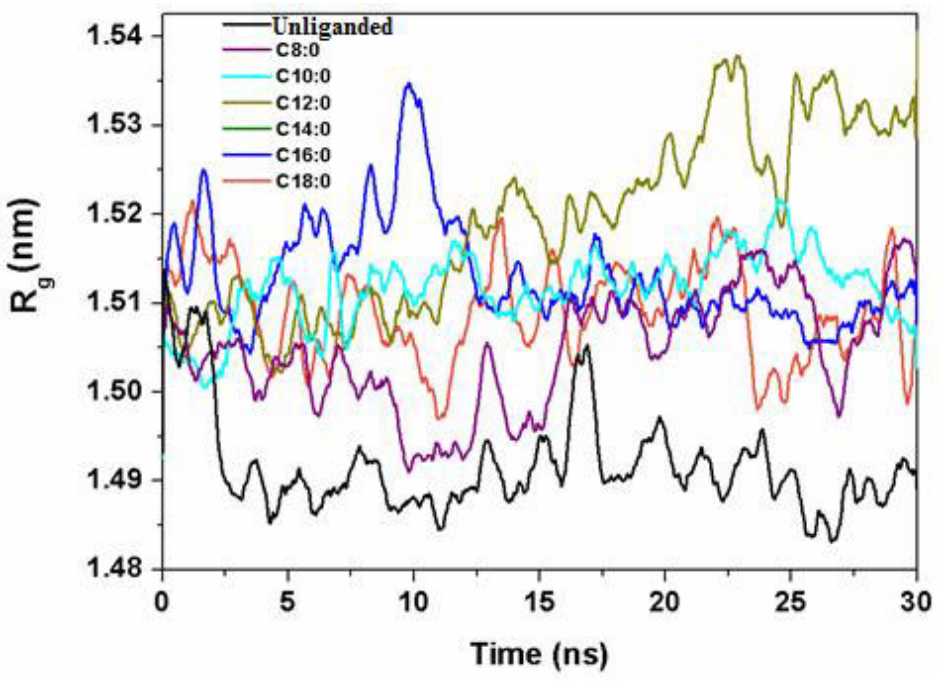


Fig. 2.2 *–Radius of gyration (R_g) with respect the time, for the unliganded and complexed βLG.*

For the unliganded βLG, an equilibrium value of 1.49 ± 0.01 nm for R_g and 0.20 ± 0.02 nm for atomic RMSD of the backbone was obtained. Similar values were found for the unliganded protein and in the presence of each different fatty acid.

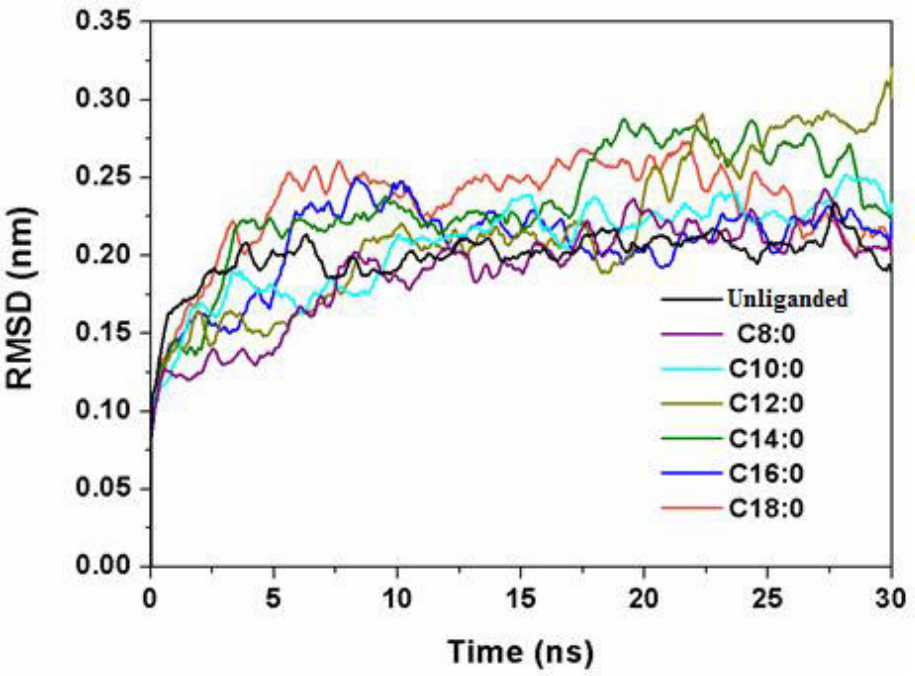


Fig. 2.3 *- Root mean square deviations (RMSD) of C^{α} atoms with respect the time, for the unliganded and complexed form βLG.*

This indicates that ligand association has little effect on the global structure of βLG, according with experimental data [Loch et al., 2012] showing a well-defined pre-formed site that does not require major conformational modifications upon fatty acid binding.

Information on the local dynamics of the main chain of βLG can be derived from the analysis of atomic RMSF as a function of residue index. In Figure 2.4, the fluctuations averaged for each element of secondary structure are reported. In particular, RMSF values obtained in multiple runs for unliganded βLG are compared with the atomic fluctuations found by combining the simulation data obtained for the protein complexed with the fatty

acids. This comparison allows to capture the main differences in the dynamics of the protein structure due to the presence of fatty acids in the binding site, regardless of the length of their hydrocarbon chain. A general increase of protein fluctuations upon binding of the ligands is evident, and major differences with respect to the unliganded form are in correspondence of loops AB, CD, EF, and in the C-terminal region of the protein backbone.

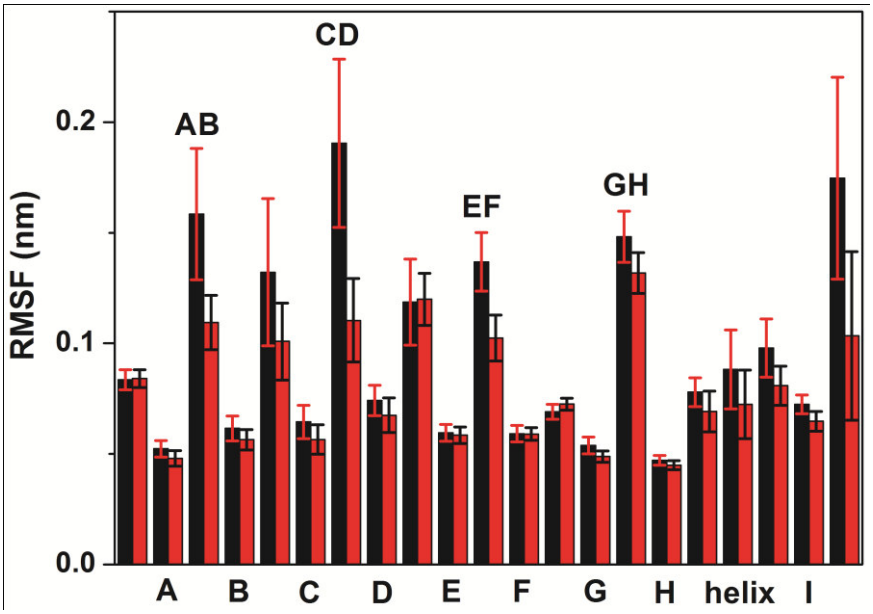


***Fig. 2.4** – Positional RMSF of the protein C^{α} atoms, averaged for each element of secondary structure, for (black) liganded and (red) unliganded βLG. Single letters indicate the protein β-strands, double letters indicate loops at the opening of the protein calyx.*

This results is interesting because the mobility of the superficial loops is an important requirement for a transport protein involved in a molecular complex [Jameson et al., 2002]. The AB, CD and EF loops are all located at the entrance of the protein calyx, thus fluctuations can be related to the biological function of ligand binding to βLG. In fact, these loops are involved in the ligand penetration process, as shown in MD simulations [Bello et al., 2012] of βLG complexed with either lauric or palmitic acid (C12:0 and C16:0, respectively). In

addition, the flexibility of these loops constitutes an intrinsic determinant of the conformational stability to maintain the large internal hydrophobic surface of the calyx, which is an empty dry cavity [Jameson et al., 2002; Qvist et al., 2008]. Moreover, the pH dependent conformation transition of the EF loop (Tanford transition) determines the ligand accessibility to the binding pocket [Qin et al., 1998], whereas the CD loop contains charged residues involved in the anchoring of the fatty acid head-group.

An increased mobility of the binding site loops was previously found in the comparative NMR study of ligand-free and palmitate-complexed βLG [Konuma et al., 2007]. Our MD results agree with this finding and additionally suggest that the same behavior should be expected for fatty acids of different length. High fluctuations in the C-terminal end of βLG can be easily explained by the high mobility that is often found in the terminal regions of a protein backbone, and are also in agreement with experiment [Konuma et al., 2007]. In contrast, our results are at variance with a recent MD simulation on βLG complexed with lauric acid and dodecyl sulfate, reporting a higher flexibility of the apo form of the protein compared to the protein-ligand complexes [Bello et al., 2012].

The conventional analysis of atomic RMSF does not allow to determine how fluctuations take place along the global degrees of freedom of the molecule and how residue displacements coordinate with each other. Thus, the Essential Dynamics technique [Garcìa, 1992; Amadei et al., 1993] was used to gain additional insights into the dynamics of the cooperative inner motions in the protein structure in the presence of a fatty acid. Figure 2.5 shows the atomic fluctuations derived from the eigenvalues that correspond to the first few eigenvectors in the conformational space of the protein, with uncertainty bars indicating the variability interval due to the presence of fatty acids of different length.

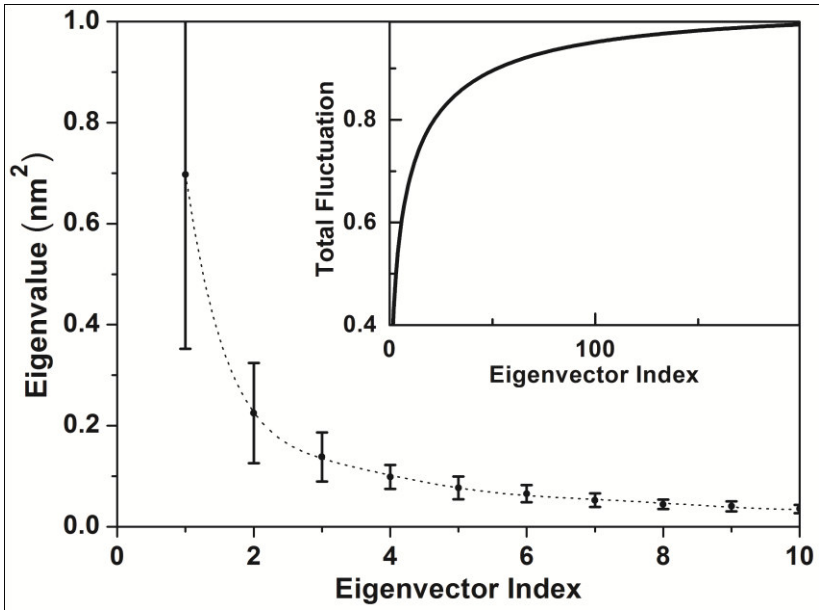


Fig. 2.5 *– Eigenvalues as a function of the first 10 eigenvectors, averaged over simulations of βLG in complex with fatty acids ranging from caprylic to stearic acid (C8:0 to C18:0). Error bars indicate standard deviations over the simulations, the dotted line is for eye guidance. Inset: total fluctuation as a function of the eigenvector index, for the first 200 eigenvectors.*

The curve indicates that the approximately 10 eigenvectors are sufficient to describe the most relevant coordinated inner motions of the protein, whereas fluctuations along the successive principal components level off rapidly. The first 2 and 3 eigenvectors contribute to, respectively, 43% and 50% of the total fluctuations of the protein, and the first 10 are sufficient to account for 69% (see inset of Fig. 2.5). After a few eigenvectors, each corresponding eigenvalue has a magnitude almost negligible in term of its individual contribution to the total fluctuations, and is also indistinguishable with respect to the closest ones (within their range of variability). These results indicate that βLG possess only a few preferential degrees of freedom that can be filtered out from the other less relevant ones.

The findings here obtained are in basic agreement with previous MD simulations of fatty acid-complexed βLG [Eberini et al., 2004; Bello et al., 2012], taking into account that

the simulation length, force field and many other simulation conditions in these studies are different. In the present analysis, compared to the previous ones, we also estimate the range of variability of atomic RMSF, which gives a measure of the structural plasticity of βLG upon association of the diverse ligands in the protein binding site. In fact, the results show that differences in (mean squared) fluctuations up to 0.5 nm^2 can be expected along the most flexible degree of freedom of the protein, constituting the first principal component.

Figure 2.6 shows the atomic RMSF of the protein backbone as a function of the residue index for the fatty acid-complexed βLG along the first three eigenvectors, as determined in the principal component analysis. Again, the fluctuation values reported are obtained by averaging the data for simulations with different chain length. Excluding the C-terminus of the protein, the highest fluctuations are found in correspondence of the protein loops AB, CD, EF and GH, all located at the entrance of the protein calyx. This behavior is particularly evident along the first eigenvector and still clearly noticeable along the second eigenvector. These results demonstrate the presence of coordinated motions among the loops that give access to the binding location, independently of the length of the fatty acid associated. Fluctuations along the successive eigenvectors are progressively lower and do not show clear differences among loops, as it is visible for the third eigenvector (see Fig. 2.6, lower panel).

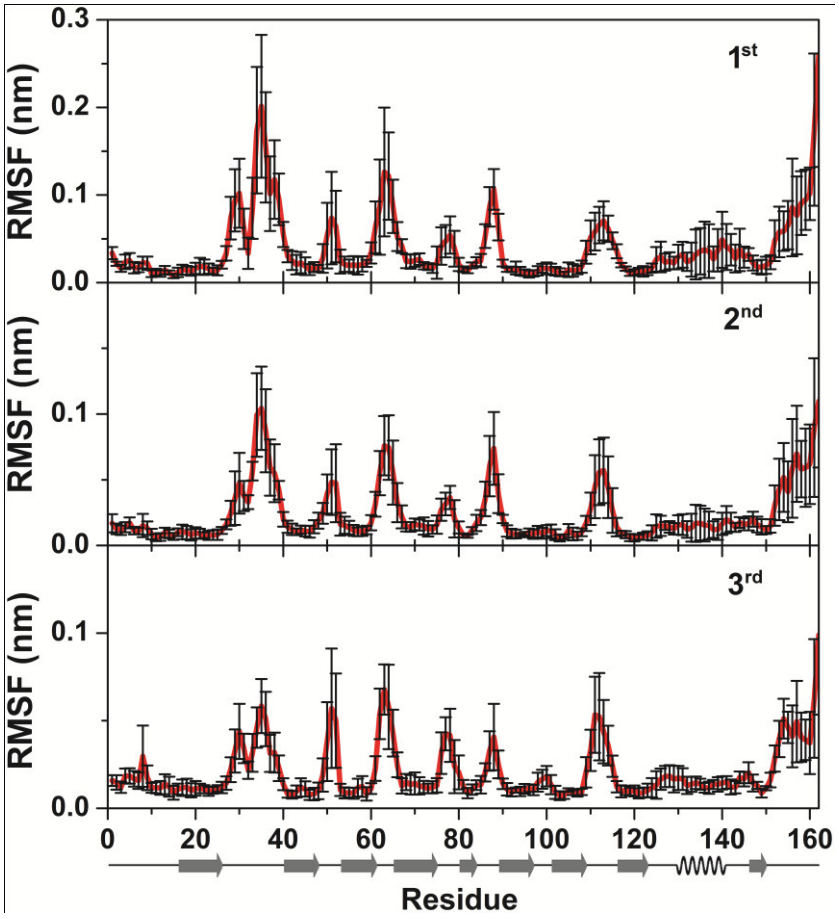


Fig. 2.6 *– Positional fluctuations of C^{α} atoms of βLG complexed with fatty acids, obtained along the first three principal components, as a function of residue number.*

It is interesting to compare our results with those recently reported by Bello and coworkers on βLG complexed with dodecyl sulfate and laurate [Bello et al., 2012]. Their simulations showed that protein fluctuations are lower in the presence of these two ligands, in contrast with NMR results [Konuma et al., 2007]. Moreover, differences in RMSF values were found in the two protein-ligand complexes, and concentrated on loops both at the entrance and on the opposite side of the binding site. Dissimilarities with our simulations can be reconciled by noting that the range of variability of the calculated RMSF values is generally significant (see error bars in Fig. 2.5 and 2.6), up to 0.15 nm. Therefore, differences

in fluctuations may become meaningful only upon averaging on several trajectories. Conversely, the direction of cooperative inner motions of the protein we found are similar to the ones previously reported [Eberini et al., 2004; Bello et al., 2012]. In particular, correlated motions in fatty acid-complexed βLG tend to displace the loops that regulate the access to the binding site in and out along a radial direction (data not shown).

2.3.2. Fatty acids: anchoring to the protein and dynamics within the calyx

It is interesting to focus on how cooperative inner motions that characterize the βLG structure, especially at the entrance of the protein calyx, can contribute to anchor the fatty acids within the protein binding site. In addition, a thorough investigation of the dynamics of the fatty acids within the binding site is complementary to the study of the dynamics of the surrounding protein matrix.

As a first step in these directions, it is noteworthy to investigate the displacements of the CD loop, which contains the residues Lys60 and Glu62, and also influences the dynamics of the nearby Lys69 located in the β-strand D. All these charged protein residues are involved in HBs and electrostatic interactions with the carboxylate group of the fatty acid, contributing to secure the ligand into the binding site of βLG. Such interactions, especially attraction with the two Lys residues, are believed to play a distinctive role in determining the degree of penetration of the fatty acids into the protein calyx, as found in crystallography [Loch et al., 2011; Loch et al., 2012]. On the other hand, comparison of the fluctuations detected in simulation for this loop indicate that it has the highest mobility overall (see Fig. 2.4), but not in terms of correlated motions (Fig. 2.6).

The simulation data show that the carboxylate group of each fatty acid is able to rotate around the C^1–C^2 axis and, consequently, the two oxygen atoms can swap their position

several times over a sub-ns timescale. This behavior is not unusual for fatty acids, for instance it was previously described for palmitic acid in the highest affinity binding site of human serum albumin [Rizzuti et al., 2010]. To account for this effect, to determine the formation of a bond between each fatty acids and the side chain of Lys60 and Lys69, the distances between the C^1 atom of the lipid head-group and either N^{ξ}-Lys60/Lys69 (and also C^{δ}-Glu62) were evaluated. Both simulated and crystallographic distances are summarized in Table 2.1.

Table 2.1 *– Distances between the carboxylate group of each fatty acid and Lys60, Glu62 and Lys69. Crystallographic results are from Loch and coworkers [Loch et al., 2011; Loch et al., 2012], except [(a)] from a previous determination of the βLG-palmitate complex [Kontopidis et al., 2002] and [(b)] referring to 12-bromododecanoic acid [Qin et al., 1998].*

	N^{ξ}-Lys60 vs. C^1 distance (nm)		N^{ξ}-Lys69 vs. C^1 distance (nm)		C^{δ}-Glu62 vs. C^1 distance (nm)	
	Simulation	Crystallography	Simulation	Crystallography	Simulation	Crystallography
C18:0 stearic acid	0.58 ± 0.13	0.57 ± 0.11	1.02 ± 0.25	0.44 ± 0.11	0.93 ± 0.16	0.54 ± 0.11
C16:0 palmitic acid	0.51 ± 0.12	0.60 ± 0.10 0.54 ± 0.12[(a)]	0.86 ± 0.31	0.39 ± 0.10 0.45 ± 0.12[(a)]	0.88 ± 0.23	0.50 ± 0.10 0.62 ± 0.12[(a)]
C14:0 myristic acid	0.58 ± 0.15	0.73 ± 0.10	1.08 ± 0.26	0.32 ± 0.10	0.99 ± 0.15	0.66 ± 0.10
C12:0 lauric acid	0.50 ± 0.15	0.73 ± 0.11 0.35 ± 0.11[(b)]	0.61 ± 0.25	0.72 ± 0.11 0.58 ± 0.11[(b)]	0.98 ± 0.20	0.87 ± 0.11 0.69 ± 0.11[(b)]
C10:0 capric acid	0.45 ± 0.12	0.87 ± 0.10	0.74 ± 0.26	0.87 ± 0.10	0.90 ± 0.18	1.03 ± 0.10
C8:0 caprylic acid	0.52 ± 0.15	0.68 ± 0.10	0.77 ± 0.30	0.50 ± 0.10	0.90 ± 0.16	0.50 ± 0.10

Crystallographic data are here reported with a range of uncertainty corresponding to the maximum resolution in the structure determination, which varies from 1.9 to 2.35 Å [Kontopidis et al., 2002; Loch et al., 2011; Loch et al., 2012]. This easily overestimates these

values for each single X-ray structure, but it is reasonable to take into account the variations due to conformational changes in the crystal formation process [Rashin et al., 2009]. For instance, in the case of palmitic acid two different X-ray structures are available [Kontopidis et al., 2002; Loch et al., 2012] and show discrepancies up to 0.12 nm in the distance between the C^1 atom of the lipid head-group and C^{δ}-Glu62. Simulated distances show uncertainties of the same magnitude of crystallographic values, but in this case the range of variation is due to the actual dynamics of both the protein loops and the ligands.

In crystallography, the distances between the ligand head-group and such key protein residues depend on the length of the hydrocarbon chain [Loch et al., 2011; Loch et al., 2012]. Both palmitic and stearic acid form HBs with Lys60, Glu62 and Lys69, which contribute to maintain the fatty acid head-group close to the calyx entrance. In contrast, shorter fatty acids penetrate 2-3 methylene groups deeper into the binding site. In particular, myristic acid forms a HB only with Lys69 and lauric acid does not form any HBs, although the 12-bromododecanoic acid (which is identical to lauric acid, with an additional Br atom at the end of the tail) forms a HB with Lys60 [Qin et al., 1998]. In simulation, the HB between the fatty acid head-group and Lys60 is present for all the fatty acids considered. Coordination with Lys69 and Glu62 is more flexible and distances are generally higher compared to the corresponding crystallographic values, thus the presence of HBs depends on the criteria (such as cut-off distance and permanence time) chosen to assess their formation. However, electrostatic interactions between all these residues and the fatty acid head-group is always present, even for a distance longer than the crystallographic values and above the one typical for a HB, and clearly contribute to fasten the ligand into the binding site.

The dynamics of the fatty acids within the calyx of βLG can be further investigated by considering the atomic fluctuations of their methylene groups along the aliphatic chain. To

this aim, the motion of a fatty acid molecule can be decomposed into two distinct components. The main one corresponds to the relative motion of the entire fatty acid molecule within the binding cavity, which includes a combination of both longitudinal and lateral displacements with respect to the protein calyx. The other component is due to internal vibrations of the hydrocarbon tail and, compared to the former, represents a smaller superimposed dynamical contribution that determines an uncertainty on the displacement of the ligand within the binding site. In this description, the ideal case of a perfectly rigid ligand free to move in a cavity would result in relatively large RMSF values with zero uncertainty.

Both components are reported in Figure 2.7 for each fatty acid, as a function of the position of carbon atoms along the chain. The RMSF values are obtained by calculating the fluctuation of the fatty acid molecule after removing the rototranslation of the protein structure, by superimposing the protein-ligand complex to the reference (average) structure, thus they indicate how the chain of the ligand fluctuates within the binding site. The error bars on the RMSF values are obtained by calculating the atomic fluctuations after removing the rototranslation of the fatty acid alone, without considering the presence of the surrounding protein structure, thus they are a measure of the intrinsic vibrations along the hydrocarbon chain.

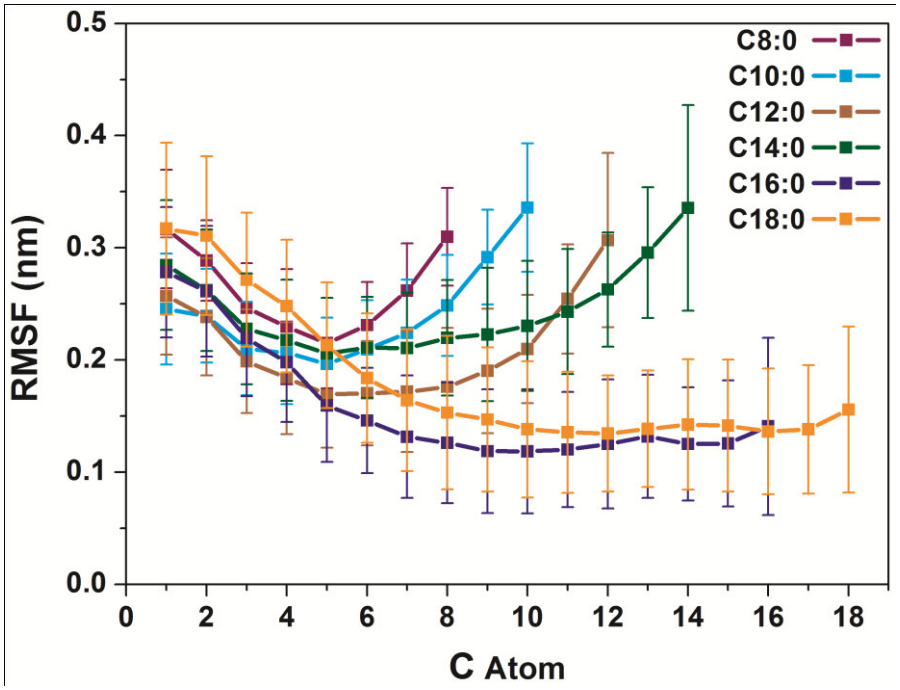


Fig. 2.7 *– Positional RMSF along the aliphatic chain of fatty acids, ranging from caprylic to stearic acid (C8:0 to C18:0), within the βLG calyx.*

The curves in Figure 2.7 indicate that, for all the fatty acids, the fluctuations are not uniform along the aliphatic chain. Fluctuations of the C^1 atom are within 0.25-0.30 nm and with an uncertainty of 0.05 nm for all the fatty acids considered. This indicates that the anchoring of the carboxylate group is identical for each fatty acid, in spite of the variability reported in Table 2.1 for both simulated and experimental distance values. In particular, this strongly suggests that differences in the depth of the fatty acid position within the calyx found in crystallography do not determine a difference in the dynamics of the ligand head-group, and may be induced by the crystallization process. In contrast, fluctuations markedly differ in the terminal portion of the tail, depending on the length of the fatty acid. RMSF values for palmitic and stearic acid are similar, 0.15 ± 0.05 nm, and considerably lower compared to the other fatty acids, which are all above 0.3 nm. All the RMSF curves are similar up to the C^5

atom, whereas proceeding further along the chain the fluctuations they either stabilize (for palmitic and stearic acid) or increase (for all the other fatty acids).

It is interesting to investigate which protein residues are the main determinant for this behavior. In this respect, two key residues are Phe105 and Met107, as shown in Figure 2.8. In particular, Met107 is located below the entrance of the binding site and provides a blocking mechanism for any long-chain fatty acid. In fact, the flexible side chain of Met107 contributes to confine the fatty acid against Val41, which is placed slightly upper on the opposite side of the protein calyx, and determines a constraint clamping the lipid chain in correspondence with (approximately) the C^5 atom. Deeper within the protein calyx, Phe105 provides an additional selective constraint. In fact, the side chain of Phe105 can assume either an 'open' conformation, when the aromatic ring is roughly parallel to the binding cavity, or a 'closed' conformation when the ring is perpendicular. For both palmitic and stearic acid, Phe105 has the 'open' conformation that maximizes the hydrophobic interactions with the ligand, securing the terminal portion of the lipid tail. In contrast, fatty acids with less than 16 carbon atoms are too short and this blocking mechanism does not operate on them.

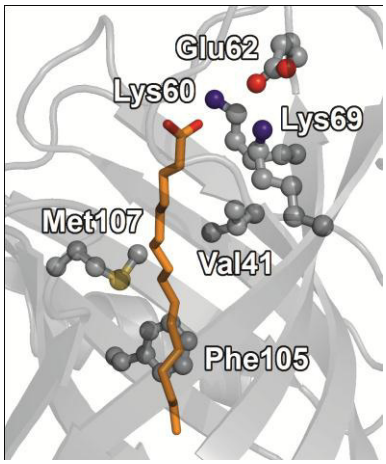


Fig. 2.8 – *Key protein residues in the anchoring of a palmitic acid molecule in the βLG calyx.*

These observations provide an explanation for a wealth of NMR [Ragona et al., 2000; Konuma et al., 2007; Sakurai et al., 2009] and crystallographic studies [Wu et al., 1999; Kontopidis et al., 2002; Loch et al., 2011; Loch et al., 2012], indicating that the side chains of Phe105 and Met107 change their conformation upon ligand binding, whereas all the rest of the core lipocalin structure is invariant [Edwards et al., 2009]. The simulation results point out that the role of these two residues is to regulate access to the binding site and assist the fastening of the chain of the ligand. In addition, fluorescence [Frapin et al., 1993] and isothermal calorimetric studies [Loch et al., 2012] show that the binding affinity increases with the length of the hydrocarbon chain and, in particular, the highest affinity is found for palmitic and stearic acid. On the basis of the simulation data, it is possible to suggest that the binding affinity of fatty acids to βLG is related to fluctuations of the molecule tail, with fatty acids that fluctuate the less having a higher binding affinity. In particular, Phe105 seems to play a major role in this respect. This residue also gives the higher contribution to the total binding free energy in MD simulations of long-chain fatty acids [Bello and Garcìa-Hernàndez, 2014; Bello, 2014].

2.3.3. Secondary binding sites for palmitic acid

The existence of one or more secondary binding sites for fatty acids in βLG was suggested on the basis of spectroscopic studies on the protein in solution [Narayan and Berliner, 1998; Wang et al., 1998; Yang et al., 2008], but never eventually revealed in crystallography [Kontopidis et al., 2002; Loch et al., 2011; Loch et al., 2012]. To explore this possibility, molecular docking was used to examine the binding of a fatty acid molecule to an external site in βLG. In particular, interaction of βLG with palmitic acid was considered, because it constitutes the main fraction of the lipid-protein complex in isolated βLG [Barbiroli

et al., 2011] and experimentally shows a binding affinity greater than the other fatty acids [Frapin et al., 1993; Loch et al., 2012].

In a first step, molecular docking was used to verify the possibility of reproducing the βLG-palmitate complex already identified in crystallography [Loch et al., 2012]. To this aim, the structure of βLG was extracted from the complex and a palmitate molecule was docked back to it. The first ten most likely energetic configurations obtained for the fatty acid are all within the protein calyx, with an estimated binding free energy of −6.0 ± 0.1 kcal/mol for the first pose. This confirms the reliability of molecular docking in detecting the hydrophobic calyx as the optimal binding site for fatty acids to βLG, in agreement with experimental data [Kontopidis et al., 2002; Loch et al., 2012].

Subsequently, it was examined whether the docking technique is able to reveal a secondary binding site, located in a position other than the protein calyx. Docking of vitamin D_3 to βLG was considered as a preliminary test case, because this ligand is the only one for which there is both crystallographic [Yang et al., 2008] and *in vivo* [Yang et al., 2009] evidence of an additional binding site distinct from the calyx. When the protein calyx is already occupied, docking of a vitamin D_3 molecule results in poses on the surface of βLG located in three distinct regions: pose 1 in between β-strand B and the helical turn 153-157, pose 2 in between β-strand I and the α-helix, and pose 3 close to the entrance of the protein calyx. The binding energies in these three poses are reported in Table 2.2.

Table 2.2 – *Binding energy of vitamin D_3 and palmitic acid associated to the external surface of βLG, obtained by molecular docking when the main binding site is already occupied. The locations are shown in Figure 2.9.*

	Binding energy (kcal/mol)	
Pose	*Vitamin D_3*	*Palmitic acid*
1	−6.5 ± 0.4	−4.1 ± 0.5
2	−6.3 ± 0.3	−3.4 ± 0.2
3	−6.4 ± 0.1	−3.4 ± 0.3

It is interesting to note that the affinity calculated for the docking of vitamin D_3 into the external site determined in crystallography [Yang et al., 2008], corresponding to pose 2, is close to the value obtained for the docking of palmitic acid in the calyx. This is in agreement with experiments [Yang et al., 2008] showing that the binding affinity for the two ligands is similar in these two locations. On the other hand, the values predicted for vitamin D_3 in the three poses are very close to each other and within the variability of the computational procedure, the binding energy for pose 2 being even slightly higher compared to the other two. These findings reveal that the docking procedure can correctly identify protein regions that are candidates for being secondary binding sites, but it may not be able to further discriminate among different alternatives.

Docking of a second palmitic acid molecule, performed to βLG already containing one in the calyx, resulted in several binding modes on the protein surface, as shown in Figure 2.9. The poses correspond to the same positions already found for docking of vitamin D_3 to βLG. In particular, in pose 1 palmitic acid has two binding modes with an approximately opposite orientation, binding either tail-first or head-first towards βLG. Similarly, pose 2 corresponds

to two slightly distinct binding modes. Finally, pose 3 corresponds to the fatty acid placed above the protein calyx, which is already occupied by the other palmitate molecule. The binding energies for palmitic acid in the three poses can be directly compared with those found for vitamin D_3, and are also reported in Table 2.2. The values obtained for the two ligands for pose 2, corresponding to the external binding site of vitamin D_3, indicate that the binding energy for palmitic acid is considerably higher. Affinity of palmitic acid in pose 1 is greater than in pose 2, but still far from being comparable to the affinity of vitamin D_3 in any pose external to the calyx.

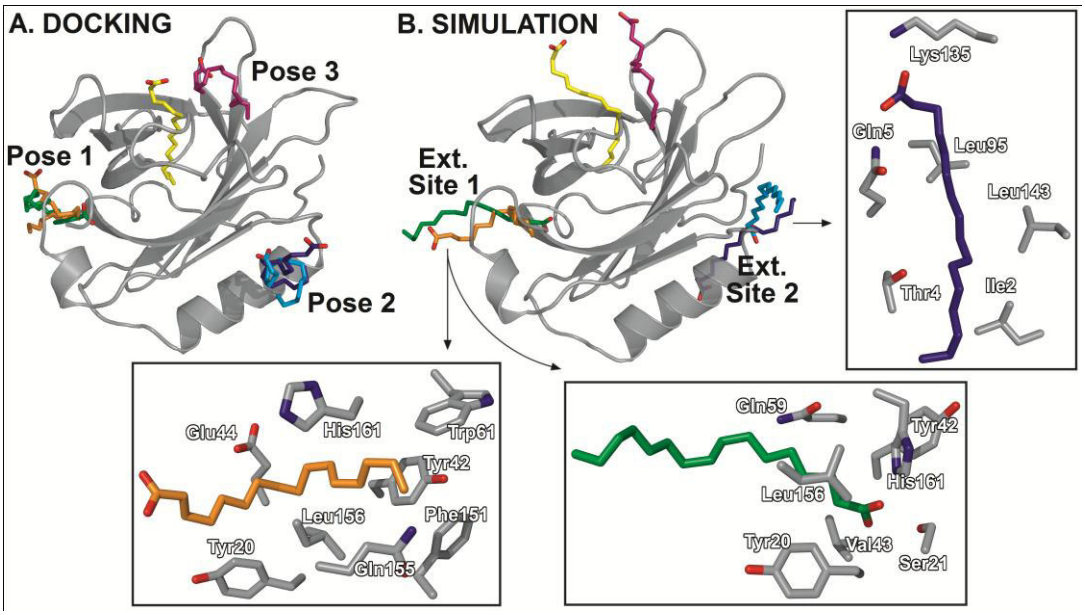


Fig. 2.9 *– Interaction of palmitic acid with βLG: (A) poses obtained by molecular docking with the crystallographic protein structure and (B) external binding sites found in simulation. The ligand in yellow corresponds to the fatty acid occupying the protein calyx, and is always present together with a second palmitate (either the one in green, orange, cyan, blue, or purple). Residues interacting with the fatty acid in the external binding site 1 (two binding modes) and site 2 are shown in framed panels.*

As noted above, on the basis of the docking calculations alone it is difficult to discriminate whether the poses found for palmitic acid correspond to non-specific sites or to genuine secondary binding sites, although with low specificity compared to the main site in the protein calyx. In particular, molecular docking does not take into account the dynamics of the protein-ligand complex, thus it cannot identify long-lived interactions, especially if accompanied by local modifications of the protein structure that helps to stabilize them. For this reason, the five different βLG-palmitic acid complexes generated in the docking procedure were used as starting structures to perform MD simulations. Each simulation was performed on βLG associated with two fatty acid molecules: one located into the protein calyx, and the other in each of the binding modes previously detected. The final positions of each palmitate molecule at the end of the MD simulations are reported in Figure 2.9.

In simulations performed with the palmitate molecule placed in pose 1, the fatty acid molecule penetrates the protein surface either tail-first or head-first (Fig. 2.9, panels with palmitic acid represented in orange and green, respectively), in accordance with its starting configuration. In both cases the hydrocarbon chain accessed the small protein pocket in correspondence of residues Tyr20, Glu44, Glu157 and Glu158, in the same binding site identified by molecular docking for epigallocatechin-3-gallate [Wu et al., 2013]. In this location the fatty acid further advanced towards the protein interior, favored by local displacement of amino acid residues resulting from βLG dynamics, in analogy with residue-assisted ligand penetration into the protein calyx [Bello and Garcìa-Hernàndez, 2014]. In the binding mode tail-first the palmitate inserts with about 10 carbon atoms and secures its tail by hydrophobic interactions with aromatic residues Tyr42, Trp61 and Phe151. In the binding mode head-first the fatty acid penetrates up to the C^6 atom and attaches its head-group to the protein interior through a strong HB with the amide nitrogen atom of Ser21 (bond length 0.22

± 0.03 nm) and two weaker HB with O^{γ}-Ser21 and N^{ε}-His161 (bond length 0.29 ± 0.03 and 0.28 ± 0.03 nm, respectively).

When the ligand is placed in pose 2, in between β-strand I and the α-helix, during the simulations the fatty acid migrates from its starting position. The palmitate found a stable location on the other side of the C-terminal region of the α-helix, either inserted into the protein crevice formed with the terminal portion of β-strand F (Fig. 2.9, palmitic acid represented in blue) or more exposed to the solvent (Fig. 2.9, cyan). Interestingly, the ending position after the simulation is compatible with the dimeric structure of the protein, whereas the starting pose is at the protein-protein interface. The fatty acid in the first binding position adopts an extended conformation, with the carboxylate group forming a HB with the side chains of either Lys135 or Lys138, and the rest of the hydrocarbon chain interacts with various protein residues. Both the starting and ending position of the palmitic acid molecule in this simulation correspond to two potential binding sites found with a different docking procedure [Yang et al., 2008]. In the other case (Fig. 2.9, colored in cyan), the fatty acid head-group forms a permanent HB with the protein main chain in correspondence with the amide nitrogen atom of Ala142 (bond length 0.30 ± 0.02 nm), whereas the tail interacts with the imidazole ring of His146.

Finally, a simulation performed with the external palmitate molecule located above the hydrophobic calyx (pose 3) provided details on the competition of two ligands for the same binding site [Collini et al., 2003]. In fact, after initial exploration of the entrance of the protein calyx, the external palmitate molecule penetrated into the upper region of the hydrophobic channel with the 5 ending carbon atoms of the tail. Simultaneously, the other palmitate molecule partially dissociated from the binding site, remaining with only the 8 ending carbon atoms into the channel. This coexistence at the entrance of the binding site (Fig. 2.9,

molecules represented in purple and yellow) is an unstable configuration that was observed for several ns, and ended up with one molecule returning into the pocket while the other continued to explore the region above the protein calyx.

In conclusion, MD simulation suggests additional sites for a stable binding of palmitic acid to βLG, outside the protein calyx. These locations do not necessarily correspond to docking poses, as in the case of simulations starting from pose 2, or involve a significant re-adaptation of the protein surface, for pose 1. For these sites is possible to estimate the binding affinity of the palmitic acid. The procedure consists in taking the structure of the complex obtained in MD simulation, extracting the palmitate molecule, and then dock it back into the corresponding binding pocket of βLG. The values of the binding energies are reported in Table 2.3.

Table 2.3 *– Binding energy of palmitic acid associated to the external surface of βLG, obtained by re-docking the ligand in ending location found by MD simulations. The positions are shown in Figure 4.9.*

starting location	**ending location**	**binding mode**	**binding energy (kcal/mol)**
pose 1	pose 1	tail-first	−5.5 ± 0.2
pose 1	pose 1	head-first	−5.2 ± 0.2
pose 2	α-helix C-terminus	crevice-inserted	−5.0 ± 0.3
pose 2	α-helix C-terminus	solvent-exposed	−3.8 ± 0.1

The binding affinity in three cases are ≤ –5.0 kcal/mol, a value comparable with the corresponding one for the binding of palmitic acid in the calyx, –6.0 kcal/mol. In contrast, the

energy value −3.8 ± 0.1 kcal/mol indicates a binding affinity similar to other non-specific regions on the βLG surface. Thus, there are two protein locations that can be considered secondary binding sites for palmitic acid, one of which with two distinct ligand insertion modes. A feature of these external sites, which is also in common with the primary binding site, is that the fatty acid is found in a fully extended conformation. In addition, these sites are compatible with the dimeric conformation of βLG at pH 7. In these locations the affinity is lower compared to the main binding site, and probably this is the reason why they are only probed by using spectroscopic techniques in solution [Kontopidis et al., 2004] and not found in crystallography [Kontopidis et al., 2002; Loch et al., 2012].

3. Absolute free energy calculations of HSA complexed with ibuprofen

3.1. Structure of HSA

HSA is the most abundant protein in the human blood plasma (30-50 g/L) and represents ~60% of the total plasma proteins. It is also present in cellular tissues, chyle, aqueous and vitreous humor, lymph, synovial and cerebrospinal fluid [Rothschild et al., 1988]. HSA is an important carrier of fatty acids [Carter and Ho, 1994; Cistola, 1998] and other endogenous substances such as bile salts, bilirubin, hematin, metal ions, steroid hormones, tryptophan, thyroxine and vitamins. In addition, HSA is also able to bind numerous drugs such as ibuprofen, warfarin and diazepam [Krag Hansen, 1990; Carter and Ho, 1994; Peters, 1996; Henrik, 1999]. The interaction HSA-drug is significant for the pharmacokinetic properties of a range of pharmaceutical compounds [Peters, 1996].

HSA is a monomeric protein formed by 585 amino acids with a complex three-dimensional structure. It has three homologous α-helix domains (I, II, III), each comprising two sub-domains (A, B) arranged in an overall heart-shaped structure (Fig. 3.1) [Carter et al., 1989; Carter and Ho, 1994].

Each domain (I, II, III) [He and Carter, 1992; Carter and Ho, 1994], is composed of ten α-helices, six for the A and four for the B domain. The following is the amino acid index for each domain: IA (1-107), IB (108-195), IIA (196-297), IIB (298-383), IIIA (384-497) and IIIB (498-585) [Sugio et al., 1999]. The protein is stabilized by 17 disulphide bonds distributed in all of the domains. SS bonds do not link different domains of the protein, only subdomains [Sugio et al., 1999]

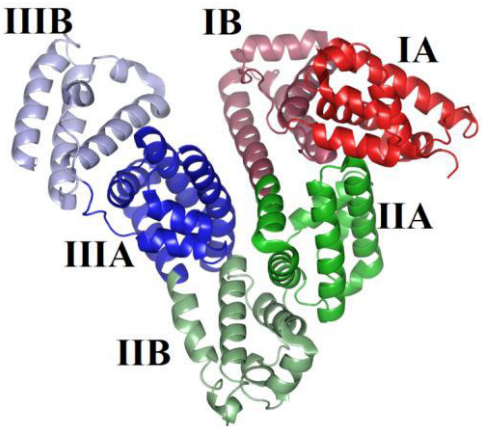


Fig. 3.1 *– Crystallographic structure [Ghuman et al., 2005] of HSA. Domains and subdomains are represented in different colors: IA (red), IB (light red), IIA (green), IIB (light green), IIIA (blue) and IIIB (light blue).*

The binding of fatty acids to HSA causes conformational changes into the protein as revealed in the crystallographic structures [Curry et al., 1998; Bhattacharya et al., 2000; Petitpas et al., 2001]. Structurally, these changes regard 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 II and III [Cuya Guizado, 2014].

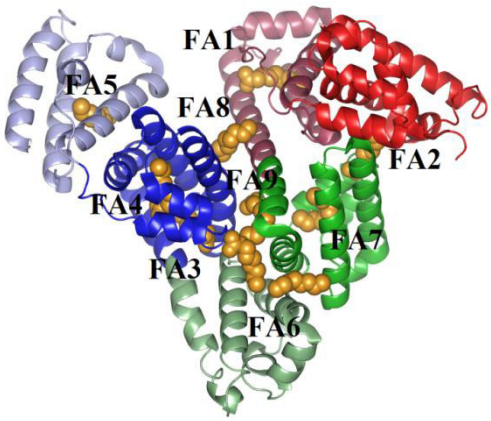


Fig. 3.2 **–** Structure of HSA complexed with capric acid (C10:0). The fatty acid is shown in sphere representation.

Long chain fatty acids such as palmitic and stearic acid (C16:0 and C18:0, respectively) can bind to HSA in seven binding sites [Bhattacharya et al., 2000] named FA1,..., FA7. In particular, FA2 is located at the interface IA-IIA, FA1 in the subdomain IB, FA7 in the subdomain IIA, FA6 at the interface of subdomains IIA-IIB, FA3 and FA4 in the subdomain IIIA, FA5 in the subdomain IIB. Shorter fatty acids such as lauric and myristic (C12:0 and C14:0, respectively) can bind a second molecule in the site FA2 [Curry 1998; Curry 1999; Bhattacharya et al., 2000], thus they possess also the site FA2′. Capric acid (C10:0) has two additional binding sites [Bhattacharya et al., 2000] FA8 and FA9 (see Figure 3.2), as well as possessing also the site FA6′. In particular, FA8 is located at the interface IIIA-IB and FA9 at the interface IIA-IIIA.

HSA can also bind a wide variety of drug molecules [Sudlow et al., 1975; Ghuman et al., 2005; Fasano et al., 2005; Simard et al., 2006]. Drugs can bind to other plasma proteins but HSA is the primary binding protein for lipophilic drugs with acid or electronegative features. Most drugs bind to one of two principal binding sites , proposed by Sudlow [Sudlow et al., 1975; Sudlow et al., 1976] and called drug site DS1 and DS2 [He and Carter, 1992; Carter and Ho, 1994]. Site DS1 is a preformed pocket located into the core of subdomain IIA, composed of all six helices of the subdomain and a loop-helix (148-154) of the subdomain IB, and it almost completely overlaps with the site FA7. The interior of this pocket is principally apolar. Site DS2 is formed by all six helices of subdomain IIIA, and it is formed by a rearrangement of sites FA3 and FA4. The topology of site DS2 is similar to DS1, but it is smaller.

Ibuprofen (IBP) (Fig. 3.3) is considered a stereotypical ligand for Sudlow sites. IBP has a pK_a = 4.91 [Sangster, 1994] and is a chiral drug, and its pharmacological properties reside almost entirely with the S(+)-enantiomer [Evans et al., 1990].

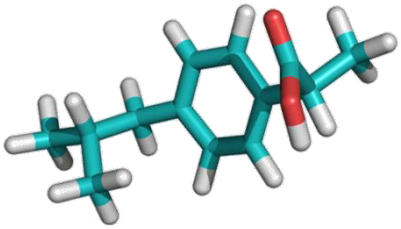


Fig. 3.3 – *Neutral structure of the S(+)-enantiomer of IBP, colored according to the atom type: C, cyan; O, red; H, white. Charged IBP has the same structure, but the H atom bonded to the O atom is not present.*

Crystallographic studies of the molecular complex demonstrate that IBP binds HSA in the site DS2 [Ghuman et al., 2005]. Moreover, the electron density map indicates that IBP can occupy a secondary site that overlaps with the fatty acid binding site FA6 [Bhattacharya et al., 2000; Petitpas et al., 2001]. Other low-affinity IBP binding sites have been only suggested, since the electron density maps are too weak to incorporate them in a model. These sites correspond to site DS1 [Ghuman et al., 2005] and FA2 [Di Masi et al., 2011].

3.2. Computational methods

3.2.1. Protein modeling and ligand docking

The structure of HSA complexed with IBP was obtained from the X-ray structure crystallized in the presence of two S(+)-IBP molecules (2BGX entry [Ghuman et al., 2005] in the Protein Data Bank (PDB)). The position of 7 missing residues, as well as other 62 missing

atoms in solvent-exposed side chains of other residues, was reconstructed *in silico* by using VMD [Humphrey et al., 1996].

AutoDockVina [Trott and Olson, 2010] was used for the docking of charged IBP to HSA. The graphical interface AutoDock Tools 1.5.6 [Morris et al., 1998] was used (a) to convert the structures of the receptor and ligand from PDB to PDBQT format, (b) to add polar hydrogens to the protein, (c) to determine the allowed torsions for the ligand, and (d) to find out the center of the grid box, for which a grid spacing of 37.5 pm was applied. The search space for the binding of the ligand to the protein was not restricted and the entire protein was considered because we're trying to understand the range of possible binding sites.

The docking procedure consisted of ten independent runs, each determining the best 20docking conformations ranked according to their binding affinity, for a total of 200binding modes in six different poses. A reduction of the number of binding modes from 200 to 35 was obtained through a clustering procedure based on distances, by excluding binding modes with a RMSD (Root Mean Square Deviation) of atomic positions below2.5 Å to obtain a wide sampling of the range of possible different binding modes.

3.2.2. Molecular dynamics

The simulation package GROMACS 4.6.3 [Hess et al., 2008; Pronk et al., 2013] was used in combination with the AMBER 99SB_ILDN [Hornak et al., 2006] and GAFF force field [Wang et al., 2004]. The software VMD [Humphrey et al., 1996] and PyMOL [DeLano, 2002] were used for molecular visualization. MD simulations were performed for HSA complexed with either neutral (protonated) or charged (deprotonated) IBP. The topologies of both the neutral and charged IBP form was built by using AmberTools13 [Wang et al., 2004;

Wang et al., 2006]. Atomic charges were assigned by using the AM1-BCC method [Jakalian et al., 2000] as implemented in Antechamber.

All simulations were performed by placing HSA at the center of a rhombic dodecahedron box with a minimum distance of 1 nm with respect to cell walls. HSA complexed with IBP was surrounded with about 32000 water molecules described by the TIP3P water model [Jorgensen et al., 1983]. The overall charge of the system was neutralized by adding 15 Na^+ counterions, for the neutral form, and 16 Na^+ counterions, for the charged one. Periodic boundary conditions were applied to avoid edge effects. The system was energy minimized by using a steepest descendent method for 1500 steps.

Initial velocities were obtained by randomly drawing them from a Maxwell-Boltzmann distribution at 300 K. The leap-frog Stochastic Dynamics integrator (SD) [Van Gunsteren and Berendsen, 1988] was used during equilibration and production phases, which were carried out for 10 ps and 5 ns, respectively. The reference temperature for the entire system was set to 300 K, and the inverse friction constant to 0.1 ps. The Berendsen barostat was used to control the pressure [Berendsen et al., 1984] with a time constant 0.5 ps and compressibility $4.5 \cdot 10^{-5}$ bar^{-1}. The electrostatic interactions were treated with the Particle-Mesh Ewald (PME) method [Darden et al., 1993; Essman et al., 1995]. Bond distances were constrained with the P-LINCS algorithm [Hess et al., 2008] and an integration time step of 2 fs was used.

At the end of the 35 MD simulations, three atoms in the HSA and three in the ligand of the first binding mode were picked to define six restraints [Boresch et al., 2003]: one for bond distances, two for bond angles and three for angle dihedrals. These restraints constitute the reference index used during the clustering process. We used the PCCA clustering method [Deuflhard and Weber, 2003; Weber, 2003] to group together conformations which

interconvert quickly in the MD simulations, and keep those which are kinetically distinct as separate clusters, based on the transition matrix. Followed this timescale-based structural clustering, we calculated the interaction energies of the different clustered binding modes starting from the combination of individual trajectories resulting from the MD simulations. The goal of this clustering procedure was to select stable or metastable ligand binding modes which make favorable interactions with the protein, and carry the most promising of these on to binding free energy calculations which we use to determine the most likely binding modes.

3.2.3. Absolute binding free energy

Simulation to calculate the free energy differences were performedby using GROMACS 4.6 [Hess et al., 2008; Pronk et al., 2013] and analyzed with the MBAR method [Shirts and Chodera, 2008].To estimate the binding free energy of IBP interacting with HSA it is necessary to evaluate the energy difference between the complex (Fig. 3.4a) and the configuration in which the ligand and the protein are separately present in solution and non-interacting (Fig. 3.4e).

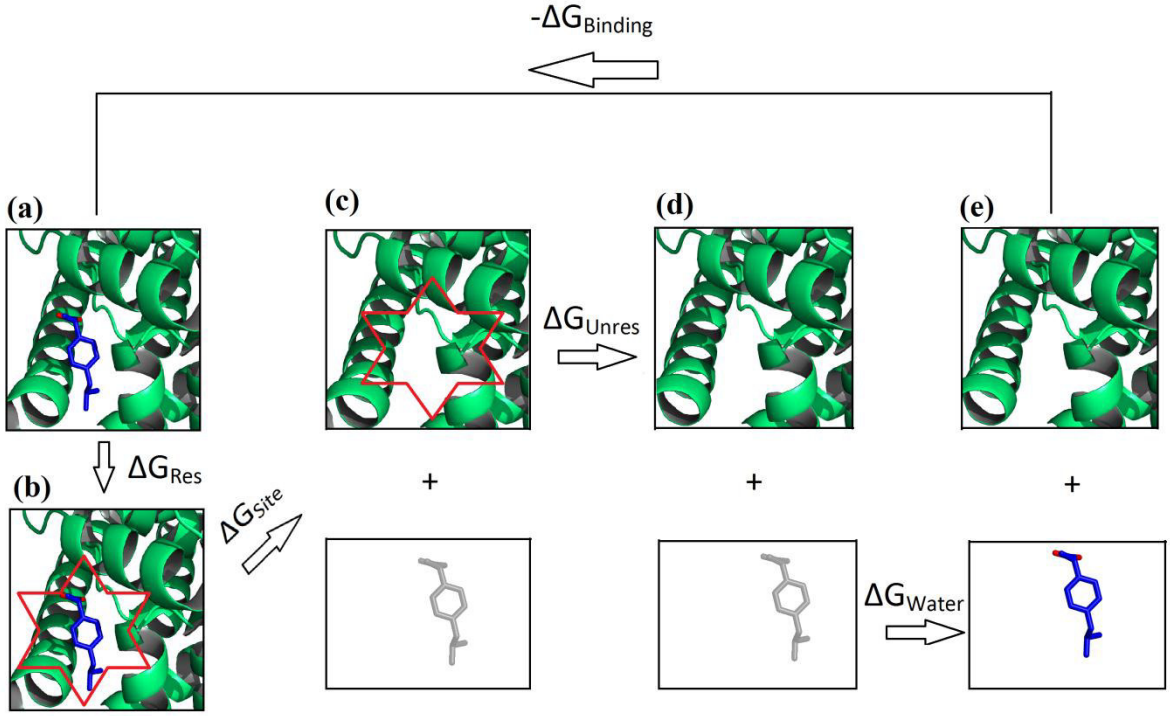


Fig. 3.4 *- Graphical representation of the HSA-IBPthermodynamic cycle. (a) Structure of HSA-IBP complex. The protein is represented in limegreen, the IBP in blue and the oxygen atoms in IBP are red. (b) HSA-IBP complex in the presence of restraints (the red star). (c) HSA and IBP non-interacting (gray), with restrains are still present. (d)HSA and IBP non-interacting (gray), restrains are no longer present, (e) non-interacting HSA and IBP, with the IBP molecule (blue) interacting with the solvent.*

Since the direct estimation of free binding energy differences between the unbounded (Fig. 3.4e) and bonded (Fig. 3.4a) states is computationally too expensive to calculate, it is necessary to construct a thermodynamic cycle [Gilson et al, 1997; Boresch et al., 2003; Mobley et al., 2006] by introducing intermediate steps of the process (Fig. 3.4b, c and d). The summation of all contributes will allow to calculate.

In the thermodynamic cycle, the starting configuration (Fig. 3.4a) consists in HSA and IBP complex simulated in a conventional MD run. A subsequent and gradual decoupling of

the IBP is guaranteed by performing 24 separate simulations of 1 ns each at increasing values, after the addiction of a set of restraints between the ligand and the protein (Fig. 3.4b), obtained as discussed above. The presence of restraints is necessary to ensures that IBP does not leave the binding pocket [Mobley and Chodera, 2006] both for practical reasons relating to sampling, and because the restraints allow rigorous definition of the standard state, which is essential for calculating standard binding free energies, our goal. An evaluation of the energy contribution due to the use of the restraints is the following step of the thermodynamic cycle, and can be calculated analytically [Boresch et al., 2003]. The final step is the calculation of the free energy of hydration of IBP, with the ending state (Fig. 3.4e) consisting in the ligand interacting again with the solvent, i.e. both electrostatic and non-electrostatic interactions are turned on again. The production run was carried out for 5 ns at each value of λ.

This thermodynamic cycle is applied to simulate each binding mode, as obtained after the clustering process.

3.1. Results

3.1.1. Docking and clustering

The existence of at least two binding sites for IBP in HSA has been revealed in crystallography [Ghuman et al., 2005]. Other possible binding sites with lower affinity have been suggested, but their precise location has not been not defined [Ghuman et al, 2005; Di Masi et al., 2011].We used molecular docking to identify all the possible candidate locations for binding sites of IBP in HSA. To this aim, the structure of HSA was extracted from the complex HSA-IBP identified in crystallography [Ghuman et al., 2005] and a charged IBP

molecule was docked back to it in 10 different simulations by using AutoDock Vina [Trott and Olson, 2010]. The 200 binding modes obtained by molecular docking were distributed in six different poses (see Table 3.1, column 3) and were reduced to 35 by using a clustering method based on RMSD between docking poses, in order to discard binding modes closer than 2.5 Å from each other. Specifically, we retained the best scoring docking pose, then the next best pose different by more than 2.5 Å RMSD from that one, and then the next best pose different by the same amount from those two, and so on.

Table 3.1 – *Summary of the number of binding modes of IBP complexed to HSA before and after the RMSD clustering procedure, and correspondent docking energy. The reported uncertainty is the maximal variation between the mean docking energy and the other reported docking energies in that pose.*

HSA Domain	**Binding Site**	**# Docking binding modes (200 total)**	**# RMSD Clustering binding modes (35 total)**	**Docking energy (kcal/mol)**
IB	FA1	1	1	-6.5 ± 0.1
IIA	DS1 (FA7)	49	11	-6.4 ± 0.2
IIA-IIB	FA6	90	10	-7.0 ± 0.1
IIIA	DS2 (FA3+FA4)	34	5	-7.5 ± 0.1
IIIA-IB	FA8	5	2	-6.1 ± 0.1
IIA-IIIA	Lower cleft	21	6	-7.1 ± 0.2

The six poses found by molecular docking correspond to well known binding sites of HSA for different ligands [Petitpas et al., 2001; Fasano et al., 2005; Simard et al., 2006], including fatty acids [Bhattacharya et al., 2000], and are located as shown in Figure 3.3.

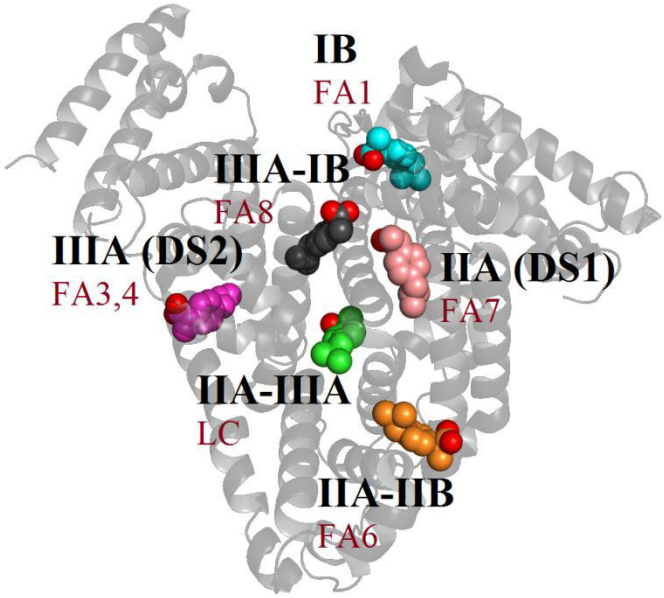


Fig. 3.3*- Structure of HSA complexed with IBP, located in six poses found by molecular docking. The protein is represented as a gray cartoon, and IBP is shown in sphere representation with oxygen atoms colored red. The subdomain and binding site name are labeled in black and purple red, respectively.*

The poses can be ranked according to their docking scores (Table 3.1, column 5), in order of decreasing affinity. From the predicted scores, the most probable site is located in the subdomain IIIA and corresponds to site DS2 [Sudlow et al, 1976], which comes out from a combination of the two fatty acid binding sites FA3 and FA4 [Battacharya et al, 2000]. The second most probable binding site is located at the interface IIA-IIIA and, although not too distant from the short-chain fatty acid site FA9 [Battacharya et al., 2000], it does not correspond exactly to the fatty acid site (see Figure 3.4), nor to any other known binding site for ligands.

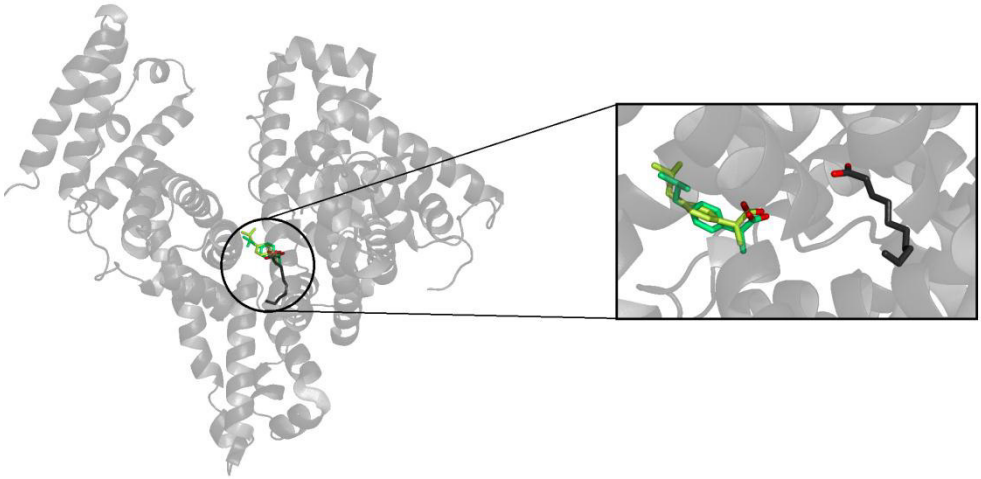


Fig. 3.4 –*Comparison between the location of two binding modes of IBP (bright and lime green) at the interface IIA-IIIA (LC) and a short-chain fatty acid (black) in the binding site FA9. Both IBP and the fatty acid are represented in stick and with oxygen atoms colored red.*

Other sites are located at the interface IIA-IIB, i.e. FA6 [Battacharya et al., 2000], and within the subdomain IB, i.e. FA1 [Battacharya et al., 2000]. A single pose is found within the subdomain IIA, which corresponds to the binding site DS1 [Sudlow et al, 1976] and mostly overlaps with the fatty acid binding site FA7 [Battacharya et al, 2000]. Finally, the binding mode with the lowest affinity is found at the interface IB-IIIA, and coincides with the binding site FA8 that is only available for short-chain fatty acids [Battacharya et al., 2000].

It is interesting to note that the subdomain IIIA (site DS2) and the interface IIA-IIB (site FA6), which represent the two sites reported by crystallography [Ghuman et al., 2005], fall within the first three most likely docking poses ranked according to their docking scores.

3.1.2. MD simulations and PCCA clustering

On the basis of the docking calculations alone it is difficult to evaluate whether the poses found for IBP correspond to genuine binding sites or to non-specific interaction sites. In fact, molecular docking does not consider the dynamics of the protein-ligand complex. Long-lived interactions can be also assisted by local rearrangements of the protein structure, which can be detected by MD simulations. For this reason, the 35 binding modes previously selected (together with the protein) were used as starting structures to perform MD simulations of the HSA-IBP complex. IBP in water under physiological conditions is expected to be in the charged state, due to its pK_a value 4.5; however, within the protein matrix it cannot be excluded that IBP could also assume a neutral state, due to the interaction with the amino acid residues. The possibility that IBP is either charged or neutral is taken into account by performing two distinct sets of simulations, for a total of 70 MD runs of 5 ns each.

At the end of the MD simulations, the trajectories obtained for both charged and neutral IBP bound to HSA were separately analyzed to the most populated kinetically stable binding modes. Particularly, for each charge, all trajectories were lumped together and then clustered based on kinetics using PCCA [Deuflhard and Weber, 2003; Weber, 2003], so that conformations which are fast to interconvert were grouped together and those that were slow to interconvert were treated separately. Each resulting cluster of conformations was then analyzed by several metrics, including its population, its short-range interaction energy with the protein, and the free energy of turning off the short-range interactions with the protein (evaluated via the Zwanzig relationship) We retained the binding modes which were the most populated and with the most favorable contributions from these energies, up to include a maximum of three binding modes for each protein site. The binding modes analyzed via this

approach were 72 and 80 for charged and neutral IBP, respectively, and distributed as reported in Table 3.2 (column 3 and 5).

Table 3.2 – *Number of binding modes of IBF complexed to HSA obtained by PCCA clustering (column 3 and 5), and further selected after an evaluation based on both occurrence frequency and interaction energy (column 4 and 6).*

		charged IBP		**neutral IBP**	
HSA Domain	**Binding Site**	**# Binding modes (72 tot.)**	**# Selected binding modes (15 tot.)**	**# Binding modes (80 tot.)**	**# Selected binding modes (12 tot.)**
IB	FA1	2	2	3	1
IIA	DS1 (FA7)	23	3	28	3
IIA-IIB	FA6	25	2	22	3
IIIA	DS2 (FA3+FA4)	11	3	8	3
IIIA-IB	FA8	6	3	4	0
IIA-IIIA	Lower cleft	5	2	15	2

Through an evaluation based on both the frequency (i.e., the number of simulation snapshots of the protein-ligand complex in which IBP corresponds to a given binding mode) and interaction energy, a maximum of 3 binding modes were selected for each protein location. This resulted in 15 and 12 binding modes being selected for charged (Fig. 3.5a) and neutral IBP (Fig 3.5b), respectively. It is worth to note that the site FA8 can be already excluded as a binding location for neutral IBP on the basis of this (preliminary) energetic selection. Furthermore, the exact location and conformation of the binding modes selected after this overall procedure is generally different compared to those obtained after molecular docking, although these binding modes are distributed in the same six main locations.

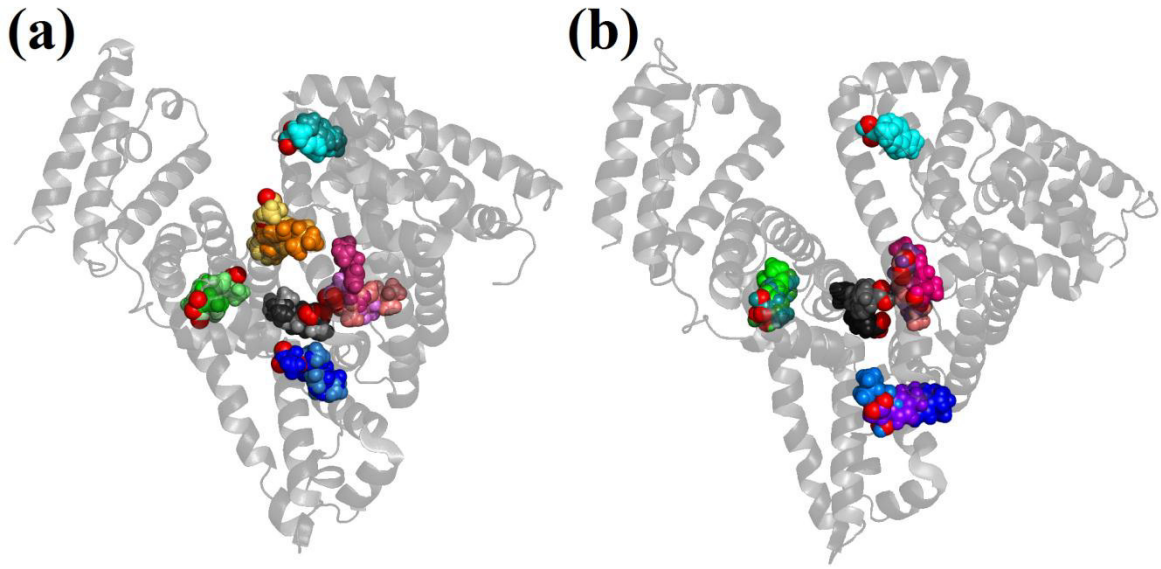


Fig. 3.5– *Structures of (a) charged and (b) neutral IBP binding modes selected after MD simulations. IBP molecules in each binding mode are colored according to their location: green variants for DS2, blue for FA6, red for DS1, gray for the lower cleft, orange for FA8 and cyan for FA1.*

3.1.3. Free energy values for charged IBP

The binding free energy of IBP to HSA can be obtained in simulation, and ligand poses ranked according to binding free energy can be compared with the locations determined in crystallography. In this way, it is possible to relate information obtained from computational methods with the binding sites already experimentally known, as well as demonstrate the existence of binding sites previously supposed but whose position is not already revealed in the experiment.

The free energy values are obtained by applying the thermodynamical cycle previously described (section *3.2.3*) and summing the following terms:

$$\Delta G_{Binding} = -\Delta G_{Res} - \Delta G_{Site} + \Delta G_{Unres} + \Delta G_{Water} \qquad (3.30)$$

With corresponding uncertainties calculated by an error propagation on this summation. Absolute binding free energy calculations were performed on the two sets of binding modes for charged and neutral IBP (15 and 12 binding modes, respectively), and the overall results are shown in Table 3.3.

Table 3.3 – *Absolute binding free energy data obtained for the simulated binding modes of IBP complexed to HSA.*

		charged IBP		**neutral IBP**	
HSA Domain	**Binding Site**	**# Selected binding modes (15 tot.)**	**Binding Free Energy (kcal/mol)**	**# Selected binding modes (12 tot.)**	**Binding Free Energy (kcal/mol)**
IB	FA1	2	-12.7 ± 0.7 -6.0 ± 0.4	1	-15.7 ± 0.6
IIA	DS1 (FA7)	3	-15.9 ± 0.3 -13.3 ± 0.3 -7.4 ± 0.5	3	-5.5 ± 0.4 +2.5 ± 0.4 +3.1 ± 0.6
IIA-IIB	FA6	2	-11.1 ± 0.4 -10.6 ± 0.4	3	-7.7 ± 0.5 -5.3 ± 0.6 +15.1 ± 0.3
IIIA	DS2 (FA3+FA4)	3	-23.5 ± 0.6 -16.7 ± 0.3 -12.3 ± 0.3	3	-10.9 ± 0.6 -9.9 ± 0.4 -9.0 ± 0.6
IIIA-IB	FA8	3	+2.6 ± 0.3 +3.1 ± 0.4 +6.3 ± 0.2	0	–
IIA-IIIA	Lower cleft	2	-4.4 ± 0.5 -2.4 ± 0.6	2	-2.0 ± 0.6 +0.8 ± 0.5

It is immediate to observe that, from a general point of view, charged IBP shows a greater affinity for HSA compared to neutral IBP. This is probably due to interactions between the carboxylate group and positively charged or polar residues into the HSA binding sites. Thus, we first analyze the results for charged IBP, which is the form normally present in solution at physiological pH [Lockhart et al., 2012] and most likely to occur within the HSA matrix.

The site DS2 is the most favorable binding site for charged IBP, with two binding modes with similar orientation and binding energy, −23.5 ± 0.6 kcal/mol and −16.7 ± 0.3 kcal/mol, and a third binding mode with opposite orientation and a smaller binding energy, -12.3 ± 0.3 kcal/mol. The first value, −23.5 ± 0.6 kcal/mol, is likely an overestimate of the real value, because it would lead to a femtomolar dissociation constant. This suggests that the computed values are qualitatively correct, but typical, unavoidable difficulties associated with MD simulations (such as intrinsic limits in both sampling and force field) might affect the quantitative estimates of the free energy values.

According to the crystallographic data [Ghuman et al., 2005], DS2 consists of a single polar patch that can also be occupied by a number of other drugs such as diflunisal, diazepam and indoxyl sulphate. In particular, these drugs interact with at least one of their O atoms in proximity to the polar patch. In this protein pocket, both the binding modes of IBP with the best affinities show a HB between their carboxylate group and both O^{γ}-Ser489 (donor-acceptor distance 0.26 ± 0.01 nm) and N^{ζ}-Lys414 (donor-acceptor distance 0.31 ± 0.01 nm). Although these binding modes (Fig. 3.6a) are very close to the location of IBP detected in crystallography, in the latter case the carboxylate group of IBP forms a HB with Arg410, whereas in simulation only an electrostatic interaction with this residue is present.

IBP can bind within DS2 also in an opposite orientation (Fig. 3.6b) compared to the crystallographic one, and in this case it forms three HBs: one with N-Gly341 (donor-acceptor distance 0.26 ± 0.01 nm) and the other two with N-Lys432, with donor-acceptor distance from the two carboxylate O atoms of 0.32 ± 0.01 nm and 0.34 ± 0.01 nm.

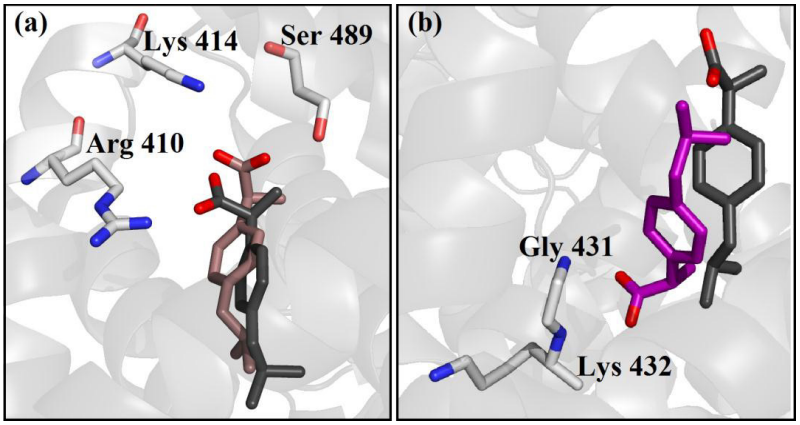


Fig. 3.6–*Crystallographic (black) and simulated (either brown or purple) charged IBP molecule in interaction with HSA within DS2. Some selected protein side chains are also shown.*

The second most likely pose for charged IBP interacting with HSA is the site DS1, with binding energy -15.9 ± 0.3 kcal/mol. X-ray data [Sugio et al., 1999; Bhattacharya et al., 2000] shows that this site is a predominantly apolar pocket, with two clusters of polar residues at the entrance and at the bottom of the protein cavity. In particular, the two key residues Lys195 and Lys199 are at the entrance of the binding pocket. In Figure 3.7the different binding modes for this pose are shown.

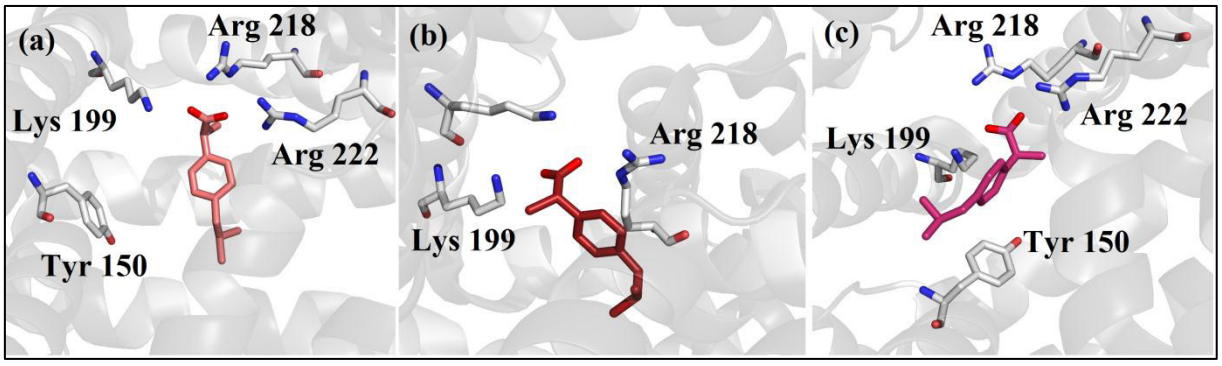


Fig. 3.7– *Charged IBP in interaction with HSA in the site DS1, shown according to decreasing affinity (see Table 3.3) from (a) to (c). Selected protein side chains are also shown.*

From the simulation results the ligand forms HBs with N^{ξ}-Lys199 and N^{η}-Arg218 in all the binding modes in DS1, with an average donor-acceptor distance from the carboxylate group of IBP of 0.28 ± 0.01 nm in both cases (Fig. 3.7). An additional HB between IBP and N^{η}-Arg222 is also present in two cases (Fig. 3.7a,c), with bond distance 0.31 ± 0.01 nm. Finally, only the binding mode showed in figure 3.7b presents an HB between N^{ξ}-Lys195 and the carboxylate group of IBP, with a donor-acceptor distance of 0.31 ± 0.01 nm.

A key residue in the binding site DS1 is Tyr150. In fact, its hydroxyl group can form HBs with several drugs bind in this pocket [Ghuman et al. 2005]. Moreover, although in general HSA shows only small rearrangement of side chains to accommodate drugs, Tyr150 and Trp214 can be an exception. In simulation, Tyr150 does not form any HB with IBP, but in two of the three binding modes studied (Fig. 3.7a,c) it undergoes a rotation with respect to its conformation in the absence of ligand, as it can be seen in Fig. 3.8.

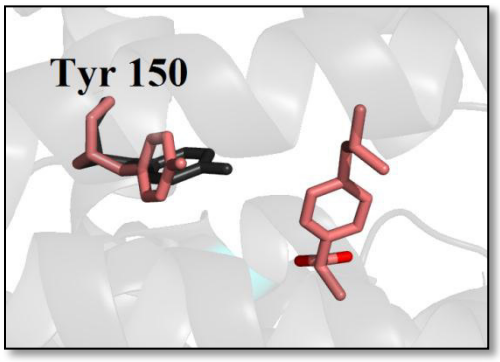


Fig. 3.8 – *Comparison of the Tyr150 conformation in the absence of ligands within the binding site DS1 in crystallography (black) and in the presence of charged IBP in simulation (pink).*

The pose corresponding to FA6 is the third most likely binding site (Fig. 3.9), with binding free energy -11.1 ± 0.4 kcal/mol. The carboxylate group of IBP interacts with this HSA pocket by forming three HBs with N^{ξ}-Lys351 (donor-acceptor distance 0.27 ± 0.01 nm),

N-Leu481 (donor-acceptor distance 0.29 ± 0.01 nm) and N-Val482 (donor-acceptor distance 0.30 ± 0.01 nm).

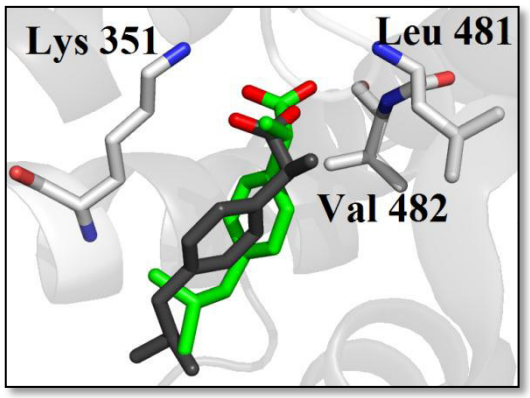


Fig. 3.9–*Comparison between crystallographic (black) and simulated (green) charged IBP in interaction with HSA at the interface IIA-IIB. Selected protein side chains are also shown.*

In the charged form, as well as in the neutral one, IBP shows a (rather unexpected) binding affinity for HSA in the FA1 site. In particular, charged IBP has binding energy -12.7 ± 0.7 kcal/mol, comparable with the one found within both DS1 and FA6 binding sites. In the FA1 site, the carboxylate group of IBP forms a weak HB with N-Arg117, with a donor-acceptor distance 0.34 ± 0.01 nm. Concurrently, as shown in Figure 3.10, the 'tail' of IBP is confined by van der Waals interactions with the two Tyr residues 138 and 161, which assume an approximately parallel orientation with each other.

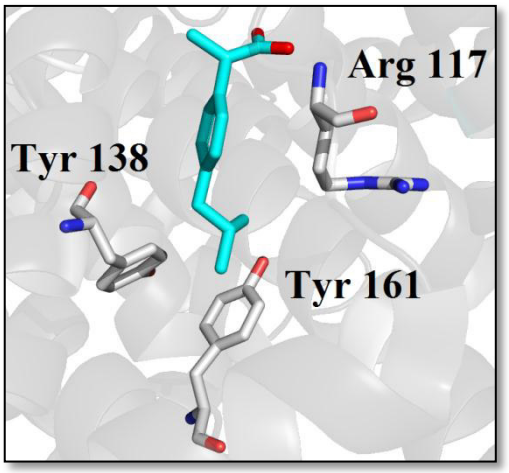


Fig. 3.10 - *Charged IBP in interaction with HSA in the subdomain IB. Selected protein side chains are also shown.*

The last two poses, corresponding to the upper (FA8 site) and lower cleft of HSA, cannot be considered binding sites on the basis of the absolute binding free energy calculations. In fact, the upper cleft has positive values of binding energies (see Table 3.1), whereas values for the lower cleft are negative but too unfavorable to be considered a real binding site.

3.1.4. Free energies values for neutral IBP

When the neutral form of IBP is considered, absolute binding free energy values are generally less favorable compared to the charged form of the ligand, and only four of the six poses have a free energy compatible with a possible formation of a molecular complex.

The most favorable pose with binding energy -15.7 ± 0.6 kcal/mol is the binding site FA1 [Bhattacharya et al., 2000]. In the absence of ligands [Sugio et al., 1999; Bhattacharya et

al., 2000], Tyr138 stacks with Tyr161 and both occlude the binding pocket. In contrast, as shown in Figure 3.11, in the presence of IBP the side chains of these residues rotate to harbor the ligand in a hydrophobic clamp. A weak HB is also formed between N-Glu119 and the unprotonated O atom of the carboxyl group of IBP, with a donor-acceptor distance 0.38 ± 0.01 nm.

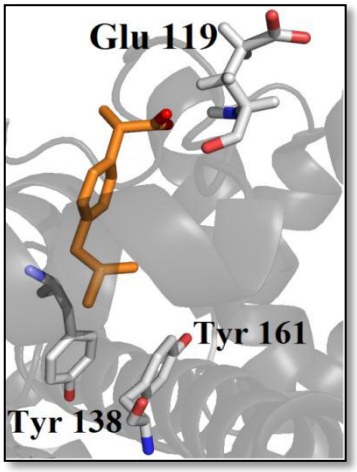


Fig. 3.11 –*Neutral IBP interacting with HSA within the subdomain IB. Selected protein side chain are also shown.*

The site DS2 is the second most likely binding site for neutral IBP and has an energy of-10.9 ± 0.6 kcal/mol. As shown in Figure 3.12, the orientation of IBP into the pocket detected in MD simulation is consistent with the one found in crystallography. In this pocket IBP forms a HB between its unprotonated O atom and N^{ξ}-Lys414 (Fig. 3.12), with a donor-acceptor distances 0.30 ± 0.01 nm.

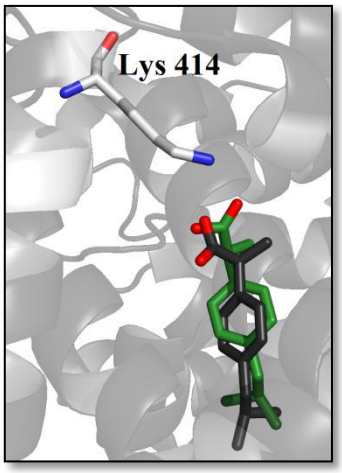


Fig. 3.12 - *Comparison between crystallographic (black) and simulated (green)neutral IBP in interaction with HSA in the binding site DS2.Selected protein side chains are also shown.*

Regarding theFA6 binding site, one of the three simulated binding modes has a positive binding energy (+15.1 ± 0.3 kcal/mol) and can be immediately excluded. The other two binding modes have a marked difference in the orientation with respect to the crystallographic structure of IBP reported for this site, as shown in Figure 3.13. In fact, compared to the crystallographic orientation of the ligand, one binding mode is perpendicular (Fig. 3.13a), whereas the other one has an opposite orientation (Fig. 3.13b).

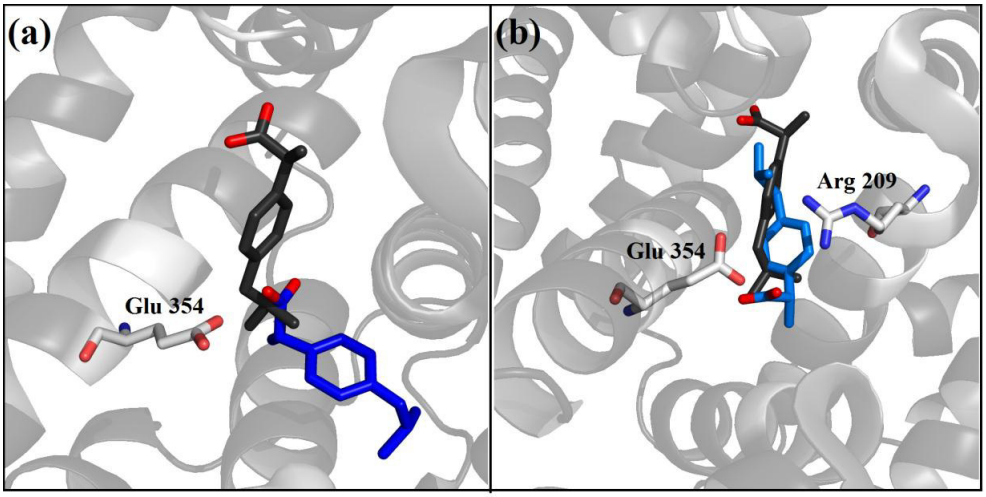


Fig. 3.13 - *Comparison between crystallographic (black) and simulated (either blue or light blue) neutral IBP in interaction with HSA within the binding site FA6. Selected protein side chains are also shown.*

In both cases IBP forms an HB between the protonated O atom of its carboxyl group and O^{ε}-Glu354, with a donor-acceptor distance 0.26 ± 0.01 nm. In addition, only in the second case a weak HB is found between the unprotonated O atom of IBP and N^{η}-Arg209, with donor-acceptor distance 0.33 ± 0.01 nm.

In the binding site DS1 (Fig. 3.14), IBP forms a HB between its unprotonated O atom and N^{ξ}-Lys199, with a donor-acceptor distance 0.29 ± 0.01 nm.

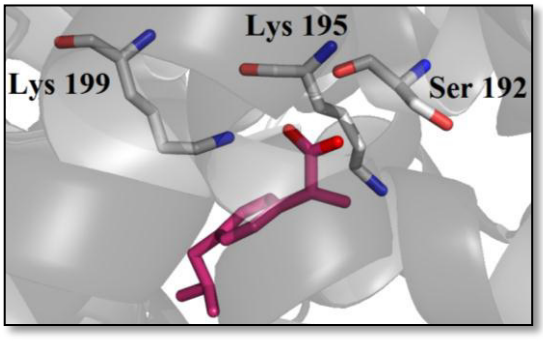


Fig. 3.14 - *Neutral IBP-HSA interaction within the binding site DS1. Selected protein side chains are also shown.*

Finally, the lower cleft can be excluded as a binding site of IBP on the basis of the binding free energy values calculated.

Conclusions

In the scientific research concerning complex systems, biomolecules have a role of considerable interest. In particular, in biophysics, one of the topics of great importance is the study of the interactions between biological molecules and ligands. These studies are also relevant for innovative applications in medicine, nanotechnology and food science.

The present thesis is in the framework of this research field and concerns the study of the two model proteins βLG and HSA in interaction with, respectively, fatty acids with different chain length and the pharmacological compound IBP, by using molecular docking and MD simulations. The topic of this work was to determine the dynamical and binding properties of such biological complexes, fundamental to understand the molecular basis of the association and release of a ligand compound.

The results obtained on βLG show that the binding of fatty acids within the calyx determine an enhancement of the dynamics of the protein compared to its unliganded form, especially for loops located at the entrance of the main protein binding site. Although these regions are unstructured, they show coordinated motions that are relevant to ensure the protein function. From a structural point of view, the key residues have been determined that contribute to anchor the bound fatty acid at the entrance of the main protein binding site, in the intermediate region of the calyx, and in the innermost part of the hydrophobic channel.

Furthermore, the binding of fatty acids is a dynamical process, with the possibility for additional lipid molecules to compete for the occupation of the βLG calyx, or to interact with the outer protein surface and relocate in additional low-affinity sites. In this respect, by combining docking and MD simulation results, we predicted the existence of two additional binding sites for fatty acids previously only hypothesized in solution. These external sites are

involved in transient and relatively short-lived βLG-fatty acid interactions, which can be detected by computational methods.

A similar approach has been also applied to investigate the interaction of IBP, either charged or neutral, with HSA, a larger macromolecule with a higher structural complexity. Docking results highlight six different candidate locations for IBP interacting with HSA. Five of these poses correspond to binding sites for fatty acids and drugs (Sudlow's sites), whereas the sixth is located in the lower region of the protein cleft.

Absolute free energies of binding have been calculated by using an alchemical free energy approach, which allows a reliable estimate and ranking of the binding affinity of the ligand in each pose. The comparison of charged and neutral IBP binding free energy values evidences that charged IBP has a greater affinity for HSA. The different behaviour can be assigned to the carboxylate group that is deprotonated at neutral pH, therefore the negatively charged form of IBP is the one expected under physiological conditions.

The drug site DS2 of HSA is the most favorable binding site for charged IBP and the ligand can associate with high affinity either in the same orientation as found in crystallography, or in an opposite orientation. The second most likely binding site for IBP is the drug site DS1, as previously suggested by X-ray data, but not reported in terms of atomic coordinates. Furthermore, the FA6 site can also a be a binding site, in agreement with experimental data. Finally, both in the charged and neutral form, IBP shows a significant binding affinity for the FA1 site in HSA. In contrast, the last two poses, corresponding to the upper (FA8 site) and lower cleft of HSA, cannot be considered binding sites on the basis of the absolute binding free energy values obtained. As regards neutral IBP, the only other possible location is the DS2 binding site, although with much lower affinity compared to charged IBP, whereas other binding sites can be excluded.

The overall results obtained in this study show that computational approaches can give deep insights into protein-ligand interactions, and allow to reproduce and extend experimental data by accurately predicting binding locations and free energy values. In particular, the possibility of exploring alternative binding locations or geometries of a ligand within a protein can be of great interest in pharmacology and drug design. The use of theoretical methods is demonstrated to be an effective tool to help in the rational engineering of innovative compounds to be used in the current scientific research.

Bibliography

1. Amadei A, Linssen ABM, Berendsen HJC. Essential dynamics of proteins. Proteins 1993;17:412–425.
2. Baldini G, Beretta S, Chirico G, Franz H, Maccioni E, Mariani P, Spinozzi F. Macromolecules 1999; 32(19):6128-6138.
3. Barbiroli A, Bonomi F, Ferranti P, Fessas D, Nasi A, Rasmussen P, Iametti S. Bound fatty acids modulate the sensitivity of bovine β-lactoglobulin to chemical and physical denaturation. J Agr Food Chem 2011;59:5729–5737.
4. Basset GJC, Quinlivan EP, Gregory JF, Hanson AD. Folate synthesis and metabolism in plants and prospects for biofortification. CS 2005;45 (2):449-453.
5. Bhattacharya AA, Grüne T, Curry S. Crystallographic analysis reveals common modes of binding of medium and long-chain fatty acids to human serum albumin. J Mol Bio 2000; 303(2):721-732.
6. Bello M, Portillo-Téllez MD, García-Hernández E. Energetics of ligand recognition and self-association of bovine β-lactoglobulin: Differences between variants A and B. Biochemistry 2011; 50:151–161.
7. Bello M, Gutiérrez G, García-Hernández E. Structure and dynamics of β-lactoglobulin in complex with dodecyl sulfate and laurate: A molecular dynamics study. Biophys Chem 2012; 165:79–86.
8. Bello M, García-Hernández E. Ligand entry into the calyx of β-lactoglobulin. Biopolymers 2014; 101:744–757.
9. Bello M. Binding free energy calculations between bovine β-lactoglobulin and four fatty acids using the MMGBSA method. Biopolymers, 2014; 101(10): 1010-8.

10. Bekker H, Van Den Berg JP, Wassenaar TA. A method to obtain a near-minimal-volume molecular simulation of a macromolecule, using periodic boundary conditions and rotational constraints. J Comput Chem 2004; 25 (8):1037-1046.
11. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J. Interaction models for water in relation to protein hydration. In: Pullman B, editor. Intermolecular forces. Dordrecht: Reidel; 1981. pp 331–342.
12. Berendsen HJC, Postma JPM, W.F. vG, Di Nola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys 1984; 81:3684−3690.
13. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissing H, Shindyalov IN, Bourne PE. The protein data bank. Nucl Acids Res 2000; 28 (1):235-242.
14. Boresch S, Tettinger F, Leitgeb M. Absolute binding free energies: a quantitative approach for their calculation. J Phys Chem B 2003; 107 (35):9535-9551.
15. Brownlow S, Cabral JHM, Cooper R, Flower DR, Yewdall SJ, Polikarpov I, North AC, Sawyer L. Bovine β-lactoglobulin at 1.8 Å resolution – still an enigmatic lipocalin. Structure 1997; 5:481–495.
16. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. J Chem Phys 2007; 126:0140101.
17. Carter DC, Ho JX. Structure of Serum Albumin. Adv. Protein Chem. 1994; 45:153-203
18. Carter DC, He XM, Munson SH, Twigg PD, Gernert KM, Broom MB, Miller TY. Three-dimensional structure of human serum albumin. Science 1989; 204 (4909):1195-1198.
19. Cistola DP, Small DM. Fatty acid dirtribution in systems modeling the normal and diabetic human circulation. A ^{13}C nuclear magnetic resonance study. J Clin Invest 1991; 87:1431-1441.
20. Collini M, D'Alfonso L, Molinari H, Ragona L, Catalano M, Baldini G. Competitive binding of fatty acids and the fluorescent probe 1-8-anilinonaphthalene sulfonate to bovine β-lactoglobulin. Protein Sci 2003; 12:1596–1603.

21. Curry S, Mandelkow H, Brick P, Franks N. Crystal structure of human serum albumin complexed with fatty acid reveals an asymmetric distribution of binding sites. Nature Struct. Bio. 1998; 5:827-835.
22. Curry S, Brick P, Franks NP. Fatty acid binding to human serum albumin: new insights from crystallographic studies. BBA-MOL CELL BIOL L 1999; 1441: 131–140.
23. Darden T, York D, Pedersen L. Particle mesh Ewald: an N·log(N) method for Ewald sums in large systems. J Chem Phys 1993; 98:10089–10092.
24. De Palma C, Di Paola R, Perrotta C, Mazzon E, Cattaneo D, Trabucchi E, Cuzzocrea S, Clementi E. Ibuprofen–arginine generates nitric oxide and has enhanced anti-inflammatory effects. Pharmacol Res 2009; 60 (4):221–228.
25. DeLano WL. The PyMOL molecular graphics system. DeLano Scientific, Palo Alto, CA; 2002.
26. Deuflhard P, Weber M. Robust Perron cluster analysis in conformation dynamics. Linear Algebra Appl. 2003; 398:161-184.
27. Di Masi A, Gullotta F, Bolli A, Fanali G, Fasano M, Ascenzi P. Ibuprofen binding to secondary sites allosterically modulates the spectroscopic and catalytic properties of human serum heme-albumin. FEBS J 2011; 278 (4):654-662.
28. Domínguez-Ramírez L, Del Moral-Ramírez E, Cortes-Hernández P, García-Garibay M, Jiménez-Guzmán J. β-lactoglobulin's conformational requirements for ligand binding at the calyx and the dimer interphase: a flexible docking study. PLoS One 2013; 8:e79530.
29. Dufour E, Roger P, Haertlé T. Binding of benzo(α)pyrene, ellipticine, and cis-parinaric acid to β-lactoglobulin: Influence of protein modifications. J Protein Chem 1992; 11:645–652.
30. Eberini I, Baptista AM, Gianazza E, Fraternali F, Beringhelli T. Reorganization in apo- and holo-β-lactoglobulin upon protonation of Glu89: Molecular dynamics and pK_a calculations. Proteins 2004; 54:744–758.

31. Edwards PB, Creamer LK, Jameson GB. Structure and stability of whey proteins. In: Thompson A, Boland M, Singh H, editors. Milk proteins – From expression to food. Academic Press: San Diego; 2009. pp 163–203.
32. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. J Chem Phys 1995; 103:8577–8593.
33. Evans AM. Enantioselective pharmacodynamics and pharmacokinetics of chiral non-steroidal anti-inflammatory drugs. Eur J Clin Pharmacol 1990; 42 (3):237-256.
34. Ewald PP. Ewald summation. Ann. Phys. 1921.
35. Farrell Jr. HM, Jimenez-Flores R, Bleck GT, Brown EM, Butler JE, Creamer LK, Hicks CL, Hollar CM, Ng-Kwai-Hang KF, Swaisgood HE. Nomenclature of the Proteins of Cows' Milk—Sixth Revision. J Dairy Sci 2004; 87 (6): 1641–1674.
36. Fasano M, Curry S, Terreno E, Galliano M, Fanali G, Narciso P, Notari S, Ascenzi P. The extraordinary ligand binding properties of human serum albumin. IUBMB Life, 2005; 57(12): 787–796.
37. Flower DR. The lipocalin protein family: structure and function. Biochem J 1996;318:1–14.
38. Frapin D, Dufour E, Haertle T. Probing the fatty acid binding site of β-lactoglobulins. J Protein Chem 1993; 12:443–449.
39. García AE. Large-amplitude nonlinear motions in proteins. Phys Rev Lett 1992; 68:2696–2699.
40. Gilson MK, Given JA, Bush BL, McCammon JA. The statistical-thermodynamic basis for computation of binding affinities: a critical review. Biophys J 1997; 72 (3): 1047–1069.
41. Ghuman J, Zunszain PA, Petitpas I, Bhattacharya AA, Otagiri M, Curry S. Structural Basis of the Drug-binding Specificity of Human Serum Albumin. J Mol Biol 2005; 353 (1): 38–52.
42. Goga, N, Rzepiela AJ, de Vries AH, Marrink SJ, Berendsen HJC. Efficient algorithms for Langevin and DPD dynamics. J. Chem. Theory Comput 2012; 8:3637–3649.

43. Cuya Guizado TR. Analysis of the structure and dynamics of human serum albumin. J Mol Model 2014; 20:2450.
44. Guzzi R, Rizzuti B, Bartucci R. Dynamics and binding affinity of spin-labeled stearic acids in β-lactoglobulin: Evidences from EPR spectroscopy and molecular dynamics simulation. J Phys Chem B 2012; 116:11608–11615.
45. He XM, Carter DC. Atomic structure and chemistry of human serum albumin. Nature 1992; 358: 209 – 215.
46. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. LINCS: a linear constraint solver for molecular simulations. M J Comput Chem 1997; 18, 1463.
47. Hess B. P-LINCS: A parallel linear constraint solver for molecular simulation. J Chem Theory Comput 2008; 4:116–122
48. Hockney RW, Goel SP, Eastwood JW. Quiet high-resolution computer models of a plasma. J Comput Chem 1974; 14 (2): 148–158
49. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins 2006; 65(3): 712–725.
50. Hu WB, Liu JA, Luo Q, Han YM, Wu K, Lv S, Xiong SX, Wang FY. Elucidation of the binding sites of sodium dodecyl sulfate to β-lactoglobulin using hydrogen/deuterium exchange mass spectrometry combined with docking simulation. Rapid Commun Mass Sp 2011; 25:1429–1436.
51. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. J Mol Graph Model 1996; 14:33–38.
52. Jain, A. N. Scoring non-covalent protein-ligand interactions: A continuous differentiable function tuned to compute binding affinities. J Comput-Aided Mol Des 1996; 10:427-440.
53. Jameson GB, Adams JJ, Creamer LK. Flexibility, functionality and hydrophobicity of bovine β-lactoglobulin. Int Dairy J 2002; 12:319–329.

54. Jarzynski C. Rare events and the convergence of exponentially averaged work values. Phys Rev 2006; E 73.

55. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983; 79, 926.

56. Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. Nat Struct Biol 2002; 9:646 – 652.

57. Kirkwood JG. Statistical Mechanics of Fluid Mixtures. J Chem Phys 1935; 3,300.

58. Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov 2004; 3:935-949.

59. Kontopidis G, Holt C, Sawyer L. The ligand-binding site of bovine β-lactoglobulin: Evidence for a function? J Mol Biol 2002; 318:1043–1055.

60. Kontopidis G, Holt C, Sawyer L. *Invited review*: β-lactoglobulin: binding properties, structure, and function. J Dairy Sci 2004; 87:785–796.

61. Konuma T, Sakurai K, Goto Y. Promiscuous binding of ligands by β-lactoglobulin involves hydrophobic interactions and plasticity. J Mol Biol 2007; 368:209–218.

62. Kragh-Hansen U. Structure and ligand binding properties of human serum albumin. Dan Med Bull 1990; 37(1):57-84.

63. Kumosinski TF, Timasheff SN. Molecular Interactions in β-Lactoglobulin. X. The Stoichiometry of the β-Lactoglobulin Mixed Tetramerization. J Am Chem Soc 1966; 88 (23):5635–5642.

64. Leach, A. R. Molecular Modelling: Principles and Applications,2nd ed.; Prentice Hall: Upper Saddle River, NJ, 2001.

65. Liang L, Subirade M. β-Lactoglobulin/Folic Acid Complexes: Formation, Characterization, and Biological Implication. J Phys Chem B 2010; 114 (19):6707–6712.

66. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE. Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins 2010; 78 (8):1950–1958.

67. Loch J, Polit A, Górecki A, Bonarek P, Kurpiewska K, Dziedzicka-Wasylewska M, Lewiński K. Two modes of fatty acid binding to bovine β-lactoglobulin – crystallographic and spectroscopic studies. J Mol Recognit 2011; 24:341–349.

68. Loch JI, Polit A, Bonarek P, Olszewska D, Kurpiewska K, Dziedzicka-Wasylewska M, Lewiński K. Structural and thermodynamic studies of binding saturated fatty acids to bovine β-lactoglobulin. Int J Biol Macromol 2012; 50:1095–1102.

69. Lockhart C, Kim S, Klimov DK. Explicit Solvent Molecular Dynamics Simulations of Aβ Peptide Interacting with Ibuprofen Ligands. J Phys Chem B 2012; 116 (43): 12922–12932.

70. Lu N, Kofke DA. Optimal intermediates in staged free energy calculations. J Chem Phys 1999; 111, 4414.

71. Lu ND, Singh JK, Kofke DA. Appropriate methods to combine forward and reverse free-energy perturbation averages. J Chem Phys 2003; 118:2977-2984.

72. Mandalari G, Mackie AM, Rigby NM, Wickham MSJ, Mills ENC. Physiological phosphatidylcholine protects bovine β-lactoglobulin from simulated gastrointestinal proteolysis. Mol Nutr Food Res 2009; 53:131–139.

73. McKee AC, Carreras I, Hossain L, Ryu H, Kleine WL, Oddo S, LaFerla FM, Jenkinsg BG, Kowall NW, Dedeoglu A. Ibuprofen reduces Aβ, hyperphosphorylated tau and memory deficits in Alzheimer mice. Brain Res 2008; 1207:225–236.

74. McKenzie HA, Sawyer, WH. The genetic variants of β-lactoglobulin undergo changes of conformation and molecular size as the pH is varied. These may help to explain some of the changes which occur in milk when it is processed. Nature 1967; 214 (5093):1101-1104.

75. McMeekin TL, Polis BD, DellaMonica ES, Custer JH. A Crystalline Compound of β-Lactoglobulin with Dodecyl Sulfate. J Am Chem Soc 1949; 71 (11):3606–3609.

76. Mobley DL, Chodera JD, Dill KA. On the use if orientational restraints and symmetry corrections in alchemical free energy calculations. J Chem Phys 2006; 125:084902.
77. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J Comput Chem 1998; 19:1639–1662.
78. Muresan S, van der Bent A, de Wolf FA. Interaction of β-Lactoglobulin with Small Hydrophobic Ligands As Monitored by Fluorometry and Equilibrium Dialysis: Nonlinear Quenching Effects Related to Protein−Protein Association. J Agric Food Chem 2001; 49 (5), 2609–2618.
79. Nanau RM, Neuman MG. Ibuprofen-induced hypersensitivity syndrome. Transl Res 2010; 155 (6): 275–293.
80. Narayan M, Berliner LJ. Mapping fatty acid binding to β-lactoglobulin: Ligand binding is restricted by modification of Cys 121. Protein Sci 1998; 7:150–157.
81. Nocedal J, Wright SJ. Numerical Optimization (Springer-Verlag), 1999.
82. Oostenbrink C, Villa A, Mark AE, Van Gunsteren WF. 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 2004; 25:1656–1676.
83. Pelletier E, Sostmann K, Guichard E. Measurement of Interactions between β-Lactoglobulin and Flavor Compounds (Esters, Acids, and Pyrazines) by Affinity and Exclusion Size Chromatography. J Agric Food Chem 1998; 46 (4), 1506–1509.
84. Pervaiz S, Brew K. Homology of beta-lactoglobulin, serum retinol-binding protein, and protein HC. Science 1985; 228(4697):335-337.
85. Pérez MD, Calvo M. Interaction of β-lactoglobulin with retinol and fatty acids and its role as a possible biological function for this protein: a review. J Dairy Sci 1995; 78:978–988.
86. Peters T Jr. All about Albumin: Biochemistry, Genetics, and Medical Applications. Academic Press, 1996. New York, NY.

87. Petitpas I, Grün T, Bhattacharya AA, Curry S. Crystal structures of human serum albumin complexed with monounsaturated and polyunsaturated fatty acids. J Mol Biol 2001; 314 (5): 955–960.

88. Piazza R, Iacopini S. Transient clustering in a protein solution. Eur Phys J 2002. E 7, 45–48.

89. Ponder JW, Case DA. Force fields for protein simulations. Adv Protein Chem 2003; 33.

90. Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, van der Spoel D, Hess B, Lindahl E. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics 2013; 29:845–854.

91. Qin BY, Bewley MC, Creamer LK, Baker HM, Baker EN, Jameson GB. Structural basis of the Tanford transition of bovine β-lactoglobulin. Biochemistry 1998; 37:14014–14023.

92. Qin BY, Creamer LK, Baker EN, Jameson GB. 12-Bromododecanoic acid binds inside the calyx of bovine β-lactoglobulin. FEBS Lett 1998; 438:272–278.

93. Qvist J, Davidovic M, Hamelberg D, Halle B. A dry ligand-binding cavity in a solvated protein. Proc Natl Acad Sci USA 2008; 105:6296–6301.

94. Ragona L, Fogolari F, Zetta L, Pérez DM, Puyol P, De Kruif K, Löhr F, Rüterjans H, Molinari H. Bovine β-lactoglobulin: Interaction studies with palmitic acid. Protein Sci 2000; 9:1347–1356.

95. Rahman A, Stillinger FH. Molecular Dynamics Study of Liquid Water. J Chem Phys 1971; 55, 3336.

96. Rashin AA, Rashin AHL, Jernigan RL. Protein flexibility: coordinate uncertainties and interpretation of structural differences. Acta Crystallogr D 2009; 65:1140–1161.

97. Rizzuti B, Pantusa M, Guzzi R. The role of Lys525 on the head-group anchoring of fatty acids in the highest affinity binding site of albumin. Spectrosc-Int J 2010; 24:159–163.

98. Rocha TL, Paterson G, Crimmins K, Boyd A, Sawyer L, Fothergill-Gilmore LA. Expression and secretion of recombinant ovine β-lactoglobulin in *Saccharomyces cerevisiae* and *Kluyveromyces lactis*. Biochem J 1996; 313:927–932.

99. Rothschild MDMA, Oratz M, Schreiber SS. Serum albumin. Hepatology 1988; 8(2): 385–401.

100. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comp Phys 1977; 23(3): 327–341.

101. Sahihi M, Heidari-Koholi Z, Bordbar AK. The interaction of polyphenol flavonoids with β-lactoglobulin: Molecular docking and molecular dynamics simulation studies. J Macromol Sci B 2012; 51:2311–2323.

102. Sakurai K, Oobatake M., Goto Y. Salt-dependent monomer–dimer equilibrium of bovine β-lactoglobulin at pH 3. Protein Sci 2001; 10(11): 2325–2335.

103. Sakurai K, Konuma T, Yagi M, Goto Y. Structural dynamics and folding of β-lactoglobulin probed by heteronuclear NMR. Biochim Biophys Acta-Gen Subj 2009; 1790:527–537.

104. Sawyer L, Brownlow S, Polikarpov I, Wu S. β-Lactoglobulin: Structural Studies, Biological Clues. Int Dairy J 1998; 8(2): 65–72.

105. Sawyer L. β-Lactoglobulin. Advanced Dairy Chemistry—1 Proteins 2003; 319-386 2003;

106. Sangster j. US EPI Estimation Program Interface (EPI) Suite (2000) KOWWIN v1.66 EPA and Syracuse Research Corporation (1994)

107. Schlick T. Molecular modeling and simulation: an interdisciplinary guide. 2nd edition. Springer 2010.

108. Shirtz MR, Pitera JW, Swope WC, Pande VS. Extremely precise free energy calculations of amino acid chain analogs: comparison of common molecular mechanics force fields for proteins. J Chem Phys 2003; 119:5740-5761.

109. Shirtz MR, Pande VS. Solvation free energies of amino acid side chains for common molecular mechanics water models. J Chem Phys 2005; 122:134508.

110. Shirtz MR, Chodera JD. Statistically optimal analysis of samples from multiple equilibrium states. J Chem Phys 2008; 129:129105.

111. Sudlow G, Birkett DJ, Wade DN. The Characterization of Two Specific Drug Binding Sites on Human Serum Albumin. Mol Pharmacol 1975; 11(6):824-832.

112. Sudlow G, Birkett DJ, Wade DN. Further Characterization of Specific Drug Binding Sites on Human Serum Albumin. Mol Pharmacol 1976; 12(6):1052-1061.

113. Sugio S, Kashima A, Mochizuki S, Noda M, Kobayashi K. Crystal structure of human serum albumin at 2.5 Å resolution. Protein Eng 1999; 12:439–446.

114. Timesheff SN, Townend R. Molecular Interactions in β-Lactoglobulin. V. the Association of the Genetic Species of β-Lactoglobulin below the Isoelectric Point. J Am Chem Soc 1961; 83(2):464–469.

115. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 2010; 31:455–461.

116. Van Gunsteren WF, Berendsen HJC. Algorithms for macromolecular dynamics and constraint dynamics. Mol Phys 1977; 34(5): 1311-1327.

117. Van Gunsteren WF, Karplus M. Effect of constraints on the dynamics of macromolecules. Macromolecules, 1982; 15(6):1528–1544.

118. Wang QWQ, Allen JC, Swaisgood HE. Protein concentration dependence of palmitate binding to β-lactoglobulin. J Dairy Sci 1998; 81:76–81.

119. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. J Comp Chem 2004; 25(9): 1157–1174.

120. Weber M. Improved Perron Cluster Analysis. ZIB-Report 03-04 (April 2003).

121. Wu SY, Pérez MD, Puyol P, Sawyer L. β-lactoglobulin binds palmitate within its central cavity. J Biol Chem 1999; 274:170–174.

122. Wu XL, Dey R, Wu H, Liu ZG, He QQ, Zeng XJ. Studies on the interaction of -epigallocatechin-3-gallate from green tea with bovine β-lactoglobulin by spectroscopic methods and docking. Int J Dairy Technol 2013; 66:7–13.

123. Wu D, Kofke DA. Phase-space overlap measures. I. Fail-safe bias detection in free energies calculated by molecular simulation. J Chem Phys 2005; 123:054103.

124. Wu D, Kofke DA. Phase-space overlap measures. II. Design and implementation of staging methods for free-energy calculations. J Chem Phys 2005; 123:084109.

125. Yang MC, Guan HH, Liu MY, Lin YH, Yang JM, Chen WL, Chen CJ, Mao SJT. Crystal structure of a secondary vitamin D_3 binding site of milk β-lactoglobulin. Proteins 2008; 71:1197–1210.

126. Yang MC, Guan HH, Yang JM, Ko CN, Liu MY, Lin YH, Huang YC, Chen CJ, Mao SJT. Rational design for crystallization of β-lactoglobulin and vitamin D_3 complex: Revealing a secondary binding site. Cryst Growth Des 2008; 8:4268–4276.

127. Yang MC, Chen NC, Chen CJ, Wu CY, Mao SJT. Evidence for β-lactoglobulin involvement in vitamin D transport *in vivo* – role of the γ-turn (Leu-Pro-Met) of β-lactoglobulin in vitamin D binding. FEBS J 2009; 276:2251–2265.

128. Zhang G, Zhao N, Wang L. Fluorescence spectrometric studies on the binding of puerarin to human serum albumin using warfarin, ibuprofen and digitoxin as site markers with the aid of chemometrics. J Lumin 2014; 13:2716-2724.

129. Zwanzig RW. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. J. Chem. Phys. 1954:22,1420.